Prediction during native and non-native language comprehension: The role of mediating factors

AINE ITO

A thesis submitted in fulfilment of requirements for the degree of

Doctor of Philosophy

to

Department of Psychology

School of Philosophy, Psychology and Language Sciences

University of Edinburgh

2016

Abstract

Psycholinguistic evidence suggests that people predict upcoming words during language comprehension. While many studies have addressed what information people predict, less is known about the role of factors that potentially mediate predictive processing. This thesis examines predictions of semantic information and word form information. It investigates whether predictive processing is mediated by availability of cognitive resources and time to generate predictions, and compares predictive processing in native (L1) speakers and non-native (L2) speakers. This thesis presents two major lines of work. Two eye-tracking studies investigate prediction of semantic and word form information using a visual world paradigm. In further two ERP studies, we address the interplay of semantic and word form information in a paradigm which combines both possibilities. Experiments 1 and 2 were an eye-tracking study conducted on L1 and L2 speakers of English. The study has demonstrated that L1 and L2 speakers predict semantic information, but their predictive eye movements are delayed when they are under a cognitive load. The effects of cognitive load on predictive eye movements suggest a role of cognitive resources in language prediction in both L1 and L2 speakers. Experiments 3 and 4 were another eye-tracking study conducted on L1 and L2 speakers. The study has shown that L1 speakers predict word form information, but L2 speakers do not. Experiments 5 and 6 were an ERP study, which investigated the interplay of prediction of semantic and word form information in L1 English speakers. Consistent with the two sets of eye-tracking experiments, L1 speakers predicted both semantic and word form information, but word form was only predicted when sentences were presented at a slower rate, while semantic information was predicted at standard and slow presentation rates. Experiments 7 and 8 used the same method as Experiments 5 and 6, conducted on L2 English speakers. L2 speakers comprehended sentences incrementally, but there was no clear evidence that they predicted semantic information or word form information. Experiments 5 – 8 suggest that prediction of word form information is mediated both by nativeness of the target language and by reading rates. To conclude, both L1 and L2 speakers make predictions, but prediction of semantic information occurs only when there are enough cognitive resources available. Prediction of word form can occur in L1 speakers, but it occurs only when there is enough time available. There is no evidence that L2 speakers predict word form, suggesting a role of nativeness of the target language. The findings are consistent with the production-based prediction model of language prediction, in that prediction of word form is less likely to occur compared to prediction of semantic information. Furthermore, the findings are also consistent with the claim that not everyone makes predictions, and predictions do not always occur. The thesis concludes that prediction is additional processing for the comprehension system, and is not always implicated in the comprehension system.

Acknowledgements

Many people have provided support that enabled me to complete this thesis. First, I would like to thank my supervisors, Prof. Martin J. Pickering, Dr. Martin Corley, Dr. Mante S. Nieuwland and Dr. Andrea E. Martin for leading me to achieve various academic and research goals as well as for allowing me to improve my inadequacy. Martin (P), thank you for all your support which helped me to expand my knowledge, to develop experimental ideas, and to write theoretical discussions in a clearer manner. Martin (C), thank you for your advice and guidance that let me develop skills to design and conduct eye-tracking experiments, for teaching me R and helping me tackle complex statistics issues, and for always being positive about my progress. Mante, thank you for teaching me the ERP technique from scratch, and for always providing me detailed feedback at an amazing speed. Andrea, thank you for always valuing my thoughts, for suggestions and advice regarding my career, and for sharing your own strategies for writing papers. These are only a small part of your support – I would like to thank you for everything. I am so lucky that I have got such a wonderful supervision team. My sincere gratitude also goes to my examiners Prof. Gerry Altmann and Prof. Holly Branigan, for giving me thoughtful comments as well advice for my future development. It was a great opportunity to have a discussion with you.

Second, the experiments in this thesis could not have been completed without support of various people. I am grateful to Dr. Madeleine Beveridge, Dr. Patrick Sturt, and Amie Fairs for help with recording the experimental stimuli and to Eleanor Drake for help with checking stimuli. I am also grateful to Dr. Sam Miller for helping me learn how to build eye-tracking experiments. I thank Dr. Ian Finlayson for advice on eye-tracking analyses, and Ronny Wiegand for technical help for conducting ERP experiments.

Third, I thank Dr. Mariana Vega-Mendoza for help with recruiting participants, Dominik Freunberger and Max Dunn for help with data collection, and all my participants. I am also grateful to all my officemates for friendly chats, discussions, encouragements and humour, and for giving me confidence about my progress. I would like to acknowledge PPLS School Research Support Grants for providing me with financial support.

Declaration

I hereby declare:

(a) that the thesis is my own composition, and

(b) that the work reported in this thesis has been carried out by myself, except where due acknowledgment is made in the text, and

(c) that the work has not been submitted for any other degree or professional qualification except as specified in text.

Aine Ito

Table of contents

1 Literature Review

During language comprehension, people can predict information about upcoming language. Hearing an utterance such as "*He went to the gym to swim in the …*," a listener may interrupt and say "*Oh, the pool?*" before the speaker completes the sentence. This commonly seen behaviour is compatible with the idea that the listener does not just passively comprehend the utterance, but also actively predicts what the speaker is likely to say next. What does such a prediction actually entail? Psycholinguistic evidence suggests that when a word is predictable, some information related to the word can be activated prior to its mention (generally called *pre-activation*). As we shall see, there is evidence that people predict semantic information (e.g., semantic category, animacy), phonological/ orthographic word form information, syntactic properties (e.g., grammatical gender), and some non-linguistic conceptual information (e.g., colour, shape). While it is well-documented what types of information can be pre-activated, it is less clear what is essential for predictions to occur. The primary aim of this thesis is to investigate how predictive processing is mediated by cognitive resources and time available when people generate predictions, and by whether people are comprehending in their native language (L1) or in a non-native language (L2).

The current chapter first provides an overview of existing findings which support the view that people make predictions during language comprehension. With regard to what types of information people predict, the current thesis investigates pre-activation of semantic and phonological/orthographic word form information. Therefore, the introduction starts with a review of studies which investigated pre-activation of these types of information (Section 1.1.1 – 1.1.2). It then briefly discusses what information can be used to generate predictions (Section 1.1.3). Section 1.2 discusses models of language prediction, and proposed mechanisms of how people make linguistic predictions. The models propose that there are several factors that could influence predictive processing. Focusing on these mediating factors, Section 1.3 reviews studies that found effects of working memory, time to generate predictions and language proficiency on predictive processing, which are closely related to the investigation of this thesis. Section 1.4 provides a methodological debate on cloze probability, which many of the past studies employed as a measure of predictability of a word, as does this thesis. It also introduces the methodologies used in this thesis. Lastly, Section 1.5 summarises the literature review and introduces the aim and scope of the current thesis.

1.1 Predicting language during sentence comprehension

When people comprehend a sentence, they do not usually wait until the end of the sentence to interpret its message. Instead, information in a sentence is processed incrementally (i.e., on a word-by-word basis), and newly encountered information is integrated into the sentence before the sentence reaches an end. A large number of studies have found evidence for incremental sentence processing across various languages. For example, people’s eye movements during reading comprehension reflect immediate integration of information that is being encountered in an unfolding sentence with information that has already been encountered in the preceding context and world knowledge that is already stored in the comprehender’s brain (Aoshima, Yoshida, & Phillips, 2009; Frazier & Rayner, 1982; Frisson, Rayner, & Pickering, 2005; Sedivy, 2002; Traxler & Pickering, 1996). Further evidence suggests that comprehenders not only process each word in a sentence as they encounter them, but also make predictions about upcoming words (Altmann & Kamide, 1999, 2007; Kamide, Scheepers, & Altmann, 2003; see Federmeier, 2007; Kamide, 2008 for reviews). Evidence for such language prediction has been shown using different experimental methods such as visual world eye-tracking (e.g., Altmann & Kamide, 1999), ERPs (e.g., Federmeier & Kutas, 1999) and self-paced reading (e.g., Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005).

Incremental sentence processing and predictive processing are thought to be tightly intertwined. Kamide (2008) explains that incremental processing forms the basis of predictive processing. If people delay their interpretations until the information of the entire sentence becomes available, there is no way that they can predict. Predictions about upcoming words are generally made using information extracted from preceding contexts (in combination with other types of information). If people are predicting upcoming words, they should be processing information conveyed in an on-going sentence rapidly while the sentence is still unfolding. Accordingly, evidence for prediction demonstrates that sentence processing is incremental. However, the opposite is not necessarily true. That is, people could comprehend sentences incrementally without making predictions, so evidence for incremental sentence processing on its own cannot demonstrate predictive processing. Nevertheless, successful prediction seems to facilitate on-going comprehension processing. For example, predictable words tend to be read faster, and they are more likely to be skipped compared to unpredictable words (Frisson et al., 2005; Hare, Elman, Tabaczynski, & McRae, 2009; Rayner & Well, 1996; Roland, Yun, Koenig, & Mauner, 2012; Smith & Levy, 2013; Staub, 2011). Furthermore, listeners tend to predict when the speaker’s turn is going to end (Magyari & De Ruiter, 2012), which is likely to make conversations more efficient. The relationship between incremental processing and predictive processing suggests that the two processes aid each other during language comprehension. Given the effects of prediction on general sentence comprehension, the mechanisms of language prediction have important implications on the models of how people process sentences. This thesis aims to discuss such implications based on the investigation on factors that could affect predictive processing. In the next section, we consider previous studies that investigated predictive processing during language comprehension.

1.1.1 Prediction of meaning

Evidence from eye-tracking studies

Altmann and Kamide (1999) recorded participants’ eye movements while showing them a scene depicting a boy, a cake, and several inedible objects (toys). Participants listened to a predictable or an unpredictable sentence: “*The boy will eat the cake*” or “*The boy will move the cake*.” When participants heard the predictable verb *eat*, there were more saccadic eye movements to the cake before the acoustic onset of the word *cake*, compared to when they heard the unpredictable verb *move*. These predictive eye movements indicate that participants immediately processed the semantic information of the verb, and predicted which of the shown objects meets the selectional restrictions of the verb (e.g., edible). Kamide, Altmann, and Haywood (2003) extended the findings to show that people make predictive eye movements not only to predictable direct objects but also to predictable indirect objects in sentences with a double-object structure (Experiment 1), and that people can also make use of their world knowledge (Experiment 2) or syntactic case markers (Experiment 3) in combination with semantic information to predict upcoming referents.

The results from these two studies (Altmann & Kamide, 1999; Kamide, Altmann, et al., 2003) are compatible with the interpretation that people predicted which of the depicted objects was likely to be mentioned and directed their eyes to what they predicted. One potential counterargument for this interpretation is that the observed results only reflect priming effects from the visual scene and context words (cf. Bar, 2007; McRae, Hare, Elman, & Ferretti, 2005). This priming-only account assumes that predictable words are pre-activated by the visual scene and lexical associates which have appeared in contexts. There is strong evidence that names of the depicted objects are rapidly activated as soon as the visual scene appears, which in turn can facilitate recognition of the names (Huettig, Rommers, & Meyer, 2011). In addition to this processing, if predictable words (e.g., ‘*cake*’) had a stronger semantic association with verbs in the predictable condition (e.g., ‘*eat*’) than with verbs in the unpredictable condition (e.g., ‘*move*’), predictable words were more likely to be primed by the preceding verb (Ferretti, McRae, & Hatherell, 2001). This priming effect may result in activation of the target words before they were mentioned. However, Knoeferle, Crocker, Scheepers, and Pickering (2005) precluded the possibility that the predictive eye movements only reflect priming from contextual words by showing that predictive eye movements can occur in the absence of semantic associations that can cause priming effects. They used pictures of characters performing atypical actions (e.g., a princess washing a pirate), so that one would expect no top-down influence from stored information (e.g., preceding words *princess* or *wash* are not likely to prime the object *pirate*). Critical sentences were German sentences with initial syntactic ambiguity between SVO and OVS structures (e.g., "*Die Prinzessin wäscht offensichtlich den Pirat*" or "*Die Prinzessin malt offensichtlich der Fechter*"; order-matched English translation: the princess (subject/object) washes apparently the pirate (object), or the princess (subject/object) paints apparently the fencer (subject)). Because the first noun phrase *Die Prinzessin* was ambiguous between the nominative case and the accusative case, it could be either the subject or the object of the sentence until the case marker of the second noun phrase (*den*/*der*) disambiguated it. Participants listened to these sentences while viewing a visual scene in which a princess is washing a pirate and is being painted by a fencer. After hearing the critical verb (*wäscht/malt*) but before hearing the disambiguating case marker, participants were more likely to fixate the appropriate object or subject (*den Pirat*/*der Fechter*), suggesting that they built a thematic relationship incrementally and predicted an upcoming referent based on the information of the critical verbs and the depicted actions. The results indicate that priming effects alone cannot explain predictive eye movements, because the thematic relations were atypical and no priming effects were expected to occur in this study. However, it is still possible that priming partly plays a role or has additive effects together with contextual information.

In an investigation of the role of priming, Kukona, Fang, Aicher, Chen, and Magnuson (2011) manipulated visual scenes so that they contained two objects that had equally strong lexical associations with preceding verbs. They presented participants with sentences such as "*Toby arrests the crook*" (predictable condition) or "*Toby notices the crook*" (unpredictable condition) with a scene containing pictures of *Toby*, *crook*, *policeman* and distractors (Kukona et al.; Experiment 1). The scene always contained a typical patient (*crook*) and a typical agent (*policeman*) of the predictable verb (*arrest*). If participants' eye movements are driven only by a priming effect and not by a prediction effect, the pictures of *crook* and *policeman* would be similarly likely to be fixated as the verb *arrest* is heard, because '*arrest*' would prime both '*crook*' and '*policeman*' to a similar degree. If a prediction effect occurs in the absence of a priming effect, the picture of *crook* would be more likely to be fixated than distractors, and looks to *policeman* would not differ from those to distractors. Upon hearing the verb *arrest*, participants were equally likely to fixate a predictable patient and a typical agent. It was only after the patient noun was mentioned that the predictable patient attracted more fixations than the agent picture. While this result seems to be more consistent with the priming account, a somewhat different result was found when passive sentences were used (Kukona et al.; Experiment 2). Participants saw the same scenes and heard passive counterparts of the previously used sentences (e.g., "*Toby was arrested by the policeman*."). In this experiment, there were more fixations on objects which the verb predicted (*policeman*) compared to unpredictable lexical associates of the verb (*crook*). This result allows the interpretation that the predictive eye movements are not a simple reflection of priming effects. The authors claimed that the boosted prediction effect (stronger fixation bias towards *policeman* versus *crook*) was probably due to (1) the presence of an additional syntactic cue *by* or (2) a longer period between the critical verb (*arrest*) and the predictable agent (*policeman*). This different result from their Experiment 1 points to the possibility that predictive processing may occur only when there is a strong cue or when there is more time to build predictions. Their Experiment 2 also found that there were more looks to thematic associates (*crook*) compared to unrelated distractors. This result suggests that thematic priming contributes, at least partly, to predictive eye movements.

In sum, predicting semantic information appears to occur via both priming effects and combinatorial use of bottom-up contextual information and top-down world knowledge.

Along with what the reviewed studies implicate, some researchers note that the role of priming should be considered when discussing models of prediction (Bar, 2007; DeLong, Troyer, & Kutas, 2014; Huettig, 2015a). A parsimonious interpretation is that priming and predictive processing both facilitate the processing of upcoming words. However, in visual world studies, listeners would activate names of objects in visual scenes immediately after the visual information becomes available. This means that a word that corresponds to an upcoming referent becomes more active than words that are unrelated to the scene purely from the visual information. This poses a question; do people predict upcoming words when there is no visual context to help them narrow down likely referents and any word could potentially be mentioned? In fact, a number of ERP studies investigated language prediction by presenting participants with sentences only, and several of them aimed to dissociate prediction effects from lexical priming effects (Amsel, DeLong, & Kutas, 2015; Brothers, Swaab, & Traxler, 2015; Lau, Holcomb, & Kuperberg, 2013; Metusalem et al., 2012; Otten & Van Berkum, 2008). While integrating contextual information, comprehenders build representations of an on-going event in their brain. This is likely to give rise to pre-activation of semantic information that is related to upcoming information (Kim, Oines, & Sikos, 2016). The following section discusses ERP studies which investigated pre-activation of semantic information.

Evidence from ERP studies

Early evidence for meaning pre-activation comes from a study by Federmeier and Kutas (1999). They created sentence contexts that predicted a specific word using a cloze test (cf. Taylor, 1953), in which participants were asked to complete sentence fragments. The predictability of a word was defined using *cloze probability*, a proportion of participants who used that word to complete the sentence (see Section 1.4.1 for detailed discussion about the cloze procedure). Participants read discourse contexts, for which the cloze pre-test revealed a high proportion of the participants produced the predictable word (e.g., “*They wanted to* *make the hotel look more like a tropical resort. So along the driveway, they planted rows of…*"). These contexts were followed by the predictable word (*palms*), an implausible word that belonged to the same semantic category as the predictable word (*pines*), or an implausible word from a different semantic category (*tulips*). Participants showed classic N400 anomaly effects (Kutas & Hillyard, 1980) for both semantically related and unrelated implausible words relative to plausible predictable words. The critical difference was that the N400 effects were smaller for semantically related words (*pines*) relative to semantically unrelated words (*tulips*). Furthermore, this N400 reduction was greater in *high-cloze* contexts (*Mean cloze probability* = 90%) than in *medium-cloze* contexts (*Mean cloze probability* = 59%) (for similar findings, see Thornhill & Van Petten, 2012).

Metusalem et al. (2012) extended the findings by Federmeier and Kutas to show that N400 reductions for words related to predicted semantic information occur in the absence of strong lexical priming from context words. Their participants read scenarios describing an event, followed by a high-cloze sentence (e.g., "*A huge blizzard ripped through town last night. My kids ended up getting the day off from school. They spent the whole day outside building a big…*"). The high-cloze sentences contained a highly predictable word (*snowman*), an implausible word that was related to the described event (*jacket*), or an implausible, event-unrelated word (*towel*). They found reduced N400 effects for event-related words relative to event-unrelated words. This N400 reduction was not found when the scenarios that conveyed the event description were removed. The lack of the N400 reduction in the latter condition means that, if lexical priming contributed to the N400 reduction, prime words must be in the removed scenarios. The authors suggested that lexical priming from these scenarios was highly unlikely, because critical words were several words distant from the scenarios, and lexical priming effects during such a discourse comprehension have been found to be weak or null (Camblin, Gordon, & Swaab, 2007). Therefore, their findings show that (event-related) semantic information can be pre-activated based on message-level information independently of lexical priming. Although Metusalem et al. did not control for plausibility, Amsel et al. (2015) replicated their findings while controlling for plausibility as well as for semantic relatedness to context words.[1]

A related argument also comes from a study by Otten and Van Berkum (2008; Experiment 1B). In their study, participants read contexts that contained a word that was either predictable or unpredictable from the preceding context. Both predictable and unpredictable contexts contained the same content words which could prime the target word. Therefore, if prediction is merely a by-product of priming effects, no difference at the target word was expected as a function of the predictability conditions. Predictability effects were examined by comparing ERPs of gender-marked adjectives preceding a critical noun in the predictable and unpredictable conditions. Adjectives that matched a predictable noun in gender elicited a positive ERP deflection between 900 and 1200 ms, relative to prediction-inconsistent adjectives. The results therefore rule out the account that priming from context words is the only cause for prediction, and instead support the notion that the message-level information of the context triggers prediction independently from priming effects.

Now, how do lexical information and contextual information interact? Boudewyn, Long, and Swaab (2015) investigated the combinatotial use of message-level information and lexical information for prediction. They had participants listen to high-cloze contexts (e.g., “*Frank was throwing a birthday party, and he had made the dessert from scratch. After everyone sang, he sliced up some*…”) which induced a prediction of a specific word (*cake*). The contexts were followed by a feature word describing a prototypical feature (*sweet*) or an atypical feature (*healthy*) of the predictable word. After the feature word and a few neutral words, participants heard either the word that the global context predicted (*cake*) or a word

---

[1] The authors did not conclude that the results exclusively show pre-activation, but suggested a possibility of an integration account or combinatory effects from pre-activation and integration.

that was unpredictable but semantically compatible with the atypical feature word (*veggies*). This manipulation assured that the prototypical feature word would support the prediction built from the preceding context, while the atypical feature word would disconfirm the prediction and might force a comprehender to revise the prediction. Participants showed N400 effects for feature words that were incompatible with the predictable words (*healthy*) relative to compatible feature words (*sweet*). Since the cloze test indicated that these feature words were not predictable, this facilitatory effect for prediction-consistent feature words fits with the account that semantic features of predictable words are pre-activated before they are encountered. Predictability effects were found at the critical words too; unpredictable words elicited larger N400s than predictable words, and this effect was greater for words preceded by an atypical feature word (*healthy* - *cake* or *sweet* - *veggies*) than for words preceded by a prototypical feature word (*sweet* - *cake* or *healthy* - *veggies*). Typicality of feature words and predictability of critical words interacted in the post-N400 time window as well. Unpredictable words elicited more positive ERPs than predictable words at frontal channels when they were preceded by a prototypical feature word (*sweet* - *cake* or *healthy* - *veggies*). In contrast, when critical words were preceded by an atypical feature word (*healthy* - *cake* or *sweet* - *veggies*), predictable words elicited more positive ERPs than unpredictable words, and this difference was broadly distributed across channels. Relative to the N400 effect, the late positivity effect is still regarded as under-investigated, but it is supposed to reflect detection of a conflict, and potentially reanalysis processing for a conflicting input (e.g., prediction-inconsistent words) (Van de Meerendonk, Kolk, Vissers, & Chwilla, 2010) or processing cost for the input that is hard to integrate into the sentence (Van Petten & Luka, 2012). The overall results of Boudewyn et al.'s study show that global contexts and local prediction-consistency have differential and interactive effects on predictive processing.

ERP studies therefore reveal evidence for pre-activation of semantic information, strengthening the evidence that was found in eye-tracking studies. Consistent with the eye-tracking studies, the findings also suggest that both contextual information and priming effects function in predictive processing independently.

Key characteristics of semantic pre-activation

The aforementioned studies demonstrate that people predict semantic information of upcoming words. These studies also suggest that pre-activation of semantic information of predictable words could occur through building up message-level information from contexts and/or via lexical priming from preceding context words (Kukona et al., 2011; Metusalem et al., 2012; Otten & Van Berkum, 2008). Through both or either of these pathways, pre-activation of semantic information does not necessarily give rise to prediction of a specific lexical item. As listening to the context "*The boy will eat…*," people may pre-activate concepts that fit with the context (i.e., things that a boy is likely to eat) or that are associable with a boy, edible things, or an occasion of eating, but may not pre-activate the phonology or orthography of the word *cake* (or pre-activation of the word *cake* may be possible only with the presence of the visual scene depicting a cake).

An alternative possibility is that a specific word is predicted first, and as the predicted word becomes active, the activation spreads to information that is semantically related to the predicted word. When a context bears strong constraints about upcoming words, it is possible that comprehenders predict a specific word. People may also predict multiple words that are highly likely to be the continuation of a sentence. This can lead to the activation of the semantic information that is shared among these words.

Current evidence for pre-activation of semantic information appears to be consistent with both accounts. A critical difference between the two accounts is whether a specific word is predicted or not. If people predict a specific word when they pre-activate semantic information of that word, other lexical information such as phonological information of that word should also be pre-activated, because activation of a word should entail all the lexical information associated with that word. Thus, it is important to investigate pre-activation of other types of linguistic information. To this end, the section below reviews studies that examined pre-activation of phonological/ orthographic word form information, and discusses how this type of pre-activation is different from pre-activation of semantic information.

1.1.2 Prediction of phonological/ orthographic word form

Evidence from eye-tracking studies

Eye-tracking studies on reading comprehension have shown that predictable words tend to be read faster or tend to be skipped relative to less predictable words (Frisson et al., 2005; Rayner, Slattery, Drieghe, & Liversedge, 2011; Roland et al., 2012; Smith & Levy, 2013). However, the facilitatory effects on reading predictable words might be related to predictable words being easier to integrate into the context than unpredictable words. Hence, the predictability effects found in these studies are compatible with both prediction accounts and integration accounts. One possible reason for an easier integration cost for predictable words is that predictable words are more plausible to fit into the context than unpredictable words. Because of this (cf. Dambacher et al., 2012), it is hard to dissociate plausibility effects from predictability effects. However, prediction-related effects observed before the occurrence of predictable words (e.g., Altmann & Kamide, 1999; Van Berkum et al., 2005) would be able to rule out the plausibility accounts, because plausibility information becomes available only when the word is accessed. Predictive eye movements were observed in the visual world paradigm investigating prediction of semantic information, but no visual world eye-tracking study seems to have investigated pre-activation of word form information. Experiments 3 and 4 in the current thesis aim to fill this gap.

Evidence from ERP studies

DeLong, Urbach, and Kutas (2005) took advantage of the English phonological rule that the article *a* precedes consonant-initial words and the article *an* precedes vowel-initial words to investigate whether people pre-activate phonological information of predictable words. Participants read high-cloze sentence contexts (e.g., "*The day was breezy so the boy* *went outside to fly…*") followed by the predictable article-noun combination (*a* followed by *kite*) or an unpredictable but plausible article-noun combination (*an* followed by *airplane*). The predictable noun began with a vowel and the unpredictable noun with a consonant, or vice versa. As expected, the N400 was reduced when the noun was predictable. Critically, the N400 reduction was also found when the preceding article was predictable (and thus matched the noun). DeLong et al. therefore argued that participants pre-activated representations (e.g., an initial consonant) of the upcoming noun before the appearance of the noun (for similar results see DeLong, Groppe, Urbach, & Kutas, 2012; Martin et al., 2013). These results therefore suggest that word form (at least the first phoneme of the word, but crucially not merely aspects of meaning) is pre-activated.

A different approach was taken by Laszlo and Federmeier (2009), who had participants read predictable contexts (e.g., "*Before lunch he has to deposit his paycheck at the…*") that were completed by a predictable word (*bank*), words that were orthographically related to the predictable word (*bark*), pseudowords that were orthographically related to the predictable word (*pank*), and illegal strings that were orthographically related to the predictable word (*bxnk*). In matched unrelated conditions, participants read other predictable contexts (e.g., "*She loves the way the leaves change colors in the…*") that were completed by the predictable word (*fall*), and orthographically unrelated sets of words (*hook*), pseudowords (*jank*), and illegal strings (*tknt*). For words, pseudowords, and illegal strings, Laszlo and Federmeier found a reduced N400 for the forms that were orthographically related to the predictable word as compared to the forms that were orthographically unrelated. They concluded that pre-activated orthographical features impact semantic processing prior to any filter on lexical status.

In a related study, Kim and Lai (2012) found no N400 effect for pseudowords that were orthographically similar to predictable words (e.g., "*She measured the flour so she could bake a ceke*"), whereas dissimilar pseudowords (*tont*) elicited a clear N400 effect.

Orthographically similar pseudowords elicited an enhanced P600, which was smaller than the P600 elicited by illegal strings (*srdt*). Similarly to Laszlo and Federmeier (2009), Kim and Lai argued that the impact of prediction occurs before visual word recognition (because pseudowords cannot be recognised as words). Both Laszlo and Federmeier (2009) and Kim and Lai (2012) interpreted the respective reduced or absent N400 effects associated with orthographical overlap as evidence for pre-activation of orthographic information.

In addition to the effects in the N400 window, both of the studies (Kim & Lai, 2012; Laszlo & Federmeier, 2009) revealed post-N400 positive ERP effects (LPC effects) for non-words that were closely related in form to predictable words, and the LPC effects seem to indicate that comprehenders consider the form of the predictable word. Newman and Connolly (2004) and Vissers, Chwilla, and Kolk (2006) found that pseudohomophones that were orthographically similar to highly predictable words (e.g., *bouks* for the predictable *books*) elicited larger LPCs than predictable words, but pseudohomophones that were orthographically similar to unpredictable words did not. Similarly, Laszlo and Federmeier (2009) and Kim and Lai (2012) reported a post-N400 LPC effect for pseudowords that were orthographically similar to the predicted words. Along with Vissers et al. (2006), both sets of authors interpreted the effect as a detection of a conflict between predicted and actually encountered words. An important distinction to be made between the LPC effect and the N400 reduction is that the LPC effect was not interpreted as evidence for prediction of word form. Encountering an input that resembles a highly predictable word can prime the predictable word due to the word form similarity (Ferrand & Grainger, 1994; Grainger & Ferrand, 1996) and boost the activation of the predictable word (whose semantic information may already be active). This way, the conflict between the two words can be detected, and word form information of the predictable word becomes active only after the form-related input is encountered. Thus, while the LPC effect appears to be related to the strength of word predictability, it does not necessarily show that word form information was pre-activated.

Another reported index of prediction-related ERP effect is early negativity, referred to as Phonological Mismatch Negativity (PMN) or N200s. This ERP component is associated with phonological mismatch between expected words and an actual input, and tends to be prominent at frontal ERP channels. Van den Brink, Brown, and Hagoort (2001) had participants listen to high-cloze sentences which varied at the sentence-final word. The sentence ended with the predictable high-cloze word, a phonologically related word whose initial phonemes were the same as the high-cloze word, or an unrelated word that did not share initial phonemes with the high-cloze word. Only the predictable words were semantically plausible to finish the sentences. Compared to predictable words, unrelated anomalous words elicited larger N200s, followed by a semantic anomaly N400 effect, but phonologically related anomalous words elicited the N400 effect only. Boudewyn et al. (2015) further showed that N200 effects were present when encountered words violated predictions of a specific word form, but not when they violated predictions of semantic features. However, like the LPC effect, the phonological mismatch effect alone does not prove prediction, because it could reflect a bottom-up selectional process, in which the encountered input is assessed with respect to the goodness of semantic fit with the context (Van den Brink et al., 2001). This N200 effects were not found in the visual modality (Connolly, Phillips, & Forbes, 1995). Therefore, the N200 effect related to phonological mismatch between a predicted word and encountered input appears to be specific to auditory modality.

In sum, results from the reviewed ERP studies suggest that people pre-activate word form information. However, there seems to be less evidence for word form pre-activation compared to evidence for semantic pre-activation. As a further exploration, Experiments 3 and 4 in this thesis investigated pre-activation of word form using eye-tracking.

Key characteristics of phonological/ orthographic word form pre-activation

As opposed to pre-activation of semantic information, pre-activation of the phonological/ orthographic word form is not likely to occur unless a specific word is predicted. This assumption comes from the difference in the process of pre-activation. Predictable words often have some semantic relations to context words (which is likely to aid pre-activation of relevant semantic information), but phonological/ orthographic relations between context words and newly-encountered words are normally not stronger than chance-level. This means that pre-activation of word form information is not expected to occur via priming from context words. Given the lack of a priming pathway, pre-activation of word form is most likely to be mediated by pre-activation of a specific word.

Thus, while pre-activation of semantic information may or may not be mediated by a prediction of a specific word, pre-activation of word form information must occur as a result of a prediction of a specific word. This critical difference leads to a hypothesis that pre-activation of word form is less likely to occur compared to pre-activation of semantic information. Experiments 5 – 8 in this thesis tested this hypothesis by investigating co-occurrence patterns of semantic and word form pre-activation. Previous studies have not examined pre-activation of semantic and word form information in a single experiment, so the relationship between the two types of pre-activation is unknown. Investigation of this relationship would allow deeper understanding of the mechanism of predictive processing, such as at which stage each level of pre-activation likely occurs.

1.1.3 Cues that trigger prediction

The sections above argued that prediction entails pre-activation of different features of upcoming words. The extant findings also suggest that various types of cues can trigger prediction of upcoming words. Such cues include semantic information of preceding verbs (Altmann & Kamide, 1999; Kukona et al., 2011) or adjectives (Tribushinina & Mak, 2016), contextual message (Federmeier & Kutas, 1999; Laszlo & Federmeier, 2009), and event knowledge built up in contexts (Knoeferle et al., 2005; Metusalem et al., 2012; Otten & Van Berkum, 2008). In addition, predictions can be triggered by syntactic cues such as case markers (Kamide, Scheepers, et al., 2003), grammatical number markings (see Kouider, Halberda, Wood, & Carey, 2006; Robertson, Shi, & Melançon, 2012 for evidence among children; but for inconsistent findings among adults, see Riordan, Dye, & Jones, 2015), syntactic dependencies (Sussman & Sedivy, 2003) or syntactic structures probabilistically associated with specific verbs (Arai & Keller, 2013). Syntactic information can also be used to predict when a speaker's turn ends in natural conversations (Magyari & De Ruiter, 2012). Furthermore, predictions can be generated by information accompanied by speech such as prosody (Hirose & Mazuka, 2015; K. Ito, Jincho, Minai, Yamane, & Mazuka, 2012; K. Ito & Speer, 2008; Nakamura, Arai, & Mazuka, 2012; Weber, Grice, & Crocker, 2006) or disfluency (Arnold, Kam, & Tanenhaus, 2007; Corley, 2010; Kidd, White, & Aslin, 2011). These types of cues are often used in combination to restrict properties of upcoming words (Altmann & Kamide, 2007; Kamide, Scheepers, et al., 2003). Altogether, these studies suggest that comprehenders are sensitive to various types of available information (e.g., semantic, syntactic, prosodic), and can efficiently make use of the information to make predictions.

1.2 Theories of language prediction

In the previous section, I have discussed what information people predict, and what information can be used to generate predictions. The current section focuses on the question how people make predictions, and discusses proposed models of language prediction.

1.2.1 Prediction vs. Integration accounts

Integration accounts for predictability related effects do not assume that comprehenders make predictions. Instead, they attribute facilitatory effects from word predictability to increased ease of integrating predictable words into the unfolding sentence relative to unpredictable words (see Kutas, DeLong, & Smith, 2011 for a discussion on the debate about the prediction view versus the integration view). For example, predictable words may be a more plausible continuation of a sentence compared to less predictable words, and greater plausibility generally results in facilitatory effects such as faster reading time (Rayner, Warren, Juhasz, & Liversedge, 2004). Even though the plausibility is controlled, predictable words may still be easier to integrate then unpredictable words due to their predictability (cf. Amsel et al., 2015). Experiments showing facilitatory effects for predictable words do not allow teasing apart effects of integration ease and effects of prediction, as the facilitatory effects could stem from either or both of them (cf. Hagoort & Indefrey, 2014). However, previous studies did address this issue by examining prediction-related effects occurring before the presentation of predictable words. For example, inhibitory effects on words that are plausible but inconsistent with predictable words (e.g., effects on '*an*' preceding a predictable word '*kite*') cannot be influenced by plausibility effects, because the prediction-inconsistent words ('*an*') and prediction-consistent words ('*a*') were equally plausible to fit into the contexts. Of course, plausibility of not-yet-encountered predictable words also cannot affect processing of preceding words, unless these words are predicted (DeLong et al., 2005). Such inhibitory effects for plausible but prediction-inconsistent words encountered before predictable words were also reported for mismatches in animacy (Szewczyk & Schriefers, 2013) and in grammatical gender (Van Berkum et al., 2005; Wicha, Bates, Moreno, & Kutas, 2003; Wicha, Moreno, & Kutas, 2004, 2003). The findings from these studies are therefore more compatible with the prediction account.

In sum, current evidence clearly shows that processing of predictable words is facilitated relative to unpredictable words. Among the findings showing effects of predictability, some are compatible with both prediction accounts and integration accounts. Yet, there are studies showing predictability effects before predictable words are encountered, and these findings can rule out integration-only accounts.

1.2.2 Production-based prediction accounts

Pickering and Garrod’s (2007, 2013) theoretical model of language comprehension proposes that the language production system is used for language prediction (see also Dell & Chang, 2014; Federmeier, 2007). In the first place, Pickering and Garrod propose that language comprehension and language production systems are interwoven, and that the language production system aids prediction during language comprehension. The assumption of the use of the production system during comprehension is in line with findings that hearing another person speaking activates a comprehender’s motor system used to produce speech (Pulvermüller et al., 2006; Scott, McGettigan, & Eisner, 2009; Wilson, Saygin, Sereno, & Iacoboni, 2004). According to the production-based prediction account, people covertly imitate a speaker’s utterance using their own production system during comprehension (Gambi & Pickering, 2013; Garrod, Gambi, & Pickering, 2013). People then make use of their comprehending experience and production experience to predict upcoming words. A related argument is made by Huettig (2015a), who agrees on the point that comprehenders use their production system to make predictions. While Pickering and Garrod (2013) propose that comprehenders do not use a fully implemented production system for prediction, Huettig’s model posits that the production system for prediction is fully implemented. The production-based prediction accounts are consistent with the findings that brain activity in areas associated with predictive processing tends to be more synchronous between a speaker and a listener when the listener is hearing predictive utterances compared to when he or she is hearing non-predictive utterances (Dikker, Silbert, Hasson, & Zevin, 2014). Another support comes from Federmeier (2007), who proposed a framework that is compatible with the production-based prediction accounts based on ERP evidence on hemispheric differences. In her framework, the left hemisphere plays a predominant role in language production, and it is the left hemisphere that subserves pre-activation of likely upcoming words (and their associated information). This suggests that the language production system and language prediction system may share the same resources.

The production-based prediction models agree on the hypothesis that people with a higher language production skill show a greater degree of prediction. Several studies tested this hypothesis by examining relationships between children's prediction performance and their production/ comprehension vocabulary. Replicating Altmann and Kamide's (1999) findings, Mani and Huettig (2012) found that two-year-old children showed predictive eye movements to likely upcoming referents. They also found that the children's degree of predictive eye movements positively correlated with their production vocabulary size, but not with their comprehension vocabulary size. The results suggest that the language production system, but not the comprehension system, may play a role in prediction (see also Mani, Daum, & Huettig, 2015). However, what Borovsky, Elman, and Fernald (2012) found was contradictory to the findings of Mani and Huettig (2012). Borovsky et al. found that both adults (aged between 18 and 28 years) and children (aged between 3 and 10 years) with a higher comprehension vocabulary size were faster to direct their looks to target objects in a predictive fashion than those with a poorer vocabulary size. Moreover, participants' sentence completion (production) test scores did not correlate with their prediction performance. Their findings suggest that people's comprehension skill rather than production skill is related to predictive processing. On the other hand, Nation et al. (2003) found similar prediction performances between children with less skilled comprehension skill and control children. These studies differed in the measurements they used to evaluate participants' production and comprehension skills, and in the tested population (e.g., the age group of the participants). Due to the inconsistency among these studies, it is hard to determine a clear relationship between people's prediction performance and their production and comprehension skill. Moreover, correlational studies by nature cannot prove a causal link, so it is unclear if a better production or comprehension skill leads to a stronger prediction performance or vice versa.

In an ERP study, Federmeier, Kutas, and Schul (2010) used a speeded language production task to examine the link between the language production mechanism and predictive processing in an elderly population (mean age of 68 years). In their Experiment 1, participants saw semantic category cues (e.g., “An insect”), and then a high typicality exemplar *ant*, a low typicality exemplar *hornet*, or an incongruent exemplar *gate*. Their ERPs showed a larger frontal positivity for low typicality exemplars compared to the other two types of words, which were manifest in the 500-900 ms post-stimulus time window. The frontal positivity effects, which are thought to be indicative of disconfirmed prediction (cf. Thornhill & Van Petten, 2012; Van Petten & Luka, 2012), correlated with the participants’ category fluency (the number of words people can generate in one minute within a specified semantic category). Notably, the frontal positivity effects did not correlate with the participants’ comprehension vocabulary sizes or their linguistic and non-linguistic working memory measures. In Experiment 2, the same set of participants came back after a five-month lag, and were asked to produce a word as quickly as possible after the same cues. The authors found that the degree of the prediction-related frontal positivity effects correlated with the participants’ standardised response time in this category word production task.[2] The relationships between prediction-related ERP effects and language production scores are consistent with the notion that production system is involved during predictive processing. However, the production tasks they used were limited to the production of semantic category exemplars, and the study did not investigate the link in the younger population. Therefore, further evidence would be needed to probe how generalizable their findings are.

Stepping aside from correlational investigation, Drake and Corley (2015b) investigated how prediction of specific word form affects picture naming (i.e., production)

---

[2] The correlation with non-standardised response time was reported as marginally significant.

latencies. Participants listened to sentence contexts that supported prediction of a specific word (e.g., *He managed to fix the drip from the old leaky…*). They then named a picture that was presented immediately after the sentence fragment. The picture names corresponded to the predictable word (*tap*), or a word phonologically related to that word (*cap*/*tan*), or a word that had no phonological overlap (*cone*). They hypothesised that if people use their production systems to make predictions, picture naming would be facilitated when the name of the picture shared phonological information with the predicted word. Against this hypothesis, they found no facilitatory effect or inhibitory effect of the phonological overlap. Picture naming latencies were shorter for pictures of a predictable word, but phonological overlap pictures were not named faster than no-overlap pictures. It appears that the findings fail to support production-based prediction accounts. In a follow-up study, however, Drake and Corley (2015a) pointed out that the picture naming latency might not be an appropriate measurement, and investigated the involvement of the language production system using ultrasound imaging, which would better reflect the involvement of the speech articulatory system. After listening to high-cloze sentence contexts, participants named a picture that corresponded to the predictable word (*tap*) or to a word that phonologically mismatched the predictable word at onset (*cap*). The ultrasound recordings in these conditions were compared with those in a pair of control conditions, in which participants named the same pictures without any sentence context. The difference in the ultrasound data between the onset mismatch condition and its control condition was greater than the difference between the prediction-match condition and its control condition. This greater change in articulatory movements for the onset mismatch condition was found before the picture naming onset. The results show that predictions influenced comprehenders’ production (articulation) system, and therefore suggest that the speech production system is employed in prediction generation.

Hintz and Meyer (2015) also investigated whether use of the production system facilitates prediction using the visual world eye-tracking paradigm. While looking at an analogue clock face, participants listened to simple mathematical equations (e.g., Dutch equivalent of "*Three plus eight is eleven*.") after which they heard the answer or produced the answer verbally. On both comprehension and production trials, participants fixated the number which corresponded to the result before the acoustic offset of the word '*is*'. Notably, these predictive eye movements were stronger when participants produced the result numbers than when they did not. Therefore, the engagement of a comprehender's production system indeed seems to facilitate predictive processing. However, it should be noted that, with regard to the experimental manipulations, the authors discussed points where questions remain. First, participants were instructed to move their eyes to mentioned numbers as quickly as possible. Although the task was not to predict the result numbers, predictions can benefit the eye movement speed, so this instruction could have led the participants to make predictions. Second, comprehending mathematical equations might involve processing that is specific to number processing or equation solving, which may not be employed in non-mathematical everyday language comprehension (Carreiras, Monahan, Lizarazu, Duñabeitia, & Molinaro, 2015). It would strengthen the evidence if the observed facilitatory effects as a function of the involvement of the production system persist when people comprehend more complex sentences without any instructions about eye movement control.

In a similar vein, Hintz, Meyer, and Huettig (2015) used a different experimental paradigm and investigated effects of production task on predictive processing. In their Experiment 2, half of the trials were self-paced reading, in which participants read predictive or non-predictive sentences word-by-word in a moving window format (i.e., as the next word appeared, the previous word was replaced with an underscore). The other half involved a production task, in which participants named a picture that was either predictable or unpredictable based on a context which was presented auditorily right before. These trials were randomly intervened to make an experimental list. The participants' reading times in the self-paced reading task were compared to the reading times from another set of participants who did not perform the object-naming (production) task. The authors found faster reading times for a spill-over region (i.e., a word after the target) after predictive contexts versus after non-predictive contexts. Critically, this facilitation effect was found for participants who performed both production and comprehension tasks, but not for participants who performed only the comprehension task. The results suggest that performing the production task encourages participants to use their production system to make predictions, and this greater involvement of the production system facilitates prediction.

To sum up, empirical studies support production-based prediction accounts by showing that people with a higher production skill exhibit a better prediction performance or by showing that employment of a production system during comprehension facilitates predictive processing. However, reported findings about the relationship between prediction performance and production/ comprehension skills are not consistent. Experiments 5 – 8 in this thesis evaluated production-based prediction accounts from a different perspective by examining whether the patterns of pre-activation of semantic and word form information would fit with the models of language production.

1.2.3 Multiple mechanisms accounts

From a slightly different perspective, Huettig (2015a) proposes that prediction is made using different types of mechanisms, including production, association, combinational, and simulation. The central argument of this account is that there are multiple systems that jointly aid predictive processing. Supportive evidence for this argument comes from findings which indicate that the degree of prediction correlate with two or more factors independently. For example, Rommers, Meyer, and Huettig (2015) showed that strength of predictive eye movements towards predictable objects was systematically related to people's vocabulary size and verbal fluency, whereas the looks towards objects that shared shape features with the predictable objects were related to people’s performance on a non-linguistic task (spatial cueing task). Based on these findings, it seems that there are linguistic and non-linguistic systems that contribute to different kinds of predictive mechanism. There is also evidence that predictions can be triggered by combinatorial mechanisms which make use of priming via lexical association and message-level information (Kukona et al., 2011). However, not many studies have directly probed roles of different mechanisms, so further investigation appears to be needed to substantiate this account.

1.3 Mediating factors for predictive performance

Theories about the mechanisms of prediction suggest that there are individual differences in people’s predictive performances. Under production-based prediction accounts (Dell & Chang, 2014; Pickering & Garrod, 2007, 2013), it is expected that people’s language prediction performance is mediated by their language production skill. Multiple mechanisms accounts (Huettig, 2015a) stipulate that people’s prediction skill is mediated by multiple different factors, including language production skill. Consistent with these accounts, past research has found effects of production skill (Mani & Huettig, 2012), age (Federmeier & Kutas, 2005) or literacy (Huettig & Brouwer, 2015; R. K. Mishra, Singh, Pandey, & Huettig, 2012). The current section discusses how people’s predictive performances are mediated by their working memory, available time to generate predictions and language proficiency.

1.3.1 Effects of working memory

People’s prediction performances appear to be mediated by their working memory capacity. For example, Huettig and Janse (2016) conducted a visual world eye-tracking experiment designed to investigate prediction. They measured participants’ working memory capacities using several different measurements including an auditory nonword repetition task and a backwards digit span task. They found a positive correlation between people’s working memory capacity and their degree of predictive eye movements. The results suggest a role of working memory in making predictive eye movements.

In an ERP experiment, Otten and Van Berkum (2009) examined predictive processing in participants with high- and low working memory capacity. Participants read high-cloze sentence contexts which predicted a specific noun. These contexts were followed by a determiner which matched or mismatched in syntactic gender with the expected noun. If participants predicted the noun, encountering a prediction-inconsistent determiner was expected to cause an inhibitory effect. Working memory capacity did not affect an early ERP component which is generally regarded as indicative of a prediction effect;[3] both groups of participants showed more negative ERPs at right frontal channels for prediction-inconsistent determiners relative to prediction-consistent determiners. However, only the low working memory capacity group showed additional late negativity for inconsistent determiners. This effect therefore appears to be associated with increased demands on their working memory during predictive processing. The authors concluded that working memory capacity did not affect the predictive processing itself, but people may differ in how they deal with disconfirmed predictions depending on their working memory capacity (but see Federmeier et al., 2010, who reported that working memory did not affect processing of disconfirmed predictions).

Federmeier and Kutas (2005) examined N400 responses for words after highly constraining contexts and after mildly constraining contexts in undergraduate students (*Mean age* = 20 years) and older adults (*Mean age* = 67 years). N400 latency of older adults correlated with their reading span (cf. Daneman & Carpenter, 1980) for highly constraining

---

[3] Otten and Van Berkum did not interpret this effect as a canonical N400 effect, because the right frontal distribution of the effect they found was different from the standard centro-parietal distribution of an N400 effect. A similar effect was reported in their previous study which also manipulated a gender disagreement between adjectives and predictable nouns (Otten et al., 2007).

sentences but not for mildly constraining sentences, showing that predictability-related N400 effects were delayed for those who had lower reading span. In contrast, there was no correlation between N400 effects and the reading span in the younger group. While the results on the older group suggests that smaller working memory capacity might cause difficulty in building prediction from sentence contexts, the lack of correlation in the younger group implies that this is not necessarily generalizable in all age groups. Nonetheless, together with the findings in Otten and Van Berkum (2009), this suggests that it is possible that effects of working memory are particularly at play in a low working memory population.

Why does working memory capacity affect predictive performance? One reason could be that cognitive resources in the working memory system are used for prediction. Therefore, limited cognitive resources would cause an interference on prediction performance. Federmeier and Kutas (2005) discussed that highly constraining sentences carry rich information, which may increase cognitive load on the working memory system. Therefore, smaller working memory capacity may suffer from the load on incremental sentence integration, and consequently interfere with predictive processing. The above-mentioned studies fit with this explanation, but a causal link has not yet been shown. Experiments 1 and 2 in this thesis investigate this causal relationship.

Federmeier and Kutas also noted that the word-by-word presentation used in their ERP study might further burden the working memory system because readers cannot adjust the reading speed to whatever allows the best comprehension performance for them. If people with low working memory capacity predict less because they are less capable of keeping up with the reading speed, it is possible that ample time to process incoming information is important to use that information to generate predictions.

1.3.2 Effects of time

Several studies investigated effects of word-by-word reading rate (i.e., stimulus onset asynchrony, or SOA) on prediction. Kutas (1993) had participants read sentences that ended with the highest cloze word, an incongruent word that was semantically related to the highest cloze word, or an incongruent word that was semantically unrelated to the highest cloze word, at four different SOAs (100, 250, 700, or 1150 ms). N400s were smaller for semantically related incongruent words than for semantically unrelated incongruent words at all SOAs. More interestingly, the N400 peak latency difference was delayed for the fastest 100 ms SOA compared to the slower (250-1150 ms) SOAs, and the N400 difference effect was smaller for the 100-250 ms SOAs than for the 700 ms SOA (albeit not smaller than for the 1150 ms SOA). The results suggest that the effect of semantic relatedness to the expected words is very robust and occurs at relatively slow or fast presentation rates.

Dambacher et al. (2012) found N400 effects for low-predictable words relative to high-predictable words in three different SOAs (280 ms, 490 ms, and 700 ms). The N400 amplitude difference was the smallest in the 280 ms SOA, with no difference between 490 ms and 700 ms SOAs, and the onset latency of the N400 effect was delayed in the 280 ms SOA relative to the 700 ms SOA. However, their high-predictable words were also more plausible than the low-predictable words, so it remains unclear whether the effects of SOA on the N400 reflect the semantic processes associated with pre-activation or plausibility.

In replication of Federmeier and Kutas (1999), Wlotko and Federmeier (2015) found that N400 effects were smaller for words semantically related to a predictable word relative to unrelated words. This N400 reduction was found at a 500 ms SOA, but not at a 250 ms SOA. Along with Dambacher et al. (2012), the impact of faster presentation on the N400 effects points to the possibility that prediction-related N400 effects may suffer from uncomfortably rapid serial presentation rates.

The implications from these studies are that prediction may not occur or may be limited when people are forced to comprehend sentences relatively rapidly. Prediction of semantic information can be attenuated during such rapid reading. However, from these studies, it is not clear how time affects other aspects of prediction, such as prediction of word form. Experiments 5 – 8 in the current thesis investigate the effects of word presentation rate on prediction in an experimental setting which induced pre-activation of both semantic and word form information individually.

1.3.3 Effects of language proficiency

The effects of working memory and time on prediction can also be linked to the role of proficiency in the target language. A better working memory capacity and less time constraints on comprehension speed would allow people room for deeper processing of ongoing language. When people comprehend a language which they have not acquired a good command of, they sometimes find it hard to catch up with the speed with which sentences unfold. Such comprehension should be cognitively demanding, and less proficient language users may have reduced cognitive resources available during comprehension. In line with this, lexical access tends to be more demanding or slower in less proficient language users such as children and late L2 learners (e.g., Lew-Williams & Fernald, 2010). Thus, people may be less likely to make predictions when comprehending a less proficient language.

Consistent with this hypothesis, several studies have found a link between children’s prediction performance and their various linguistic skills, but not all the linguistic measures were found to be related to their prediction performance (Borovsky et al., 2012; Mani & Huettig, 2012, 2014). Among adults, category verbal fluency and (comprehension) vocabulary were shown to be correlated with predictive eye movements (Rommers et al., 2015), suggesting that both production and comprehension aspects of linguistic skills are related to prediction performance. Mishra et al. (2012) found that people with low literacy failed to show anticipatory eye movements unlike high literate control group (see also Huettig, Singh, & Mishra, 2011). Huettig and Brouwer (2015) found that people with dyslexia were slower than control adults to make anticipatory eye movements. These studies suggest that a higher reading skill leads to a better prediction performance. However, some of the findings should be interpreted with caution. For instance, literacy may be distinct from general proficiency for its relation to non-linguistic cognitive ability (Huettig, 2015b), and it may be related to people's socioeconomic status, which was found to influence some aspects of predictive eye movements (Troyer & Borovsky, 2015).

On the other hand, Tribushinina and Mak (2016) found that three-year-old children and university undergraduate students, who were highly likely to differ in their reading skill, were similarly quick to make predictive eye movements. However, the task was made easy for children by using only two pictures and ensuring a lag of at least three seconds between onsets of a prediction-cue adjective and a predictable noun, so it is possible that both groups of participants were performing at the ceiling level. Taken together, the conclusions from these studies are inconsistent, and it is unclear what types of linguistic skill are actually related to prediction.

Another approach to examining the effects of language proficiency on prediction is a comparison of predictive performances between L1 and L2 speakers. At the moment, it is not clear whether L2 speakers make predictions in similar ways as L1 speakers do. In fact, there are only a few published studies that have investigated predictive processing during L2 comprehension, and current evidence still lacks consistency. Martin et al. (2013) conducted an ERP study using DeLong et al.'s (2005) experimental design, and examined whether L2 English speakers whose native language was Spanish pre-activate phonological information of predictable words like L1 English speakers. Martin et al. had participants read high-cloze contexts (e.g., "*She has a nice voice and always wanted to be …*"), which supported prediction of a specific noun phrase (*a singer*). These contexts were completed with either the expected noun phrase or another unexpected but plausible noun phrase (*an artist*). Crucially, when the expected noun began with a consonant, the unexpected noun began with a vowel, or vice versa, so that the preceding indefinite article was always different for expected nouns and unexpected nouns. Based on DeLong et al.'s finding, Martin et al. expected to find an N400 effect for both unexpected articles and nouns. If the phonological information of predictable words is pre-activated, encountering an article that phonologically mismatches the predictable word should cause an effect related to prediction mismatch. Consistent with DeLong et al. (2005), Martin et al. (2013) found an N400 effect for unexpected articles in L1 English speakers, suggesting that L1 speakers pre-activated phonological information of the noun (whether the first phoneme was a consonant or a vowel). However, L2 speakers did not show any effect for unexpected articles. Notably, the critical nouns were similarly highly predictable for both groups of participants, as revealed in an offline cloze test. The offline cloze test also revealed that the L2 participants were aware of the phonological rule of the *a/an* usage. Moreover, in an additional ERP pre-test, L2 participants with a similar L2 competence showed a P600 effect when a noun was incongruent with the preceding indefinite article, suggesting their sensitivity to the violation of the phonological rule. Therefore, the lack of an N400 effect at articles in L2 speakers appears to be related to the nature of online predictive processing in L2. Martin et al. (2013) proposed two possible explanations for their findings. First, people may predict more slowly in L2 comprehension than in L1 comprehension, so the effect of prediction may not be reflected in online language comprehension. Second, L2 comprehension may only involve passive integration processing, resulting in no predictive processing in L2 comprehension. In addition to these accounts, it is also possible that the types of information that L2 speakers can predict is limited. Their results show that the first sound of predictable words was pre-activated by L1 speakers but not by L2 speakers. However, it is possible that L2 speakers still pre-activate other aspects of predictable words, such as their semantic information or syntactic properties.

Following this work, Foucart, Martin, Moreno, and Costa (2014) investigated whether a syntactic property, namely grammatical gender in Spanish, was pre-activated as a result of prediction. They hypothesised that the lack of L2 prediction effect in Martin et al. (2013) might be caused by the fact that the phonological rule for *a/an* does not exist in Spanish. Based on this hypothesis, they tested L2 Spanish speakers whose native language was French, because French shares similar syntactic rules with Spanish in terms of the grammatical gender. Participants were presented with high-cloze sentences that contained either an expected noun phrase or an alternative unexpected but plausible noun phrase. Expected and unexpected nouns always differed in gender. If participants predict the expected nouns, encountering an article that does not match the expected noun in gender would violate their prediction and result in differential ERPs. They found that both L1 Spanish speakers and L2 Spanish speakers showed N400 effects for prediction-inconsistent articles, as well as for unexpected nouns. Importantly, this effect did not differ between the language groups. The authors concluded that both L1 and L2 sentence processing involves similar predictive processing.

As they had hypothesised, Foucart et al. attributed the inconsistent findings between their study and Martin et al.'s (2013) study to a lack of the manipulated feature (*a*/*an* distinction) in the L2 speakers' native language, Spanish. They proposed that similarity between L1 and L2 (Spanish and French are similar, compared to Spanish and English) might facilitate predictive processing in L2. That is possible, but it should be noted that, apart from the cross-linguistic similarity, there are other differences between the two studies which could account for their different findings. An alternative account is that the prediction difficulty might be related to the cloze probabilities in their sentences. The cloze probability in both studies did not differ between L1 and L2 participant groups, but critically, the mean cloze probability for the predictable word was higher in Foucart et al. (81%) than in Martin et al. (65%). Thus, the predictable words were simply more predictable in Foucart et al.'s study. Another possibility is that linguistic features that can be pre-activated during L2 processing are limited. It is possible that participants pre-activate syntactic features such as grammatical gender but not phonological or orthographical features. This assumptions is compatible with production-based prediction accounts discussed in Section 1.2.2, because selection of syntactic properties precedes retrieval of phonological information in production models (cf. Levelt, 1999).

In a visual world eye-tracking study, Mitsugi and MacWhinney (2016) adopted the experimental design of Kamide, Altmann, et al. (2003; Experiment 3), who showed that L1 Japanese speakers were able to utilise a case marker to predict upcoming pre-verb referents. Mitsugi and MacWhinney presented participants with Japanese sentences with a dative structure (e.g., *weitoresu-ga kyaku-ni tanosigeni hanbaagaa-o hakobu*; order-matched English equivalent: waitress-nominative customer-dative merrily hamburger-accusative bring, “the waitress will merrily bring the hamburger to the customer”). The nominative-dative case markers supported prediction of the occurrence of the accusative object *hamburger*. Upon hearing the nominative-dative sequence, participants were more likely to fixate predictable objects (*hamburger*) before hearing the word. Mitsugi and MacWhinney replicated Kamide, Altmann, et al.’s findings with L1 Japanese speakers, but not with L2 Japanese speakers (whose native language was English), despite the fact that their L2 participants exhibited good knowledge of Japanese case markers in an offline grammar test. The authors proposed that the grammatical knowledge of L2 speakers might not be readily accessible for use in prediction during online comprehension. However, as the authors noted, the results do not preclude the possibility of L2 prediction. Their study required participants to rely on rather complex cues (combinations of semantic and syntactic information) to make predictions, and L2 speakers might not be able to use syntactic information on the fly to predict upcoming words. Moreover, their L2 participants were intermediate learners of Japanese, and prediction might be limited to highly proficient L2 speakers.

Using sentences with a simple syntactic structure, Hintz and Meyer (2015) found similar predictive eye movements between L1 Dutch speakers and L2 Dutch speakers whose native language was German. The sentences they used comprised mathematical equations, where predictions of the result numbers are based on pre-encountered numbers. Since all the numbers used in the study were cognates in Dutch and German, the task probably had little demand. The results are consistent with the view that L2 speakers can predict when the target language is not syntactically complex.

With regards to the effects of L2 proficiency on predictive eye movements, Peters, Grüter, and Borovsky (2015) compared bilinguals with a high proficiency and those with a low proficiency. Bilinguals in the high proficiency group regarded English as one of their native languages, and those in the low proficiency group did not.[4] Participants listened to sentences such as “*The pirate hides the treasure*,” while viewing images of the target object (*treasure*), an agent-related object (*ship*), an action-related object (*bone*), and an unrelated object (*cat*). Predictive eye movements to target objects occurred earlier in the high proficiency group than in the low proficiency group. Proficiency also had an effect on looks to action-related objects relative to unrelated distractors. The fixation bias towards action-related objects was greater in the low proficiency group than in the high proficiency group. This result suggests that bilinguals with a lower L2 proficiency were more likely to be distracted by less-likely but locally coherent objects. The authors noted that their inferential statistics on the predictive eye movements were not strong enough for a strong claim, so further evidence would be needed to be certain of the proficiency effects on the time course of prediction.

---

[4] The two groups differed significantly in the self-rated English proficiency scores, the age of acquisition of English and the length of exposure to English.

Inconsistent results from existing studies which specifically target prediction suggest that L2 speakers predict only under limited circumstances. However, research on L2 sentence processing more generally is certainly compatible with prediction. As discussed in Section 1.1, prediction is of course consistent with incremental sentence processing, and past studies discussed evidence for prediction in relation to incremental sentence processing (Altmann & Kamide, 1999; Altmann & Mirković, 2009). For instance, Altmann and Mirković (2009) argued that predictive eye movements reflect incremental mapping of contextual information onto a concurrently presented visual scene. Notably, research suggests that L2 sentence processing is incremental (Jackson & Dussias, 2009; Trenkic, Mirković, & Altmann, 2014; Williams, 2006). For example, Trenkic et al. (2014) examined whether L2 English speakers whose native language was Mandarin comprehend English articles incrementally and utilise the information conveyed by the articles in online sentence processing. Participants listened to sentences referring to a goal object of the verb *put* either with *the* or *a* (e.g., "*The pirate will put the cube inside the/a can*."), while examining a visual scene containing one or two compatible goal(s) (e.g., one open can and one unopened can/ two open cans). They were instructed to indicate the goal where a referent was moved to. The definite noun phrase *the can* signals that the noun is uniquely identifiable (i.e., one compatible goal), while the indefinite counterpart *a can* implies that there are two or more cans that could be a potential goal. In an earlier study, Chambers, Tanenhaus, Eberhard, Filip, and Carlson (2002) found that a goal object was identified faster when the linguistic expressions pragmatically matched the visual scene. Trenkic et al. (2014) replicated their results in L2 English speakers. Although Mandarin does not have equivalent of English articles *a* or *the*, the Mandarin-native L2 English speakers showed online sensitivity to the morpho-syntactic structure, providing converging evidence that L2 sentence processing is also incremental. Therefore, it is likely that L2 speakers access available cues for prediction during online sentence processing, which makes it more likely that they can also make predictions about upcoming words. However, consistent with Mitsugi and MacWhinney

(2016), Grüter et al. (2012) and Lew-Williams and Fernald (2010) failed to find online utilisation of L2-unique Spanish grammatical gender in English-native L2 Spanish speakers. The findings seem to be against the notion that L2 sentence processing is incremental, but notably it was shown that online utilisation of grammatical gender, which these researchers manipuated, was mediated by participants' L2 proficiency and partly by the presence of the linguistic feature in their L1 (Dussias, Valdés Kroff, Guzzardo Tamargo, & Gerfen, 2013). Overall, the results about L2 incremental processing suggest that L2 speakers may have a more limited capacity for incremental processing than L1 speakers. Such limitations could mean that their ability to make predictions is also limited.

In sum, the inconsistent findings on prediction in L2 speakers suggest that predictive processing during L2 processing might be limited relative to that during L1 processing. However, how prediction in L2 is limited is unclear. For example, do L2 speakers predict limited types of information, or do they predict under limited circumstances? For the former question, Experiments 7 and 8 in this thesis investigated whether L2 speakers predict semantic information and phonological/orthographic word form information. For the latter question, Experiment 2 investigated effects of cognitive load and Experiments 7 and 8 investigated effects of word presentation rate on predictive processing during L2 processing.

1.4 Methodologies

1.4.1 Cloze probability as a measure of predictability

A word's cloze probability refers to the proportion of participants who provided that word to complete a given sentence context (cf. Bloom & Fischler, 1980). This measurement is commonly used to assess how predictable a word is in a particular context. A cloze probability test is usually an offline language production task, carried out as a norming test. Participants usually read incomplete sentences and provide a word or words to complete the sentences under no time pressure (but see Staub, Grant, Astheimer, & Cohen, 2015, discussed below).

The cloze probability has been found to be a reliable predictor of reading times in eye-tracking (Frisson et al., 2005; Rayner et al., 2011) and self-paced reading (Roland et al., 2012; Smith & Levy, 2013) studies, as well as of N400 effects in ERP studies (Kutas & Hillyard, 1984; Thornhill & Van Petten, 2012). Despite its significant informativeness, there appears to be no clear definition of what cloze probability actually represents. One potential reason is that the instructions used in a cloze probability test vary across studies, and these instructions could affect how people complete sentences. Some researchers instruct participants to simply complete each sentence fragment (Otten, Nieuwland, & Van Berkum, 2007) or complete it with one word (Altarriba, Kroll, Sholl, & Rayner, 1996). Others instruct participants to fill in the first word or noun that comes to their minds (Frisson et al., 2005; Martin et al., 2013; Rommers, Meyer, Praamstra, & Huettig, 2013; Thornhill & Van Petten, 2012), the word that they would expect to come next (Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Rayner et al., 2011), the word most likely to appear next (Ashby, Rayner, & Clifton, 2005; Dambacher et al., 2012; Staub, 2011), a meaningful short continuation (Stafura & Perfetti, 2014), or the best continuation (DeLong et al., 2005; Kleinman, Runnqvist, & Ferreira, 2015) for each given context. These instructions seem to differ in the extent how much they encourage participants to give a predictable word. Therefore, participants might use different strategies depending on the instructions given, or simply due to their individual preference. For example, some people may imagine themselves producing the sentence and provide a word that they would continue the sentence with. Others may imagine another individual producing the sentence to provide a word that they expect to hear or read after the sentence fragment. A clear difference in the two example strategies is that one employs the language production system, whereas the other the language comprehension system. It is not clear from existing literature whether people use such strategies during the cloze task, and how the use of strategy may affect the cloze probability outcome.

Questioning what cloze probability represents, Staub et al. (2015) conducted a timed cloze test, in which participants were asked to read sentences presented in Rapid Serial Visual Presentation (RSVP) format and say out loud the next word that they thought should be in each sentence, while their response times were measured. They found that participants' responses correlated highly with offline cloze responses (r = .91) for items with high cloze probabilities in the offline cloze test. This high correlation warrants the reliability of the offline cloze test, or at least precludes potential effects associated with the lack of constraint on response time, which is conventional in paper-and-pencil or Internet-based cloze tests. Additionally, they found that higher cloze responses were given more quickly. Based on their findings, the authors proposed a model of what happens when people perform a cloze task, in which all possible responses for a given sentence fragment compete for activation. They thus cast doubt on cloze probability as predictability, and instead claimed that cloze probability may be a measure of relative activation level of the word. This interpretation posits that high-cloze words are easier to process than low-cloze words because high cloze words are more strongly activated at the time they are processed. The critical argument is that even if the highest cloze word is the most active, this does not necessarily mean that comprehenders predict that word. On the other hand, some psycholinguists regard prediction as equal to pre-activation (e.g., Kutas et al., 2011). There appears no clear way to distinguish the two in empirical studies. In this thesis, I regard pre-activation as a part of processes involved in prediction, wherein predicting a certain information would spread activation to related information, and I define pre-activation as the activation of predictable and related information prior to the appearance of the predictable information.

Chou, Huang, Lee, and Lee (2014) demonstrated another implication of the relationship between cloze probability and prediction-related effects by examining effects of cloze probability and semantic constraints individually. They utilised a fixed phrase "*numeral* + *classifier* + *noun*" in Chinese (e.g., 一[one] + 頂 [*classifier*] + 帽子[hat]), and manipulated the constraint strength of classifiers and the cloze probability of the matched noun. Chinese classifiers and nouns have some systematic relationships based largely on semantic properties of the noun, so nouns can be predictable when classifiers are semantically constraining (i.e., when not many nouns can plausibly occur after the classifier). Strongly constraining classifiers had fewer nouns that people could think of as an appropriately matched noun, compared to weakly constraining classifiers, which could be followed by a large number of nouns. The strength of constraint was fully crossed with cloze probability[5] of the following critical words. Critical words in the high cloze condition had a cloze probability of at least 50%, those in the low cloze condition had a cloze probability of 6.9% or less, and those in the implausible condition had zero cloze probability (the preceding classifier was inappropriate). ERPs were measured while participants saw a classifier, and then a critical word. After a strongly constraining classifier, high-cloze words elicited smaller N400s than implausible words, but low-cloze words did not differ from implausible words. In contrast, after a weakly constraining classifier, N400s were smallest (least negative) for high-cloze words, largest for implausible words, and intermediate for low-cloze words. These results suggest that predictability-related ERP effects do not solely depend on cloze probability, but instead can be mediated conjointly by cloze probability and semantic constraints. However, it is not clear whether similar results can be found when participants read complete sentences rather than classifier-noun pairs. If participants made predictions about the forthcoming noun in this experiment, the classifier was the only cue for prediction, whereas more factors, such as semantic associations with contextual information, would come into play during sentence comprehension.

---

[5] In their cloze test, participants completed "*numeral* + *classifier* + _____" with the first noun that came to mind.

In sum, reliable effects of cloze probability on sentence processing have been shown in both eye-tracking and ERP studies (Frisson et al., 2005; Kutas & Hillyard, 1984; Rayner et al., 2011; Thornhill & Van Petten, 2012). However, it appears that little consideration is given to how different instructions in a cloze test could affect the responses, and there seems to be no agreement on what cloze probability represents. Effects of cloze probability could be interpreted differently depending on what instructions were used to collect the cloze probability data. In this thesis, I regard cloze probability of a word as a proxy for predictability of that word, following its conventional usage (DeLong et al., 2005; Federmeier & Kutas, 1999).

1.4.2 The visual world eye-tracking paradigm

The visual world eye-tracking paradigm is often used in the investigation of predictive processing and incremental processing during language comprehension along with others. In a standard set-up, participants hear an utterance while viewing a semi-realistic scene containing several objects. This paradigm makes use of the sensitivity of people's eye movements to the information in the spoken language they are listening to. Cooper (1974) demonstrated that people tended to fixate pictures that were referred to in concurrently presented speech, even if they were told that they were free to look anywhere. Furthermore, participants were more likely to fixate pictures that were semantically related to the spoken words compared to semantically unrelated pictures. For example, participants tended to fixate pictures of *lion*, *zebra* and *snake* upon hearing the word *Africa* (for similar effects, see Huettig & Altmann, 2005; and for effects when words were not embedded in a sentence, see Yee & Sedivy, 2006). Allopenna, Magnuson, and Tanenhaus (1998) showed that participants also tended to fixate pictures whose name shared the initial or final phonemes with a spoken target word relative to objects with no phonological overlap with the target word. Additionally, people's eye movements are also sensitive to non-linguistic information of depicted objects that are shared with the referents of spoken words, such as shape (Dahan &

Tanenhaus, 2005; Huettig & Hartsuiker, 2008) or colour (Huettig & Altmann, 2011). The results from these studies suggest that participants activated related concepts or linguistic information associated with spoken words, and they mapped the activated information (e.g., meaning or sound) onto the visual scene. In this regard, the visual world paradigm excels in examining relative activation strength of certain types of information.

Effects of contexts on visual world competitor effects

Effects of information overlap between mentioned target objects and other depicted objects are broadly called '*competitor effects,*' as the referred object and an object that is related to that object are thought to compete for activation due to the shared information. This competitor effect was found to be dependent on contextual constraints. Dahan and Tanenhaus (2004) found that when a target word was presented in a neutral context, its phonological onset competitor and semantic competitor objects were more likely to be fixated than distractor objects that had no phoneme-onset or semantic overlap with the target word. However, when the competitors were no longer an appropriate thematic fit to the preceding context (i.e., the mention of these competitors instead of the target word would make the sentence implausible), they were not fixated more than the distractors. The decreased activation of phonologically or semantically related words after semantically constraining contexts indicates that listeners incrementally process the contextual information and utilise it to build constraints about upcoming referents (but see Huettig & Altmann, 2007, for investigation on visual-shape competitor effects).

Types of visual stimuli

The visual scenes used in this paradigm are not limited to simple line drawings, pictures or real objects. Andersson, Ferreira, and Henderson (2011) used photographic scenes and narratives referring to four objects in succession in the scene. Despite the complexity of the scenes, participants tended to track all the target objects. The application of this paradigm can also be extended to printed words. McQueen and Viebahn (2007) depicted words instead of pictures to examine spoken word recognition. They replicated Allopenna et al. (1998) and showed that printed words that phonologically overlapped with target words attracted more fixations than unrelated words. Using words is useful especially if critical words are too abstract to depict, or if the investigation is on orthographic processing (Salverda & Tanenhaus, 2010). Although the logic is similar between the two versions, it appears that effects related to phonological or orthographic manipulations are more robust in the printed word version than in the picture version. Huettig and McQueen (2007) found that a phonological competitor effect was stronger when the visual scene depicted the names of critical words than when it depicted their pictures. In contrast, pictures that shared semantic features or shape features with target words attracted more fixations than unrelated pictures, but neither the semantic competitor effect nor the shape competitor effect was replicated when the pictures were replaced with their names. The results suggest that the degree of competitor effects seems to differ depending on the type of visual scene.

Rapidity of language-mediated eye movements

The early studies by Cooper (1974) and Allopenna et al. (1998) showed that the observed fixation bias towards pictures that were related to spoken language was closely time-locked to the on-going speech. The rapidity of the language mediated eye movements is also argued by Altmann (2011), who has shown that spoken language can mediate people's eye movements within 100 ms (for a slightly longer estimate of 200 ms, see Saslow, 1967), critically, when they were under no instructions to move their eyes to a particular object as quickly as possible. However, eye movements at an early stage were found to be mediated by word frequency (Dahan, Magnuson, & Tanenhaus, 2001), so the rapidity of saccadic movements could depend on lexical characteristics.

Limitation of the visual-world paradigm

Despite its temporal advantage, there are limitations in the method that are important to be noted. First, effects of interest can depend on the mode of presentation or visual information. Clear examples of such a caveat are that eye movements to competitor objects that are related to mentioned words can be affected by the type of visual input (printed words or drawings; Huettig & McQueen, 2007), or the depicted colour of objects (Huettig & Altmann, 2011). In addition, participants would retrieve names of pictures in a visual scene when the pictures are shown, which leads to activation of the words corresponding to the pictures in their mental lexicon. This is particularly worth noting for research on prediction, as it should be assumed that picture names become already active (relative to words that are not related to the scene) regardless of the spoken stimuli. Although the effects of the visual stimuli could be minimised by reducing the preview time (cf. Rommers et al., 2013), it is impossible to completely get rid of the effect.

Summary of the visual world paradigm

The current section explained the use of the visual world paradigm in studies on language comprehension. Language-mediated eye movements tend to occur without any task that requires meta-linguistic judgements or visual search, and the eye movements reflect what information is activated in a tightly time-locked manner (for a review, see Huettig, Rommers, & Meyer, 2011). Owing to this nature, this paradigm is widely used to investigate activation of linguistic features such as meaning or sound (Huettig & McQueen, 2007; Yee & Sedivy, 2006), word recognition in L1 or L2 (Allopenna et al., 1998; Magnuson, Dixon, Tanenhaus, & Aslin, 2007; Spivey & Marian, 1999), or processing of non-linguistic features such as colour or shape that are related to spoken language (Dahan & Tanenhaus, 2005; Huettig & Altmann, 2007, 2011; Huettig & McQueen, 2007). In this thesis, I used the visual world paradigm to investigate predictive processing during language comprehension (cf. Altmann & Kamide, 1999; Rommers et al., 2013). The paradigm is suited for investigation on prediction, because the timing of predictability effects is crucial in studies investigating prediction. As discussed in Section 1.2.1, effects emerging after the occurrence of a predictable target may be explained by non-predictive accounts (e.g., integration accounts).

1.4.3 Event-Related Potentials (ERPs)

This section briefly introduces the ERP technique used in Experiments 5 – 8, and outlines the characterisation of two prediction-related ERP components focused on in this study, namely N400 and Late Positive Complex (LPC/ P600). The ERP technique involves a measurement of electrical activity of the human brain called electroencephalogram (EEG) using electrodes placed on the scalp. The amplified EEG is filtered to reject trials containing artefacts (e.g., eye blinks) that may contaminate the EEG, and then it is averaged in order to remove noise in EEG and extract responses to specific events, called event-related potentials (ERPs).

ERPs are well-suited to investigate predictive processing mainly for two reasons. First, unlike behavioural measures, ERPs can be collected without participants performing any task, so the collected data are not influenced by a meta-linguistic task which people usually do not do during natural comprehension. Second, due to their high temporal resolution, it is possible to determine which stages of processing are affected by an experimental manipulation (Luck, 2005).

N400 as an index of predictability and overlap with pre-activated information

An N400 is an ERP component, a negativity with a peak at around 400 ms post-stimulus. Kutas and Hillyard (1980) discovered that semantically incongruent words elicit larger N400s than semantically congruent words in the concurrent context. This ERP difference is termed an N400 effect. The distribution of the N400 effect tends to be widespread, but it tends to be the most prominent at posterior channels (Kutas, Van Petten, & Besson, 1988). N400 is regarded to be broadly sensitive to semantic aspects of stimuli, and it is supposed to be a reliable measure of how easy it is for a word to be integrated with the mental representations built from the on-going sentence (Holcomb, 1993; Kutas & Federmeier, 2011; Van Berkum, Hagoort, & Brown, 1999). In the context of research on prediction, it has been found that (1) N400 amplitude is systematically related to word predictability, and (2) the semantic anomaly N400 effect is reduced for words that share some features with predictable words. The design of the ERP study in this thesis utilises the latter characteristics. To make the distinction clear, I explain both characteristics below briefly.

First, N400 amplitude was shown to inversely correlate with cloze probability. Kutas and Hillyard (1984) had participants read sentences which ended with a high-cloze, a medium-cloze or a low-cloze word, and found that the lower the cloze probability, the higher was the N400 amplitude (more negative). Since their discovery, this graded predictability-related N400 effect has been replicated in a number of studies (Dambacher, Kliegl, Hofmann, & Jacobs, 2006; DeLong et al., 2005; Wlotko & Federmeier, 2013). Making use of the N400 amplitude being larger for unpredictable words than for predictable words, several studies have shown that N400s are larger for words that are plausible but inconsistent with predictable words compared to prediction-consistent words (DeLong et al., 2005; Otten et al., 2007; Wicha et al., 2004).

Second, N400 effects were found to be reduced for implausible words that were related to predictable words compared to control words that were implausible and unrelated to predictable words (e.g., Federmeier & Kutas, 1999; Kim & Lai, 2012; Laszlo & Federmeier, 2009; Rommers et al., 2013). These reduced N400 effects are usually interpreted as a facilitation effect caused by a feature match between information pre-activated in the semantic memory and the actual input. Therefore, this effect is regarded as evidence for pre-activation of the information associated with predictable words.

It is important to recall here that N400 is also susceptible to various semantic aspects, such as plausibility, because more plausible words tend to have a smaller integration cost (Matsuki, Chow, & Hare, 2011; Traxler & Pickering, 1996). Owing to a likely linear relationship between cloze probability and plausibility (i.e., high-cloze words tend to be more plausible words; e.g., Dambacher et al., 2012), it can be difficult to dissociate N400 effects from predictability and those from integration ease, unless plausibility and other semantic aspects are controlled or critical investigation is on the words preceding predictable words (Amsel et al., 2015; DeLong et al., 2005; Otten et al., 2007; Wicha et al., 2004).

LPC as an index of prediction inconsistency

A Late Positive Complex (LPC) or a P600 is a positivity peaking around 600 ms post-stimulus, and has posterior scalp distribution. LPC effects have been reported for various types of syntactic violations or syntactic ambiguity (Frisch, Schlesewsky, Saddy, & Alpermann, 2002; Hagoort & Brown, 2000; Osterhout & Holcomb, 1992), and are supposed to reflect syntactic reanalysis (Friederici, Hahne, & Mecklinger, 1996) or syntactic integration difficulty (Kaan, Harris, Gibson, & Holcomb, 2000). However, it has been found that semantic anomaly could also elicit an LPC effect, so the LPC effect is now more largely associated with detection of conflict, which could also involve reanalysis processing or disconfirmation of an encountered input (Frenzel, Schlesewsky, & Bornkessel-Schlesewsky, 2011; Kuperberg, 2007; Van Herten, Chwilla, & Kolk, 2006; Vissers et al., 2006). In studies using high-cloze sentences, LPC effects were found in response to stimuli that were phonologically and/or orthographically similar to the high-cloze word relative to dissimilar control stimuli (Kim & Lai, 2012; Laszlo & Federmeier, 2009; Newman & Connolly, 2004; Vissers et al., 2006). It appears that the late posterior positivity reflects a neural response to semantically implausible words or nonwords that are related to a highly predictable word (for a review, see Van Petten & Luka, 2012),[6] or more in general, when a highly strong

---

[6] This late posterior positivity is distinguished from a late frontal positivity, which is thought to be elicited when the prediction-inconsistent word is a plausible continuation of the sentence. Prediction-inconsistent plausible words tend to elicit larger frontal positivity relative to highly predictable words (DeLong, Quante, et al., 2014; Thornhill & Van Petten, 2012; Van Petten & Luka, 2012). The distinct scalp topographies suggest that different neural mechanisms may underlie these effects. They could reflect different approaches in dealing with a prediction-inconsistent input. If the encountered input is implausible, comprehenders may disregard this input and maintain their prediction as an intended input. If

prediction is violated (Xiang & Kuperberg, 2015). In sum, a parsimonious account for the LPC effect would be that it indicates a conflict detection between the encountered input and the highly predictable information (which may or may not involve reanalysis).

Implications of N400 and LPC effects on predictive processing

While both N400 and LPC effects can be elicited in response to prediction-inconsistent stimuli, their implications differ in terms of the underlying processes that trigger these effects. A critical difference lies in whether the process involves pre-activation of relevant information. N400 reductions are supposed to be related to a degree of overlap between the encountered information and pre-activated information. In contrast, LPC effects show a difference between the encountered information and information of predictable (highly supported by the sentence context but not necessarily *predicted*) words. Therefore, the presence of an N400 reduction assumes that some information of predictable words is pre-activated (if other factors such as plausibility are controlled), whereas an LPC effect does not. On this ground, I treat LPC effects in relation to high predictability of a specific word, but do not regard the effects as evidence for prediction per se.

1.5 Summary and the current thesis

The review of current research on language prediction provides rich evidence that people predict upcoming language during comprehension. I reviewed studies that showed pre-activation of semantic information and word form information, theories on the mechanisms of prediction, and factors that mediate people's prediction performances. At the same time, I pointed out several unanswered questions from past studies. The current thesis aims to fill these gaps of the existing findings, focusing on three factors that arguably play an important role in predictive processing: working memory, time and language proficiency. A

---

the encountered input is plausible, comprehenders may revise their prediction and accept the unexpected input as a correct input.

larger working memory capacity was found to be associated with greater predictive processing, but this finding is based on correlational studies, and the causal link between the two has not yet been shown. Experiments 1 and 2 examined this in L1 and L2 speakers by testing effects of a cognitive load on their predictive eye movements using the visual world paradigm. Experiments 3 and 4 investigated whether L1 and L2 speakers predict phonological information, and compared the time-course of predictions between L1 and L2 speakers. Experiments 5 – 8 investigated how prediction of semantic and word form information is mediated by word presentation rate, and compared these predictions in L1 and L2 speakers.

2 Study 1. Effects of cognitive load on prediction in L1 and in L2

This chapter is based on a manuscript submitted for publication as Ito, A., Corley, M. & Pickering, M. J. A cognitive load delays predictive eye movements similarly during L1 and L2 comprehension.

2.1 Introduction

During comprehension, people construct representations that help them predict what may be mentioned next (Kuperberg & Jaeger, 2016). Are these predictions resource-intensive? To explore how predictions are affected by factors that can reduce available cognitive resources, we used the visual world eye-tracking paradigm to compare predictions between people who performed a concurrent working memory task and those who did not. In addition, we considered predictions in L1 speakers and L2 speakers, as they may differ in their resources and therefore in the extent to which they may be affected by the working memory load manipulation.

People make predictions rapidly (Altmann & Kamide, 1999; Kukona et al., 2011; Van Berkum et al., 2005) and even toddlers can predict (Borovsky et al., 2012; Mani & Huettig, 2012). One possibility for why predictions are so efficient is that they are made automatically. For example, Huettig (2015a) proposed a simple lexical association route to prediction where lexical priming pre-activates related information. Consider priming from a verb to its typical agents and patients (which occurs in isolation; Ferretti, McRae, & Hatherell, 2001). Kukona et al. (2011) had participants hear sentences that predicted a verb’s patient (e.g., “*Toby arrests the crook.*”). They found that verbs (e.g., *arrests*) can lead to predictive eye movements to both their typical agents (e.g., *policeman*) and patients (e.g., *crook*). This finding suggests that predictions can be (at least partly) driven by semantic association between verbs and nouns. Thus, predictive eye movements that are mediated by lexical priming may occur automatically.

However, the automaticity in predictive processing may depend on the comprehender’s linguistic competence. Some eye-tracking and ERP studies provide evidence for prediction in L1 but not in L2 speakers (Martin et al., 2013; Peters et al., 2015), whereas other studies have found similar predictive performances in L1 and L2 (Foucart et al., 2014; Hintz & Meyer, 2015). It is unclear what factors led to these differences.

The current study investigated predictive processing in L1 and L2 comprehension. Following the design of Altmann and Kamide (1999), we investigated verb-related prediction in L1 (Experiment 1) and L2 (Experiment 2). To the extent that prediction depends on linguistic competence, we expected stronger evidence for prediction in L1 than in L2. Each experiment additionally included a cognitive load manipulation. Since L2 processing tends to be more cognitively demanding than L1 processing (Segalowitz & Hulstijn, 2009), we hypothesised that L2 speakers might be particularly subject to any effects of cognitive load. We now review evidence for predictive processing during L1 and L2 comprehension, and then discuss effects of working memory capacity on predictive processing before introducing our study.

2.1.1 Prediction during L1 and L2 comprehension

Visual world eye-tracking experiments studying prediction have found that participants make use of contextual information to direct their eyes to objects that are likely to be mentioned (Altmann & Kamide, 1999; Chambers, Tanenhaus, & Magnuson, 2004; Kamide, Scheepers, et al., 2003; Knoeferle et al., 2005; Kukona et al., 2011). For instance, Altmann and Kamide (1999) presented L1 English speakers with sentences such as “*The boy will eat the cake*” together with a scene depicting a cake and some inedible objects. The participants were more likely to fixate the cake before it was mentioned, compared to when they heard “*The boy will move the cake*.” The predictive eye movements suggest that people process sentences on a word-by-word basis, and integrate information extracted from each word to build predictions about upcoming words.

Chambers and Cooke (2009) found predictive eye movements in L2 speakers in a similar experiment to Altmann and Kamide (1999). Late English-French bilinguals who had relatively high French proficiency listened to sentences in French, such as “*Marie va nourrir la poule*” (*Marie will feed the chicken*) or “*Marie va décrire la poule*” (*Marie will describe the chicken*), while viewing a scene where all the depicted objects could plausibly be described but only the chicken could plausibly be fed. Participants were more likely to look at the chicken when they heard the verb *nourrir* (*feed*) relative to when they heard *décrire* (*describe*) (and before hearing *poule*, chicken). Together with Altmann and Kamide (1999), the findings suggest that both L1 and L2 speakers are able to predict upcoming words based on the meaning of the preceding verb.

However, there is evidence that predictions can differ for L1 and L2 speakers. Kamide, Altmann, et al. (2003; Experiment 3) showed that L1 Japanese speakers were able to utilise a case marker to predict an upcoming word. Their participants heard sentences with a dative structure (e.g., “*weitoresu-ga kyaku-ni tanosigeni hanbaagaa-o hakobu*”; order-matched English equivalent: waitress-nominative customer-dative merrily hamburger-accusative bring, meaning the waitress will merrily bring the hamburger to the customer). The nominative-dative case markers supported prediction of the occurrence of the direct object *hamburger*. Upon hearing the nominative-dative sequence, participants were more likely to fixate predictable objects (a hamburger) before hearing the word. Mitsugi and MacWhinney (2016) replicated their findings with L1 Japanese speakers, but not with L2 Japanese speakers, despite the fact that their L2 participants exhibited good knowledge of Japanese case markers in an offline grammar test. The authors proposed that L2 speakers’ grammatical knowledge might not be readily accessible for use in prediction during online comprehension, as their stimuli required participants to rely on rather complex cues (combinations of semantic and syntactic information) to make predictions. These findings might indicate that L2 speakers predict less well than L1 speakers, and that they do not predict when predictions are to be made via relatively complex linguistic computation. (Note that the L2 participants were intermediate learners, and prediction might be limited to high proficiency L2 speakers.)

Taken together, it appears that L2 speakers do not always make predictions like L1 speakers. However, it is possible that predictions in L1 and L2 are similar when predictive processing does not involve complex linguistic computations. Considering this possibility, our study used sentences with a simple syntactic structure (cf. Altmann & Kamide, 1999).

2.1.2 Effects of cognitive load on predictive processing

Predictions can be made through integration of information encountered in the on-going sentence, information from the (visual world) environment, and information in the comprehender’s memory (Slevc & Novick, 2013). Given that this integrative mechanism requires a memory retrieval process, these predictions are likely to be affected by working memory load. Consistent with this hypothesis, Huettig and Janse (2016) found a positive correlation between people’s working memory capacity and their predictive eye movements in the visual world paradigm. People with greater working memory capacities made stronger predictions, using grammatical gender information conveyed by Dutch articles. Huettig and Janse’s findings suggest that some of the cognitive resources that are used for making predictive eye movements are also used for performing a working memory task. But to be confident of this, it is of course necessary to show a causal relationship – that a high cognitive load interferes with predictive eye movements.

Such an interference effect of cognitive load on prediction may be particularly strong during L2 processing. L2 speakers sometimes fail to use complex cues for prediction (e.g., Mitsugi & MacWhinney, 2015), perhaps because cognitive resources are likely to be reduced during L2 processing relative to L1 processing (Segalowitz & Hulstijn, 2009).

Therefore, if cognitive resources that are used for L2 comprehension is shared by working memory resources that are used for remembering words, an additional cognitive load during L2 comprehension may increase the effects of cognitive load.

However, predictions that do not rely on complex linguistic computations may occur automatically. If so, predictions may be intact under a cognitive load. If a working memory load does not impact an aspect of processing, this suggests that the processes underlying the working memory task and making predictive eye movements do not share resources (e.g., Waters, Caplan, & Yampolsky, 2003). On the other hand, if making predictive eye movements is not automatic, this suggests that resources for the working memory task are also used for making predictive eye movements. Making use of this logic, we tested the effects of a working memory task of remembering words (cf. Gordon et al., 2002) on eye movements while people listened to predictive and non-predictive utterances.

2.1.3 The current study

We investigated whether predictions that do not involve complex syntactic computations are affected by a cognitive load. The experiments manipulated whether the verb was semantically restrictive or not. We recruited L1 English speakers (Experiment 1) and advanced L2 speakers of English (Experiment 2). In both experiments, we recorded participants' eye movements as they listened to sentences containing a predictive verb that was compatible with an upcoming patient that referred to one of the depicted objects or a non-predictive verb that was compatible with all the depicted objects. It therefore followed the design of Altmann and Kamide (1999), but (1) we ensured that critical objects had relatively high frequency and low Age of Acquisition (AoA) (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012), so that none of the words would be new to L2 speakers, and (2) we added an additional picture that was semantically related to the target in order to examine whether semantic features of predictable words are pre-activated (cf. Yee & Sedivy, 2006). We were interested in whether semantic competitor effects also occurred under our experimental conditions, with participants directing some looks to an object that was semantically related to the target object.

If participants predict upcoming words, we expected that they would be more likely to fixate a predictable than an unpredictable object before it was mentioned. Half the participants performed the same task under cognitive load (they had to remember a list of words). If predictions hinge on available cognitive resources (i.e., are not automatic), predictive eye movements might be delayed or eliminated under these conditions. If a cognitive load has a greater effect on L2 speakers' predictive performances, L2 speakers may be less likely to predict than L1 speakers. If semantic competitor effects occur under predictive sentences, we might also find that semantic competitor effects would be reduced under cognitive load. This would in turn mean that pre-activating semantic information does not happen completely automatically.

2.2 Experiment 1

2.3 Method

2.3.1 Participants

Forty-eight native English speakers participated in the experiment. All the participants had normal vision, and none reported any language disorder.

2.3.2 Stimuli

The auditory stimuli comprised 16 sentence pairs, each of which had two conditions, differing only at the critical verb (see Appendix 8.1 for the full set of items). In the *predictable* condition, the target object was the only appropriate patient of the verb among four depicted objects (e.g., "*The lady will fold the scarf.*"). In the *unpredictable* condition, any of the depicted objects could plausibly be the patient of the verb (e.g., "*The lady will find* *the scarf.*”). The sentences were recorded by a female native British English speaker, and sampled at 48 kHz with a format of 32-bit float. The speaker read the sentences at a rate of 1.3 syllables per second with some space between phrases (following Altmann & Kamide, 1999).

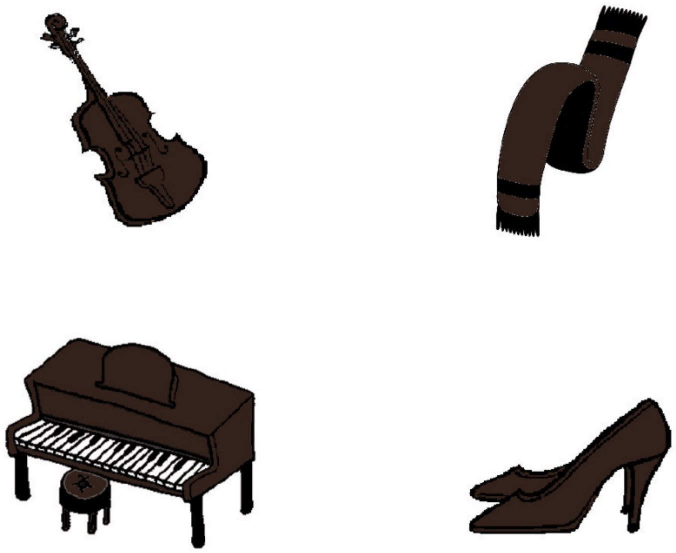


Figure 1. An example picture (for the sentences “*The lady will fold/ find the scarf.*”).

The visual stimuli were 16 pictures, each with four objects, one depicted in each quadrant. All the objects were presented in one of seven colours (grey, pink, purple, yellow, green, blue, and brown) in either dark or light shade.[7] Each of the target objects was matched with a semantic competitor that was in the same semantic category (according to Van Overschelde, Rawson, & Dunlosky, 2004). The two semantically related distractors in each item set were also in the same category as each other but from a different category to the target and semantic competitor, to prevent participants determining that the target object would be one that had a category coordinate in the array. For the sentences “*The lady will*

---

[7] In another experiment that is not reported here, we used the same visual stimuli with colour terms in order to increase the time available to make predictions. Therefore, the pictures are always coloured. No colour terms were used in the current experiments, but they were used in the predictability pre-test.

find/fold the scarf', the scene had the target object *scarf*, a semantic competitor object *high-heels* and two distractors, *violin* and *piano* (Figure 1). The names of target, semantic competitor, and distractor objects did not differ in CELEX frequency (Baayen, Piepenbrock, & Gulikers, 1995), $F(2, 56) = .36$, $p = .70$, or AoA, $F(2, 54) = 1.7$, $p = .19$. The mean frequency (per 17.9 million) and AoA for object names were 1094 (*SD* = 1633) and 5.0 years (*SD* = 1.4), respectively. The positions of all objects were counterbalanced across items.

We tested the experimental items for predictability. Twenty-one native English speakers were presented with the coloured pictures and sentences without target words, but with two-word shade and colour modifiers (e.g., "*The lady will fold/find the dark brown* _____."), and told to give the name of one of the depicted objects to complete the sentence. After excluding unclear answers (0.9%), which could refer to more than one object, participants selected the target 92% of the time in the predictable condition and 26% of the time in the unpredictable condition (with the other responses split among the three distractors). Thus, target objects were generally considered to be the most plausible continuation in the predictable condition, and no more (or less) plausible than other three objects in the unpredictable condition.

The experimental items additionally included 16 fillers. The filler sentences were similar in length and syntax to the critical sentences. Accompanying pictures depicted four objects in one of the colours and shades used in critical items, and between one and four of the objects could serve as a plausible patient of the verb. The pictures also comprised two pairs of semantically related objects.

For the working memory task, 160 mid-frequency words were selected from low-concreteness (concreteness < 3, on the scale of 1-5) words in the corpus of Brysbaert, Warriner, and Kuperman (2014). All the words had the maximum of three syllables. Each picture-sentence pair was matched with a set of 5 words. The words were unrelated to the picture or sentence. They were semantically unrelated to one another and did not share onset or offset syllables.

2.3.3 Procedure

We created two experimental conditions (no-load, and load), and participants were randomly assigned to one of the conditions. Participants were seated in front of a computer screen and tested individually in a quiet room. The participants were instructed that they would hear a sentence and see a picture at the same time, and were asked to click a mentioned object. The presentation order was randomised, and every participant saw items in a different order. No participant saw more than two items in the same condition successively. Eye movements were recorded using Eyelink 1000 tower mount eye tracker sampling at 500 Hz. Participants placed their chin on a chin rest, and the eye-tracker was calibrated using a nine-point calibration grid. The pictures were presented on a computer monitor at a resolution of 1024×768 pixels. Before every trial, drift correction was performed, followed by a 500 ms blank screen. Pictures were presented for 1000 ms before the sentence onset in order to give participants a preview (cf. Huettig, Rommers, & Meyer, 2011). The pictures disappeared when the participants clicked an object.

Participants first clicked the mouse when they were ready. Participants in the load condition then saw five words (presented together) on the screen for eight seconds. All participants then saw a 500 ms blank screen followed by the pictures. After 1000 ms, they heard the sentence. Participants then clicked on the picture that they judged to correspond to the final word. Participants in the load condition then attempted to list the words in any order within eight seconds.

No feedback was given during the experiment. The position of the mouse pointer was corrected to the centre of the screen after every trial. The experiment started with two practice trials, and lasted for about 15-25 minutes.

2.4 Results

2.4.1 Behavioural task accuracy

The mouse-clicking responses were not recorded for participants who were assigned to the no-load condition. The accuracy for the target clicking task in the load condition was 100%. The mean percentage of correctly recalled words for the working memory task was 73% (*SD* = 20%; range = 48-91%).

2.4.2 Eye-tracking data analyses

The eye-tracking data were analysed using linear mixed models with the lme4 package (D. M. Bates, Maechler, & Dai, 2008) in R (R Development Core Team, 2015). The proportion of time spent fixating on target and semantic competitor objects was calculated separately for each 50 ms bin relative to the target noun onset (following Altmann & Kamide, 1999). We constructed two linear mixed-effects models, which evaluated the fixation probability on target objects and on semantic competitor objects as predicted by Predictability (predictable vs. unpredictable), Load (no-load vs. load), and the interaction of Predictability by Load. The model included random intercepts and slopes for Predictability by participants and by items and for Load by items (Barr, 2008). The variable Predictability was centred. Because we were interested in the time-course of prediction, this model was run repeatedly for every 50 ms bin from 1500 ms before to 500 ms after the target word onset (Borovsky et al., 2012; Ellis, Borovsky, Elman, & Evans, 2015). The effect of Predictability was evaluated by assessing whether the absolute *t*-value exceeded 2 (Baayen, Davidson, & Bates, 2008); the differences reported show consistently reliable effects over multiple bins.

One of the participants in the load condition failed to complete two trials because of a technical problem. These trials were treated as missing in the eye-tracking analyses.

2.4.3 Effects of prediction and load

Figure 2 shows fixation probabilities on target objects and mean fixation probabilities on distractor objects in the predictable and unpredictable conditions, separately for the load condition and for the no-load condition. The time was synchronised to target noun onset, with verb onset and offset being the means of all the critical items. The graphs show the time window from 2000 ms before to 500 ms after the target noun onset.

Visual inspection of the graphs suggests that differences in the fixation proportions on target objects in the predictable versus the unpredictable condition began to emerge later in the load condition than in the no-load condition. In support of this, the linear mixed-effects model showed an interaction of Predictability by Load ($|t|$s > 2) in every 50 ms window from 900 ms before the noun onset until 500 ms after the noun onset. The interactions indicate that predictive eye movements were delayed by load.

To understand the interaction in more detail, we ran another model for no-load and load conditions separately. The model evaluated the fixation probability on target objects as predicted by Predictability (predictable vs. unpredictable), including random intercepts and slopes for Predictability by participants and by items. As the upper panel of Figure 2 indicates, participants in the no-load condition were more likely to look at target objects in the predictable condition than in the unpredictable condition from 1050 ms before the noun onset onwards (shown as ● in Figure 2). This corresponded almost exactly to mean target verb offset. The result suggests that participants in the no-load condition predicted upcoming objects that were predictable. As the lower panel of Figure 2 shows, participants in the load condition were also were also more likely to look at target objects in the predictable condition than in the unpredictable condition, but this effect did not emerge until 250 ms before the noun onset. To sum up, the analyses show that predictive eye movements occurred in both conditions, but that they began about 800 ms earlier in the no-load than the load condition.

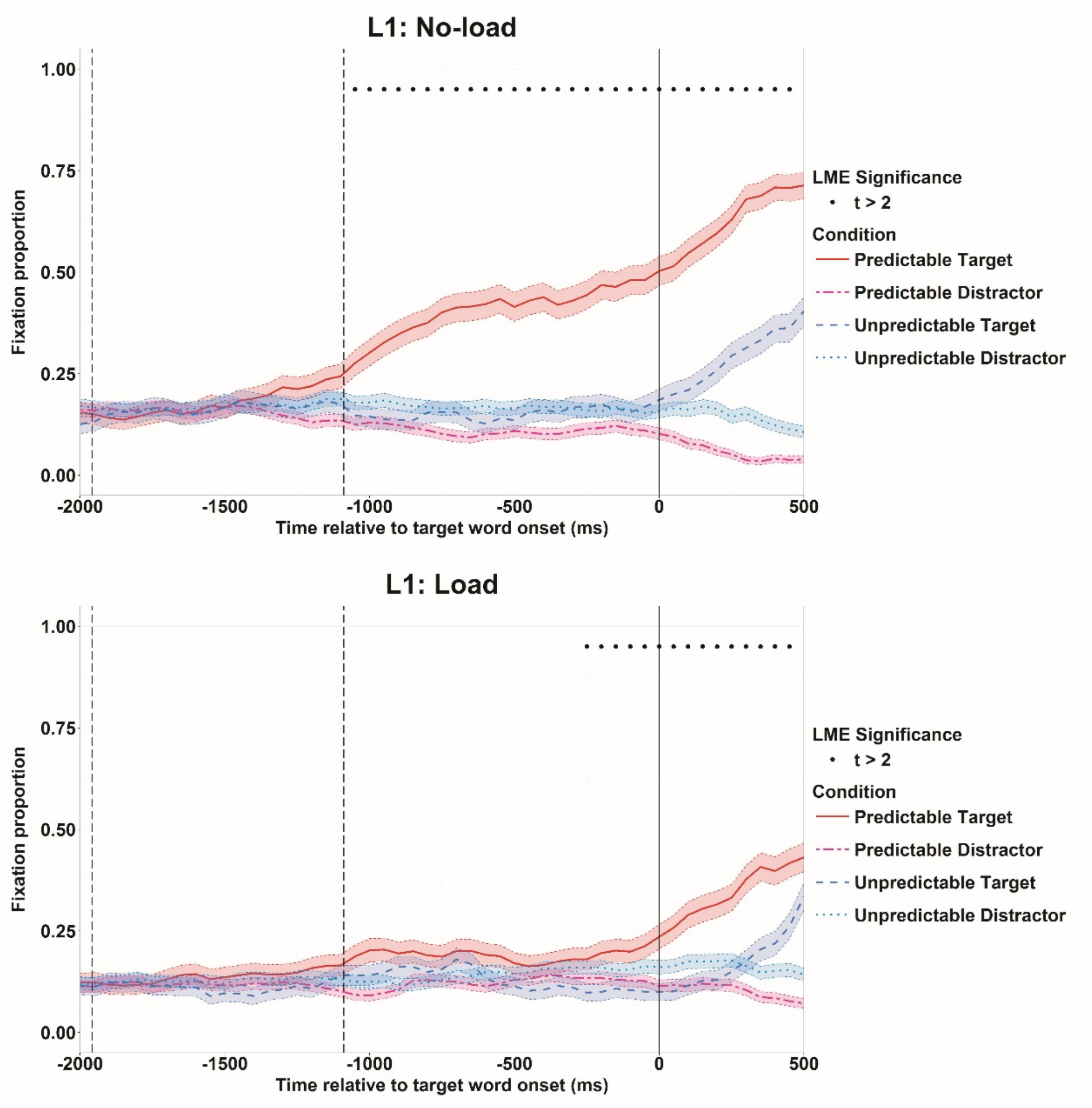


Figure 2. Fixation probabilities on target objects and mean fixation probabilities on distractor objects in the predictable and unpredictable conditions in the no-load condition (top) and in the load condition (bottom) in Experiment 1. Time 0 ms shows target word onset. The left-most dashed line on the y-axis direction (y = -1959 ms) indicates mean verb onset; the next dashed line (y = -1090 ms) indicates mean verb offset. Standard error bars are represented using transparent thick lines. The significance of the model (|*t*|>2) is shown on the top of the graphs, with a solid circle (●) showing a significant effect of Predictability.

Figure 2 suggests that identification of unpredictable target objects is delayed by the cognitive load. We tested this with a linear mixed-effects model examining the fixation probability on target objects in the unpredictable condition predicted by Object Type (target vs. distractor) and by Load (no-load vs. load). The fixation probabilities on the distractors were averaged between the two distractors. The model run in each 50 ms window showed an interaction of Object Type by Load from 50 ms until 500 ms after the target word onset ($|t|$s > 2; except the 100-150 ms bin). Therefore, cognitive load influenced identification of unpredictable target objects.

2.4.4 Semantic competitor effect

We conducted analyses on the semantic competitor that were parallel to the analyses on the target. The linear mixed-effects model showed no effects of Predictability or Load, nor an interaction of Predictability by Load, on the proportion of looks to the semantic competitor objects in any of the time windows ($|t|$s < 2). We also conducted the analyses parallel to the analyses in the unpredictable condition in order to test whether there is a semantic competitor effect in neutral sentence contexts. The model showed no effects of Competitor (semantic competitor vs. distractor) or Load, nor an interaction of Competitor by Load in any of the time windows ($|t|$s < 2). Therefore, there was no indication of a semantic competitor effect in Experiment 1.

2.5 Discussion

Experiment 1 investigated whether making successful predictive eye movements during language comprehension is affected by cognitive load. We found that predictive eye movements in L1 speakers occurred whether or not those speakers were faced with additional cognitive load. However, cognitive load led to those eye movements being delayed. It seems that the additional load caused participants to have fewer cognitive resources that could be allocated for making predictive eye movements. Thus making predictive eye movements takes up resources and is not therefore an automatic process.

2.6 Experiment 2

Experiment 2 addressed similar questions to Experiment 1, but using L2 speakers of English. It asked whether predictive eye movements in L2 speakers occurred under conditions of load and no load, and whether load caused any predictive eye movements to be delayed. Given the results of Experiment 1, we hypothesised that predictive eye movements in L2 speakers would also be delayed under a load. Alternatively, L2 speakers may not make predictive eye movements at all under a load, due to fewer resources available during L2 comprehension.

2.7 Method

2.7.1 Participants

Forty-eight L2 English speakers participated in Experiment 2. Native languages of the L2 participants were Chinese (20), Polish (3), Spanish (3), Romanian (2), Norwegian (2), German (2), Lithuanian (2), Malay, French, Czech, Dhivehi, Greek, Bulgarian, Swedish, Russian, Urdu, Catalan, Slovak, Dutch, Hindi, and Armenian. They filled in a language background questionnaire before the experiment. Their mean length of their stay in the UK was 13.3 months (*SD* = 24.7, range = 3-159 months), and their mean length of exposure to English was 12.5 years (*SD* = 4.8, range = 1-21 years). They also rated their English proficiency on a scale from 1 (not good at all) to 10 (very good), and the mean self-rated English proficiency score was 8.1 (*SD* = 1.1, range = 5.5-10).

2.7.2 Stimuli and procedure

The stimuli and the procedure in Experiment 2 were identical to those in Experiment 1.

2.8 Results

2.8.1 Behavioural task accuracy

The mouse-clicking responses were not recorded for participants who were assigned to the no-load condition. The accuracy for the target clicking task for participants in the load condition was 98%. Incorrectly answered trials were excluded from the eye-tracking analyses. In the working memory task, the mean percentage of correctly recalled words was 68% (*SD* = 26%; range = 40-94%).

2.8.2 Eye-tracking data analyses

The eye-tracking data were analysed as in Experiment 1.

2.8.3 Effects of prediction and load

Figure 3 shows the fixation probabilities on target objects and averaged fixation probabilities on distractor objects in the predictable and unpredictable conditions for participants in the load condition and in the no-load condition separately. The time was synchronised at the target noun onset, and verb onset and offset are the means of all the critical items. The graphs show the time window from 2000 ms before to 500 ms after the target noun onset. The model testing the fixed effects and interaction of Predictability (predictable vs. unpredictable) and Load (no-load vs. load) showed a significant interaction of Predictability by Load (|*t*|s > 2) in every 50 ms window from 850 ms before the target noun onset all until 500 ms after onset, except in the 300-350 ms time window. The significant interactions in the time window before target word onset indicate that participants showed more predictive eye movements when they were not under cognitive load than when they were under cognitive load.

To explore the interaction, we ran a model evaluating the fixation probability on target objects as predicted by Predictability (predictable vs. unpredictable), including random intercepts and slopes for Predictability by participants and by items. We ran this model for no-load and load conditions separately. Participants in the no-load condition were more likely to look at target objects in the predictable condition than in the unpredictable condition from 950 ms before the noun onset onwards (shown as ● in Figure 3). In contrast, participants in the load condition did not show a significant effect of condition in any of the time windows before the noun onset. As in Experiment 1, we can conclude that predictive eye movements were significantly diminished as a result of the additional cognitive load.

We further examined the relationship between the extent of L2 participants' prediction and their English proficiency. We computed the arcsine-transformed target fixation proportion difference between the predictable and the unpredictable conditions from 200 ms after the mean verb offset (= 890 ms before the target noun onset) until the target noun onset, and used this as a proxy for the extent of prediction. We computed the correlation between this measure and L2 proficiency measures for participants in the load condition and for those in the no-load condition separately. In both groups of participants, the extent of prediction did not correlate with participants' self-rated proficiency scores (no-load condition, $r(22) = -.092$, load condition, $r(22) = .15$, *p*s > .1), with their lengths of stay in the UK (no-load condition, $r(21)$[8] = .052, load condition, $r(22) = .11$, *p*s > .1), or with their length of exposure to English (no-load condition, $r(22) = -.064$, load condition, $r(21)$[9] = .058, *p*s > .1).

---

[8] One participant did not provide this information.
[9] One participant did not provide this information.

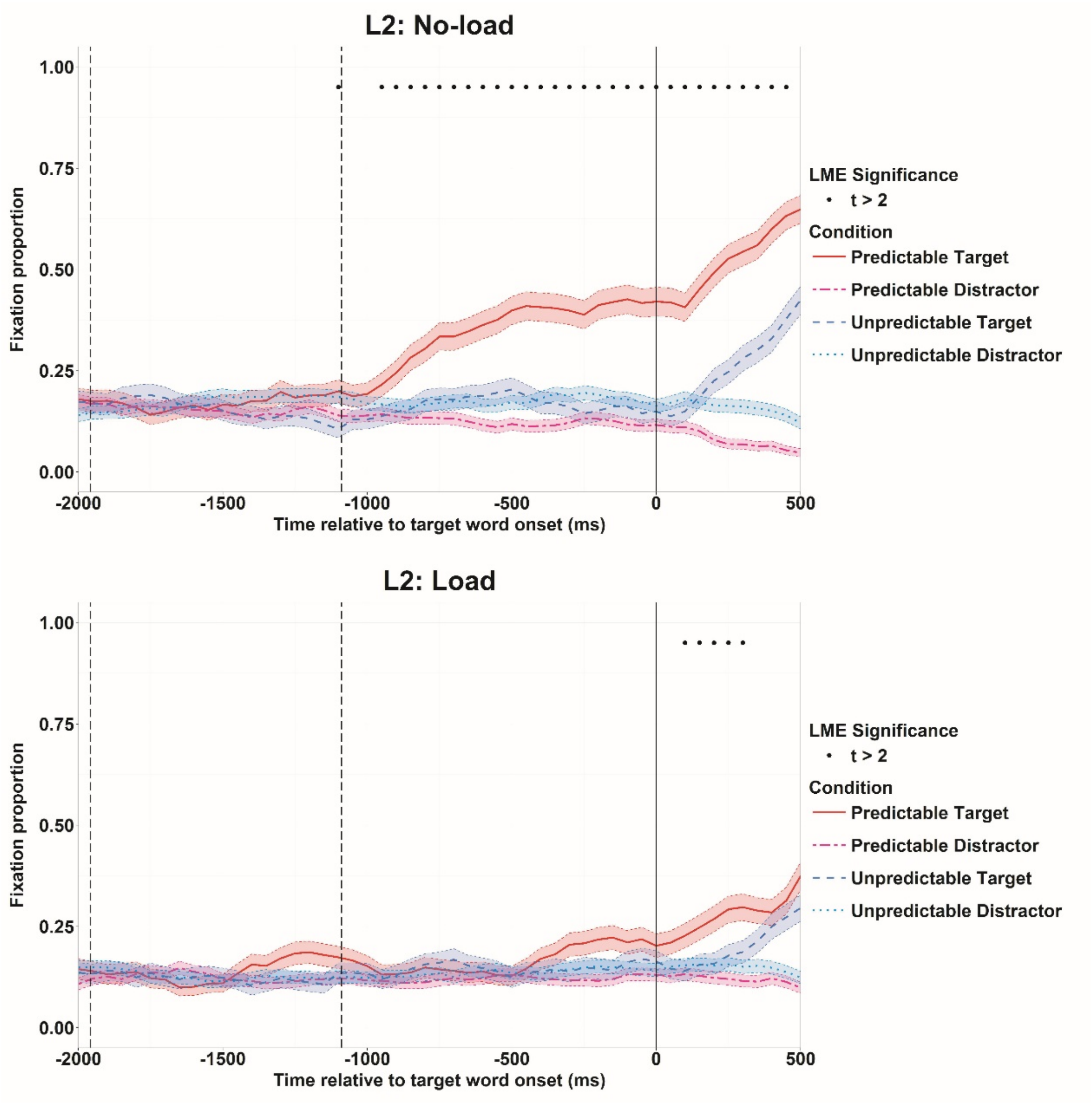


Figure 3. Fixation probabilities on target objects and mean fixation probabilities on distractor objects in the predictable and unpredictable conditions in the no-load condition (top) and in the load condition (bottom) in Experiment 2. Time 0 ms shows target word onset. The left-most dashed line on the y-axis direction (y = -1959ms) indicates mean verb onset; the next dashed line (y = -1090 ms) indicates mean verb offset. Standard error bars are represented using transparent thick lines. The significance of the model (|*t*|>2) is shown on the top of the graphs, with a solid circle (●) showing a significant effect of Predictability.

2.8.4 Effects of cognitive load in the unpredictable condition

Experiment 1 found that cognitive load may interfere with general identification of target objects. We examined if the cognitive load affected eye movements in the unpredictable condition in Experiment 2 as well. The same linear mixed-effects model as in Experiment 1 examined the fixation probability on target objects predicted by Object Type (target vs. distractor) and by Load (no-load vs. load). The model did not show effects of Object Type, Load, or their interaction in any time window between the target word onset and 500 ms after the target word onset. Therefore, cognitive load did not influence identification of unpredictable target objects after target word onset in L2 participants.

2.8.5 Semantic competitor effect

The linear mixed-effects model run for the semantic competitor objects did not show any significant effect of Predictability in any of the time windows ($|t|$s < 2). We also ran a model parallel to the model in the unpredictable condition on semantic competitor objects, which did not show any effects of Competitor (semantic competitor vs. distractor) or Load, nor an interaction of Competitor by Load in any of the time windows ($|t|$s < 2).

2.9 General Discussion

We investigated effects of cognitive load on predictive eye movements in L1 and L2 speakers. In Experiment 1, L1 English speakers listened to predictive and non-predictive sentences and clicked on an object that was mentioned in the sentence. Half the participants performed an additional working memory task of remembering word lists. In Experiment 2, fairly advanced L2 speakers of English were tested under the same conditions (i.e., either under a load or under no load). The results showed that both L1 and L2 participants directed their eyes to a predictable target object before it was mentioned (and did not show such predictive looks to the same object when the sentence was non-predictive), which suggests that they made predictions about upcoming referents. Participants who were under a cognitive load showed increased looks to predictable objects much later compared to those who did not perform the concurrent working memory task. This pattern of results was similar for L1 and L2 participants. Taken together, the results indicate that predictive eye movements do not happen automatically, but require cognitive resources that are used to remember words.

2.9.1 Evidence for prediction in L2

The current findings suggest that L2 speakers can make use of the information extracted from each word to predict a likely referent like L1 speakers when there is no additional cognitive load. This conclusion is consistent with Chambers and Cooke (2009), but inconsistent with Mitsugi and MacWhinney (2016). However, Mitsugi and MacWhinney's study was different in several respects from Chambers and Cooke's study and from our study. In Mitsugi and MacWhinney's study, L2 speakers had to use syntactic information of case markers in addition to the meaning of encountered words to make predictions. But this combinatorial utilisation of the cues might have been particularly difficult in L2, because the manipulated syntactic rules were specific to the L2. In Chambers and Cooke's and in our study, experimental sentences were syntactically simple (no double-object structure), and L2 speakers did not have to rely on L2-specific cues for predictions, so predictions were probably easier.

Another explanation for the inconsistency with Mitsugi and MacWhinney (2016) relates to proficiency. Mitsugi and MacWhinney's L2 participants were intermediate learners, having studied the L2 for 4.3 years on average. Our participants had been exposed to English for more than 12.5 years on average (albeit this had a wider distribution), and participants in Chambers and Cooke (2009) for 11.9 years. A higher proficiency may underlie successful prediction for our participants who were not under a cognitive load and in Chambers and Cooke. Finally, the sentences in our study were spoken slowly with pauses, so our participants had longer time to process contextual information and to generate predictions compared to participants in Mitsugi and MacWhinney. Consistent with this explanation, a recent study has shown that a slower reading rate enhanced predictive processing in L1 speakers (A. Ito, Corley, Pickering, Martin, & Nieuwland, 2016). In sum, the inconsistent results between Mitsugi and MacWhinney and our studies could be explained by different types of cues, speech rate, or proficiency.

We did not find any relationship between L2 participants’ proficiency scores and their predictive eye movements for the robust prediction effect. This is surprising, given that Chambers and Cooke (2009) used a similar set of proficiency measures and found a robust correlation between L2 proficiency and the extent of predictive eye movements. One possibility is that our participants who did not perform the working memory task were predicting at the ceiling level. Although the experiment design is similar in Chambers and Cooke’s study and our own, time between the critical verb onset and the predictable noun onset was longer in our study (1959 ms) than in Chambers and Cooke (1220 ms). This may have made predictions easier for less proficient L2 speakers in our study.

2.9.2 Cognitive load affects predictive eye movements

We found that cognitive load affects predictive eye movements for both L1 and L2 participants. This suggests that cognitive resources are required for making predictive eye movements across different groups of participants. It appears that participants who were under a cognitive load had to allocate the cognitive resources to the working memory task, additionally to prediction, whereas those who were not under a cognitive load could focus more on prediction. Our results are therefore compatible with Huettig and Janse (2016), who showed that people with better working memory capacities made more predictive eye movements. People with a larger working memory capacity have more resources available compared to those with a smaller working memory capacity. Hence, both studies found that predictive eye movements are stronger when there are more resources available.

It is possible that making predictive eye movements is not fully automatic. But predictive eye movements involve multiple stages of processing from identifying objects in the visual scene, building constraints about an upcoming referent, judging which object fits in the constraint, and finally moving the eyes to the predicted object. It is not clear which of the involved processes was affected by the cognitive load.

One possibility is that cognitive load delays prediction itself. Huettig (2015a) discusses two-systems accounts of prediction, comprising one automatic and one non-automatic system. The automatic system involves simple associative mechanisms, wherein semantic relationships between words in a sentence context are compared with the information stored in long term lexical semantic memory. Computing this association appears to lead to pre-activation of information that is semantically related to context words. The non-automatic system involves building up message-level information from information extracted from each word in a sentence, which leads to prediction of information of words (e.g., meaning) that are likely to occur. This account assumes that the exploitation of semantic information of preceding verbs to generate predictions is non-automatic. Predictions cued by a preceding verb as in our study can use both automatic (semantic associations) and non-automatic (computing syntactic relations between words, or the use of production system; cf. Dell & Chang, 2014; Pickering & Garrod, 2007, 2013) pathways. Therefore, the non-automatic route might have been influenced by cognitive load.

Alternatively, the cognitive load may have affected the integration of the prediction with the depicted objects. This explanation assumes that prediction of semantic properties (e.g., something a lady would fold) occurs automatically, but the process of making an inference about which objects would meet the predicted properties (e.g., that a scarf but not a piano can plausibly be folded by a lady) is not automatic. Fixations on an appropriate object (here, the scarf) would be related to this process of inference-making and would therefore be affected by cognitive load.

Finally, cognitive load might have interfered with the process of memorizing the location of each object (e.g., the scarf is right top), since short-term memory is responsible for storing temporal information (Baddeley, 2012). In accord with this, Ferreira, Foucart, and Engelhardt (2013) showed that a complex scene with more distractor objects delayed online identification of referents, a finding which indicated a difficulty in integrating linguistic information with visual information. Alternatively, participants in the load condition may well have been rehearsing words to be recalled in their head. This could have interfered with prediction, particularly if the predictions relied on the production system (cf. Dell & Chang, 2014; Pickering & Garrod, 2007, 2013). However, such mental rehearsal could also interfere with spoken word recognition regardless of prediction. These explanations, which assume the effect of load on general (not prediction-related) processing, are compatible with our additional finding that cognitive load led to delayed fixations on target objects in the unpredictable condition in L1 participants. However, we did not find any effect of cognitive load in the unpredictable condition in L2 participants. The results suggest that at least in L2 participants, diminution of language-mediated eye movements alone cannot explain the effect of cognitive load. But the working memory load in L1 participants may have interfered with the process of remembering information that is relevant to the process of prediction.

According to past research, working memory plays a role for successful L2 sentence comprehension (Dussias & Pinar, 2010), so the working memory task in our study may have interfered with L2 sentence processing in general. This would consequently hinder prediction, as incremental processing is needed for predictions (Kamide, Altmann, et al., 2003). However, our data suggest that predictive eye movements were affected similarly in L1 participants and in L2 participants. Hence, the finding of non-automaticity in L2 speakers is unlikely to reflect increased cognitive demand in L2 comprehension.

2.9.3 No semantic competitor effects

Our study found no evidence that semantic competitor objects were more likely to be fixated than distractors. This lack of effect may be due to the presence of the target object, which had a very strong tendency to attract fixations. Our study therefore contrasted with Rommers et al. (2013), in which a predictable target and its shape competitor were never co-present, and so fixations on shape competitors could not have been overridden by fixations on target objects. Yee and Sedivy (2006) did observe a semantic competitor effect when target and competitor objects were co-present, but their study differed from our study in some aspects that might have led to the inconsistent results. Importantly, their critical words were not embedded in a sentence, which means that there were no contextual constraints that could make the objects a more or a less likely referent (i.e., all the depicted objects were equally likely to be mentioned). In our study, semantic competitor objects could not serve as plausible referents for the upcoming word; therefore, participants may have reduced their looks to the competitor objects. The results are consistent with Dahan and Tanenhaus (2004; Experiment 1), who did not find more looks to phonological competitor objects than to non-competitor objects when reference to those objects would be implausible continuations of the preceding context.

However, we did not find a semantic competitor effect in the unpredictable sentences either. It is unlikely that the semantic relatedness was weak in the current study. We selected semantic competitors from the same semantic category as the target words. This selection was also used in Huettig and Altmann (2005), who did obtain a semantic competitor effect in neutral sentence contexts. The inconsistency could be due to a task demand. Our participants were asked to click on a mentioned object, while participants in Huettig and Altmann (2005) performed no explicit task. The presence of the task may have made our participants focus more on target objects, reducing looks to unmentioned semantic competitor objects. Another possibility relates to the characteristics of the distractor objects: the two distractors in our study were semantically related to each other as well as target and sematic competitor objects, while the distractors were semantically unrelated in Yee and Sedivy's (2006) and Huettig and Altmann's (2005) studies. This means that, in their studies, participants could have predicted a reference to one or other of the related objects on the basis of the visual scene alone. Alternatively, fixations on one related object might have preferentially led to fixations on its related partner – something that could not happen for the unrelated objects.

2.10 Conclusion

We reported two experiments that investigated whether L1 and L2 speakers' predictions are subject to processing limitations. We found predictive eye movements in L1 and L2 speakers, but these predictive eye movements were delayed for participants who performed a working memory task of remembering words concurrently. To conclude, making predictive eye movements is not a fully automatic process, as it requires cognitive resources that are used for remembering words.

3 Study 2. Prediction of phonological information in L1 and in L2

This chapter is based on a manuscript submitted for publication as Ito, A., Pickering, M. J. & Corley, M. Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study.

3.1 Introduction

Studies have shown that people make predictions about upcoming words during language comprehension, and that such predictions can entail specific phonological or orthographic word forms. However, as we shall see below, it is not clear when these predictions occur. People may be able to predict as soon as relevant information becomes available, but when resources are limited, predictions may be delayed or may be weakened. In order to explore the time-course of phonological prediction, we used a visual world eye-tracking paradigm to investigate phonological predictions. In order to explore effects of resources, we tested L1 and L2 speakers.

3.1.1 Prediction of phonological information in L1

In an ERP study, DeLong et al. (2005) found that people can predict phonological aspects of highly predictable words during reading comprehension. As expected, unpredictable (but plausible) words elicited larger N400 amplitudes than predictable words. But most importantly, an article (*a*/*an*) that was incompatible with the form of the predictable word (e.g., *an* when the word *kite* was predictable) also elicited a larger N400 effect relative to an article that was compatible with the predictable word (in this case, *a*). The authors argued that this effect for prediction-inconsistent articles could not be explained by integration, and indicated that people predicted an element of the phonological form of predictable words (whether it began with a vowel or a consonant).

Other ERP studies have also found evidence that people predict word form. Words or nonwords that are highly similar in form to predictable words elicit smaller N400s than words or nonwords that are dissimilar to predictable words (A. Ito et al., 2016; Kim & Lai, 2012; Laszlo & Federmeier, 2009). In one study, participants read contexts such as "*The student is going to the library to borrow a…*", followed by the predictable word *book*, an unpredictable word whose form was related to the predictable word (*hook*), or an unpredictable word whose form was unrelated to the predictable word (*sofa*). The unpredictable words *hook* and *sofa* both showed larger N400s compared to the predictable word *book*, but the N400 was reduced for the form-related word *hook* compared to the unrelated word *sofa* (A. Ito et al., 2016). The findings are compatible with the view that readers pre-activate the forms of predictable words.

A limitation of these ERP studies is that they cannot reveal when the predictions occurred. On the other hand, a visual world experiment can continuously record eye movements as participants listen to a sentence. Thus, eye movement evidence is closely time-locked to when predictions happen, and predictive eye movements to form-related distractors should reveal the specific point at which word forms are predicted, if at all.

Whereas we expect word form to be predicted in L1, such predictions may be less likely to be made by L2 speakers. According to production-based prediction accounts (e.g., Pickering & Garrod, 2007, 2013), word form predictions occur at a later stage of predictive processing relative to predictions of other linguistic information (e.g., meaning). Such accounts assume that word form predictions may not occur when available resources are limited, such as when people comprehend in a language in which they are less proficient. We tested this hypothesis by investigating phonological predictions in L2 speakers. Note that, in common with other studies, we make no distinction between the prediction of phonology or of orthography. We return to this point in the general discussion; for now, we assume that predictions of 'form' are driven by phonology.

3.1.2 Prediction of phonological information in L2

L2 speakers may show delayed or weaker predictions relative to L1 speakers, presumably as a result of having reduced resources available. Compared to the evidence in L1 speakers, it is less clear whether L2 speakers predict phonological information. Martin et al. (2013) used a paradigm similar to that of DeLong et al. (2005), and found an increased N400 for articles that were incompatible with predictable words in L1 speakers. But they did not find this effect in L2 speakers. Like their L1 counterparts, L2 speakers showed an N400 effect for unpredictable words (nouns) relative to predictable words, but their N400 responses did not differ at the preceding articles (even though they were familiar with the *a*/*an* rule in English). The results suggest that L2 speakers do not predict phonological information like L1 speakers.

However, the cloze probabilities were not particularly high in this study (69% in L1 speakers and 65% in L2 speakers), and it is possible that L2 speakers predict phonological information only when the relevant word is highly predictable. In fact, there is evidence that L2 speakers can predict some features of upcoming words, including semantic information (Chambers & Cooke, 2009) or specific lexical information (Foucart et al., 2014; Foucart, Ruiz-Tada, & Costa, 2016). Although these studies do not provide evidence for phonological prediction, it is possible that L2 speakers pre-activate phonological information when contextual constraints are strong enough that they can confidently predict a specific lexical item.

3.1.3 The design of the current study

To specifically investigate prediction of phonological information, we conducted two experiments using a phonological competitor paradigm. This paradigm exploits the fact that when people hear a word, they show a tendency to fixate objects whose name is phonologically related to the mentioned word relative to other objects (Allopenna et al., 1998). This paradigm is suitable for an investigation of the pre-activation of phonological information.

However, phonological competitor effects may disappear if a context is predictive and listeners detect that competitor words would not plausibly fit into the on-going sentence (Dahan & Tanenhaus, 2004; Weber & Crocker, 2012). Taking this into account, the current study adopted a target-absent design, where the target object and its phonological competitor were never co-present. In an analogous experiment, Rommers et al. (2013) investigated the prediction of physical aspects of mentioned items. Their participants heard highly constraining sentences and saw three unrelated distractor pictures together with one of a picture of the target object (i.e., corresponding to the predictable word), an object of a similar shape to the target object, or an unrelated object. Participants fixated the similar-shaped object more than the unrelated objects before the target word could be processed (assuming a 200 ms delay to initiate eye movements; Saslow, 1967). These findings support pre-activation of shape information. We therefore adopted their design, but used objects corresponding to phonologically related words rather than similar-shape objects. The primary advantage of this design is that it should prevent looks to the competitor object being swamped by looks to the target object. In other words, the absence of the predictable object should give participants more opportunity to fixate on the competitor object.

3.2 Experiment 3

Experiment 3 investigated whether L1 English speakers (who reported no knowledge of Japanese) pre-activate phonological information when a specific word is highly predictable. Participants listened to sentences which contained a highly predictable word, while viewing a scene depicting one of four critical objects: a *target* object whose English name corresponded to the predictable word [*cloud*; Japanese: *kumo*], an *English competitor* object whose English name was phonologically related to the predictable word [*clown*; *piero*], a *Japanese competitor* object whose Japanese name was phonologically related to the Japanese translation of the predictable word [*bear*; *kuma*], or an object that was *unrelated* to the predictable word [*globe*; *tikyuugi*]. They also saw three unrelated distractor objects. The Japanese competitor was included as Experiment 4 tested whether Japanese-English bilinguals pre-activate Japanese phonology and therefore activate phonological information of the Japanese translation of the target words; Experiment 3 also included this condition so that the stimuli could be identical in the two experiments, so that we could conduct a between-experiment comparison. The English name for this Japanese competitor object was unrelated to the predictable word, and therefore there was no reason to expect a difference between the Japanese competitor condition and the unrelated condition in this experiment.

If L1 speakers predict highly predictable words, they should fixate more on the target object than on unrelated objects before hearing the target word. Such predictive looks would not demonstrate that participants predict phonological information, since the effect could occur as long as participants predict some information about target words (e.g., meaning). The critical hypothesis concerns the English competitor condition. If participants predict phonological information, they should fixate on objects corresponding to English competitors more than unrelated objects.

3.3 Methods

3.3.1 Participants

Twenty-four native English speakers who reported no knowledge of Japanese participated in the experiment. Two further participants were excluded from the analyses because they almost never (less than 3% of the time) fixated the depicted objects (experimental items and filler objects); cf. Hintz and Meyer (2015). All participants had normal vision and reported no language disorders.

3.3.2 Stimuli

Experimental stimuli consisted of 16 sentences, each paired with one of four visual scenes (see Appendix 8.2 for the full set of items). The experimental sentences each contained a highly predictable word (e.g., *cloud* in "*The tourists expected rain when the sun went behind the cloud, but the weather got better later.*") at varied positions in the sentence (range = 9th-20th word, *M* = 13.7, *SD* = 2.6) but never sentence-final. The sentences consisted of a mean of 17.6 words (*SD* = 1.4, range = 16-21 words). There were an additional 16 filler sentences, of similar length to the experimental sentences. The sentences were recorded by a male native British English speaker, and sampled at 48 kHz with a format of 32-bit float. The speaker read the sentences at a rate of approximately 2.6 syllables per second with some space between phrases. The mean length of experimental sentences was 10.1s.

The predictability of the target words was assessed using a cloze probability test. Twelve further native English speakers who did not participate in the eye-tracking experiment read sentences truncated before the target, and completed each sentence fragment using the first word that came to mind. The mean cloze probability of the predictable word was 97.5% (*SD* = 3.7, range = 91.7-100%).

Each of the visual scenes contained four objects: a critical object and three distractors. In the target condition, the critical object corresponded to the predictable word (e.g., *cloud* [Japanese: *kumo*]). In the English competitor condition, the English name of the critical object phonologically overlapped at onset with the predictable word (e.g., *clown* [*piero*]). In the Japanese competitor condition, the Japanese name of the critical object phonologically overlapped at onset with the Japanese translation of the predictable word (e.g., *bear* [*kuma*]). The mean number of phonemes shared between predictable words and English competitor words was 2.9 (*SD* = .83) out of a mean of 4.4 phonemes (66.2%), and that between Japanese translations of predictable words and Japanese competitor words was

2.6 (*SD* = .60) out of a mean of 4.9 phonemes (53.8%). English names and Japanese names of the Japanese competitor objects were both unrelated to any of the English names of the target, English competitor and unrelated objects. English and Japanese names of each critical object were also unrelated to each other. In the unrelated condition, the name of the critical object did not have phonological onset overlap with the predictable word or its Japanese translation (e.g., *globe* [*tikyuugi*]). All four objects were semantically unrelated to each other.

We conducted a picture naming test to assess name agreement for the depicted objects. Native English speakers who did not participate in the eye-tracking experiment saw pictures of objects and gave the first word that came to mind when they saw each picture. Some of the items were changed and re-tested, and every picture in the final set of stimuli was tested by at least 12 participants. The naming agreement for objects was 94.2% (*SD* = 6.3, range = 83.3-100%) in the target condition, 86.6% (*SD* = 13.2, range = 61.1-100%) in the English competitor condition, 92.8% (*SD* = 9.4, range = 66.7-100%) in the Japanese competitor condition, and 92.2% (*SD* = 8.4, range = 75-100%) in the unrelated condition.

All of the visual stimuli were shown twice, once in an experimental trial and once in a filler trial. Each experimental list comprised two half-lists, each made up of the 16 visual stimuli paired with 8 experimental and 8 filler recordings. Matched visual stimuli contained the same objects, but the quadrants in which these objects appeared were varied. Visual stimuli which were paired with experimental items in one half-list were paired with fillers in the other half-list, and vice versa. Experimental pictures were counterbalanced in the full lists, resulting in 4 different sets of items, or 8 experimental lists in total.

Critical objects appeared at each of the four quadrants equally frequently. Filler sentences mentioned one of the three distractor objects in the visual scene 75% of the time, so together with the experimental sentences (which mentioned one of the four objects 25% of the time), 50% of sentences referred to an object in the visual scene. An example item with four conditions is shown in Figure 4.

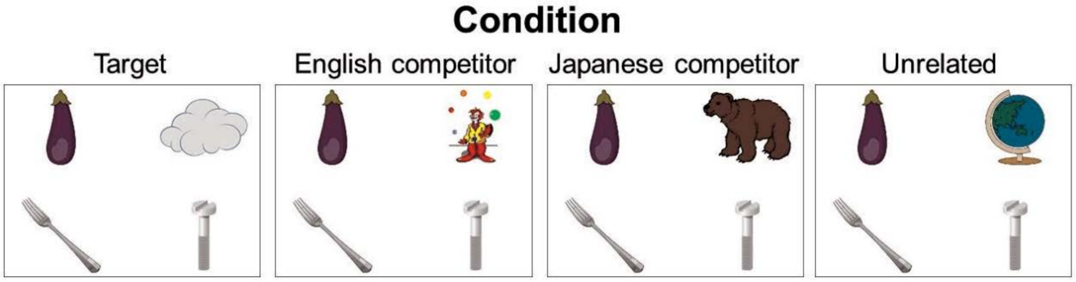


Figure 4. Example visual scene in four conditions for the experimental sentence "*The tourists expected rain when the sun went behind the* ***cloud****, but the weather got better later*." The object depicted at the top right corner is the critical object for this item. The visual stimuli were also paired with the filler sentence "*The waiter immediately came over to the table when the woman carelessly dropped her* ***fork***."

3.3.3 Procedure

The experiment started with a picture familiarisation task. First, participants saw the 64 experimental objects one by one with their English name presented both visually and auditorily at the same time (the names were recorded by the same speaker in the same way as the experimental sentences). Participants were instructed to associate the images with the words, so that they would be able to name them later. After that, they were asked to name each object using the word given earlier. Incorrectly named objects were repeated until participants named them correctly.

In the eye-tracking experiment, participants were seated in front of a computer screen, and they were asked to listen to the sentences and judge whether each sentence mentioned any of the objects in the display. After the instructions, each participant placed their chin on a chin rest, and the eye-tracker was calibrated using the nine-point calibration grid. The experiment started with two practice trials, after which participants were given a chance to ask questions. The pictures were presented on a viewing monitor at a resolution of 1024×768 pixels. Each trial started with a drift correction, which was followed by a 500 ms blank screen. The visual scene was presented 1000 ms before the onset of predictable words in experimental trials. On filler trials, the presentation was 1000 ms before the onset of a word that referred to a distractor or at a random mid-sentence position when the sentence did not mention anything in the scene. The picture stayed on the screen for 750 ms after the offset of the spoken sentence. After the picture disappeared, there was always a same comprehension question "*Did the sentence mention any of the pictures?*". The next trial started after participants gave their answer using a keyboard. No feedback was given during the experiment. The session took about 30 minutes.

3.4 Results

3.4.1 Data analyses

The eye-tracking data were analysed using linear mixed-effects models with the lme4 package (D. M. Bates et al., 2008) in R (R Development Core Team, 2015). The proportions of time spent fixating on target, English competitor, Japanese competitor, and unrelated objects were calculated separately for each 50 ms bin relative to the target noun onset (following Altmann & Kamide, 1999). A linear mixed-effects model evaluated the fixation probability on critical objects as predicted by Condition (Target vs. unrelated, English competitor vs. unrelated, Japanese competitor vs. unrelated). The model included random intercepts by participants and by items (Barr, 2008). Random slopes were not included because the model with them did not converge for several bins. This model was run repeatedly for every 50 ms bin from 1000 ms before to 1000 ms after the target word onset. Such multiple comparisons could give rise to a statistical significance by chance, but our conclusion will be based on the results where a series of consecutive bins showed a significant difference (cf. Borovsky et al., 2012), and a growth curve analysis that did not involve multiple comparisons. Differences between unrelated and other levels of Condition were evaluated by assessing whether the associated *t*-value had absolute values which exceeded 2 (Baayen et al., 2008). Two items were excluded from the eye-tracking analyses because the English competitor object in these items attracted significantly more looks than the unrelated object within 1000 ms after the picture onset when the pictures were presented with a neutral sentence that was unrelated to the English competitor objects in filler trials. This left 14 items for the analyses.[10]

3.4.2 Picture naming accuracy

We calculated the proportion of trials where participants named the pictures with a correct name in the first instance. The mean picture naming accuracy across participants was 99.1% (*SD* = 1.3%). The high accuracy suggests that the pictures were relatively easy to associate with the intended names.

3.4.3 Comprehension question accuracy

The mean accuracy for the comprehension questions in the experimental trials was 100%.

3.4.4 Eye-tracking data

Figure 5 shows the proportion of fixations on target, English competitor, Japanese competitor, and unrelated objects. The linear mixed-effects model showed that participants were more likely to fixate target objects than unrelated objects 600 ms before the acoustic onset of the predictable target word up until 1000 ms after the target word onset. This bias towards the target objects indicates participants' sensitivity to the target word predictability. Critically, participants were also more likely to fixate English competitor objects than unrelated objects between 500 ms before the target word onset and 350 ms before the target word onset. As predicted, participants did not show any bias towards Japanese competitor objects relative to unrelated objects in any of the time windows. The results suggest that participants predicted target words and pre-activated their phonological information.

---

[10] We also analysed the data including all the 16 items. The same linear-mixed model showed significant differences between English competitor and unrelated conditions in the same time window.

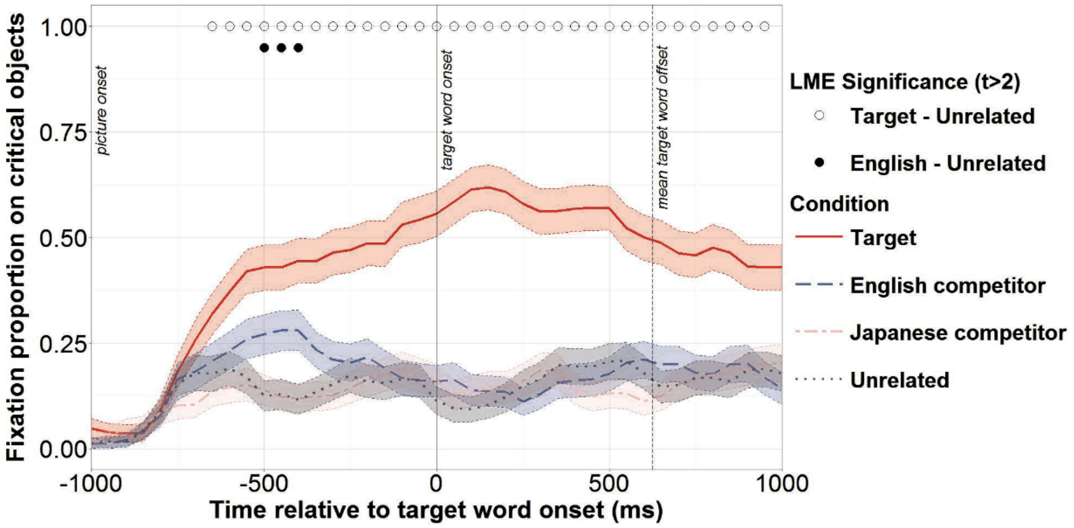


Figure 5. Eye-tracking results in Experiment 3. Time-course graph showing fixation proportion on target, English competitor, Japanese competitor, and unrelated objects. Time 0 ms shows target word onset. The dashed vertical line (y = 625 ms) indicates the mean target word offset. Circles at the top of the graph show significant differences (|*t*|>2) between the target and unrelated conditions (open circle, ○), and between the English competitor and unrelated conditions (solid circle, ●), corresponding to the time on the x-axis. Transparent thick lines are error bars representing standard errors.

We also analysed the filler trials in order to examine whether there was any visual bias towards critical objects irrespective of the predictive contexts. As reported above, two items were excluded after an initial analysis of the filler items. In the remaining 14 items, the linear mixed-effects model did not show any fixation proportion differences between conditions, except that Japanese competitor objects attracted more fixations than unrelated objects in a single 550-600 ms window after the picture onset, and that English competitor objects attracted more fixations than unrelated objects from 1850 ms to 2000 ms after the picture onset.[11] Since these biases do not pattern with the data in experimental trials, the predictive English phonological competitor effect obtained in experimental trials cannot be explained by any visual biases towards the competitor objects.

---

[11] The late advantage for competitors in fillers may reflect the fact that they had been seen before in 50% of trials, and competitor objects had attracted more fixations than unrelated objects in those previous trials.

3.4.5 Discussion

Before the mention of predictable words, participants were more likely to fixate objects whose names were phonological competitors of predictable words than phonologically unrelated objects. The results pattern with the data reported in Rommers et al. (2013), who found that shape competitor objects of predictable words attracted more looks than unrelated objects prior to the mention of predictable words. While the two sets of results are consistent in that both effects started to emerge about 500 ms after the picture appeared, the shape competitor effect in Rommers et al. lasted for much longer than the phonological competitor effect in our study. We will return to this point in the General Discussion (Section 3.9). Consistent with the ERP reading studies that found pre-activation of word form in L1 speakers (DeLong et al., 2005; A. Ito et al., 2016; Kim & Lai, 2012; Laszlo & Federmeier, 2009), Experiment 3 showed that L1 speakers pre-activated phonological information of predictable words during listening comprehension. This phonological prediction effect emerged 500 ms before the onset of predictable words. The onset of this effect roughly corresponded to the acoustic offset of a word that was two words before the predictable word (= 469 ms before the onset of predictable words). We therefore suggest that participants had predicted phonological information of the predictable word by this point.

3.5 Experiment 4

As we reviewed in the introduction, it is not clear whether L2 speakers predict word forms associated with predictable words. Since L2 processing tends to be more cognitively demanding than L1 processing (e.g., Segalowitz & Hulstijn, 2009), predictions in L2 may happen less automatically. If the lack of evidence for L2 phonological prediction in the past research (e.g., Martin et al., 2013) was due to the words not being sufficiently predictable, we expected to find evidence for phonological prediction in the current study, as our experimental sentences were highly predictable. However, if L2 speakers predict less automatically than L1 speakers, we may not find evidence for phonological prediction, or may find a delay in their prediction. Experiment 4 tested whether people ever predict phonological information during L2 comprehension, and additionally compared the time-course of phonological prediction in L2 speakers with that in L1 speakers.

Experiment 4 additionally investigated whether speakers comprehending in their L2 pre-activate phonological information in their L1. There is some evidence for such cross-language activation in non-predictive contexts. For example, Mishra and Singh (2014) found that Hindi-English bilinguals activated the L1 (Hindi) translation-equivalent word form of an L2 (English) target word which was embedded in a non-predictive sentence context after hearing that target word. Other studies report cognate facilitation effects, in which participants process L2 words that share phonological or orthographic form with an L1 word faster than words that do not (A. M. B. De Groot & Nas, 1991; Dijkstra, Van Jaarsveld, & Brinke, 1998; Libben & Titone, 2009). These results suggest that L2 speakers may activate translation equivalents (including their phonology) in their L1 during L2 comprehension.

Another study had Dutch-English late bilinguals listen to English sentences and found no evidence that they activated the Dutch translation of a target word (FitzPatrick & Indefrey, 2010). In this study, the target word was moderately predictable (*mean cloze probability* = 47%). However, there is no clear evidence that cross-linguistic phonological activation occurs *predictively*. Experiment 4 therefore investigated whether L1 Japanese – L2 English speakers pre-activate L1 phonological information that corresponds to the translation equivalent of a highly predictable word in L2.

3.6 Methods

3.6.1 Participants

Twenty-four L2 English speakers whose L1 was Japanese participated in this experiment. Two further participants who fixated all of the depicted objects less than 3% of the time were excluded. All participants had normal vision, and none reported any language disorders. None of them had participated in cloze probability and picture naming pre-tests, which are described in the Stimuli and procedure section below. Their mean age of first exposure to English was 10 years (range = 5-15 years), and the mean length of exposure to English was 13 years (range = 4-20 years)[12]. Participants also self-rated their English proficiency on a scale from 1 (not good at all) to 10 (very good), and the mean self-rated proficiency was 7 (range = 3-10).

3.6.2 Stimuli and procedure

The stimuli and the procedure in Experiment 4 were identical to Experiment 3, except that participants filled in a language background questionnaire before the experiment. Twelve Japanese-English late bilinguals who were studying in the UK and did not participate in the eye-tracking experiment took part in the same cloze probability test as in Experiment 3. The mean cloze probability for L2 speakers was 88.6% (*SD* = 7.1, range = 81.8-100%), which was significantly lower than the L1 cloze probability in Experiment 3 (97.5%), $t(30) = 3.9$, $p < .001$.

Two further groups of Japanese-English late bilinguals from a similar population participated in the picture naming pre-test used in Experiment 3, and the Japanese version of the same pre-test. Each of the pictures in the final set was tested by at least 12 participants. In the English naming pre-test, the naming agreement for critical objects was 93.2% (*SD* = 8.4, range = 76.9-100%) in the target condition, 86.7% (*SD* = 10.8, range = 66.7-100%) in the English competitor condition, 93.8% (*SD* = 9.9, range = 75.0-100%) in the Japanese competitor condition, and 94.4% (*SD* = 9.9, range = 66.7-100%) in the unrelated condition. These name agreement scores were very similar to those in Experiment 3.

---

[12] The length of exposure to English was defined as the total length of any form of regular exposure to English, including both classroom and non-classroom situations. IELTS scores (www.ielts.org) were reported by 15 participants (*M* = 7, range = 6.5-8).

In the Japanese naming pre-test, the instructions were translated into Japanese and English was not used throughout the test. The naming agreement for critical objects was 91.7% (*SD* = 12.5, range = 58.8-100%) in the target condition, 87.9% (*SD* = 16.6, range = 41.2-100%) in the English competitor condition, 90.1% (*SD* = 13.7, range = 64.7-100%) in the Japanese competitor condition, and 97.4% (*SD* = 4.3, range = 88.2-100%) in the unrelated condition.

3.7 Results

3.7.1 Data analyses

The eye-tracking data were analysed in the same way as in Experiment 3. In the filler trials, L2 speakers showed no preference to fixate any given item over another, so no items were excluded.

3.7.2 Picture naming accuracy

The mean accuracy in the picture naming task across participants was 93.4% (*SD* = 7.2%). It appears that the pictures were fairly easy for the L2 speakers to associate with the intended names (though the accuracy was slightly lower than L1 speakers, 99.1%).

3.7.3 Comprehension question accuracy

The mean accuracy for the comprehension questions in the experimental trials was 99.2% (*SD* = 2.1%). Incorrectly answered trials were excluded from the eye-tracking analysis.

3.7.4 Eye-tracking data

Similarly to Experiment 3, we plotted fixation proportions on target, English competitor, Japanese competitor, and unrelated objects (Figure 6). We ran the same linear mixed-effects model as in Experiment 3. The model indicated that target objects attracted significantly more fixations than unrelated objects between 800 ms before the target word onset and 700 ms before the target word onset, and from 350 ms before the target word onset up until 1000 ms after the target word onset. The predictive looks to target objects suggest that L2 speakers predicted some information about target words. In contrast, the fixation proportions on English competitor objects differed from the fixation proportions on unrelated objects from 600 ms until 1000 ms after the target word onset.[13] This late effects of the English phonological competitor suggest that they did not predict phonological information, but that they did activate phonological information associated with the target word after encountering it. At no point from 1000 ms before the target word onset to 1000 ms after the target word onset were there any differences between the Japanese competitor condition and the unrelated condition. Thus, there was no evidence that L2 speakers ever activated the phonology of the Japanese translations of predictable words.

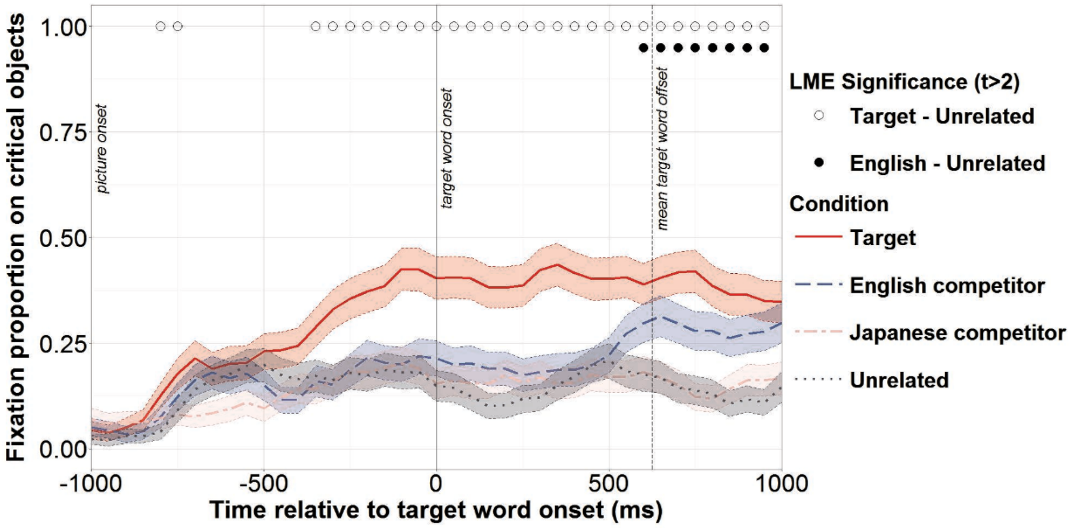


Figure 6. Eye-tracking results in Experiment 4. Time-course graph showing fixation proportion on target, English competitor, Japanese competitor, and unrelated objects. Time 0 ms shows target word onset. The dashed vertical line (y = 625 ms) indicates the mean target word offset. Circles at the top of the graph show significant differences (|*t*|>2) between the target and unrelated conditions (open circle, ○), and between the English competitor and unrelated conditions (solid circle, ●), corresponding to the time on the x-axis. Transparent thick lines are error bars representing standard errors.

---

[13] The difference between the English competitor and unrelated conditions was not significant from 1000 ms onwards.

We further explored the relationship between the English competitor effects in L2 speakers and their L2 proficiency. For each L2 participant, we calculated a difference in the mean arcsine-transformed fixation proportions between the English competitor and unrelated conditions in a time window from 600 ms to 1000 ms relative to the target word onset. We used this as a measure of the English competitor effect, and computed a correlation with the L2 participants' length of exposure to English. As shown in Figure 7, we found a positive correlation between the two measures, $r(22) = .55$, $p < .01$; L2 speakers who had been exposed to English for longer showed a stronger English phonological competitor effect.

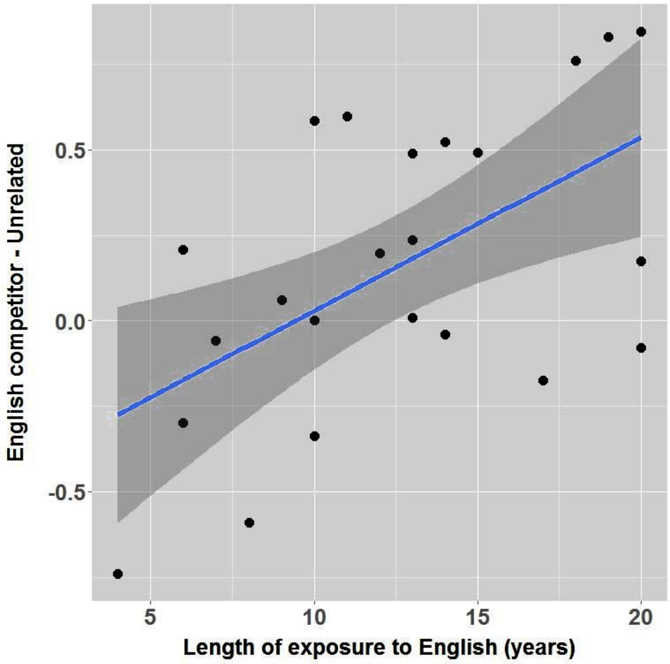


Figure 7. Correlations between the arcsine-transformed fixation proportion difference in the English competitor condition and in the unrelated condition (in the 600 ms to 1000 ms window) and the length of exposure to English in L2 speakers in Experiment 4.

We analysed the 16 filler items using the same linear mixed-effects model as in Experiment 3. Fixation proportion did not differ between any of the condition pairs. Therefore, the English competitor effect in the late time window in L2 speakers cannot be attributed to visual attractiveness of the competitor objects.

3.7.5 Between-experiments analysis

To test whether L1 speakers and L2 speakers differed in predictive eye movements associated with phonological pre-activation, we tested for an interaction of language nativeness with the English phonological competitor effect during the prediction time window (i.e., the 1000 ms window from picture onset to target word onset). 14 items that were used for the analysis in Experiment 3 were included in the data from L1 participants, and all 16 experimental items were included in the data from L2 participants. Figures 5 and 6 suggest that the changes in fixation proportions in this time-course are not linear. To capture this data pattern, we used growth curve analysis with a third-order (cubic) orthogonal polynomial (Mirman, Dixon, & Magnuson, 2008). We coded fixations binomially depending on whether the object was fixated or not in each 50 ms bin. The coded data were then translated to log odds (suitable to test effects on a categorical variable). We constructed a model evaluating the fixation proportions predicted by fixed effects of condition (English competitor vs. unrelated) and language group (L1 vs. L2), and the interaction of the two on all time terms. The model also included participant random effects on all time terms and participant-by-condition random effects on all time terms except the cubic (estimating random effects is “expensive” in terms of the number of observations required, so this cubic term was excluded because it tends to capture less-relevant effects of the tails).

The by-participant analysis model revealed a significant effect of condition on the intercept term, $\beta = -.62$, $SE = .29$, $t = -2.1$, indicating more looks to English competitor objects relative to unrelated objects overall. The interaction of condition by language group was significant on the quadratic term, $\beta = -3.1$, $SE = 1.4$, $t = -2.3$, and on the cubic term, $\beta = -1.7$, $SE = .62$, $t = -2.8$. An inspection of Figure 8 suggests that this effect is because the fixation difference between the conditions increased over time and decreased after reaching a clear peak in L1 speakers, whereas the conditions did not differ throughout the entire time window in L2 speakers (Figure 8, top). The by-item analysis model also revealed a significant effect of condition on the intercept term, β = -.67, *SE* = .32, *t* = -2.2, and a significant interaction of condition by language group on the intercept term, β = .57, *SE* = .21, *t* = 2.7, and on the quadratic term, β = -4.1, *SE* = .93, *t* = -4.4. The interaction on the quadratic term captured the pattern wherein L1 speakers showed the largest difference between the conditions approximately in the middle of the time window, but L2 speakers showed the smallest difference between the conditions in the corresponding time window (Figure 8, bottom).

Since the by-participant and by-item analyses showed slightly different patterns, Figure 8 presents a graph for each analysis. Despite the different patterns in the time-course, the interaction of condition by language group was found in both analyses, and showed that the English competitor effect in the prediction time window was evident in L1 speakers but not in L2 speakers.[14]

---

[14] We also ran a generalized linear mixed-effects model using the glmer() function in R to analyse the data in the prediction time window including both participants and items as random effects. The model assessed raw fixation (fixated vs. not fixated) predicted by condition, language group, and their interaction, including a random intercept by participant and by item (random slopes or the variable Time was not included because the models including these did not converge). The model showed a significant effect of condition, β = .16, *SE* = .023, *z* = 7.0, $p < .001$, and a significant condition by language group interaction, β = -.049, *SE* = .023, *z* = -2.2, $p < .05$.

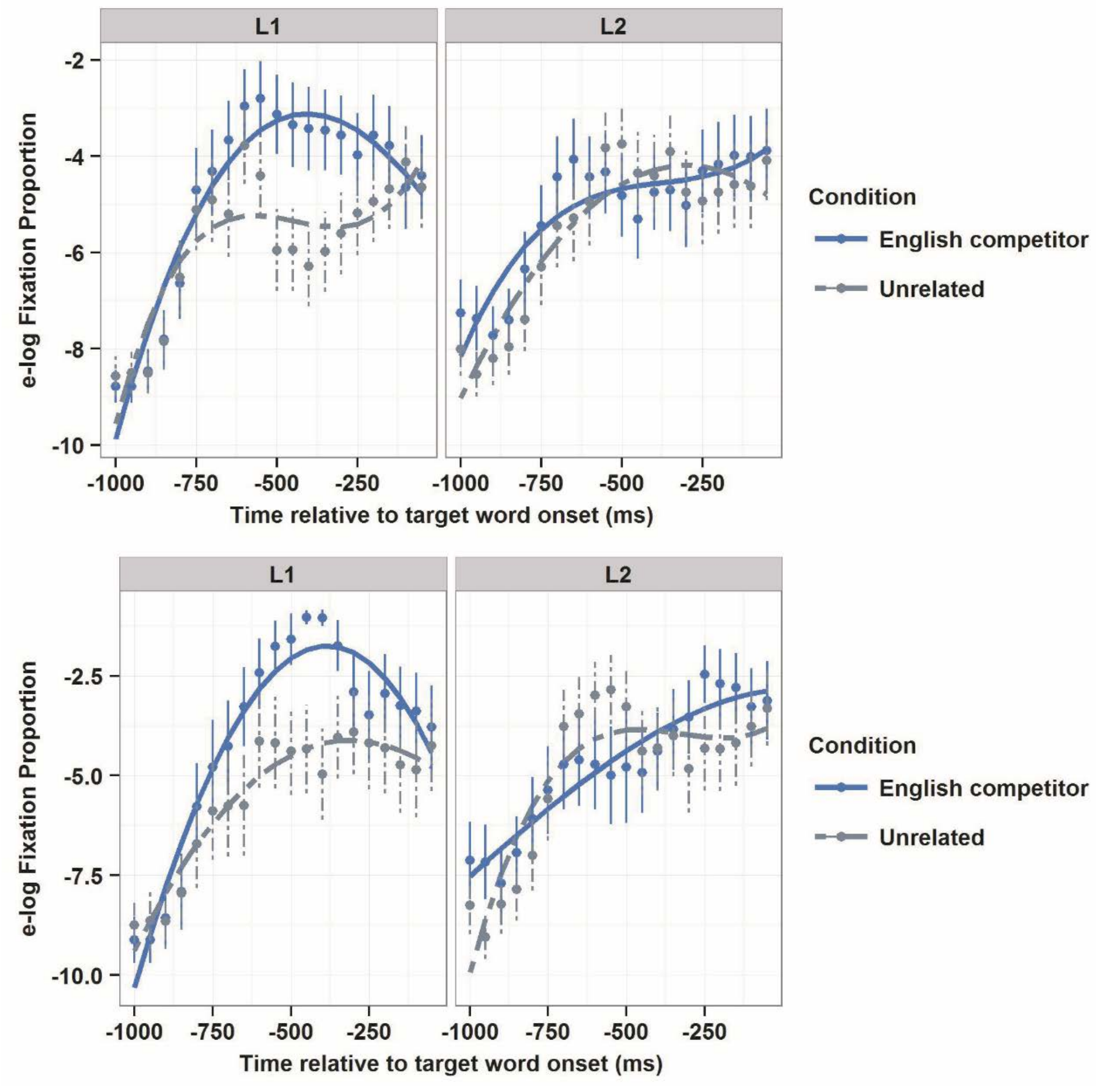


Figure 8. Growth curve analysis model fits (lines) of the fixation data in the English competitor and unrelated conditions in the L1 group (left) and the L2 group (right). The top graph shows the by-participant analysis and the bottom graph shows the by-item analysis. Error bars represent ±1 *SE*.

3.8 Discussion

L2 speakers showed increased looks to target objects before they were mentioned, suggesting that L2 speakers predicted some information about target words. However, we did not find evidence that L2 speakers predict phonological information. L2 speakers showed a tendency to fixate phonological competitor objects relative to unrelated objects, but this tendency did not manifest itself until well after the predictable word onset. Therefore, the phonological competitor effect is more compatible with the interpretation that hearing target word (e.g., *cloud*) spread activation to its phonological competitor words (e.g., *clown*) via phonological priming.

3.9 General discussion

We investigated the time-course of prediction of the phonological information associated with highly predictable words in L1 and L2 speakers. Both groups of participants showed increased looks to predictable target objects well before they were mentioned. L1 speakers were more likely to fixate objects whose name was phonologically related to the predictable word relative to objects whose name was phonologically unrelated to the predictable word from 500 ms before the predictable word onset (hereafter, *competitor prediction effect*). However, L2 speakers did not show such a tendency until 600 ms after the predictable word onset (hereafter, *competitor priming effect*). Their tendency to fixate phonologically related objects over unrelated objects positively correlated with their L2 proficiency (as indexed by the length of exposure). In addition, they did not fixate objects whose name was phonologically related to the Japanese translation of the predictable word.

3.9.1 The timing of phonological prediction in L1

L1 speakers' predictive looks to target objects suggest that they predicted some aspects of target words, and their predictive looks to the phonological competitor demonstrate that they specifically predicted aspects of word form. As noted in the introduction, we cannot be certain whether participants predicted phonology or orthography. The fact that we had no written stimuli (and that we specifically selected target pictures whose names were phonologically related to the spoken target words) suggests that prediction is likely to be largely phonological but we cannot of course rule out orthographic effects.

Interestingly, the phonological competitor prediction effects in L1 speakers were very short-lived (about 150 ms duration) compared to the shape competitor prediction effects obtained in Rommers et al. (2013) (about 1000 ms duration, based on visual inspection). Our study was closely modelled on Rommers et al.'s experimental design, except that we changed the preview from 500 ms to 1000 ms. The difference in the time-course between phonological and shape predictions may be because shape competitor effects are generally stronger and more sustained than phonological competitor effects (Hintz & Huettig, 2015; Huettig & McQueen, 2007). Shape predictions relate to semantics (rather than phonology). According to an account in which predictions make use of the production system (Pickering & Garrod, 2007, 2013; see also Dell & Chang, 2014), phonological predictions may occur after semantic predictions (cf. A. Ito et al., 2016) and this may have contributed to longer-lasting effects of shape prediction compared to effects of phonological prediction.

Alternatively (or additionally), the stronger competitor effects for shape information may be due to the visual world setting, in which the task (e.g., look-and-listen, visual search) and the dependent measure (i.e., eye movements) are of course heavily visual (F. De Groot, Huettig, & Olivers, 2016). Alternatively, retrieval of phonological information from a visual scene is likely to be more costly than retrieval of shape information, because phonological information, unlike shape information, is not present in the visual scene. This reduced cost for the shape information retrieval could explain stronger shape competitor effects.

3.9.2 L2 speakers predict some information but not phonological information

L2 speakers' predictive eye movements to target objects suggest that they predicted some information about target words. This result was qualitatively similar to that of L1 speakers. However, there was no indication that L2 speakers predicted phonological information of highly predictable words. This was consistent with the results of Martin et al. (2013), even though word predictability was higher in our study (89%) than Martin et al. (65%). This finding fits with the proposal that language understanding does not require prediction at all levels of representation (Huettig & Mani, 2016). It may be that L2 comprehenders' resources are limited and so they tend to predict less detailed information about predictable words. The less detailed information entailed in L2 speakers' prediction may partly account for increased effort in L2 comprehension compared to L1 comprehension, because successful prediction can reduce the processing cost of the predicted word, resulting in faster comprehension of that word (Frisson et al., 2005; Rayner et al., 2011; Smith & Levy, 2013).

It is possible to explain these findings in terms of a prediction-by-production account (Pickering & Garrod, 2013). Predictions related to phonology take place after predictions related to other aspects of words such as their meaning, and, more important, may fail to occur because the production system may be interrupted before a representation of phonology is constructed. The additional difficulty associated with L2 production means that phonological representations are less likely to be constructed than is the case in L1 production. These proposals are compatible with the findings of A. Ito et al. (2016), who found evidence from ERPs that phonological but not semantic prediction failed to occur under conditions involving a fast presentation rate.

A possible limitation of our study is that the cloze probability for the sentences used in the current study was higher for L1 speakers (98%) than for L2 speakers (89%). This means that target words were somewhat more predictable for L1 speakers than for L2 speakers. This might conceivably underlie group differences in the phonological competitor effect. But it is extremely unlikely, because even the L2 cloze probability was extremely high in comparison to other studies (Foucart et al., 2014; Martin et al., 2013).

3.9.3 Stronger phonological priming for more proficient L2 speakers

L2 speakers showed English phonological competitor priming effects after hearing target words. This finding rules out the possibility that the lack of phonological prediction in L2 speakers was because they were insensitive to the phonological overlap between target words and their competitor words. We found that L2 speakers' English competitor priming effects correlated with their length of exposure to English: L2 speakers with longer exposure to English showed a stronger English competitor priming effect. Our data fit with studies that found a stronger phonological competitor effect for more proficient L2 speakers relative to less proficient L2 speakers (Blumenfeld & Marian, 2007, 2013).

3.9.4 No evidence for L1 activation during L2 comprehension

We did not find any evidence that L2 speakers ever activated L1 translations of the English target words. Since our picture naming pre-test on L2 speakers showed that the name agreement for the pictures was similarly high in English and in Japanese, the lack of evidence for Japanese activation cannot be explained by a difference in name agreement. One possibility is that Japanese-English bilinguals did not activate Japanese at all during the experiment, because all the experimental setting was in English. However, this explanation would contradict with rich evidence suggesting that L2 speakers activate both their L1 and L2 during L2 comprehension (Marian, Spivey, & Hirsch, 2003; Spivey & Marian, 1999). The inconsistency could be because the high predictability of the target words in the current study, where participants could tell that Japanese phonological competitor words would be unlikely to be mentioned in the sentence. Another explanation of our results is that a Japanese phonological competitor effect may have been too weak to affect eye movements. One reason to assume this is that the English competitor effect was relatively weak both in terms of magnitude (about 10% difference in fixation proportion) and in terms of duration (the effect lasting for about 150 ms). Therefore, activation of Japanese from English, even if it occurred, was likely to be relatively weak.

Another explanation for the lack of a Japanese competitor effect may be related to the picture naming task conducted before the eye-tracking experiment. We conducted the picture naming in English only in order to keep the experiment comparable between L1 and L2 speakers. Picture naming might have boosted word form activation or facilitated retrieval of picture names, for example via lexical priming from production to comprehension (Wheeldon & Monsell, 1992). This may have strengthened activation of English phonology as a result.

Another possibility (albeit speculative) is that the lack of Japanese competitor effect was due to the lack of orthographic overlap. Translations of a target word and its Japanese competitor word were phonologically related (*kumo* – *kuma*) but not orthographically related (雲 – 熊) in Japanese. Further investigation would be needed to dissociate phonological effects and orthographic effects.

3.10 Conclusion

Our visual world study found that both L1 and L2 speakers made predictions about upcoming words. However, L1 speakers appeared to predict specific phonological information associated with highly predictable words, while L2 speakers do not. The results suggest a limitation in prediction in L2. The lack of phonological prediction in L2 speakers may be one of the factors that cause L2 online comprehension to be more difficult than L1 comprehension.

4 Study 3. Predicting form and meaning in L1

This chapter is based on a paper published as Ito, A., Corley, M, Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language, 86, 157-171.*

4.1 Introduction

People regularly use contextual information and world knowledge to predict aspects of language that are likely to be mentioned as a sentence or discourse unfolds (e.g., Altmann & Kamide, 1999; Federmeier, 2007; Huettig, 2015a; Kutas et al., 2011). Prediction is often hypothesized to occur via a so-called *pre-activation* mechanism, whereby some aspects of word meaning, grammar or form are activated before the onset of the predicted word (e.g., DeLong et al., 2005; Federmeier & Kutas, 1999; Laszlo & Federmeier, 2009; Otten et al., 2007; Van Berkum et al., 2005). But how these types of linguistic information are pre-activated is still unclear. The production-based prediction account proposes prediction via a comprehender’s production system (Pickering & Garrod, 2007, 2013). Under this account, pre-activation of form does not occur in the absence of pre-activation of meaning, because the language production system first accesses meaning, and then maps the meaning information onto form information. This chapter reports two ERP experiments that investigate pre-activation of meaning and form of predictable words during language comprehension to explore the relationship between meaning and form pre-activation. We investigate pre-activation, as indexed by N400 ERP modulations (Kutas & Federmeier, 2011), at a word presentation rate that is standard in reading ERP studies (Experiment 5; 500 ms per word) and at a slower presentation rate (Experiment 6; 700 ms per word) which allows more time to generate online predictions. Below, we first discuss the production-based prediction theory, and then outline existing evidence for the pre-activation of meaning and of form before introducing the current study.

4.1.1 Production-based prediction accounts

Pickering and Garrod (2007, 2013) proposed that people use the language production system when predicting upcoming words during comprehension. According to this account, when people comprehend sentences, they covertly imitate those sentences and implement their production systems to predict upcoming words. Lexical prediction is thought to involve pre-activation of linguistic information (e.g., word form, meaning) of predictable words. Linguistic information associated with predictable words is pre-activated using the same mechanisms that are used to produce words.

A most parsimonious possibility is that comprehenders make direct use of the mechanisms involved in language production – a version of prediction-by-production that we call *prediction-with-implementation.* Although language production models (e.g., Dell & O'Seaghdha, 1992; Levelt, Roelofs, & Meyer, 1999) differ in many important respects, they agree on the view that people produce a word by first activating its semantic information and then proceeding through stages that lead to activation of its phonological or orthographic information (its word form). These stages take several hundred milliseconds according to most estimates (see Indefrey & Levelt, 2004). According to prediction-with-implementation, comprehenders also pre-activate semantic information before form information, following roughly the same time-course. It is of course possible for the comprehender to actually complete the speaker's utterance, simply by continuing the process of production until the stage of articulation – this is exactly what happens in a cloze test.

However, full implementation of the production system for prediction requires time and resources. When these are lacking, only a part of the production system may be used for prediction. As activation of form information follows activation of semantic information in the language production system, a partly engaged production system might lead to pre-activation of semantic information but not of form information. This means that a comprehender might pre-activate meaning without pre-activating form under conditions of difficulty, but would not pre-activate form without pre-activating meaning.[15]

However, we note that a pattern wherein meaning pre-activation is more likely to occur than form pre-activation could also be compatible with an alternative account involving cascaded pre-activation. Cascaded pre-activation has not previously been hypothesized to underlie prediction, but cascaded activation is a common mechanism in theories of spoken word recognition (Marslen-Wilson, 1987; Norris, 1994) and in theories of language production (Caramazza, 1997; Dell, 1986). Pre-activation of meaning may cascade into pre-activation of word form, whether or not predictions are generated by the production system. The ramifications of this account will be further discussed in the General Discussion (Section 4.10).

4.1.2 Predicting meaning

Classic findings from Kutas and colleagues have shown that anomalous words lead to increased N400 ERPs in comparison to plausible words in the same sentence contexts (Kutas & Hillyard, 1980). This N400 effect has been found to be smaller for words that are semantically related to the plausible word, and this N400 reduction is greater for high-cloze

---

[15] Pickering and Garrod (2013) in fact proposed a different type of production-based prediction that they called *prediction-by-simulation*. To summarize briefly, there is good evidence that people predict their own utterances using so-called forward models, based on associations between their intention (e.g., to talk about a kite) and aspects of the word they would use to describe that intention (e.g., the phoneme /k/). These forward models are ready before the utterance itself (thus allowing self-monitoring), and there is no reason that predictions of meaning need be ready before predictions of form. They can then use such forward models to predict during comprehension, again before the speaker produces the utterance. This form of prediction makes no claim that prediction should depend on time or resources, and in particular does not assume that prediction of form is less likely to occur than prediction of meaning. However, Pickering and Garrod's model is compatible with the occurrence of both prediction-by-simulation and prediction-with-implementation.

sentences than for medium-cloze sentences (Federmeier & Kutas, 1999; Kutas & Hillyard, 1984; Thornhill & Van Petten, 2012).

Federmeier and Kutas (1999) found that implausible words that were in the same semantic category as predictable words elicited a greater N400 reduction in high-cloze sentences than in medium-cloze sentences relative to implausible words that were not in the same semantic category. However, those within-category words were rated as less plausible in high- than in medium-cloze sentences. Crucially then, because the N400 reduction did not pattern with the plausibility pre-test data, Federmeier and Kutas could rule out an integration account in which the observed N400 reductions reflected within-category words being more plausible sentence continuations (and therefore easier to integrate) than between-category words. They concluded that, prior to the onsets of the target words, participants had activated semantic features of the expected sentence continuations. This in turn implied activation of some of the within-category words' semantic features, resulting in facilitation of the within-category words relative to those which didn't share a semantic category, as indexed by N400 reduction.

Federmeier and Kutas' (1999) findings are indeed consistent with an account of prediction that operates via pre-activation of semantic category features. However, a remaining inconsistency comes from the fact that a pre-activation account also strongly suggests that high-cloze target words themselves should show a reduced N400 effect compared to medium-cloze target words. But, surprisingly, Federmeier and Kutas (1999) did not find this basic effect of cloze probability.

A possible alternative explanation of Federmeier and Kutas' data is that the N400 reduction for implausible within-category words (*pines*) occurred because within-category words also receive substantial lexical priming from the individual words in a sentence context that is highly supportive of the target word (*palms*). If, in sentence contexts that are high-cloze for the target word, within-category words also have a strong semantic relationship with the words in the sentence context (compared to the same within-category words in low-cloze contexts), a larger N400 reduction for within-category words would be observed. But it would reflect facilitation of the within-category word through lexical priming[16] rather than through semantic pre-activation of the target word.

Importantly, Metusalem et al. (2012) found that semantic pre-activation is not limited to semantic category features of expected words, while ruling out an explanation in terms of lexical priming. They conducted a pre-test to establish words that were commonly associated with the discourses presented in their experiment. They reported an N400 reduction for anomalous words that were associated with the events described in the discourses, but not to the expected target word itself (e.g., *jacket* is related to the event of building a snowman in the winter, but not to the concept 'snowman' itself), relative to event-unrelated anomalous words (*towel*). Because they controlled for degree of semantic association, this N400 difference between event-related and event-unrelated anomalous words could not be attributed to lexical priming of event-related words by context words. Though Metusalem et al. (2012) only used high-cloze sentences, and did not examine effects of cloze probability, their findings suggest that pre-activation of general or event-based knowledge relevant to the described event forms a basis of prediction (see also Nieuwland, 2015). In conclusion, people do appear to pre-activate semantic features of highly predictable upcoming words and semantic information that is more broadly relevant to the discourse context.

---

[16] Within-category words were never lexical associates of the target words, but association norms include only strongly semantically related items. Importantly, lexical priming can also occur from semantically related or lexically co-occurring words which are non-associated (e.g., Hare et al., 2009).

4.1.3 Predicting form

In contrast to pre-activation of meaning, it seems that pre-activation of form (i.e., what upcoming words will sound or look like) requires the prediction of a specific lexical item (as context words would not usually be related in form to a predictable word). A lexical prediction might pre-activate particular form features, which could in turn facilitate the processing of form-related words. In this study, we do not distinguish prediction of sound (phonological form) and shape (orthographic form).

The evidence for form pre-activation is quite complex. As discussed in Section 1.1.2, some studies found that plausible but prediction-inconsistent articles elicited larger N400s relative to plausible and prediction-consistent articles (DeLong et al., 2012, 2005). DeLong et al. argued that participants pre-activated form representations (e.g., an initial consonant) of the upcoming noun before the appearance of the noun.

However, another possibility is that participants predicted the articles themselves, rather than predicting the noun and using the phonology of the noun to compute the article.[17] Recent evidence indicates that frequently occurring word sequences are comprehended more quickly than would be expected on the basis of their individual frequencies (Arnon & Snider, 2010; Tremblay, Derwing, Libben, & Westbury, 2011), suggesting that common sequences are represented (alongside individual words), in the mental lexicon. People may thus store article-noun sequences and use context to predict such sequences, rather than word form.

Another set of evidence for form-related N400 reductions come from Laszlo and Federmeier (2009) and Kim and Lai (2012), which are also discussed in Section 1.1.2. They showed that words and pseudowords that shared orthographic information with predictable

---

[17] Importantly, the N400 effect for unexpected articles in Delong et al. (2005) was correlated with the predictability of the article, rather than the predictability of the subsequent noun. Hence, effects at the articles can be observed regardless of the cloze value of the noun.

words elicited reduced or no N400 effects, relative to orthographically unrelated control words or pseudowords. The effects were interpreted as evidence for pre-activation of orthographic information. However, both studies used high proportions of pseudowords and nonwords (54% in Laszlo & Federmeier, 2009; 75% in Kim & Lai, 2012). It remains unknown whether the reported effects generalize to settings involving only lexical items. Second, participants might learn to predict the occurrence of a nonword or a real word, depending entirely on the proportion of each type of stimulus and the nature of the design. What stimuli participants may learn to track can then affect the component that is elicited (Holcomb, 1988). This concern is particularly important for experiments concerned with prediction.

Another important concern is that these studies required participants to make a judgement or perform a task in addition to natural reading comprehension. For example, participants in Laszlo and Federmeier (2009) judged whether a stimulus was a "normal English sentence," and most non-predictable conditions were correctly answered with a "no" response (75% of the responses). Critical words can elicit positive deflections, such as the P300, as a function of extended task-related processing of these words (e.g., Newman, Connolly, Service, & Mcivor, 2003). These components could obscure N400 modulations via component overlap due to summation of positive and negative potentials at the scalp. Given that whether words were orthographically related or not was task-relevant (task difficulty increases if a critical word looks like the target word), the reported effects might reflect differences in task-related ERPs rather than, or in addition to, N400 differences.

Studies using non-words that are closely related in form to predictable words have also revealed post-N400 positive ERP effects (Late Positive Component or LPC effects) that seem to indicate that comprehenders consider the form of the predictable word. LPC effects were found for pseudohomophones that were orthographically similar to highly predictable words, but there was no LPC effect for pseudohomophones that were orthographically similar to unpredictable words (Newman & Connolly, 2004; Vissers et al., 2006). Similar LPC effects for pseudowords that were orthographically similar to the predicted words were reported by Laszlo and Federmeier (2009) and Kim and Lai (2012), which was interpreted as a detection of a conflict between predicted and actually encountered words.

It is not yet clear whether the N400 and LPC effects previously interpreted as being due to form overlap would occur in the absence of a task that requires explicit evaluation of critical words, using a design with only real words.

4.1.4 The current study

We examined pre-activation of form and meaning as participants read for comprehension. To examine the effects of prediction, we assessed the N400 effects for high-cloze items and medium-cloze items (cf. Federmeier & Kutas, 1999). Participants read constraining sentences with the predictable word (*predictable condition*), an anomalous word sharing form features (i.e., having phonological/orthographic overlap) with the predictable word (*form condition*), an anomalous word that was semantically related to the predictable word (*semantic condition*), or an anomalous unrelated word (*unrelated condition*; see Figure 9). The current study minimized potential artefactual effects by using real words only and employing no task related to critical words, while controlling for relevant variables. Moreover, we controlled the form-similarity of the semantically related words and the semantic relatedness of form-related words, to show that any demonstration of pre-activation of form cannot be wrongly ascribed to pre-activation of meaning, and vice versa.

We investigated whether there were N400 reductions for semantically related words and for form-related words, relative to the unrelated baseline. We expected the N400 reduction for both types of related words to be larger when predictable words were more strongly predicted (high cloze) than less strongly predicted (medium cloze). We hypothesized that even if pre-activation of form features was weak or absent, form-related words might impact comprehension if people detect the conflict between actual input and predictable words. If so, we expected to find a post-N400 LPC effect, which should be strongest in highest cloze sentences, because the conflict should be greater when expectation for a specific word is stronger.

We conducted two experiments that differed in presentation rate. In Experiment 5, the stimulus onset asynchrony (SOA) was 500 ms; in Experiment 6, it was increased to 700 ms. Assuming that comprehenders make predictions by going through some of the stages that they use to produce utterances, then they might be unable to predict both meaning and form in Experiment 5, in part because of the relatively short time-lag and in part because comprehension would be rendered difficult by having to integrate all the words in the prior context. In contrast, we hypothesized that they would be able to predict both meaning and form in Experiment 6, given the longer SOA.

4.2 Experiment 5

4.3 Methods

4.3.1 Participants

Twenty-four English monolinguals (6 males and 18 females, age $M = 21.4$ years, $SD = 2.8$) took part in the experiment, having given informed consent. All participants were right-handed and free from neurological or language disorders.

4.3.2 Stimuli and experimental design

We constructed 160 items (from a candidate set of 200 items) that consisted of a context that strongly predicted a specific word (e.g., *The student is going to the library to borrow a…*), followed by a critical word and a sentence-final word. In the *predictable* condition, the critical word was the predictable word (e.g., *book*). In the *form* condition, the critical word was phonologically and orthographically related to the predictable word (e.g., *hook*). The overlap could occur at word-onset (*card-cart*, 15% of the items), word-offset (*luck-duck*, 53.8%), or both (*age-ace*, 23.8%), or involved single-letter addition (*air-hair*, 5.6%), or single-letter deletion (*cold-old*, 1.9%). In the *semantic* condition, the critical word was semantically related to the predictable word (e.g., *page*). In the *unrelated* condition, the critical word was not related in terms of form or meaning to the predictable word (e.g., *sofa*).

We validated our items in four ways. In a cloze probability pre-test, 36 further participants completed each of the context fragments from our candidate set (e.g., *The student is going to the library to borrow a* -) with the first word that came to mind. We excluded items if the predictable word was not the most frequent completion or if it had a cloze probability of less than 30%. Selected items had a mean cloze value of 80% (range 31-100%; see Figure 9 for example items; the full set of items with cloze values and plausibility ratings are in Appendix 8.3). We then added an additional word to each item so that ERP responses to the critical words would not be affected by sentence wrap-up.

A further 48 native English speakers judged plausibility of the sentences excluding the post-target word on a scale from 1 (completely implausible) to 5 (completely plausible) for 173 candidate items, together with 64 further sentences that were designed to be plausible. The candidate items were placed in four lists, each containing one version of each item and 43 or 44 sentences from each condition. We excluded items in which the predictable condition had a mean plausibility rating below 3.5 or any other condition had a mean plausibility rating over 3. For the remaining 160 items, the semantic condition was more plausible than the form condition (*Mean Difference* = .16, *SD* = .56), $t(159) = 3.66$, $p < .001$, or the unrelated condition (*Mean Difference* = .15, *SD* = .52), $t(159) = 3.75$, $p < .001$, whereas the form and unrelated conditions were equally implausible.

Lexical characteristics of the critical words in each condition are shown in Table 1. We evaluated form similarity by computing the Levenshtein distance from the predictable word (the minimum number of single-letter edits including addition, deletion, and substitution needed to transform one word into the other). In the semantic condition, the critical word was semantically related to the predictable word. We assessed this similarity by pairwise Latent Semantic Analysis (LSA) (Landauer & Dumais, 1997). In the unrelated condition, the critical word was related neither in form nor in meaning to the predictable word, relative to the form or semantic conditions. Among the non-predictable conditions, the form condition had a smaller Levenshtein distance to predictable words than the semantic condition, *t*(159) = -29.4, *p* < .001, and the unrelated condition, *t*(159) = -37.1, *p* < .001. The semantic condition had a larger Levenshtein distance than the unrelated condition, *t*(159) = 5.2, *p* < .001. The direction of this difference means that any effect of semantic similarity could not in fact be due to form similarity. The semantic condition had a higher LSA than the form condition, *t*(158)[18] = 26.1, *p* < .001, and the unrelated condition, *t*(159) = 27.2, *p* < .001. The form condition had a higher LSA than the unrelated condition, *t*(158) = 2.4, *p* < .05. As will become clear in our Results section (Section 4.4), the difference in Levenshtein distance between the semantic and the unrelated condition, and the LSA difference between the form condition and the unrelated condition, cannot explain our results.

Table 1. Lexical characteristics of critical words used in Experiments 5 – 8 (*SD*s in parentheses).

Condition	**Length**	**Frequency**	**Phonological density**	**LSA**	**Levenshtein Distance**	**Concreteness**	**Familiarity**
Predictable	4.4 (1.1)	4.8 (0.5)	8.5 (6.0)	-	-	510 (108)	566 (40)
Form	4.5 (1.1)	4.3 (0.7)	9.0 (6.1)	0.07 (0.1)	1.2 (0.5)	497 (105)	526 (52)
Semantic	5.3 (1.7)	4.3 (0.7)	5.2 (5.5)	0.46 (0.2)	4.8 (1.6)	494 (101)	546 (50)
Unrelated	4.4 (1.0)	4.4 (0.6)	7.0 (6.43)	0.06 (0.1)	4.1 (1.0)	502 (108)	539 (45)

The word frequency was taken from Subtlex (http://zipf.ugent.be/open-lexicons/interfaces/subtlex-uk/). Phonological density represents orthographic neighbourhood size from MCWord (http://www.neuro.mcw.edu/mcword/). LSA shows the results of pair-wise comparison of semantic similarity scores between predictable words and words in each condition (http://lsa.colorado.edu/). Distance represents Levenshtein distance from corresponding predictable words. Concreteness and familiarity ratings are taken from MRC Psycholinguistic Database (http://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm)

---

[18] One item in the form condition did not yield an LSA value, which is why the comparisons involving this condition have 158 degrees of freedom.

4.3.3 Procedure

The 160 sentences were divided into four counterbalanced lists so that each list contained only one condition per sentence, but that across the four lists each condition for each sentence occurred equally often. They were combined with 64 additional plausible filler sentences and presented in the same randomized order for every participant with the constraint that no more than three items from the same condition appeared consecutively. Each participant thus saw a total of 104 plausible and 120 implausible sentences.

Participants silently read sentences from a computer display, presented word by word at a regular pace (300 ms word duration, 200 ms inter-word interval; sentence-final words had a 600 ms duration). A fixation-cross followed each sentence, at which point participants could start the next sentence by a button-press. Yes-No comprehension questions appeared on 25% of the trials (mean accuracy across participants, *M* = 96.1%, *SD* = 3.6, range = 86.0-100%, 6.6% of the responses are excluded due to time outs). The experiment took approximately 40 minutes.

4.3.4 Electroencephalogram (EEG) recording and data processing

The electroencephalogram (EEG) was recorded at a sample rate of 512 Hz and with 24-bit AD conversion using the Biosemi ActiveTwo system (BioSemi BV, Amsterdam, The Netherlands). This system’s hardware is completely DC coupled and applies digital low pass filtering through its ADC's decimation filter (the hardware bandwidth limit), which has a 5th order sinc response with a -3 dB point at 1/5th of the sample rate (i.e., approximating a low-pass filter at 100 Hz). Data was recorded from 64 EEG, 4 EOG, and 2 mastoid electrodes using the standard 10/20 system (for details, see Nieuwland, 2014). Offline, the EEG was re-referenced to the mastoid average and filtered further (0.019–20 Hz plus 50 Hz Notch filter). Data was segmented into 1200 ms epochs (-200-1000 ms relative to critical word onset), corrected for eye-movements using the Gratton and Coles regression procedure as implemented in BrainVision Analyzer (Brain Products ©), baseline-corrected to -100-0 ms, automatically screened for movement- or electrode-artefacts (minimal/maximal allowed amplitude = -75/75 μV), and averaged per condition per participant. The mean number of artefact-free trials per condition was 37, with no difference across conditions.

4.3.5 Statistical analysis

Mean amplitude was computed per condition at 16 EEG electrodes (F1/F3/FC1/FC3/CP1/CP3/P1/P3 plus right-hemisphere equivalents), in the N400 time window (350-450 ms) and the LPC time window (600-1000 ms). Effects of condition and scalp distributions effects were tested with a 4 (Condition: Predictable, Form, Semantic, Unrelated) by 2 (Hemisphere: left, right) by 2 (Anteriority: Frontal-Central, Central-Parietal) repeated-measures ANOVA. When appropriate, Greenhouse–Geisser corrections and corrected *F*-values are reported. Only statistical results with $p < .1$ are reported. Additionally, we divided the items into high- and medium-cloze probability sets to test an effect of cloze probability with a condition by cloze ANOVA, focusing on relevant conditions.

4.4 Results

Visual inspection of the data indicates that all implausible conditions elicited larger N400s than the predictable condition (see Figure 9). These N400 effects were widely distributed and visible at most channels. Figures showing all channels are in Appendix 8.4. The form condition also showed a post-N400 enhanced positive deflection compared to the other conditions, starting from about 600 ms and lasting until about 1000 ms, which was most prominent at posterior channels.

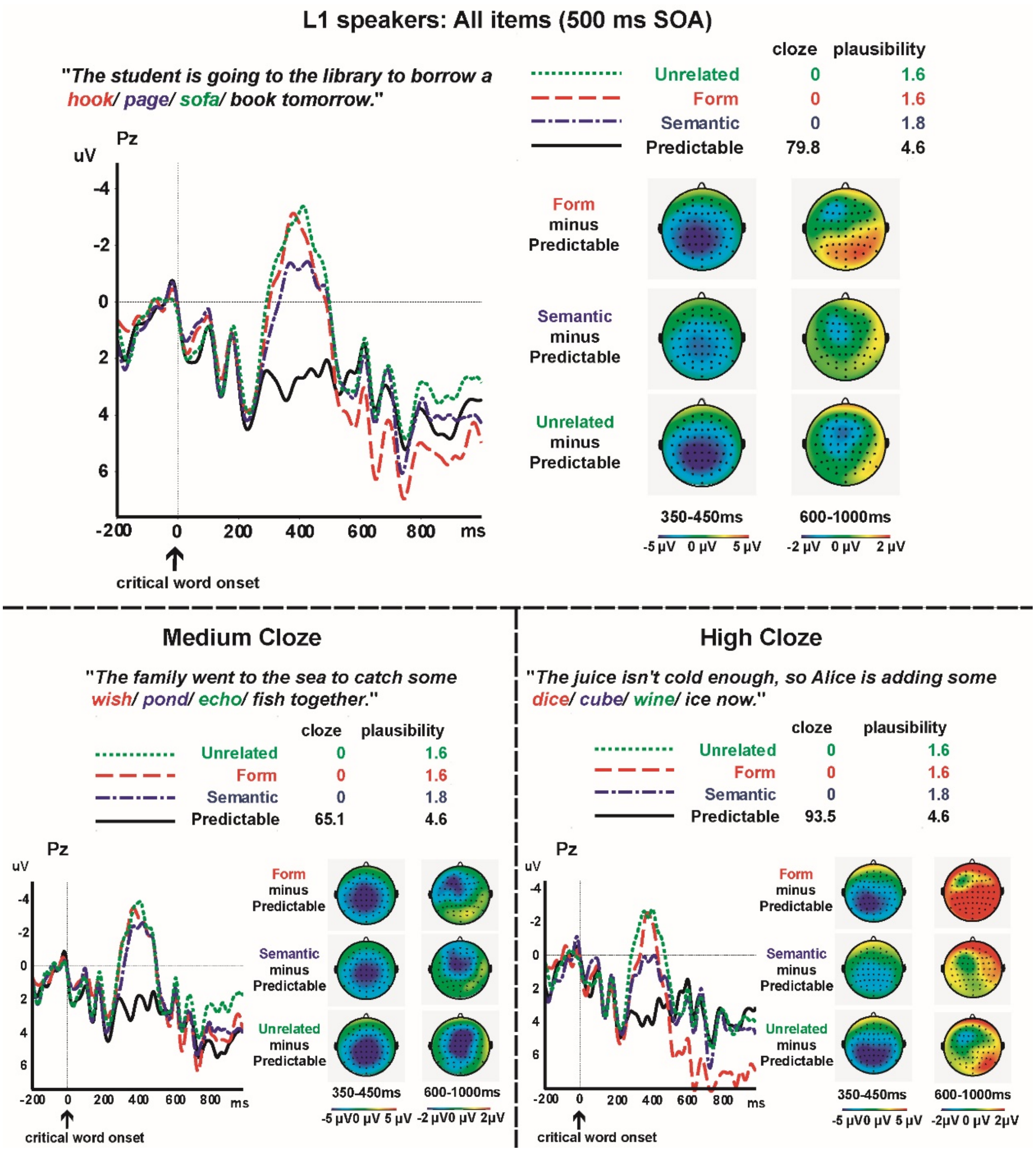


Figure 9. Results from Experiment 5 (500 ms SOA). ERPs elicited by each condition at Pz in across all items (top panel), in medium-cloze items (left lower panel) and in high-cloze items (right lower panel). Scalp distributions of the ERP effects (implausible condition minus predictable condition) in the N400 time window and LPC time window are shown on the right in each panel.

4.4.1 The N400 time window

N400 analysis revealed a significant effect of condition, $F(3, 69) = 24.6$, $MSE = 14.5$, $p < .001$, and a significant interaction of condition by anteriority, $F(2.1, 48.1) = 11.3$, $MSE = 1.9$, $p < .001$, indicating that effects of condition were more robust at posterior channels, *F*(3, 69) = 29.6, *MSE* = 4.4, *p* < .001, than anterior channels, *F*(3, 69) = 16.0, *MSE* = 3.6, *p* < .001. For further analysis, we therefore performed pairwise comparisons between conditions at the posterior channels where N400 modulations were largest (see Table 2, top-right cells). All three non-predictable conditions elicited larger (more negative) N400s than the predictable condition. Critically, the semantic condition elicited reduced N400s compared to the unrelated conditions.

Table 2. Pairwise t-test results for Experiment 5 (500 ms SOA) on mean ERP amplitude per condition at posterior channels in the N400 350-450 ms time window (top-right half) and in the LPC 600-1000 time window (bottom-left half).

Time window		N400			
Condition		Predictable	Form	Semantic	Unrelated
LPC	Predictable		-4.8 (3.2) -10.2***	-3.7 (3.1) -8.4***	-5.0 (3.8) -9.2***
	Form	-1.2 (2.9) -3.0**		1.1 (2.2) 3.3**	-0.2 (3.0) -0.5
	Semantic	0.006 (2.9) 0.01	1.2 (3.3) 2.6**		-1.3 (2.4) -3.7***
	Unrelated	0.4 (3.4) 0.9	1.6 (2.6) 4.4***	0.4 (2.3) 1.2	

The values in each cell correspond to the mean voltage difference (row condition values were subtracted from column conditions); *SD* (in parentheses); *t*-value (*df* = 47); significance level, represented as * < .1, ** < .05, *** < .001

4.4.2 The late positivity time window

This analysis revealed a significant condition by anteriority interaction, *F*(3, 69) = 9.9, *MSE* = 1.1, *p* < .001, a marginally significant interaction of condition by hemisphere, *F*(2.2, 50.8) = 2.6, *MSE* = 0.8, *p* = .08, and a marginally significant three-way interaction of condition by anteriority by hemisphere, *F*(3, 69) = 2.3, *MSE* = 0.3, *p* = .09.

Since previously reported LPC has been largest at posterior channels, we followed up on the condition by anteriority interaction with one-way ANOVAs at anterior and posterior channels separately. The effect of condition was marginally significant at posterior channels only, *F*(2.4, 55.6) = 3.0, *MSE* = 5.1, *p* = .05, and was not significant at anterior channels, *F* < 1.6. Table 2 lists the follow-up pairwise comparisons performed at posterior channels (bottom-left cells). Form-related words elicited an enhanced positivity compared to all other conditions.

4.4.3 Effects of cloze probability

We tested whether the observed N400 and LPC modulations were dependent on the cloze probability of predictable words. To do this, we compared effects in high-cloze and medium-cloze items. We used a median split (*Mdn* = 86) to form a high-cloze subset (83 items, cloze *M* = 93.5, *SD* = 4.7) and a medium-cloze subset (77 items, cloze *M* = 65.1, *SD* = 15.3). Importantly, the two subsets did not differ in plausibility ratings, in frequency, in word length, in Levenshtein distance, in word LSA nor in context LSA, all *F*s < 1.1 (for details, see Table 3). ERP waveforms for high-cloze and medium-cloze items separately are shown in Figure 9.

Table 3. Lexical characteristics of critical words in high- and medium cloze item sets. For these variables, the only robust difference between high and medium cloze sets (pair-wise *t*-tests) was found for unrelated words, which had higher LSA values in the medium cloze set than in the high cloze set, $t(142.6) = -2.3$, $p < .05$.

Condition	**Cloze set**	**Length**	**Frequency**	**Phonological density**	**LSA**	**Distance**	**Context LSA**	**Cloze**	**Plausibility**
Predictable	High	4.37	4.83	8.57			0.22	93.50	4.58
	Medium	4.48	4.79	8.40			0.20	65.12	4.59
Form	High	4.51	4.28	8.65	0.07	1.30	0.10	0	1.58
	Medium	4.48	4.28	9.38	0.08	1.18	0.09	0	1.65
Semantic	High	5.23	4.29	5.08	0.48	4.70	0.17	0	1.76
	Medium	5.43	4.40	5.34	0.44	4.88	0.16	0	1.78
Unrelated	High	4.36	4.34	6.90	0.05	4.07	0.09	0	1.61
	Medium	4.51	4.40	7.09	0.07	4.21	0.10	0	1.63

We used only the posterior channels, where, consistent with the findings in previous literature (e.g., Kim & Lai, 2012; Laszlo & Federmeier, 2009), N400 and LPC effects had been maximal. For the semantic prediction reduced-N400 effect, we tested the effect of cloze value on the crucial difference between the semantic condition and the unrelated condition. A 2-way condition by cloze ANOVA revealed a significant effect of condition, $F(1, 23) = 8.4$, $MSE = 6.6$, and of cloze, $F(1, 23) = 9.0$, $MSE = 7.1$, *p*s < .05, and a marginally significant interaction of condition by cloze, $F(1, 23) = 4.0$, $MSE = 3.5$, $p = .06$. Follow-up *t*-

tests comparing the semantic and the unrelated conditions revealed that the N400 reduction was robust in the high-cloze subset (*M* = 2.3 µV, *SD* = 3.3), *t*(47) = 4.9, *p* < .001, but not in the medium-cloze subset (*M* = .75 µV, *SD* = 3.2), *t*(47) = 1.6, *p* = .1. Absence of N400 reduction for the form condition (relative to the unrelated condition) was observed in the high-cloze and medium-cloze subsets alike, *p*s > .3.

For the form prediction LPC analysis, all four conditions were included to test whether the LPC effect was observed only for the form condition. A condition by cloze ANOVA revealed a significant effect of condition, *F*(3, 69) = 4.0, *MSE* = 9.2, *p* < .05, a marginally significant main effect of cloze, *F*(1, 23) = 4.2, *MSE* = 10, *p* = .05, and a significant interaction of condition by cloze, *F*(2.1, 48.9) = 4.5, *MSE* = 9.6, *p* < .05. Follow-up pairwise comparisons showed that, for high-cloze items, the form condition elicited an LPC effect compared to the predictable condition (*M* = 3.4 µV, *SD* = 4.9), *t*(47) = 4.8, *p* < .001, semantic condition (*M* = 2.8 µV, *SD* = 4.7), *t*(47) = 4.1, *p* < .001, and unrelated condition (*M* = 2.6 µV, *SD* = 5.4), *t*(47) = 3.4, *p* = .001. In contrast, for medium-cloze items, the unrelated condition showed a negative going shift relative to predictable condition, *t*(47) = -3.8, *p* < .001, form condition, *t*(47) = -2.8, *p* < .05, and semantic condition, *t*(47) = -2.1, *p* < .05.

4.5 Discussion

We investigated whether readers pre-activate semantic features and (phonological/orthographic) form features in a high-cloze sentence context where strong lexical predictions can be made. Critical words that were semantically related to high-cloze target words elicited a diminished N400 effect compared to unrelated words. Form-related words showed no N400 reduction, but elicited a post-N400 enhanced positivity (posterior LPC) relative to other conditions. The N400 result suggests that participants pre-activated semantic but not form information, whereas the LPC effect suggests that participants detected the form similarity with predictable words. Both effects were robust only in high-cloze sentences, even though the medium-cloze and high-cloze sentences were matched on plausibility and other relevant variables. We consider this strong evidence that both effects arise from prediction of target word meaning. However, one remaining question from Experiment 5 is whether form pre-activation never occurs, or whether it occurs under some experimental conditions but not others.

4.6 Experiment 6

Experiment 5 did not generate evidence of form pre-activation. It is possible that readers simply do not pre-activate the form of predictable words. However, it is also possible that form is pre-activated only when time or resources allow, as production-based prediction accounts predict. Do readers pre-activate form when they have more time to generate predictions, for example, when sentences are presented at a slower rate? Experiment 6 tested this possibility by increasing the SOA between words from 500 ms to 700 ms. A 700 ms SOA has also been used as a long-SOA condition in an investigation of presentation rate on prediction during sentence processing (Dambacher et al., 2012) and in word-priming studies (Hill, Ott, & Weisbrod, 2005; Luka & Van Petten, 2014).

There is some reason to believe an increased SOA might enhance pre-activation. Several previous studies have showed an effect of SOA on processing associated with prediction. However, to date, SOA manipulations appear to have been chiefly investigated with respect to semantic prediction. For example, SOA has a clear impact on semantic priming N400 effects. Semantic priming N400 effects suggest that people activate a set of words that are associated with the prime word, which facilitates the processing of the target word (Roland et al., 2012). Hill et al. (2005) reported a larger N400 priming effect for a longer SOA (700 ms) than a shorter SOA (150 ms), and suggested that the longer SOA led to deeper semantic processing. Luka and Van Petten (2014) presented participants pairs of words that were strongly, moderately, or weakly associated with each other, either simultaneously or with a 700 ms SOA. They found that stronger semantic association was associated with smaller N400s. Critically, this N400 effect was delayed in the simultaneous presentation, especially for strongly associated word pairs, suggesting that more time enhances pre-activation via semantic association (i.e., priming). These studies investigated the effect of SOA only on the word-to-word semantic relatedness effect, and it is unclear whether similar effects occur during sentence reading when predictions for a specific word can be made based on more constraining contextual information.

Studies that are discussed in Section 1.3.2. in this thesis (Dambacher et al., 2012; Kutas, 1993; Wlotko & Federmeier, 2015) investigated effects of SOA on prediction-related N400 effects. They suggest that manipulations of timing may have some effects on pre-activation,[19] but there have been no investigations of the effects of timing on form pre-activation. In our Experiment 5, participants pre-activated meaning but not form, as indicated by a reduced N400 for semantically related words relative to unrelated words, but not a reduced N400 for form-related words relative to unrelated words. Experiment 6 used a longer SOA and examined whether form pre-activation occurs when there is more time available during sentence comprehension to generate predictions. If form features are pre-activated when people read at this slower presentation rate, we would expect to see reduced N400s for form-related words relative to unrelated words.

4.7 Methods

4.7.1 Participants

Twenty-four English monolinguals (6 males and 18 females, age *M* = 20.3 years, *SD* = 2.9) took part in the experiment, having given informed consent. The participants were

---

[19] An important caveat to an interpretation of SOA-based peak latency differences, however, is that at short SOAs the N1-P2 ERP complex elicited by the subsequent word occurs in the N400 time range of the critical word. Through component overlap, ERPs elicited by the subsequent word at short SOAs (but not at a 500 ms or slower SOA) can thus ‘cut short’ the N400 component before it reaches its full peak.

from the same population as in Experiment 5, but had not participated in Experiment 5. All participants were right-handed and free from neurological or language disorders.

4.7.2 Stimuli and experimental design

The stimuli and experimental design were identical to those used in Experiment 5.

4.7.3 Procedure

The procedure was identical to Experiment 5, except that the SOA was changed to 700 ms (500 ms word duration, 200 ms inter-word interval; sentence-final words had an 800 ms duration). Mean accuracy for comprehension questions was 90.8% (*SD* = 3.9, range = 83.9-96.4%).

4.7.4 Electroencephalogram (EEG) recording and data processing

The data were processed in the same way as in Experiment 5. The mean number of artefact-free trials in Experiment 6 per condition was 36, with no difference across conditions.

4.7.5 Statistical analysis

The same statistical analysis was conducted as for Experiment 5.

4.8 Results

Visual inspection of the data indicates that all implausible conditions elicited larger N400s than the predictable condition, and the N400 effect was reduced for the semantic condition (see Figure 10), as had also been observed in Experiment 5. The form condition elicited the largest LPC at posterior channels among all the conditions.

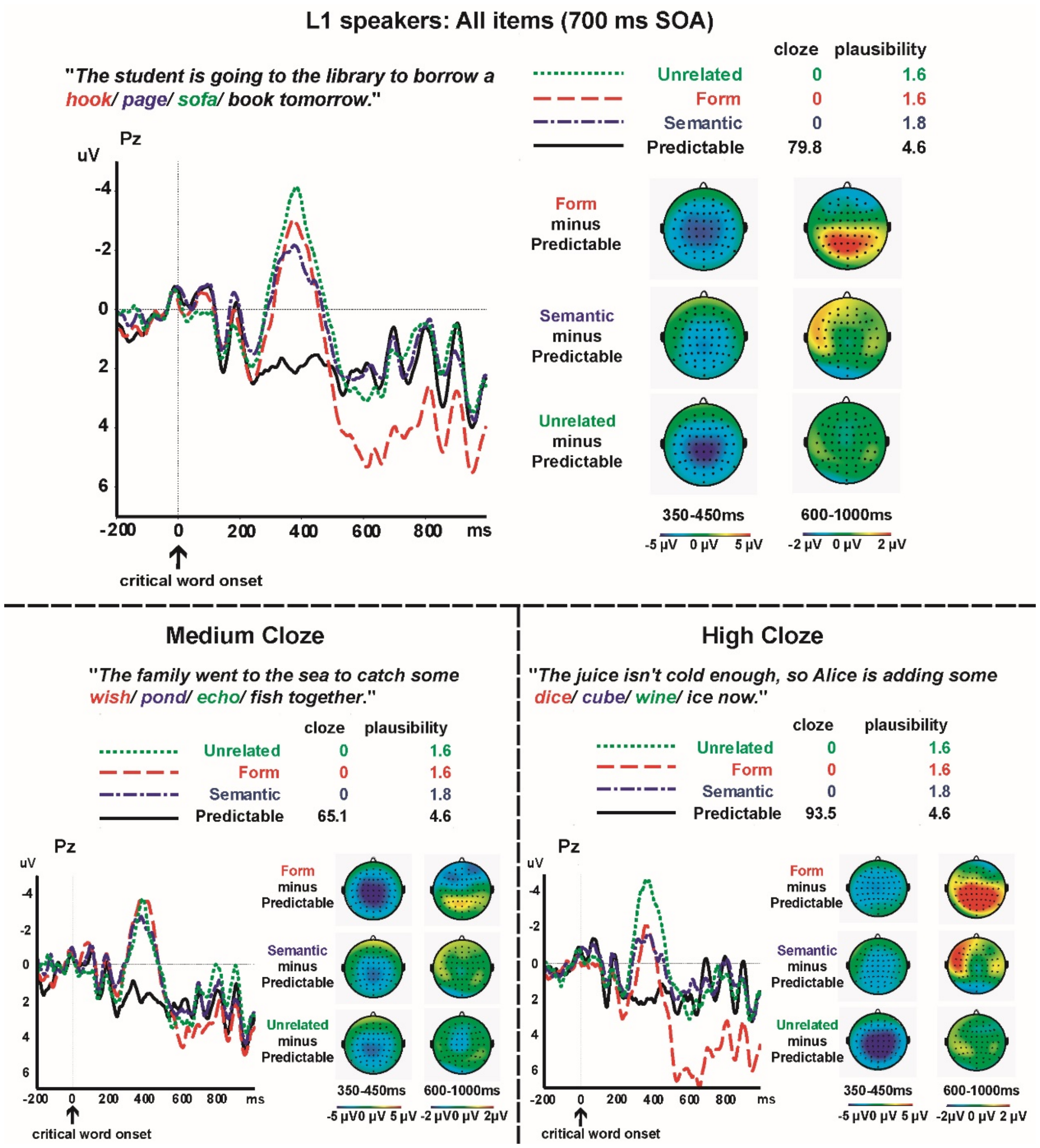


Figure 10. Results from Experiment 6 (700 ms SOA). ERPs elicited by each condition at Pz in across all items (top panel), in medium-cloze items (left lower panel) and in high-cloze items (right lower panel). Scalp distributions of the ERP effects (implausible condition minus predictable condition) in the N400 time window and LPC time window are shown on the right in each panel.

4.8.1 The N400 time window

The analysis revealed a significant effect of condition, *F*(3, 69) = 18.5, *MSE* = 17.2, *p* < .001, and a significant interaction of condition by anteriority, *F*(3, 69) = 4.3, *MSE* = 1.8, *p* < .05, which was due to a stronger effect of condition at posterior channels, *F*(3, 69) =

19.2, *MSE* = 5.0, *p* < .001, than at anterior channels, *F*(3, 69) = 15.0, *MSE* = 4.5, *p* < .001. Following the same analysis steps in Experiment 5, pairwise comparisons were performed at posterior channels. All the implausible conditions elicited larger N400s than the predictable condition, but the semantic condition elicited reduced N400s relative to the form and the unrelated conditions (see Table 4, top-right cells).

Table 4. Pairwise *t*-test results for Experiment 6 (700 ms SOA) on mean ERP amplitude per condition at posterior channels in the N400 350-450 ms time window (top-right half) and in the LPC 600-1000 time window (bottom-left half).

Time window		N400			
Condition		Predictable	Form	Semantic	Unrelated
LPC	Predictable		-4.0 (3.6) -7.6***	-3.1 (2.6) -8.4***	-4.4 (3.5) -8.9***
	Form	-1.8 (3.5) -3.6***		0.81 (2.8) 2.0**	-0.47 (3.1) -1.1
	Semantic	-0.02 (2.9) 0.06	1.8 (3.2) 3.9***		-1.3 (3.5) -2.6**
	Unrelated	-0.04 (2.5) -0.1	1.76 (2.1) 5.7***	-0.06 (2.9) -0.14	

The values in each cell correspond to the mean voltage difference (row condition values were subtracted from column conditions); *SD* (in parentheses); *t*-value (df = 47); significance level, represented as * < .1, ** < .05, *** < .001

4.8.2 The late positivity time window

The analysis revealed a significant interaction of condition by anteriority, *F*(3, 69) = 24.1, *MSE* = 1.1, *p* < .001, which was driven by the fact that the effect of condition was significant at posterior channels, *F*(2.4, 55) = 4.7, *MSE* = 5.1, *p* < .05, but not at anterior channels, *F* < 1. Similarly to Experiment 5, the pairwise comparisons at posterior channels revealed that the form condition elicited enhanced positivity relative to all the other conditions (see Table 4, bottom-left cells).

4.8.3 Effects of cloze probability

We compared high-cloze and medium-cloze items using the same median split as in Experiment 5 in order to test the effect of predictability. For the N400 window, we performed a 2-way condition by cloze ANOVA at posterior channels including the form, semantic, and unrelated conditions in order to allow investigation of pre-activation of both form and meaning. The analysis revealed a significant interaction of condition by cloze, $F(1.6, 37) = 6.0$, $MSE = 9.0$, $p < .05$, which arose from the effect of condition being significant in the high-cloze items, $F(2, 46) = 5.8$, $MSE = 9.3$, $p < .05$, but not in the medium-cloze items, $F < 1.5$. For the high-cloze items, N400s were reduced both for the form condition, ($M = 2.6$ µV, $SD = 3.6$), $t(47) = 5.0$, $p < .001$, and for the semantic condition, ($M = 2.7$ µV, $SD = 4.6$), $t(47) = 4.0$, $p < .001$, relative to the unrelated condition (see Figure 10, right lower panel).

For the LPC analysis, a 2-way condition by cloze ANOVA at posterior channels with all the conditions revealed a significant effect of condition, $F(2.4, 54) = 5.4$, $MSE = 10.3$, $p < .05$. However, neither the effect of cloze nor the interaction of condition by cloze was significant, *F*s < 1.6.

4.8.4 Between-experiment comparisons: effects of SOA

The critical difference between the results of the two experiments was the presence of the N400 modulation for form-related words in high-cloze items in Experiment 6 but not in Experiment 5. To specifically test this effect, we conducted a between-experiment comparison at posterior channels in high-cloze items, using a difference value between the form and the unrelated conditions. An independent samples *t*-test revealed a significant effect of SOA, $t(94) = -2.5$, $p < .05$. In the LPC time window, an independent samples *t*-test comparing the same 2 conditions showed no significant effect of SOA, $p = .8$.

4.9 Discussion

Experiment 6 investigated whether pre-activation of form features would occur when sentences are presented at a slower presentation rate (700 ms SOA) than in Experiment 5 (500 ms SOA). Experiment 6 partially replicated Experiment 5; semantically related words elicited reduced N400s compared to unrelated words, and form-related words elicited an enhanced later positivity relative to all the other conditions. However, unlike in Experiment

5, form-related words showed an N400 reduction for high-cloze items, and elicited an LPC effect in medium-cloze items as well as in high-cloze items. The results suggest that participants pre-activated the forms of highly predictable words, but not of moderately predictable words. The results also suggest that our participants detected conflict between form-related words and predictable words, irrespective of whether the form features were pre-activated or not.

4.10 General Discussion

Two ERP experiments examined pre-activation of form and meaning during sentence reading, and whether pre-activation of semantic and form features depends on the time constraints on reading. Participants read high-cloze sentences that were completed with the predictable word, an anomalous word that was either semantically related or form-related to the predictable word, or an unrelated word. The rate of the word-by-word presentation was 500 ms in Experiment 5 and 700 ms in Experiment 6. Anomalous words in all conditions elicited an N400 effect compared to the predictable word, but, at both SOAs, N400s for semantically related words were reduced compared to unrelated words. In contrast, form-related words elicited reduced N400s only at the 700 ms SOA, and only in the high-cloze item subset. However, form-related words elicited an enhanced post-N400, Late Positive Component (LPC) at both SOAs. This LPC effect occurred irrespective of whether form-related words elicited reduced N400 effects, and was elicited only by high-cloze items at the 500 ms SOA, but by medium-cloze items and high-cloze items alike at the 700 ms SOA. The main novel contributions of this work are that (1) both meaning and form can be pre-activated, but pre-activation of form is more influenced by time constraints than pre-activation of meaning, and (2) whether or not the form of predicted words is pre-activated, form similarity to predicted words incurs additional post-N400 processing costs, suggestive of an interpretation conflict between expected input and encountered input.

4.10.1 Pre-activation of semantic features

The results of both our experiments strongly suggest that lexical prediction entails the pre-activation of semantic features, by ruling out the effects that could be associated with an account in terms of ease of integration (see Federmeier & Kutas, 1999, for discussion). Under an integration account, the facilitation of semantically related words, reflected in the reduced N400, occurs primarily because these words are more plausible sentence continuations. Although semantically related words in the complete set of items were rated as slightly less implausible than the form/unrelated conditions (1.8 compared to 1.6, on a plausibility scale of 1 to 5), the currently observed N400 reduction for semantically related words cannot straightforwardly be explained in terms of plausibility or other factors (e.g., semantic priming from context, lexical characteristics including frequency and word length). This conclusion is based on the fact that N400 reduction was found only for high-cloze items and not for medium-cloze items, while the high-cloze and medium-cloze items were matched on plausibility and other relevant variables (see Table 4). In other words, the N400 reduction for the semantically related condition was not dependent on plausibility, but depended on the cloze probability of the predictable word. We take this as strong evidence for pre-activation of semantic information, at least in highly constraining sentences.

Our study did not find clear evidence for semantic pre-activation in the medium-cloze sentences (*mean cloze* = 65%). This result is inconsistent with other studies that found an N400 reduction for semantically related words despite relatively low cloze probabilities (Federmeier & Kutas, 1999; Thornhill & Van Petten, 2012). Thus, the semantic relatedness of the related and predictable words may have been stronger in these previous studies than in our study. Our findings suggest that prediction may not always occur, even when cloze probability is relatively high and when there is plenty of time for interpreting the previous context and lexical processing. This may be incompatible with models which regard prediction as a fundamental aspect of language processing (e.g., Dell & Chang, 2014).

However, if engagement of online prediction varies during language processing, it may be likely that factors besides cloze predictability (i.e., semantic relatedness, experimental design factors such as the inclusion of nonwords and differences in task instructions) may affect the detection of online prediction. Furthermore, it could be the case that the pre-activation of particular features during prediction is highly contextually mediated, such that the system only pre-activates those features that are relevant or strongly associated with the particular situation that unfolding evidence supports (Metusalem et al., 2012).

Semantically related words elicited identical N400 reductions at shorter and longer SOAs, suggesting that longer SOA did not lead to stronger pre-activation of semantic features. This does not necessarily mean that semantic pre-activation is an automatic process that is unaffected by time constraints. Our results are consistent with those reported by Dambacher et al. (2012), where SOAs of 490 ms and 700 ms elicited similar N400 effects of cloze probability. However, they additionally found that when sentences were presented at 280 ms SOA, the N400 effect was smaller than at the other two SOAs, and the onset latency of the N400 effect was delayed compared to the 700 ms SOA. In an earlier study, Kutas (1993) also found a delay in the onset and peak latencies of an N400 anomaly effect when using 100 ms SOA, which was faster than a normal reading speed. Such findings suggest that N400 effects for unexpected or semantically anomalous words can be affected by time constraints (i.e., word presentation rate), and appear to show a delay at relatively short SOAs. The 500 and 700 ms SOAs in the current study may have been too long to generate such patterns.

4.10.2 Pre-activation of form features

While our results show that form features can be pre-activated, form pre-activation depended on the time that was available to generate predictions during reading. At a 500 ms SOA, there was no sign of pre-activation of form features, even in the most constraining sentences (cloze value 94%) at a relatively high level of form-similarity (as reflected by a relatively low Levenshtein distance of 1.3). This result appears inconsistent with previous studies by Laszlo and Federmeier (2009) and Kim and Lai (2012), who found facilitation effects for pseudowords that were orthographically similar to predictable words using 500 ms SOA and 550 ms SOA, respectively. This inconsistency could indicate that readers process real words that resemble predictable words differently from pseudowords or non-words that resemble predictable words.

However, the discrepancy between our findings and those of Lazlo and Federmeier (2009) may also have to do with the different task instructions. Participants in the present study answered comprehension questions after some trials, whereas participants in Lazlo and Federmeier (2009) were asked to judge each sentence on whether it was a "normal English sentence". This explicit judgment task may have drawn extra attention to the included nonwords and increased the task-relevance of form-related non-words. This would have been exacerbated by the fact that the form-related targets were more similar to the correct words (i.e., had lower Levenshtein distances), than those in our study. Such issues complicate a direct comparison of our findings with those of previous studies.

In our study, form-related words showed facilitation effects at the 700 ms SOA, reflected in an N400 reduction, but this effect was limited to high-cloze sentences. It thus appears that very high predictability is critical for the pre-activation of form features. This explanation fits with related studies: All the reviewed studies that found an N400 modulation for words sharing orthographic or phonological features with predictable words used critical sentences with cloze probability of about 90% (Kim & Lai, 2012; Laszlo & Federmeier, 2009; Vissers et al., 2006). Importantly, Experiment 6 replicated the previously reported N400 reductions using real words only, and without a secondary task that required explicit judgments about the critical words. Moreover, the high-cloze items and medium-cloze items did not differ in form similarity (see Table 4, Levenshtein distance), plausibility, semantic relatedness to preceding contexts or other lexical characteristics that might explain the N400 effect difference. Therefore, we interpret the predictability-dependent facilitation effect for the form-related words as reflecting pre-activation of form features as a consequence of prediction of a specific word.

A comparison of the Experiments 5 and 6 suggests that pre-activation of form features is more dependent on time constraints than pre-activation of semantic features. This finding suggests that pre-activation of semantic features is more likely to occur compared to pre-activation of form features. Strictly speaking, we cannot rule out that the form-prediction SOA effects occur because form-predictions take slightly longer to develop from the presentation of the pre-critical word. In the latter case, we expect pre-activation of form not to be a function of general SOA but only of the time between the critical word and the pre-critical word. However, rather than making such a strong claim about the absolute time course of the unfolding form-prediction, we think that slower SOAs might in principle benefit all aspects of prediction. After all, people are more likely to finish the sentences of someone who speaks slowly or hesitantly than of someone who speaks fast and fluently (see Gambi & Pickering, 2011).

4.10.3 Pre-activation pattern and production-based prediction accounts

Our findings are compatible with an account in which comprehenders use the production system to make predictions during comprehension (Pickering & Garrod, 2007, 2013; see also Federmeier, 2007). According to this proposal, comprehenders covertly imitate what they are hearing, so that they generate a production-based representation. They then engage some of the mechanisms of language production to predict upcoming words (indicating, roughly, what they would themselves say next at that point). Normally, comprehenders do not have time or resources to construct a full “implemented” representation of what they would say next (and instead construct a forward model, as discussed in Pickering & Garrod, 2013). But when time and resources allow, they run through the stages involved in language production, which involve semantic representations followed by form representations (e.g., Levelt et al., 1999). We therefore propose that comprehenders in Experiment 5 had the time to construct meaning (thus leading to a reduced N400 in the semantic condition) but not form (thus leading to no N400 reduction in the form condition). In contrast, comprehenders in Experiment 6 had the time to construct both meaning and form (thus leading to N400 reductions in both conditions).[20]

If similar effects of meaning and form pre-activation had been obtained at both SOAs, it would have suggested that participants pre-activated a specific lexical item (i.e., lemma) first, from where the activation spread across semantically and form-related lemmas. If this were the case, the pre-activation pattern would have been incompatible with a prediction-with-implementation account.

A caveat to this claim, though, is that we did not have a condition where the critical word was related to the predictable word both in form and meaning. (Because of the very limited number of such lexical pairs, it is unclear whether enough items for ERP signal-to-noise requirements could be constructed.) Even with such a condition, it would be impossible to compare the strength of semantic relatedness and form relatedness, as they are not quantified in the same way. Hence, we cannot ensure that our semantically related words and form related words were equally strongly ‘related’ to the predictable words.

Furthermore, our observed pattern of results is also consistent with a comprehension system in which activation cascades from the semantic to the form level, regardless of engagement of the production system during prediction. It is possible that participants first pre-activated semantic information, and this activation cascaded to form information, purely within the comprehension system. As this suggests that semantic pre-activation occurs prior

---

[20] An interesting point is that the form-related LPC did occur in Experiment 5. It may be that some form-related pre-activation did occur in Experiment 5, but it was not ready at the point at which the N400 was elicited, or there was a need for concurrent activation from the form-related target word.

to form pre-activation, the effect of SOA in our study can be explained by the assumption that the SOA was slow enough for the cascading to the form level to occur in Experiment 6 but not in Experiment 5. Evidence for such cascaded lexical activation has been found in comprehension (Apfelbaum, Blumstein, & McMurray, 2011; Huettig & McQueen, 2007) as well as in production (Morsella & Miozzo, 2002). Although well-established models of language comprehension and language prediction do not yet clearly formalize the notion of cascaded processing, one might reasonably assume that a cascading architecture in comprehension could work in the following basic way: Activation cascades from word form level to semantic level. In contrast, production models that entail cascading (cf. Levelt, 1999) posit that cascading occurs from semantic level to word form level. Minimally, our results clearly support a cascaded processing architecture, whatever the nature of the representations that are being activated. Given that our data follow the cascaded pattern assumed in production models (i.e., semantic followed by form), we find the directionality of activation as most consistent with the predictions of a production-based prediction account. But our findings do not offer conclusive evidence to the point of exclusion of alternative accounts, and are consistent with both production-based prediction and cascaded lexical activation accounts.

4.10.4 Monitoring and reanalysis processing for form-related words

In both experiments, the form-related condition elicited a post-N400, posterior positive deflection (LPC) that depended on the cloze probability of predictable words. Critically, this LPC effect occurred only for the form condition, in comparison to the predictable words but also to the other two implausible conditions. Therefore, this effect cannot be explained in terms of implausibility under high-constraint conditions, as proposed by Van Petten and Luka (2012) in a review of post-N400 LPCs (see also DeLong, Quante, & Kutas, 2014). We can identify three possible accounts for this effect. According to a monitoring account, form similarity increases monitoring processes and triggers a general reanalysis to check for processing errors (Van de Meerendonk, Kolk, Chwilla, & Vissers, 2009; Van Herten et al., 2006). According to a misspelling account, participants may have considered the form-related words as being misspellings of the predicted word, triggering a repair of the surface feature that differed between the predicted and encountered input (e.g., Kim & Osterhout, 2005). Importantly, both these accounts assume that comprehenders detect a difference between the encountered input and the input that they predicted, but the accounts differ in terms of whether comprehenders lay the blame on errors in their own comprehension processes or on an error in the written input. A third interpretation is that the LPC effect only reflects the detection of similarity to the predicted form, without comprehenders considering any input or process to be erroneous.

The combination of an N400 effect and subsequent LPC effect suggests that participants in our experiment did not take the form-related words purely as a misspelling. It suggests that semantic information associated with form-related words was indeed accessed. Moreover, we observed the LPC effects whether or not the N400 for form-related words was reduced (i.e., in Experiments 5 and 6), indicating that pre-activation of form was not necessary for the occurrence of the subsequent processes reflected in the LPC. The pre-activation of form information and the detection of form similarity thus appeared to be fairly independent of each other. Detection of form similarity of encountered input with predicted input may thus arise via a bottom-up process of feature activation (Federmeier, 2007), rather than pre-activation.

4.11 Conclusions

Current neurobiological accounts of language comprehension assume lexical prediction through pre-activation (Federmeier, 2007; Kutas et al., 2009). However, less is known about how linguistic information is pre-activated. We examined the patterns of co-occurrence of form and meaning pre-activation to test whether the patterns would be consistent with production-based prediction accounts. Our study investigated prediction of form and meaning while participants read grammatical sentences without having to explicitly evaluate the critical words. Anomalous words that were semantically related to predictable words elicited reduced N400 effects compared to unrelated words, and this effect was not influenced by SOA. In contrast, highly predictable form-related words elicited a reduced N400 effect in the slower SOA, suggesting that people pre-activate the semantics of predictable words more strongly than the form. Form-related words also elicited an enhanced, post-N400 posterior positivity at both SOAs, indicating that form similarity between expected and encountered input was detected via a bottom-up mechanism, regardless of whether form features are pre-activated or not. Our results demonstrate that pre-activation of the form of upcoming words depends on the time that readers have to predict, which we suggest is in line with production-based accounts of linguistic prediction.

5 Study 4. Predicting form and meaning in L2

This chapter is based on a manuscript to be published as Ito, A., Martin, A. E., & Nieuwland, M. S. (in press). On predicting form and meaning in a second language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*

5.1 Introduction

What does it mean to understand a language? At the very least, it means to know the meaning of its words and to know which sentences are allowed by its grammatical rules. Yet those things alone are clearly not enough to engage in a regular conversation. Beyond achieving intricate and often non-literal meanings, a listener needs to access words and apply rules to establish the intended meaning in a timely manner, in order to keep up with the pace of the speaker. That is, the listener needs to process language incrementally, by using relevant information as soon as possible, and sometimes perhaps even predictively, by anticipating information where possible. L1 speakers typically do all of this without any conscious effort. However, it does not follow that incremental and predictive processing is necessary for successful comprehension (for discussion, see Huettig & Mani, 2016). For example, L2 speakers may show slower and/or less predictive processing compared to L1 speakers, while ultimately understanding the sentence correctly. We therefore tested whether fluent L2 speakers exhibit incrementality, or to what degree they pre-activate information during sentence processing.

5.1.1 Incremental processing during L2 comprehension

There is mixed evidence as to whether L2 speakers comprehend sentences incrementally like L1 speakers. Several visual world eye-tracking studies have found evidence for incrementality in L2 comprehension. For example, Trenkic et al. (2014) found that late Mandarin-English bilinguals utilised information conveyed by English articles to constrain referential domains and resolve reference (although they were overall slower than

L1 speakers). In contrast, Lew-Williams and Fernald (2010) found that, whereas L1 Spanish speakers rapidly used the grammatical gender of articles preceding a noun to identify an upcoming referent, late English-Spanish bilinguals did not take advantage of grammatical gender information. Evidence from reading studies is also not clear-cut. There is evidence for incremental processing in L2 speakers, but native-like incrementality in L2 speakers has been shown to depend on their proficiency (Hopp, 2006), their working memory span (Dussias & Pinar, 2010) or experimental task (Williams, 2006).

ERP studies also show mixed findings. L2 speakers often show qualitatively similar N400 effects to L1 speakers in response to semantic anomaly (Kotz, 2009). Native-like N400 effects have also been reported for plausible, relatively unexpected words relative to highly expected words (Foucart et al., 2014; Martin et al., 2013), suggesting that L2 speakers can comprehend meaning incrementally like L1 speakers. However, the observed N400 effects are sometimes delayed or smaller in L2 speakers relative to effects in L1 speakers (Martin et al., 2013; Weber-Fox & Neville, 1996), suggestive of impaired semantic processing in L2 speakers.

In sum, findings from multiple methodologies suggest that differences between L1 and L2 speakers can manifest themselves in the time-course or size of observed effects. However, these differences do not occur consistently in the literature on L2 comprehension.

5.1.2 Predictive processing in L2 speakers

Incremental comprehension is a prerequisite for generating predictions about upcoming words. If L2 speakers comprehend less incrementally than L1 speakers, they are also less likely to predict upcoming information. In fact, while evidence for prediction in L1 speakers is abundant, evidence for prediction in L2 speakers is sparse and mixed. Some visual world eye-tracking studies report predictive looks to upcoming referents in L2 and L1 speakers alike (Chambers & Cooke, 2009; Hopp, 2013), whereas some studies did not observe predictive looks in L2 speakers but only in L1 speakers (Mitsugi & MacWhinney, 2016). As we will discuss in more detail below, inconsistent results may be due to the proficiency of L2 speakers or due to differences in the experimental manipulations.

Two recent ERP studies also reported inconsistent findings. In Martin et al. (2013), L2 speakers did not show an N400 effect for articles that mismatched predictable nouns relative to prediction-matched articles (e.g., *an* versus *a* when readers could expect *a singer*). However, Foucart et al. (2014) found that French-Spanish late bilinguals and L1 Spanish speakers elicited the same N400 effect for gender-marked articles that mismatched predictable nouns relative to prediction-matching articles (e.g., *la* [feminine] versus *el* [masculine] when readers could expect *el tesoro*).

Several factors could account for the inconsistent findings in the literature, including differences in participants and experimental manipulations. Another explanation would be that L2 speakers can engage in predictive processing, but perhaps do not do this as routinely or automatically as L1 speakers. If this were the case, why not? If we explore this possibility, we can see that L2 speakers must overcome several obstacles in their efforts to understand language with the richness and depth that an L1 speaker naturally achieves. First, for example, L2 speakers may unintentionally co-activate words from their L1 (Chambers & Cooke, 2009; Shook & Marian, 2013). This can lead to delays in accessing words in their L2 when co-activated words differ in meaning from the intended word (Kroll, Gullifer, & Rossi, 2013). Second, L2 speakers appear to generate less robust or less detailed syntactic structures during comprehension (e.g., Clahsen & Felser, 2006), and rely more on lexical-semantic relationships instead. Under this *shallow structure account* of Clahsen and Felser (2006), L2 comprehension is particularly impaired in sentences with relatively complex or infrequent syntactic information (e.g., resolving long-distance dependencies in English; Marinis, Roberts, Felser, & Clahsen, 2005).

Beyond these two specific obstacles, however, a more general reason that L2 comprehension may be slower or less accurate than L1 comprehension is that L2 speakers are typically less proficient and/or less experienced than L1 speakers (Kroll & De Groot, 2005). Kaan (2014) recently argued that there are no qualitative differences in incremental and predictive comprehension between L1 and L2 speakers, only quantitative differences. Moreover, she suggested that those quantitative differences result from the same factors that underlie individual differences in L1 speaker proficiency, such as quality of lexical representations, the ability to entertain low-frequent word meanings or competing structural parses, as well as task-induced factors. Quality of lexical representations may be particularly important as it could lead to slower lexical access and weaker semantic networks (Ivanova & Costa, 2008). Hence, proficiency may be more important for incremental processing (Kotz & Elston-Güttler, 2004), and therefore for predictive procesing, than nativeness per se. However, proficiency levels may vary stronger in L2 speakers than in L1 speakers (e.g., Blumenfeld & Marian, 2007), so that effects of predictive processing appear sporadically in reported studies. In any case, proficiency and nativeness are co-extended, if not confounded.

If the mixed results on L2 predictive processing stem from proficiency differences, then studies with highly proficient participants would show effects of incremental or predictive processing (e.g., Hopp, 2013) and studies with less proficient participants would show no such effects (e.g., Mitsugi & MacWhinney, 2015). However, other factors may matter too. Evidence for prediction may be obtained more readily if the linguistic features of the critical manipulation (i.e., *a*/*an* distinction or grammatical gender) are also present in the L1 (e.g., Foucart et al., 2014). Yet, this role of language-similarity may not be independent of language proficiency, and on its own cannot explain that prediction effects are sometimes observed under the manipulation of grammatical features that are not shared between L1 and L2 (e.g., Hopp, 2013).

Details of the experimental paradigm are also relevant, and between-experiment comparisons are usually confounded by several factors. For example, it may be the case that prediction effects can be more reliably elicited by some linguistic manipulations than others.

If the status of a linguistic feature makes it function as a stronger cue to meaning or a higher level of representation, it might have more influence on processing than a feature that only sometimes is a cue or that does not lead to higher level of representation (E. Bates & MacWhinney, 1989). For example in Spanish, the grammatical gender of a noun is not only an important feature or cue during L1 acquisition and L2 learning of nouns, but it is also diagnostic during the establishment of adjacent and nonadjacent syntactic and referential dependencies in routine production and comprehension. In contrast, the *a*/*an* distinction in English is a phonotactic rule that does not signal any lexical feature, nor cue a non-local dependency, rather it signals what the upcoming syllable or phoneme across the boundary between determiner and noun will be. Furthermore, the rule can be violated during natural speech, due to disfluency or pauses, without eliciting judgements of ungrammaticality in the same way as a grammatical gender violation would (Pullum & Zwicky, 1988). Thus, variation in the information status of the linguistic cues that are manipulated may also contribute to the mixed pattern of results across the literature.

In addition to representational differences, the time constraints placed upon participants during processing may also shape the observed effects. For example, the visual world eye-tracking studies that have reported evidence for prediction in L2 speakers allowed more time to generate predictions (Hopp, 2013) compared to studies that did not observe prediction effects (Lew-Williams & Fernald, 2010). This is an important factor, because people are more likely to predict upcoming information when they have more time (A. Ito et al., 2016). Another important factor, in ERP studies particularly, is critical word predictability (as established in a sentence completion test or cloze probability – nota bene that cloze is a proxy for online predictability). The only ERP study that reported prediction effects in L2 speakers (Foucart et al., 2014) had critical words that were more predictable than those used in the one study that did not find prediction effects (Martin et al., 2013).

In sum, the mixed evidence for predictive processing in L2 speakers may be due to multiple factors, in particular language proficiency, experimental manipulation, the status of the linguistic cues being manipulated, time constraints, and sentence predictability. Therefore, we focus not on whether or not L2 speakers predict upcoming information, but rather on the circumstances in which they do or do not exhibit signals consistent with prediction. In the current study, we employed a design that does not principally rely on a single type of grammatical information or linguistic cue, instead we asked whether L2 speakers predict the form and meaning of upcoming words in constraining contexts and whether any effects we might find would depend on time constraints.

5.1.3 The present study

The present study adopted the experimental design used in Experiments 5 and 6. In these experiments, L1 speakers read highly constraining sentence contexts (e.g., ‘‘*The student is going to the library to borrow a*...”), followed by the predictable word (*book*), an implausible word that was form-related (*hook*) or semantically related (*page*) to the predictable word, or an unrelated implausible word (*sofa*). Participants read sentences at 500 ms or 700 ms SOA. All types of implausible words elicited a classic N400 effect relative to predictable words (Kutas & Hillyard, 1980). Semantically related words elicited a smaller N400 than unrelated words at both SOAs. Form-related words also elicited smaller N400s than unrelated words, but for sentences with a very high cloze probability and at 700 ms SOA only. Form-related words additionally elicited a post-N400 posterior positivity (LPC effect) at both SOAs. Both the observed N400 reduction and the LPC effect did not depend on the plausibility of the anomalous word itself, but were dependent on the cloze probability of the predictable word.

The observed N400 results suggest that L1 speakers pre-activated semantic and form information, but only pre-activated form when they were reading relatively slowly and had very strong predictions for the upcoming word. In addition, our results suggest that, whether form was pre-activated or not, participants activated the predictable word after accessing the meaning of the form-related word, leading to a conflict in interpretation reflected in the observed post-N400 LPC effect (Van de Meerendonk et al., 2009). Following Experiments 5 and 6, we do not regard this LPC effect as evidence for pre-activation of form information, because people can detect such a conflict without predicting word forms. Instead, this pattern arises from a combination of top-down activation of contextual meaning and bottom-up activation of word form information due to form-similarity. For example, the context ''*The student is going to the library to borrow a...*" can activate meaning of the word *book* more strongly compared to other words that are unrelated to the context. Reading a form-related word *hook* will spread activation (via priming) to the word form of the predictable word *book,* which fits highly plausibly (and much more plausibly compared to *hook*) into the given context. This will then trigger a conflict between the encountered word and the (unseen) highly predictable and plausible word.

Our current predictions are derived from these L1 speaker results. If L2 speakers can pre-activate form and meaning of upcoming words, we expect to see reduced N400s to form-related and meaning-related critical words compared to unrelated words. Such effects would need to be independent of plausibility but dependent on the cloze value of the predictable word to constitute evidence for pre-activation. Like in L1 speakers, we hypothesized that pre-activation of form in L2 speakers is less likely to occur and more dependent on reading rate than pre-activation of meaning. Relevant to this hypothesis is the fact that RSVP studies on L2 comprehension typically use a slower SOA than studies on L1 comprehension, because the standard 500 ms SOA can be an uncomfortably fast pace for L2 speakers. We therefore expected to find no evidence of pre-activation of form or meaning in Experiment 7, which used a 500 ms SOA. However, participants may nevertheless activate the predictable word after encountering the form-related word, as would be evidenced by an LPC effect to the form-related condition.

5.2 Experiment 7

5.3 Methods

5.3.1 Participants

Twenty-four Spanish-English late bilinguals (8 males and 16 females, age *M* = 26.3 years, *SD* = 4.6) participated in the experiment. One participant was excluded from analysis due to a large portion of data artefacts. All participants were right-handed and reported no neurological or language disorders. The mean age of their first exposure to English was 11.0 years (*SD* = 7.6), and the mean length of their exposure to English was 12.5 years (*SD* = 5.9). Their overall self-rated English proficiency score was 8.2 (*SD* = 0.7; rated on a scale from 1 to 10).

5.3.2 Stimuli and experimental design

The stimuli and experimental design were identical to those used in Experiments 5 and 6. We collected new cloze probability values from a group of similar pool of participants as for the ERP experiments (12 Spanish-English bilinguals who did not take part in the ERP experiment). The mean L2 cloze value was 61% (range = 0-100%; *SD* = 25.4), which was lower than the mean L1 cloze value of 80% (range = 31-100%).

We collected new plausibility values from another group of similar pool of participants as for the ERP experiments (a total of 20 Spanish-English bilinguals who did not take part in the ERP experiment or in the cloze pre-tests), following the procedure described in Experiments 5 and 6. The mean plausibility ratings for each condition are in Figure 11. The semantic condition was rated more plausible than the form condition (*Mean Difference* = .26, *SD* = .56), *t*(159) = 3.7, *p* < .001, or the unrelated condition (*Mean Difference* = .36, *SD* = .83), *t*(159) = 5.5, *p* < .001, which was consistent with the L1 speaker ratings as in Experiments 5 and 6. The form condition was rated marginally more plausible than the unrelated condition (*Mean Difference* = .10, *SD* = .68), *t*(159) = 1.9, *p* = .06, while the form condition and the unrelated condition did not differ in the L1 ratings. The full set of experimental sentences with cloze values and plausibility ratings in L2 speakers is in Appendix 8.3.

5.3.3 Procedure and data analysis

The procedure and data analysis were identical to Experiments 5 and 6, except that participants completed a language background questionnaire before the experiment. Mean accuracy for comprehension questions during the ERP experiment was 93.6% (*SD* = 6.4; 8.6% of the data was excluded due to time-outs). The mean number of artefact-free trials in Experiment 7 per condition was 37 (*SD* = .24), with no difference across conditions.

5.4 Results

Visual inspection of Figure 11 indicated that all implausible conditions elicited larger N400s than the predictable condition. These N400 effects had a broad central-posterior distribution. Unlike the findings in L1 speakers, we did not observe an LPC effect for form-related words. Figures showing all channels are in Appendix 8.5.

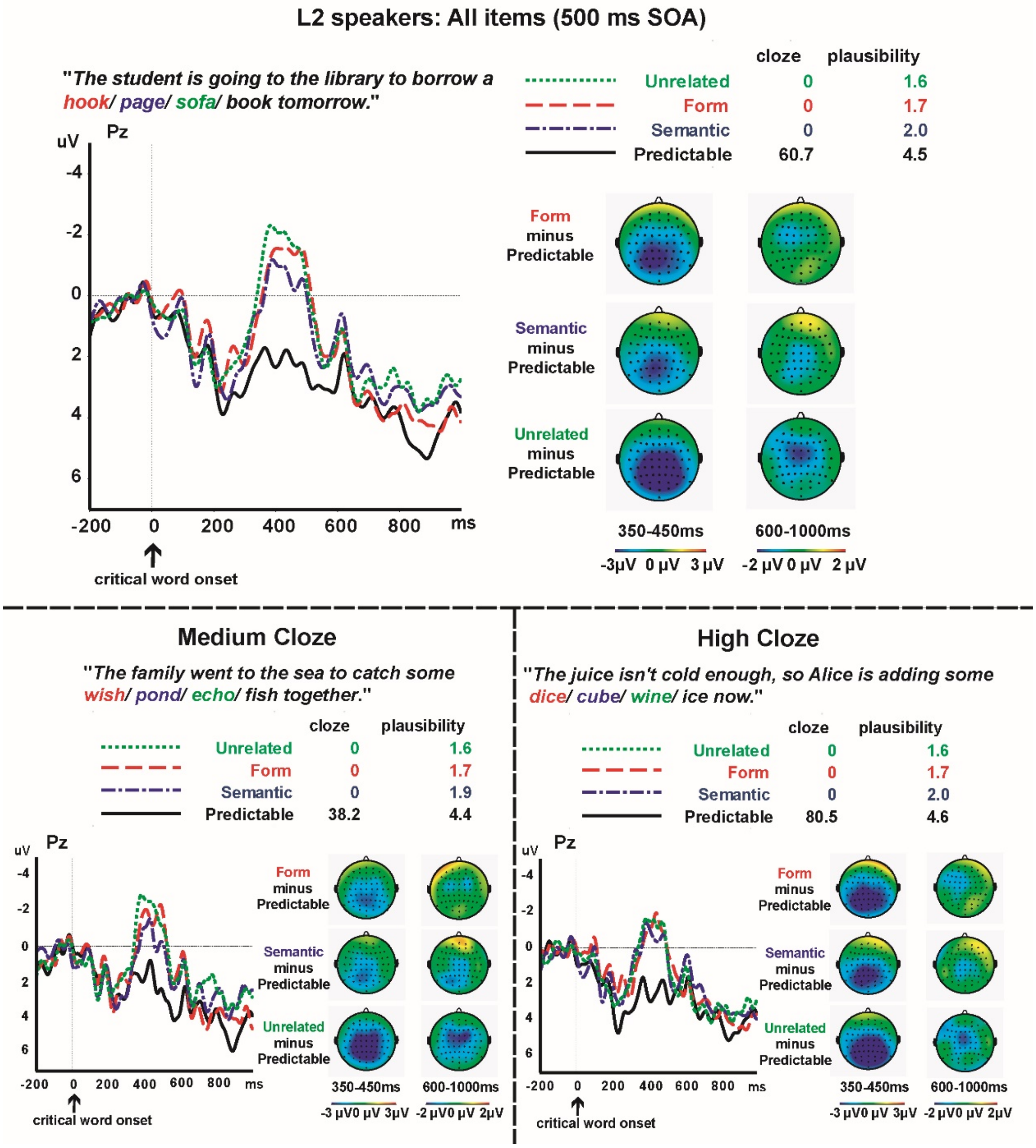


Figure 11. Results from Experiment 7 (500 ms SOA). ERPs elicited by each condition at Pz in all items (top panel), in medium-cloze items (left lower panel) and in high-cloze items (right lower panel). Scalp distributions of the ERP effects (implausible condition minus predictable condition) in the N400 time window and LPC time window are shown on the right in each panel.

5.4.1 The N400 time window

N400 analysis revealed a significant effect of condition, *F*(3, 66) = 12.0, *MSE* = 12.1, *p* < .001, and a significant interaction of condition by anteriority, *F*(3, 66) = 6.8, *MSE* = 1.5, *p* < .001, indicating that effects of condition were more robust at posterior channels, *F*(3,

66) = 16.8, *MSE* = 3.4, *p* < .001, than at anterior channels, *F*(3, 66) = 6.2, *MSE* = 3.5, *p* < .001. Given the stronger effects at posterior channels, we proceeded with pairwise *t*-tests between conditions focused on posterior channels. All the anomalous conditions elicited larger N400s than the predictable condition, *p*s < .001. The semantic condition elicited smaller N400s than the unrelated condition, (*M* = -1.1 µV, *SD* = 2.7), *t*(45) = -2.8, *p* < .01. There was no difference between the form and semantic conditions, *p* > .1.

5.4.2 The post-N400 time window

The post-N400 window analysis revealed a significant condition by anteriority interaction, *F*(3, 66) = 4.0, *MSE* = 1.7, *p* = .01, and a marginally significant condition by hemisphere, *F*(3, 66) = 2.3, *MSE* = .28, *p* = .08. We followed this up with one-way ANOVAs conducted at anterior and posterior channels separately, which revealed that the effect of condition was marginally significant at anterior channels only, *F*(3, 66) = 2.7, *MSE* = 3.7, *p* = .05.

5.4.3 Effects of cloze probability

We compared effects of condition on N400 and LPC in high-cloze and medium-cloze items, following the same procedure used in Experiments 5 and 6. Using the median cloze probability collected from the L2 speaker group (*Mdn* = 66), the items were split into a high-cloze subset (85 items, cloze *M* = 80.5, *SD* = 10.3) and a medium-cloze subset (75 items, cloze *M* = 38.2, *SD* = 17.6). Plausibility ratings in the predictable condition were higher in the high-cloze than for the medium-cloze items, *t*(146) = 2.4, *p* < .05, but plausibility in the anomalous conditions did not differ, *p*s >1. Frequency, word length, Levenshtein distance, word LSA or context LSA did not differ between the high-cloze and the medium-cloze subsets, *p*s > .05.

For the N400 amplitudes per condition, a two-way condition by cloze ANOVA revealed a significant effect of condition, *F*(3, 66) = 16.7, *MSE* = 6.6, *p* < .001, and a marginally significant effect of cloze, *F*(1, 22) = 3.0, *MSE* = 6.5, *p* = .1, but no interaction of condition by cloze, *F* < 1. For the LPC time window, we followed Experiments 5 and 6 and included posterior channels only. A condition by cloze ANOVA revealed no significant effect or interaction involving cloze, *F*s < 2.4.

5.5 Discussion

In Experiment 7, implausible words that were semantically related to predictable words elicited reduced N400s compared to unrelated control words. The observed N400 reduction did not depend on cloze probability of the predictable word. We do not take the N400 reduction as evidence for pre-activation, because it could reasonably be explained in terms of plausibility (e.g., the semantically related condition was more plausible than the unrelated condition). The results of Experiment 7 thus suggest that participants did not pre-activate semantic or form information of upcoming words.

5.6 Experiment 8

We considered the possibility that the 500 ms SOA in Experiment 7 was too fast for L2 speakers to read comfortably. It is well-established that L2 speakers, even when highly proficient, read more slowly than L1 speakers (Hopp, 2009). Probably for this reason, ERP studies on L2 sentence comprehension typically use an SOA that is slower than the standard 500 ms SOA used with L1 speakers (700 ms SOA in Martin et al., 2013 and Foucart et al., 2014; 725 ms in Tanner, Mclaughlin, Herschensohn, & Osterhout, 2013; 650 ms in Tokowicz & MacWhinney, 2005). L1 speakers who read at an uncomfortably fast pace also typically show diminished prediction effects (Dambacher et al., 2012; Wlotko & Federmeier, 2015). Therefore, Experiment 8 used a 700 ms SOA to investigate whether L2 speakers would show evidence of form or meaning pre-activation when reading more slowly.

5.7 Methods

5.7.1 Participants

Participants were 24 Spanish-English late bilinguals (10 males and 14 females, age *M* = 27.4 years, *SD* = 4.8), who did not participate in any of the previously described experiments. All participants were right-handed and reported no neurological or language disorders. The mean age of their first exposure to English was 10.8 years (*SD* = 6.7), and the mean length of their exposure to English was 13 years (*SD* = 8.2). Their overall self-rated English proficiency score was 7.8 (*SD* = 1.2; rated on a scale from 1 to 10).

5.7.2 Stimuli and experimental design

The stimuli and experimental design were identical to those used in Experiment 7.

5.7.3 Procedure and data analysis

The procedure, EEG data processing and statistical analysis were all identical to Experiment 7, except that the SOA was changed to 700 ms (500 ms word duration, 200 ms inter-word-interval; sentence final words had an 800 ms duration). Mean accuracy for comprehension questions was 88.9% (*SD* = 11.0; 10.4% of the data was excluded due to time-outs). The mean number of artefact-free trials in Experiment 8 per condition was 38 (*SD* = .23), with no difference across conditions.

5.8 Results

As in Experiment 7, the implausible conditions elicited visibly larger N400s than the predictable condition. Notably, unlike Experiment 7, the form condition elicited a larger LPC at posterior channels compared to the other conditions. Figures showing all channels are in Appendix 8.5.

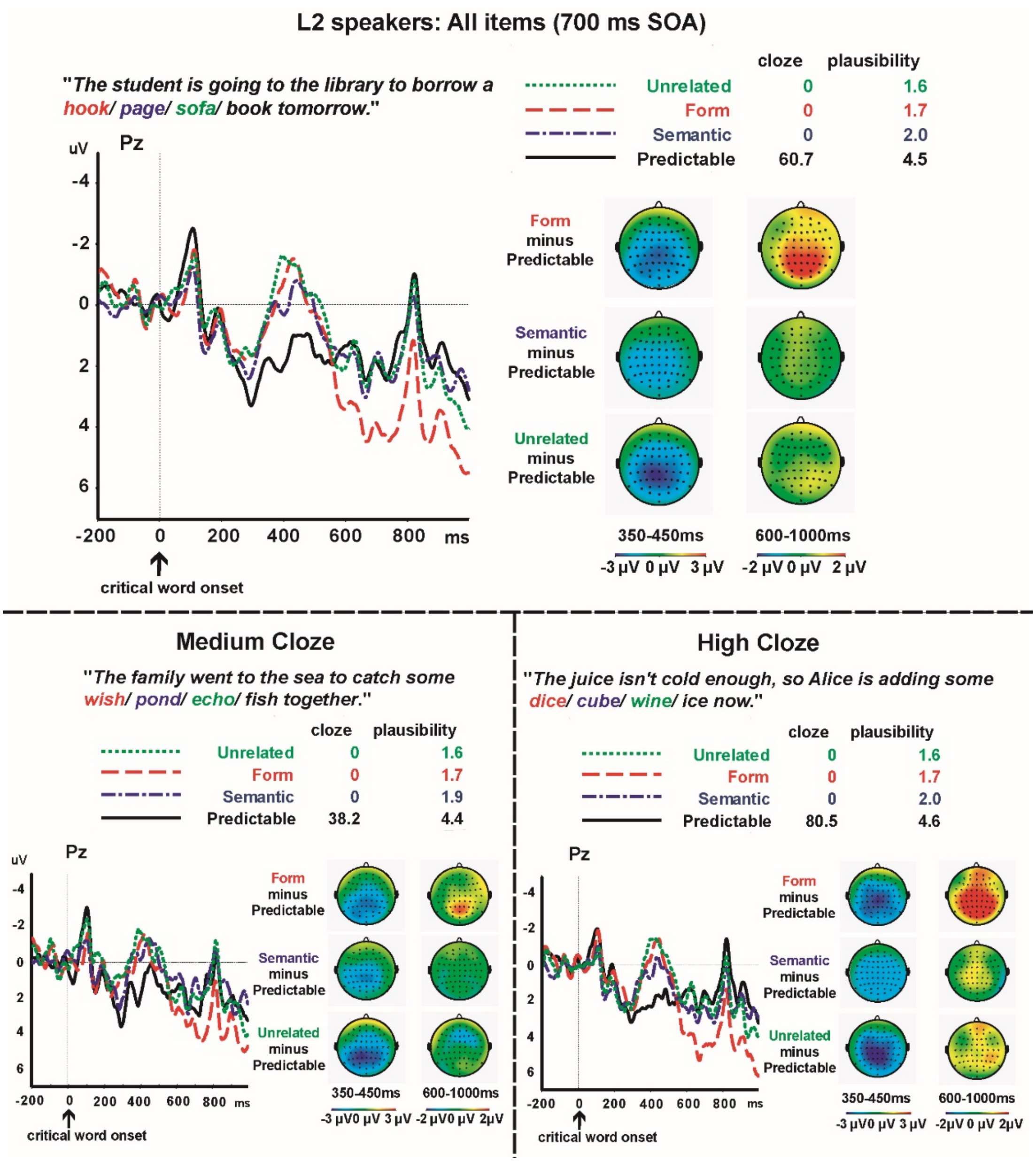


Figure 12. Results from Experiment 8 (700 ms SOA). ERPs elicited by each condition at Pz in all items (top panel), in medium-cloze items (left lower panel) and in high-cloze items (right lower panel). Scalp distributions of the ERP effects (implausible condition minus predictable condition) in the N400 time window and LPC time window are shown on the right in each panel.

5.8.1 The N400 time window

The analysis revealed a significant effect of condition, *F*(3, 69) = 6.0, *MSE* = 15.5, *p* = .001, and a significant interaction of condition by anteriority, *F*(3, 69) = 3.2, *MSE* = 1.0, *p* < .05, which stemmed from a stronger effect of condition at posterior channels, *F*(3, 69) =

7.4, *MSE* = 4.4, *p* < .001, than at anterior channels, *F*(3, 69) = 4.1, *MSE* = 3.9, *p* = .01. Pairwise comparisons between the conditions at posterior channels revealed that all the implausible conditions elicited larger N400s than the predictable condition, *p*s < .001. Critically, the semantic condition elicited a smaller N400 than the unrelated condition, (*M* = .82 µV, *SD* = 2.5), *t*(47) = 2.3, *p* < .05. There were no differences between any other condition pairs, *p*s >.1.

5.8.2 The post-N400 time window

The analysis revealed a significant effect of condition, *F*(3, 69) = 3.6, *MSE* = 13.6, *p* < .05, and a significant interaction of condition by anteriority, *F*(3, 69) = 5.4, *MSE* = 1.1, *p* < .01, which was driven by the effect of condition being significant at posterior channels only, *F*(3, 69) = 5.1, *MSE* = 4.0, *p* < .01. Pairwise comparisons at posterior channels revealed that the form condition elicited a larger positive ERP compared to the predictable condition, (*M* = 2.1 µV, *SD* = 3.0), *t*(47) = 4.8, *p* < .001, the semantic condition, (*M* = 1.7 µV, *SD* = 2.4), *t*(47) = 5.0, *p* < .001, and the unrelated condition, (*M* = 1.4 µV, *SD* = 3.0), *t*(47) = 3.3, *p* < .01.

5.8.3 Effects of cloze probability

Effects of cloze probability were tested using the same median split approach as in Experiment 7. A two-way condition (semantic vs. unrelated) by cloze ANOVA in the N400 window revealed a marginally significant effect of cloze, *F*(1, 23) = 2.9, *MSE* = 5.4, *p* = .1, but no interaction of condition by cloze, *F* < 1. Since the effect of condition was greater at posterior channels, the analysis in the post-N400 time window focused on the posterior channels only. A two-way condition (4 conditions) by cloze ANOVA revealed no effect or interaction involving cloze, *F*s < 1.4.

Unlike the results for L1 speakers, the current LPC effect for L2 speakers seemed independent of the cloze value of the predictable word. However, we noted the possibility that this absence of a cloze-effect was driven by earlier differences (in the 0-200 ms time window) in medium-cloze items. In particular, the form-related condition elicited more positive-going ERPs in this early time window compared to the predictable condition and the unrelated condition.[21] Such early differences may reflect differences in accidental background fluctuations (‘noise’ that is not related to the manipulation itself). No such early differences were observed in the high-cloze items. To investigate the effect of cloze without the potential confound of this earlier difference, we re-analysed our data using a baseline of 0 to 200 ms after word onset (effectively minimizing the impact of the earlier effect in this window on the later ERP differences). Moreover, in this additional analysis, we used cloze value as a continuous regressor instead of splitting the items into medium-cloze and high-cloze items. Such an analysis provides a more sensitive measure of the graded impact of cloze value.

5.8.4 Additional analyses with cloze value as a continuous regressor

The only change in data pre-processing was the use of a 200 ms post-stimulus time window for baseline correction. For comparability, we plotted the resulting grand-average ERP waveforms in Figure 13. While the overall N400 patterns did not visibly change, it led to a more substantial difference between the form-related LPC effect in the high-cloze items and in the medium-cloze items. Statistical tests revealed that the overall N400 and LPC effects were not different from our original analysis, and our additional analysis therefore focused only on the impact of cloze value on the LPC effect of form-similarity.

---

[21] A two-way condition (predictable vs. form) by cloze ANOVA at posterior channels in 0-200 ms time window revealed a significant effect of condition, $F(1, 23) = 8.3$, $MSE = 8.4$, $p < .01$, and a significant interaction of condition by cloze, $F(1, 23) = 4.9$, $MSE = 5.7$, $p < .05$.

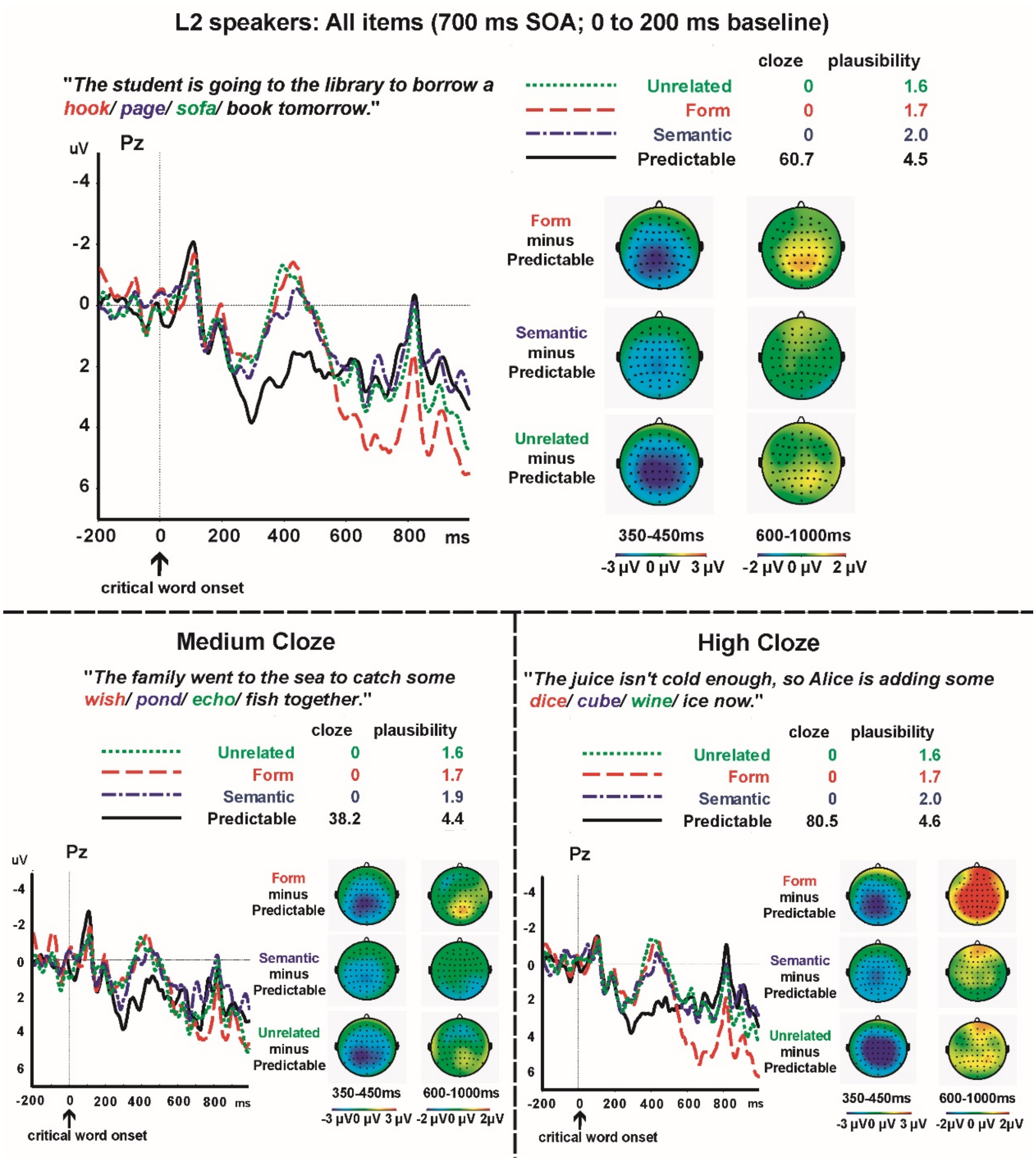


Figure 13. Results from Experiment 8 (700 ms SOA) with a 200 ms post-stimulus baseline. ERPs elicited by each condition at Pz in all items (top panel), in medium-cloze items (left lower panel) and in high-cloze items (right lower panel). Scalp distributions of the ERP effects (implausible condition minus predictable condition) in the N400 time window and LPC time window are shown on the right in each panel.

5.8.5 Effects of cloze probability

We tested effects of cloze probability on the LPC effect using linear mixed-effects models, using cloze probability as a continuous predictor. For this analysis, we exported single trial data in the 600-1000 ms time window at medial-posterior channels (CP1, CP2, CP3, CP4, CPz, P1, P2, P3, P4, Pz). We constructed two linear mixed-effects models; the first model evaluated ERP amplitudes as predicted by condition (predictable vs. form), and the second model tested the interaction of condition by cloze in addition to the main effects of condition and cloze probability.[22] The models included random intercept by participant and by item. We assessed the significance of the effects by whether the associated absolute *t*-value exceeded 2 (Baayen et al., 2008). The first model showed a significant effect of condition, $\beta = 1.7$, $SE = .45$, $t = 3.7$. The second model showed a significant interaction of condition by cloze probability, $\beta = .046$, $SE = .018$, $t = 2.6$. Effects of condition or cloze were not significant, *ts* < 1.1. Including the interaction term significantly improved the model fit of the data, $\chi^2(2) = 7.8$, $p < .05$. As shown in Figure 14, higher cloze values were associated with stronger LPC effects.

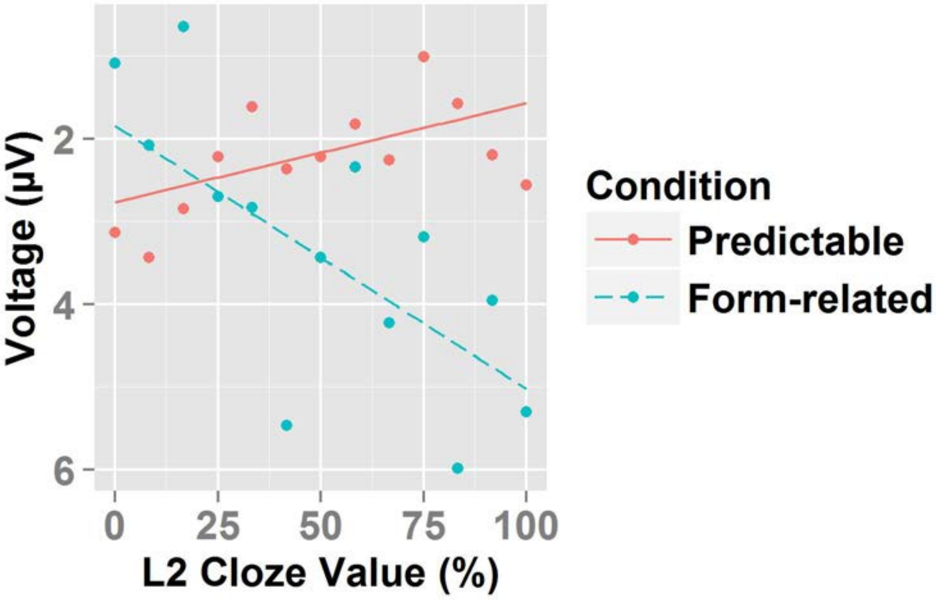


Figure 14. Fitted responses for each of the used cloze values associated with the LME analysis.

---

[22] Model 1: LPC amplitudes ~ condition + (1|participant) + (1|item)
Model 2: LPC amplitudes ~ condition * cloze + (1|participant) + (1|item)

5.8.6 Between-experiment comparisons: effects of SOA

We performed a between-experiment comparison to specifically test the impact of SOA on the post-N400 amplitudes in the form condition. We constructed two linear mixed-effects models, one that only used condition (predictable vs. form) as a predictor, and a second model that also included an interaction term of condition by SOA.[23] The first model showed a significant effect of condition, $\beta = -.86$, $SE = .33$, $t = -2.6$, and the second model showed a significant interaction of condition by SOA, $\beta = -1.8$, $SE = .65$, $t = -2.8$. Main effects were not significant in the second model, *t*s < 1. An ANOVA comparing the two models showed that the second model was a better fit for the data, $\chi^2(2) = 9.1$, $p < .05$.

5.8.7 Between-study comparisons: effects of language nativeness

We also ran linear mixed-effects models to test effects of language nativeness in the N400 (350-450 ms) and LPC (600-1000 ms) time windows. For this analysis, we used high-cloze items in the 700 ms SOA data from L1 speakers (Experiment 6) and from L2 speakers (Experiment 8), in which both the N400 reductions and the LPC effect were most robust.[24] The model in the N400 time window evaluated N400 amplitudes as predicted by main effects of condition (predictable vs. unrelated, form vs. unrelated, semantic vs. unrelated) and of language group (L1 vs. L2), and the interaction of the two.[25] The model revealed a significant interaction for the predictable vs. unrelated condition comparison, $\beta = -2.6$, *SE*

---

[23] Model 1: LPC amplitudes ~ condition + (1|participant) + (1|item)
Model 2: LPC amplitudes ~ condition * SOA + (1|participant) + (1|item)
We also ran a model with random slopes for condition by participant and for condition and SOA by item, but this model did not converge.

[24] For the N400 analysis, we used the original 100 ms pre-stimulus baseline for both L1 and L2 speakers. For the LPC analysis, we used the 200 ms post-stimulus baseline for the L2 speakers only, following the analyses in Sections 5.8.5 – 5.8.6. Using the original pre-stimulus baseline showed the same pattern of results.

[25] N400 model: N400 amplitudes ~ condition * language group + (1|participant) + (1|item)
We also ran a model with random slopes for condition by participant and by item, but this model did not converge.

= .98, *t* = -2.6, and for the form vs. unrelated condition comparison, β = -2.1, *SE* = .99, *t* = -2.1. These interactions suggest that the standard N400 effect for implausible words (which were unrelated to predictable words) relative to predictable words was greater in L1 speakers than in L2 speakers, and that the N400 reduction for form-related words relative to unrelated words was found in L1 speakers but not in L2 speakers. The interaction for the semantic vs. unrelated condition comparison was not significant, $|t| < 2$. This lack of interaction suggests that L1 and L2 speakers showed smaller N400s for semantically related words than for unrelated words alike. However, we do not interpret the N400 reduction in L2 speakers as evidence for prediction of semantic information, because the N400 reduction in L2 speakers did not depend on cloze value. If the N400 reduction had stemmed from predictions, the degree of the N400 reduction would have been larger in more predictable (higher-cloze) sentences than in less predictable (medium-cloze) sentences, as found in L1 speakers. Given that this was not the case, the N400 reduction likely reflects ease of integration for semantically related words. The plausibility ratings for semantically related words were higher than for unrelated words in high-cloze and medium-cloze sets alike, so the integration account can plausibly explain the results. Main effects of condition on all three comparisons and of language group were also significant, $|t|$s > 2. The model in the LPC time window evaluated LPC amplitudes as predicted by main effects of condition (predictable vs. form) and of language group, and the interaction of the two.[26] The model revealed a significant main effect of condition, β = -2.6, *SE* = .68, *t* = -3.7. Neither the main effect of language group nor the interaction of condition by language group was significant, $|t|$s < 1. The result indicates that the LPC effect was similar in L1 and L2 speakers.

---

[26] A model with random slopes for condition by participant and by item did not converge.

5.9 Discussion

In Experiment 8, semantically related words elicited smaller N400s than unrelated words, but this difference did not depend on cloze value. Form-related words elicited a post-N400 positivity at posterior channels (LPC effect). In an additional analysis that corrected for ERP differences in a very early (0-200 ms) time window, we found that this LPC effect gradually increased in size with increasing cloze probability of the predictable word. Whereas we did not find any evidence for pre-activation in L2 speakers in Experiment 8, the LPC results suggest that L2 speakers activated the predictable word after encountering the form-related word, leading to a conflict in interpretation.

5.10 General Discussion

We used ERPs to investigate whether L2 speakers pre-activate semantic and form information when reading sentences word-by-word at a 500 ms SOA or at a slower 700 ms SOA. These sentences contained a predictable word, an implausible word that was semantically or form-related to that word, or an implausible unrelated word. At both SOAs, predictable words elicited smaller N400s compared to implausible words (i.e., the classic N400 effect). Semantically related words elicited smaller N400s than unrelated words, but unlike the pattern observed in L1 speakers (Experiments 5 & 6), this N400 reduction did not depend on cloze probability of predictable words. Thus, we did not obtain clear evidence that L2 speakers pre-activated form or meaning, although the observed N400 effects do suggest that participants were sensitive to sentence plausibility. The second main observation was that form-related words elicited an LPC effect at 700 ms SOA, but not at 500 ms SOA. We suggest that this LPC effect reflect the activation of predictable words due to the combined top-down influence of sentence meaning and bottom-up influence of form-similarity. We will discuss the N400 and LPC effects in turn.

5.10.1 No pre-activation of form or meaning in L2 speakers

Our N400 results contrast with previously reported evidence for prediction in L2 speakers (Chambers & Cooke, 2009; Foucart et al., 2014; Hopp, 2013), adding to a body of evidence suggesting that L2 speaker do not engage in predictive processing as routinely as L1 speakers do (e.g., Martin et al., 2013).

Because our manipulation did not rely on observing effects on noun-preceding articles that code grammatical gender or the vowel/consonant of the upcoming word, we can rule out an explanation purely based on weaker sensitivity of L2 speakers to grammatical or phonotactic information (e.g., Lew-Williams & Fernald, 2010). An explanation purely in terms of time constraints does not suffice either, as we did not observe pre-activation effects at a relatively slow pace with which a previous ERP study observed pre-activation (Foucart et al., 2014).

The specific role of predictability (quantified as cloze probability) and that of proficiency is more elusive, however. L2 speakers found our predictable words less predictable (average cloze probability of about 60%) than L1 speakers did (average cloze probability of about 80%). In our study with L1 speakers, no pre-activation N400 effects were observed for items with an average cloze probability of about 65%. Thus, our results could reflect that the sentences were not constraining enough towards a particular word for L2 speakers.[27] Why sentence constraint was lower for L2 speakers than for L1 speakers remains an open question. Such constraint differences may arise from different (use of)

---

[27] To date, evidence for anticipatory processing in L2 speakers has mostly come from visual world eye-tracking (but see Foucart et al., 2014). This may well have to do with sentence constraint, because in those studies the upcoming referents are already visible on the screen, therefore the sentences have a very high constraint towards a particular continuation (i.e., the equivalent of 100% cloze probability). In this regard, evidence for predictive processing in visual world paradigms reflects the incremental mapping of incoming language onto given visual information, as opposed to evidence for predictive processing in ERP studies which typically examine brain responses to information associated with words that have not been seen before.

world knowledge but could also be a function of language proficiency. An important caveat to explaining our results directly in terms of predictability, however, is that there is no one-to-one correspondence between cloze probability and online prediction effects (e.g., A. Ito et al., 2016). Cloze probability is an offline task without time constraint, and therefore gives an estimate of the likelihood that online prediction occurs under generous time constraints (i.e., slow presentation) (Staub et al., 2015). Whereas high cloze values are associated with higher likelihood of online predictive processing than lower cloze values, high cloze values are not a guarantee that online prediction occurs. One potential implication of this is that cloze values from L2 speakers are not as good of an estimation of online prediction as cloze values from L1 speakers. Alternatively, there might always be pre-activation in proportion to cloze probability in L2 speakers, but L2 speakers may yield more noise compared to L1 speakers, which makes it harder to detect effects using ERPs due to their limited sensitivity.

Following Kaan (2014), we think that the absence of pre-activation effects in our study is primarily due to lower proficiency of L2 speakers. Even when relatively proficient L2 participants read high-constraint sentences at a relatively slow pace, build-up of sentence meaning may not have been fast enough or strong enough in order to pre-activate relevant semantic or form information in advance of the critical words.

It may be that L2 speakers, compared to L1 speakers, rely more strongly on specific lexical cues to predict upcoming information than on the compositional meaning of the context. Future research should thus tease apart the effects of lexical-associative priming from words in the context and of the exact message conveyed by that context on predictive processing (e.g., Otten & Van Berkum, 2008).

5.10.2 Activation of predictable words based on form-similarity

At 700 ms SOA, form-related words elicited a larger post-N400 positivity, relative to the predictable condition, which increased in size with higher cloze probability of the predictable word. This cloze-dependent LPC effect was similar to what was found in L1 speakers at both the 500 and 700 ms SOA (Experiments 5 & 6), and suggest a conflict in interpretation due to the activation of the predictable word. This LPC effect may therefore reflect a certain degree of incremental processing, because it requires a representation of sentence meaning that is sufficiently strong for bottom-up form-similarity to have its effect. This may work through cascaded activation, where the activation of sentence meaning in combination with the form-similar word leads to activation of the predictable word.

Absence of an LPC effect at the 500 ms SOA is consistent with such an interpretation. This SOA is standard for ERP studies on L1 comprehension but unusually fast for L2 sentence comprehension (e.g., Foucart et al., 2014; Tanner et al., 2013), and therefore our participants may not have enough time to construct a sufficiently strong or detailed contextual representation. The impact of presentation rate appears to be greater on L2 comprehension than on L1 comprehension.

What factors might explain why we did not find evidence of pre-activation in L2 speakers, but indeed LPC evidence for native-like incremental processing? One possibility is that while sentence comprehension is lagging in L2 speakers compared to L1 speakers, this lag is only short and therefore the representation of sentence meaning in L2 speakers is sufficiently strong by the time that the semantic information associated with the form-related word has been retrieved, giving rise to the conflict in interpretation. This also suggests that incremental processing can, to some extent, proceed normally without prediction via pre-activation, or without prediction at all. This is in line with recent views on the role of prediction in language comprehension (e.g., Huettig & Mani, 2016). Huettig and Mani argued that sentences in everyday conversations are normally less predictive than experimental contexts developed to study prediction. While cloze probability serves a proxy or estimate of processing situations where contextual constraint is particularly high, real language data, and certainly any model of language processing, must span the range of possible cloze values or constraining contexts. If prediction were required for truly incremental language comprehension then real-world language comprehension would not be the incremental process it is known to be.

5.11 Conclusion

We did not find N400 evidence for pre-activation of the form or meaning of upcoming words in L2 speakers. We conclude that although L2 comprehension can be highly incremental and even predictive (Foucart et al., 2014; Hopp, 2013), L2 speakers may not engage in predictive processing as routinely or robustly as L1 speakers. However, we observed evidence for the activation of unseen words in an LPC effect to words that were similar in form to a predictable word, which increased gradually with cloze value of the predictable word. We suggest that the combined effects of top-down activation (contextual meaning) and bottom-up activation (form similarity) result in activation of unseen words that fit the context well, thereby leading to a conflict in interpretation reflected in the LPC. This shows that L2 speakers can use bottom-up and top-down information to constrain incremental interpretation in much the same way that L1 speakers do.

6 General Discussion

The thesis examined what factors play a key role for predictive processing during language comprehension. A large number of psycholinguistic studies have shown that people predict upcoming language (Kamide, 2008, for a review). While linguistic predictions tend to occur fairly easily and rapidly, studies have also found that not all people make predictions to a similar extent (e.g., Mani & Huettig, 2012). Such findings suggest that there seem to be some limitations in people's prediction performances, but little is known about what factors can mediate predictive processing. To this end, the thesis investigated how available cognitive resources and time to generate predictions mediate predictive processing in L1 and L2 speakers of English, and how those factors interact. The following sections reconcile the findings in the thesis with existing models of language prediction and models of L2 sentence comprehension and L2 prediction.

6.1 Summary of empirical findings

This section summarises the main empirical findings in this thesis to answer the research questions.

6.1.1 Effects of cognitive load

Experiments 1 and 2 found that the predictive eye movements in L1 and L2 speakers were significantly reduced when they listened to sentences under a cognitive load of remembering words. Consistent with previous research (Huettig & Janse, 2016), these experiments implicate a critical role of working memory during language prediction. The findings suggest that working memory resources that are used for remembering words are used for making predictive eye movements, so those predictive eye movements tend to be reduced when there are insufficient working memory resources. The findings therefore suggest that some or all parts of making predictive eye movements are not automatic. Thus, whether people can make predictive eye movements depends on cognitive resources available, both during L1 and L2 comprehension. Making predictions in Experiments 1 and 2 was likely to be relatively easy, as participants were able to predict a single referent (with almost 100% certainty) based on the meaning of a single verb (see Section 6.2.3 for discussion of differences in those experimental contexts from reading comprehension). In such an experimental setting, even children and L2 speakers have been found to make similar predictive eye movements to L1 adult speakers. However, even this easy type of prediction does not appear to occur automatically but requires cognitive resources.

6.1.2 Effects of time

Experiments 5 – 8 used ERPs and demonstrated effects of SOA on processes associated with prediction of word form. L1 speakers showed reduced N400s for words that were form-related to highly predictable words relative to unrelated words. This effect hinged on the cloze probability of predictable words, while being independent of the plausibility of those words that participants read. This suggests that the reduced N400 effect is more compatible with a prediction account than with an integration account. This evidence for word form prediction was found at a relatively slow 700 ms SOA but not at the standard 500 ms SOA.

Although there was no evidence that L2 speakers predict form, L2 speakers activated predictable words using top-down contextual information and form-similarity from bottom-up input. This activation of predictable words from the combined use of top-down and bottom-up information occurred only at the slower 700 ms SOA. These findings suggest that L2 speakers can process sentences in a highly incremental fashion like L1 speakers when the reading rate was slow enough. As prediction would not occur without incremental comprehension, the results suggest that more time to process sentences would make prediction more likely to occur.

The effects of SOA in L1 and L2 speakers both indicate a critical role of time in predictive processing. Previous studies have found similar effects of SOA on pre-activation of semantic information in L1 speakers, indicating that semantic pre-activation is more likely to occur when there is more time available to generate predictions (Dambacher et al., 2012; Wlotko & Federmeier, 2015). The findings in Experiments 5 – 8 further revealed that prediction of word form is more subject to presentation rates than prediction of meaning by investigating the co-occurrence patterns of the two types of prediction. Taken together, prediction of word form and prediction of semantic information both depend on the time that comprehenders have to make predictions, at least in L1 comprehension. Furthermore, more time to comprehend each word makes L2 processing more native-like in the degree of incrementality, which is essential to make predictions.

6.1.3 Effects of language nativeness

The current thesis compared predictive performances in L1 and L2 speakers. Experiments 1 and 2 did not find any clear difference between L1 and L2 speakers in their predictive eye movements, irrespective of when they comprehended sentences under an additional cognitive load or when under no additional load. However, Experiments 3 and 4 found that L1 speakers predicted phonological information of highly predictable words, whereas L2 speakers did not predict phonological information. The results from Experiments 5 – 8 were consistent with Experiments 3 and 4 with regard to prediction of word form; L1 speakers predicted the form of highly predictable words, but L2 speakers did not. Evidence for prediction during L2 comprehension was not consistent across experiments; the eye-tracking experiment (Experiment 2) showed that L2 speakers predicted upcoming referents, but the ERP experiments (Experiments 7 & 8) revealed no evidence that they made any prediction.

The difference in the methodology might have caused this inconsistency. In the eye-tracking study, participants saw objects, and the sentence context supported prediction of one of the shown objects. In such a typical visual world setting, the visual scene has already provided a quite restrictive context, and makes it clear that objects in the scene are potential referents. At least, participants likely assumed this, especially because Experiment 2 asked participants to judge whether anything in the visual scene was mentioned (and one of the objects was mentioned 50% of the time). Thus, participants would use the visual context in combination with the sentence context to generate predictions about upcoming referents. This created the 100% predictability, practically, because when the context was predictive, only the predictable referent was the plausible continuation of the sentence, making predictability of the other objects 0%. In the ERP experiments, participants read sentences without any such contexts. Thus, there are far more words that could plausibly continue the sentence, so participants would narrow predictions down from a much wider set of candidates. Thus, the predictability here was considerably lower than 100% (the offline cloze probabilities, which we used as a proxy for predictability in the ERP experiments, were 80% in L1 speakers and 61% in L2 speakers).

Alternatively, the modality of the stimulus presentation might be at play too. Spoken sentences contain articulatory information that may signal certain phonological properties about an upcoming word, as pronunciation of a preceding word often changes according to the next word (see Altmann, 2011 for discussion). Such subtle co-articulatory information may provide cues that help listeners to predict information about the following word (Van Berkum et al., 2005). Moreover, prosodic information has been shown to aid prediction or efficient ambiguity resolution (Nakamura et al., 2012; Snedeker & Trueswell, 2003; Weber et al., 2006). These cues are naturally absent in reading comprehension. In addition, the word-by-word presentation used in the ERP study does not allow a preview of following words, and this rather unnatural reading manner might make predictions or incremental interpretations more difficult or effortful. Although the comparison between different methodologies cannot distinguish potential effects of these factors, the overall findings in the thesis suggest that L2 speakers' prediction can be similar to L1 speakers' prediction under some circumstances. In addition to the important role of time, further research would be able to determine what other factors make L2 prediction more native-like.

While L2 speakers can show native-like predictions under some circumstances (e.g., prediction of meaning when there is a rich context), L2 speakers consistently did not predict word form unlike L1 speakers. As shown in the results from L1 speakers in Experiments 5 and 6, word form prediction is less likely to occur relative to semantic prediction. Therefore, it is possible that L2 speakers do not predict word form information at all. However, the current results do not rule out the possibility that L2 speakers predict word form under some circumstances. The L2 speakers recruited for the experiments in this thesis were all late learners, whose first exposure to the L2 (English) was at the age of five years or later. In addition, the L2 groups in experiments testing word form prediction (Experiments 4, 7, & 8) had either Spanish or Japanese as their L1, both of which are rather dissimilar to English. The age of first exposure to an L2, L2 proficiency and L1-L2 similarity are all known to affect L2 online processing (Foucart & Frenck-Mestre, 2011; Sagarra & Herschensohn, 2010; Steinhauer, White, & Drury, 2009). Therefore, prediction of word form may occur for L2 speakers whose L2 competence is at a near-native level or whose L1 is linguistically similar to their L2. However, a clear relationship between the L2 proficiency and L2 prediction has not yet been shown (but see Chambers & Cooke, 2009, who reported a significant correlation between L2 speakers' predictive eye movements and their L2 proficiency). The question whether L2 proficiency, L1-L2 similarity and age of L2 acquisition affect L2 predictive processing independently (or possibly additively) must be left for future research.

6.2 Implications for the models of language prediction

6.2.1 Role of prediction in language comprehension

With growing evidence that people make predictions during comprehension, prediction has been argued to be an important aspect of language comprehension or human cognition (Altmann & Mirković, 2009; Bar, 2007; Kuperberg & Jaeger, 2016). People tend to be faster to read predictable words compared to unpredictable words (Frisson et al., 2005). In addition, a listener's ability to predict when the speaker's turn ends is likely to make conversations more efficient (De Ruiter, Mitterer, & Enfield, 2006; Magyari & De Ruiter, 2012). While these predictions can indeed be beneficial for the communication system, the findings in the current thesis indicate clear limitations in the circumstances under which people can make predictions. These findings support the proposal that prediction is not necessary to understand language (Huettig & Mani, 2016). As a basis of this argument, Huettig and Mani make several claims: (1) Prediction is largely dependent on the contextual constraints and does not always occur; (2) Not everyone makes prediction; and (3) Prediction depends on available resources. The findings in the current thesis are compatible with all three points.

The first point is consistent with the finding that prediction can depend on cloze probability of predictable words. Most of the studies investigating language prediction have used highly predictive sentences, but contexts in natural conversations tend to be much less predictive than those experimental sentence contexts (Huettig, 2015a). Since failed predictions seem to incur additional processing costs to disconfirm the predicted information (DeLong, Urbach, Groppe, & Kutas, 2011; Van Petten & Luka, 2012), it is not necessarily efficient for a brain to make predictions constantly (Jackendoff, 2002; Kuperberg & Jaeger, 2016). The second point is consistent with the finding that L2 speakers do not always predict like L1 speakers. The third point is consistent with the findings that prediction can be delayed or eliminated when time or resources available are limited.

The data in this thesis indicate that, even in the conditions when there was no evidence for prediction, participants understood the language they were listening to or reading, as reflected in their relatively high accuracy for comprehension questions. Therefore, the data can be interpreted as that people can understand language without making predictions about upcoming words. This would in turn suggest that predictive processing is not automatically implicated in the comprehension system, and can be seen as additional processing that occurs when required conditions are met. The current work has shown that predictions require time and resources, but there are potentially other mediating factors, and those factors can mediate different aspects of predictive processing (see Rommers et al., 2015 for related discussion). More empirical data about what mediates which predictive processes will be able to elucidate what conditions are fundamental to predictions.

6.2.2 Production-based prediction models

The current data also support production-based prediction models (Dell & Chang, 2014; Pickering & Garrod, 2007, 2013), which stipulate that people use their own language production system to make predictions during comprehension. These models argue that prediction of semantic information precedes prediction of word form, because semantic information is accessed before word form information in the production system (Levelt, 1999). Experiments 5 and 6 have shown that prediction of semantic information is more likely to occur compared to prediction of word form. This pattern of the results is compatible with what the models propose. However, as discussed in Section 4.10.3, the results do not exclusively support the production-based prediction accounts, although there appear to be no other models that clearly predict the pattern of results that we found. The other experiments in this thesis do not allow evaluation of this model. One way to test the model more directly could be to compare the time-course of different levels of pre-activation (semantic, syntactic, phonological and orthographic forms) and test whether it also is compatible with production models.

6.2.3 Automaticity of prediction

Experiments 1 and 2 investigated whether relatively simple, meaning-triggered predictions (which even toddlers can make; Mani & Huettig, 2012) can be made automatically. To test the automaticity of prediction, the experiments investigated whether predictions are affected by an additional working memory load during comprehension. If the predictions are automatic, predictive eye movements were expected to be intact under the working memory load. But this was not what we found; the working memory load delayed predictive eye movements, and it made the degree of predictive eye movements smaller in L1 and L2 speakers alike. Therefore, predictions appear to require working memory resources that are used to remember words. This suggests that making predictive eye movements is not an automatic process. However, as discussed in Section 2.9.2, it is unclear which processes involved in predictive eye movements were influenced by the working memory load. Further research would be needed to address this question.

6.3 Implications for the models of L2 sentence processing

As discussed in Section 1.1.1, evidence for prediction ultimately demonstrates that people comprehend sentences incrementally. Current evidence about incrementality in L2 processing is inconsistent. As we have noted, some studies found that L2 comprehension is incremental (e.g., Trenkic et al., 2014), but other studies failed to find such evidence (e.g., Lew-Williams & Fernald, 2010). The findings of this thesis strongly support the notion that L2 comprehension is incremental, at least under some circumstances. Experiment 2 provided evidence that L2 speakers make predictions about upcoming words. L2 speakers utilised semantic information conveyed by verbs rapidly and predicted an upcoming referent that met the semantic restrictions of the verbs. This result demonstrates that L2 processing is incremental; L2 speakers extract and process information of each word as they encounter it. Consistently across the eye-tracking study and ERP study, there was no evidence that L2 speakers predict word form. Nevertheless, the data in both studies support incrementality in L2 processing. The eye-tracking study (Experiment 4) found that L2 speakers were sensitive to phonological overlap between a mentioned word and its phonological competitor word. The ERP study (Experiment 8) found that L2 speakers used top-down contextual information and bottom-up word form information in combination, and activated predictable words that were not seen. Both sets of findings suggest that word form information of an encountered word was processed as it became available. Therefore, the results are overall in line with the idea that L2 speakers process sentences incrementally in a broadly similar way to L1 speakers.

One of the remaining questions about prediction in L2 is the role of L2 proficiency. The experiments in this thesis examined the effects of language proficiency by comparing L1 and L2 speakers, but it is unclear how highly proficient L2 speakers and less proficient L2 speakers differ. Experiment 2 did not find any relationship between L2 proficiency and the degree of L2 speakers' prediction, which was inconsistent with Chambers and Cooke (2009). Given that L2 proficiency affects both semantic processing (Kotz & Elston-Güttler, 2004) and syntactic processing (Ojima, Nakata, & Kakigi, 2005), it would be reasonable to assume a link between L2 proficiency and prediction performance in L2 as well. Further research would be able to provide a clear answer to this end.

6.4 Conclusion

Many psycholinguistic findings suggest that people use various types of cues to predict upcoming language during comprehension. This thesis has provided evidence that successful predictions depend on the conditions under which people comprehend language, and the nativeness of the target language. People may not make predictions when time or resources available are not sufficient, or when they comprehend a non-native language. While successful predictions are beneficial for the comprehension system, by allowing comprehenders to access upcoming information at an early stage, the roles of these mediating factors highlight clear limitations in the prediction system. Hence, predictive processing appears to be additional processing for the comprehension system, and it can be left out when the comprehension is cognitively burdened.

7 References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*, 419–439.

Altarriba, J., Kroll, J. F., Sholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory & Cognition*, *24*(4), 477–492.

Altmann, G. T. M. (2011). Language can mediate eye movement control within 100 milliseconds, regardless of whether there is anything to move the eyes to. *Acta Psychologica*, *137*(2), 190–200.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264.

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518.

Altmann, G. T. M., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, *33*(4), 583–609.

Amsel, B. D., DeLong, K. A., & Kutas, M. (2015). Close, but no garlic: Perceptuomotor and event knowledge activation during language comprehension. *Journal of Memory and Language*, *82*, 118–132.

Anderson, R., Ferreira, F., & Henderson, J. M. (2011). I see what you're saying: The integration of complex speech and scenes during language comprehension. *Acta Psychologica*, *137*(2), 208–216.

Aoshima, S., Yoshida, M., & Phillips, C. (2009). Incremental processing of coreference and binding in Japanese. *Syntax*, *12*(2), 93–134.

Apfelbaum, K. S., Blumstein, S. E., & McMurray, B. (2011). Semantic priming is affected by real-time phonological competition: Evidence for continuous cascading systems. *Psychonomic Bulletin & Review*, *18*(1), 141–149.

Arai, M., & Keller, F. (2013). The use of verb-specific information for prediction in sentence processing. *Language and Cognitive Processes*, *28*(4), 525–560.

Arnold, J. E., Kam, C. L. H., & Tanenhaus, M. K. (2007). If you say thee uh you are describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(5), 914–30.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*(1), 67–82.

Ashby, J., Rayner, K., & Clifton, C. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *58*(6), 1065–1086.

Baayen, H. R., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Baayen, H. R., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database (CD-ROM). Philadelphia, PA.

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1–29.

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*(7), 280–289.

Barr, D. J. (2008). Analyzing “visual world” eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474.

Bates, D. M., Maechler, M., & Dai, B. (2008). Lme4: Linear mixed-effects models using S4 classes.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The Crosslinguistic Study of Sentence Processing* (pp. 3–76). New York: Cambridge University Press.

Bloom, P. A., & Fischler, I. (1980). Completion norms for 329 sentence contexts. *Memory & Cognition*, *8*(6), 631–642.

Blumenfeld, H. K., & Marian, V. (2007). Constraints on parallel activation in bilingual spoken language processing: Examining proficiency and lexical status using eye-tracking. *Language and Cognitive Processes*, *22*(5), 633–660.

Blumenfeld, H. K., & Marian, V. (2013). Parallel language activation and cognitive control during spoken word recognition in bilinguals. *Journal of Cognitive Psychology*, *25*(5), 547–567.

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one’s age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, *112*(4), 417–436.

Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2015). Graded expectations: Predictive processing and the adjustment of expectations during spoken language comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(3), 607–624.

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition*, *136*, 135–149.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–11.

Camblin, C. C., Gordon, P. C., & Swaab, T. Y. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *Journal of Memory and Language*, *56*(1), 103–128.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*(1), 177–208.

Carreiras, M., Monahan, P. J., Lizarazu, M., Duñabeitia, J. A., & Molinaro, N. (2015). Numbers are not like words: Different pathways for literacy and numeracy. *NeuroImage*, *118*, 79–89.

Chambers, C. G., & Cooke, H. (2009). Lexical competition during second-language listening: Sentence context, but not proficiency, constrains interference from the native lexicon. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(4), 1029–1040.

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language*, *47*(1), 30–49.

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *30*(3), 687–696.

Chou, C. J., Huang, H. W., Lee, C. L. Y., & Lee, C. L. Y. (2014). Effects of semantic constraint and cloze probability on Chinese classifier-noun agreement. *Journal of Neurolinguistics*, *31*, 42–54.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*(1), 3–42.

Connolly, J. F., Phillips, N. A., & Forbes, K. A. K. (1995). The effects of phonological and semantic features of sentence-ending words on visual event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, *94*(4), 276–287.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *107*(1), 84–107.

Corley, M. (2010). Making predictions from speech with repairs: Evidence from eye movements. *Language and Cognitive Processes*, *25*(5), 706–727.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*(4), 317–367.

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *30*(2), 498–513.

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, *12*(3), 453–459.

Dambacher, M., Dimigen, O., Braun, M., Wille, K., Jacobs, A. M., & Kliegl, R. (2012). Stimulus onset asynchrony and the timeline of word recognition: Event-related potentials during sentence reading. *Neuropsychologia*, *50*(8), 1852–70.

Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research*, *1084*(1), 89–103.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466.

De Groot, A. M. B., & Nas, G. L. J. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language*, *30*(1), 90–123.

De Groot, F., Huettig, F., & Olivers, C. N. L. (2016). When meaning matters: The temporal dynamics of semantic influences on visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(2), 180–196.

De Ruiter, J.-P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, *82*(3), 515–535.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321.

Dell, G. S., & Chang, F. (2014). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1634), 20120394.

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, *42*, 287–314.

DeLong, K. A., Groppe, D. M., Urbach, T. P., & Kutas, M. (2012). Thinking ahead or not? Natural aging and anticipation during reading. *Brain and Language*, *121*(3), 226–39.

DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia*, *61*, 150–162.

DeLong, K. A., Troyer, M., & Kutas, M. (2014). Pre-processing in sentence comprehension: Sensitivity to likely upcoming meaning and structure. *Language and Linguistics Compass*, *12*, 631–645.

DeLong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology*, *48*(9), 1203–1207.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–21.

Dijkstra, T., Van Jaarsveld, H., & Brinke, S. Ten. (1998). Interlingual homograph recognition: Effects of task demands and language intermixing. *Bilingualism: Language and Cognition*, *1*(1), 51–66.

Dikker, S., Silbert, L. J., Hasson, U., & Zevin, J. D. (2014). On the same wavelength: Predictable language enhances speaker-listener brain-to-brain synchrony in posterior superior temporal gyrus. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(18), 6267–72.

Drake, E., & Corley, M. (2015a). Articulatory imaging implicates prediction during spoken language comprehension. *Memory & Cognition*, *43*(8), 1136–1147.

Drake, E., & Corley, M. (2015b). Effects in production of word pre-activation during listening: Are listener-generated predictions specified at a speech-sound level? *Memory & Cognition*, *43*(1), 111–120.

Dussias, P. E., & Pinar, P. (2010). Effects of reading span and plausibility in the reanalysis of wh-gaps by Chinese-English second language speakers. *Second Language Research, 26*(4), 443–472.

Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand. *Studies in Second Language Acquisition, 35*(2), 353–387.

Ellis, E. M., Borovsky, A., Elman, J. L., & Evans, J. L. (2015). Novel word learning: An eye-tracking study. Are 18-month-old late talkers really different from their typical peers? *Journal of Communication Disorders*, *58*, 43–157.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44*(4), 491–505.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*(4), 469–495.

Federmeier, K. D., & Kutas, M. (2005). Aging in context: Age-related changes in context use during language comprehension. *Psychophysiology*, *42*(2), 133–141.

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, *115*(3), 149–161.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75–84.

Ferrand, L., & Grainger, J. (1994). Effects of orthography are independent of phonology in masked form priming. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 47*(2), 365–382.

Ferreira, F., Foucart, A., & Engelhardt, P. E. (2013). Language processing in the visual world: Effects of preview, visual complexity, and prediction. *Journal of Memory and Language*, *69*(3), 165–182.

Ferretti, T., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language*, *44*(4), 516–547.

FitzPatrick, I., & Indefrey, P. (2010). Lexical competition in nonnative speech comprehension. *Journal of Cognitive Neuroscience*, *22*(6), 1165–1178.

Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Bilingualism: Language and Cognition*, *14*(3), 379–399.

Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1461–1469.

Foucart, A., Ruiz-Tada, E., & Costa, A. (2016). Anticipation processes in L2 speech comprehension: Evidence from ERPs and lexical recognition task. *Bilingualism: Language and Cognition, 19*(1), 213–219.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*(2), 178–210.

Frenzel, S., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2011). Conflicts in language processing: A new perspective on the N400-P600 distinction. *Neuropsychologia*, *49*(3), 574–579.

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: early and late event-related brain potential effects. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *22*(5), 1219–1248.

Frisch, S., Schlesewsky, M., Saddy, D., & Alpermann, A. (2002). The P600 as an indicator of syntactic ambiguity. *Cognition*, *85*(3), 83–92.

Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(5), 862–877.

Gambi, C., & Pickering, M. J. (2011). A cognitive architecture for the coordination of utterances. *Frontiers in Psychology*, *2*, 275.

Gambi, C., & Pickering, M. J. (2013). Prediction and imitation in speech. *Frontiers in Psychology*, *4,* 340.

Garrod, S., Gambi, C., & Pickering, M. J. (2013). Prediction at all levels: Forward model predictions can enhance comprehension. *Language and Cognitive Processes*, *29*(1), 46–48.

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological Science*, *13*(5), 425–430.

Grainger, J., & Ferrand, L. (1996). Masked orthographic and phonological priming in visual word recognition and naming: Cross-task comparisons. *Journal of Memory and Language*, *35*(5), 623–647.

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, *28*(2), 191–215.

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: The P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia*, *38*, 1531–1549.

Hagoort, P., & Indefrey, P. (2014). The neurobiology of language beyond single words. *Annual Review of Neuroscience*, *37*, 347–362.

Hare, M., Elman, J. L., Tabaczynski, T., & McRae, K. (2009). The wind chilled the spectators, but the wine just chilled: Sense, structure, and sentence comprehension. *Cognitive Science*, *33*(4), 610–628.

Hill, H., Ott, F., & Weisbrod, M. (2005). SOA-dependent N400 and P300 semantic priming effects using pseudoword primes and a delayed lexical decision. *International Journal of Psychophysiology*, *56*(3), 209–221.

Hintz, F., & Huettig, F. (2015). The complexity of the visual environment modulates language-mediated eye gaze. In R. Mishra, N. Srinivasan, & F. Huettig (Eds.), *Attention and Vision in Language Processing* (pp. 39–55). Berlin: Springer.

Hintz, F., & Meyer, A. S. (2015). Prediction and production of simple mathematical equations: Evidence from visual world eye-tracking. *PLoS ONE, 10*(7), e0130766.

Hintz, F., Meyer, A. S., & Huettig, F. (2016). Encouraging prediction during production facilitates subsequent comprehension: Evidence from interleaved object naming in sentence context and sentence reading. *The Quarterly Journal of Experimental Psychology*, *69*(6), 1056–1063.

Hirose, Y., & Mazuka, R. (2015). Predictive processing of novel compounds: Evidence from Japanese. *Cognition*, *136*, 350–358.

Holcomb, P. J. (1988). Automatic and attentional processing: An event-related brain potential analysis of semantic priming. *Brain and Language*, *35*, 66–85.

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology*, *30*(1), 47–61.

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, *22*(3), 369–397.

Hopp, H. (2009). The syntax–discourse interface in near-native L2 acquisition: Off-line and on-line performance. *Bilingualism: Language and Cognition*, *12*(4), 463–483.

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*(1), 33–56.

Huettig, F. (2015a). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–135.

Huettig, F. (2015b). Literacy influences cognitive abilities far beyond the mastery of written language. In I. Van de Craats, J. Kurvers, & R. Van Hout (Eds.), *Adult literacy, second language, and cognition. LESLLA Proceedings 2014*. Nijmegen: Centre for Language Studies.

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*, 23–32.

Huettig, F., & Altmann, G. T. M. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition*, *15*(8), 37–41.

Huettig, F., & Altmann, G. T. M. (2011). Looking at anything that is green when hearing “frog”: How object surface colour and stored object colour knowledge influence language-mediated overt attention. *The Quarterly Journal of Experimental Psychology*, *64*(1), 122–145.

Huettig, F., & Brouwer, S. M. (2015). Delayed anticipatory spoken language processing in adults with dyslexia - Evidence from eye-tracking. *Dyslexia*, *21*(2), 97–122.

Huettig, F., & Hartsuiker, R. J. (2008). When you name the pizza you look at the coin and the bread: Eye movements reveal semantic activation during word production. *Memory & Cognition*, *36*(2), 341–360.

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, *31*(1), 80–93.

Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition, and Neuroscience*, *31*(1), 19–31.

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, *57*(4), 460–482.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–71.

Huettig, F., Singh, N., & Mishra, R. K. (2011). Language-mediated visual orienting behavior in low and high literates. *Frontiers in Psychology*, *2*, 285.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*, 101–144.

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language*, *86*, 157–171.

Ito, A., Martin, A. E., & Nieuwland, M. S. (in press). On predicting form and meaning in a second language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Ito, K., Jincho, N., Minai, U., Yamane, N., & Mazuka, R. (2012). Intonation facilitates contrast resolution: Evidence from Japanese adults and 6-year olds. *Journal of Memory and Language*, *66*(1), 265–284.

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, *58*(2), 541–573.

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, *127*(2), 277–88.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

Jackson, C. N., & Dussias, P. E. (2009). Cross-linguistic differences and their impact on L2 sentence processing. *Bilingualism: Language and Cognition*, *12*(1), 65–82.

Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism*, *4*(2), 257–282.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. J. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*(2), 159–201.

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Linguistics and Language Compass*, *2*(4), 647–670.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156.

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive procesing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, *32*(1), 37–55.

Kidd, C., White, K. S. K., & Aslin, R. N. (2011). Toddlers use speech disfluencies to predict speakers' referential intentions. *Developmental Science*, *14*(4), 925–34.

Kim, A. E., & Lai, V. (2012). Rapid interactions between lexical semantic and word form analysis during word recognition in context: Evidence from ERPs. *Journal of Cognitive Neuroscience*, *24*(5), 1104–1112.

Kim, A. E., Oines, L. D., & Sikos, L. (2016). Prediction during sentence comprehension is more than a sum of lexical associations: The role of event knowledge. *Language, Cognition and Neuroscience*, *31*(5), 597–601.

Kim, A. E., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*, 205–225.

Kleinman, D., Runnqvist, E., & Ferreira, V. S. (2015). Single-word predictions of upcoming language during comprehension: Evidence from the cumulative semantic interference task. *Cognitive Psychology*, *79*, 68–101.

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, *95*, 95–127.

Kotz, S. A. (2009). A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain and Language*, *109*, 68–74.

Kotz, S. A., & Elston-Güttler, K. E. (2004). The role of proficiency on processing categorical and associative information in the L2 as revealed by reaction times and event-related brain potentials. *Journal of Neurolinguistics*, *17*, 215–235.

Kouider, S., Halberda, J., Wood, J., & Carey, S. (2006). Acquisition of English number marking: The singular-plural distinction. *Language Learning and Development*, *2*(1), 1–25.

Kroll, J. F., & De Groot, A. M. B. (2005). *Handbook of Bilingualism: Psycholinguistic Approaches*. New York: Oxford University Press.

Kroll, J. F., Gullifer, J. W., & Rossi, E. (2013). The multilingual lexicon: The cognitive and neural basis of lexical comprehension and production in two or more languages. *Annual Review of Applied Linguistics*, *33*, 102–127.

Kukona, A., Fang, S.-Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition*, *119*(1), 23–42.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*(1), 23–49.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language Cognition & Neuroscience*, *31*(1), 32–59.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30 thousand English words. *Behavior Research Methods*, *44*(4), 978–990.

Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes*, *8*(4), 533–572.

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In *Predictions in the Brain: Using Our Past to Generate a Future* (pp. 190–207).

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621–647.

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

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.

Kutas, M., Van Petten, C., & Besson, M. (1988). Event-related potential asymmetries during the reading of sentences. *Electroencephalography and Clinical Neurophysiology*, *69*(3), 218–33.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240.

Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language*, *61*(3), 326–338.

Lau, E., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience*, *25*(3), 484–502.

Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Sciences*, *3*(6), 223–232.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *The Behavioral and Brain Sciences*, *22*(1), 1–38; discussion 38–75.

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, *63*(4), 447–464.

Libben, M. R., & Titone, D. a. (2009). Bilingual lexical access in context: Evidence from eye movements during reading. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(2), 381–390.

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique*. MIT Press.

Luka, B. J., & Van Petten, C. (2014). Prospective and retrospective semantic processing: Prediction, time, and relationship strength in event-related potentials. *Brain and Language*, *135*, 115–129.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, *31*(1), 133–56.

Magyari, L., & De Ruiter, J.-P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Frontiers in Psychology*, *3*, 376.

Mani, N., Daum, M. M., & Huettig, F. (2016). "Proactive" in many ways: Developmental evidence for a dynamic pluralistic approach to prediction. *The Quarterly Journal of Experimental Psychology*, 1–13.

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake — But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 843–847.

Mani, N., & Huettig, F. (2014). Word reading skill predicts anticipation of upcoming spoken language input: A study of children developing proficiency in reading. *Journal of Experimental Child Psychology*, *126*, 264–279.

Marian, V., Spivey, M., & Hirsch, J. (2003). Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging. *Brain and Language*, *86*(1), 70–82.

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, *27*, 53–78.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*, 71–102.

Martin, C. D., Thierry, G., Kuipers, J.-R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, *69*(4), 574–588.

Matsuki, K., Chow, T., & Hare, M. (2011). Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 913–934.

McQueen, J. M., & Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *The Quarterly Journal of Experimental Psychology*, *60*(5), 661–671.

McRae, K., Hare, M., Elman, J. L., & Ferretti, T. (2005). A basis for generating expectancies for verbs from nouns. *Memory & Cognition*, *33*(7), 1174–1184.

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., & Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language*, *66*(4), 545–567.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475–494.

Mishra, R. K., & Singh, N. (2014). Language non-selective activation of orthography during spoken word processing in Hindi-English sequential bilinguals: An eye tracking visual world study. *Reading and Writing*, *27*(1), 129–151.

Mishra, R. K., Singh, N., Pandey, A., & Huettig, F. (2012). Spoken language-mediated anticipatory eye- movements are modulated by reading ability - Evidence from Indian low and high literates. *Journal of Eye Movement Research*, *5*(1).

Mitsugi, S., & MacWhinney, B. (2016). The use of case marking for predictive processing in second language Japanese. *Bilingualism: Language and Cognition*, *19*(1), 19–35.

Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 555–563.

Nakamura, C., Arai, M., & Mazuka, R. (2012). Immediate use of prosody and context in predicting a syntactic structure. *Cognition*, *125*(2), 317–23.

Nation, K., Marshall, C. M., & Altmann, G. T. M. (2003). Investigating individual differences in children’s real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology, 86*(4), 314–329.

Newman, R. L., & Connolly, J. F. (2004). Determining the role of phonology in silent reading using event-related brain potentials. *Cognitive Brain Research*, *21*(1), 94–105.

Newman, R. L., Connolly, J. F., Service, E., & Mcivor, K. (2003). Influence of phonological expectations during a phoneme deletion task: Evidence from event-related brain potentials. *Psychophysiology, 40,* 640–647.

Nieuwland, M. S. (2014). “Who’s he?” Event-related brain potentials and unbound pronouns. *Journal of Memory and Language*, *76*, 1–28.

Nieuwland, M. S. (2015). The truth before and after: Brain potentials reveal automatic activation of event knowledge during sentence comprehension. *Journal of Cognitive Neuroscience*, *27*(11), 2215–2228.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*(3), 189–234.

Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience*, *17*(8), 1212–1228.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31,* 785–806.

Otten, M., Nieuwland, M. S., & Van Berkum, J. J. A. (2007). Great expectations: Specific lexical anticipation influences the processing of spoken language. *BMC Neuroscience, 8*(89).

Otten, M., & Van Berkum, J. J. A. (2008). Discourse-based word anticipation during language processing: Prediction or priming? *Discourse Processes*, *45*(6), 464–496.

Otten, M., & Van Berkum, J. J. A. (2009). Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain Research*, *1291,* 92–101.

Peters, R. E., Grüter, T., & Borovsky, A. (2015). Anticipatory and locally coherent lexical activation varies as a function of language proficiency. In *Proceedings of the Cognitive Science Society*.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, *11*(3), 105–10.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *The Behavioral and Brain Sciences*, *36*(4), 329–47.

Pullum, G. K., & Zwicky, A. M. (1988). The syntax-phonology interface. In F. J. Newmeyer (Ed.)*, Linguistics: The Cambridge Survey; Volume I, Linguistic Theory: Foundations* (pp. 255–280). Cambridge: Cambridge University Press.

Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(20), 7865–7870.

R Development Core Team. (2015). A language and environment for statistical computing. R Foundation for Statistical Computing.

Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(2), 514–528.

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *30*(6), 1290–1301.

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, *3*(4), 504–509.

Riordan, B., Dye, M., & Jones, M. N. (2015). Grammatical number processing and anticipatory eye movements are not tightly coordinated in English spoken language comprehension. *Frontiers in Psychology*, *6*, 590.

Robertson, E. K., Shi, R., & Melançon, A. (2012). Toddlers use the number feature in determiners during online noun comprehension. *Child Development*, *83*(6), 2007–2018.

Roland, D., Yun, H., Koenig, J. P., & Mauner, G. (2012). Semantic similarity, predictability, and models of sentence processing. *Cognition*, *122*(3), 267–279.

Rommers, J., Meyer, A. S., & Huettig, F. (2015). Verbal and nonverbal predictors of language-mediated anticipatory eye movements. *Attention, Perception, & Psychophysics*, *77*(3), 720–730.

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

Sagarra, N., & Herschensohn, J. (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua*, *120*(8), 2022–2039.

Salverda, A. P., & Tanenhaus, M. K. (2010). Tracking the time course of orthographic information in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1108–1117.

Saslow, M. G. (1967). Latency for saccadic eye movement. *Journal of the Optical Society of America*, *57*(8), 1030–1033.

Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action--candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience*, *10*(4), 295–302.

Sedivy, J. C. (2002). Invoking discourse-based contrast sets and resolving syntactic ambiguities. *Journal of Memory and Language*, *46*(2), 341–370.

Segalowitz, N., & Hulstijn, J. (2009). Automaticity in bilingualism and second language learning. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 371–388). Oxford University Press.

Shook, A., & Marian, V. (2013). The bilingual language interaction network for comprehension of speech. *Bilingualism: Language and Cognition*, *16*(2), 304–324.

Slevc, L. R., & Novick, J. M. (2013). Memory and cognitive control in an integrated theory of language processing. *The Behavioral and Brain Sciences*, *36*(4), 373–4.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302–319.

Snedeker, J., & Trueswell, J. C. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, *48*(1), 103–130.

Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science*, *10*(3), 281–284.

Stafura, J. Z., & Perfetti, C. a. (2014). Word-to-text integration: Message level and lexical level influences in ERPs. *Neuropsychologia*, *64*, 41–53.

Staub, A. (2011). The effect of lexical predictability on distributions of eye fixation durations. *Psychonomic Bulletin & Review*, *18*(2), 371–376.

Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language*, *82*, 1–17.

Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research*, *25*(2009), 13–41.

Sussman, R. S., & Sedivy, J. (2003). The time-course of processing syntactic dependencies: Evidence from eye movements. *Language and Cognitive Processes*, *18*(2), 143–163.

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.

Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition*, *16*(2), 367–382.

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

Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, *83*(3), 382–392.

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition*, *27*(2), 173–204.

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, *35*, 454–475.

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, *61*(2), 569–613.

Trenkic, D., Mirković, J., & Altmann, G. T. M. (2014). Real-time grammar processing by native and non-native speakers: Constructions unique to the second language. *Bilingualism: Language and Cognition*, *17*(2), 237–257.

Tribushinina, E., & Mak, W. M. (2016). Three-year-olds can predict a noun based on an attributive adjective: Evidence from eye-tracking. *Journal of Child Language*, *43*(2), 425–441.

Troyer, M., & Borovsky, A. (2015). Childhood SES affects anticipatory language comprehension in college-aged adults. *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, 2428–2433.

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(3), 443–467.

Van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*, 657–671.

Van de Meerendonk, N., Kolk, H. H. J., Chwilla, D. J., & Vissers, C. T. W. M. (2009). Monitoring in language perception. *Linguistics and Language Compass*, *3*(5), 1211–1224.

Van de Meerendonk, N., Kolk, H. H. J., Vissers, C. T. W. M., & Chwilla, D. J. (2010). Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, *22*(1), 67–82.

Van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, *13*(7), 967–985.

Van Herten, M., Chwilla, D. J., & Kolk, H. H. J. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience*, *18*(7), 1181–97.

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language*, *50*(3), 289–335.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176–190.

Vissers, C. T. W. M., Chwilla, D. J., & Kolk, H. H. J. (2006). Monitoring in language perception: The effect of misspellings of words in highly constrained sentences. *Brain Research*, *1106*(1), 150–163.

Waters, G., Caplan, D., & Yampolsky, S. (2003). On-line syntactic processing under concurrent memory load. *Psychonomic Bulletin & Review*, *10*(1), 88–95.

Weber, A., & Crocker, M. W. (2012). On the nature of semantic constraints on lexical access. *Journal of Psycholinguistic Research*, *41*(3), 195–214.

Weber, A., Grice, M., & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition*, *99*(2), 63–72.

Weber-Fox, C. M., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience*, *8*(3), 231–256.

Wheeldon, L. R., & Monsell, S. (1992). The locus of repetition priming of spoken word production. *The Quarterly Journal of Experimental Psychology*, *44*(4), 723–61.

Wicha, N. Y. Y., Bates, E., Moreno, E. M., & Kutas, M. (2003). Potato not Pope: Human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters*, *346*(3), 165–168.

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2003). Expecting gender: An event related brain potential study on the role of grammatical gender in comprehending a line drawing within a written sentence in Spanish. *Cortex*, *39*(3), 483–508.

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, *16*(7), 1272–1288.

Williams, J. N. (2006). Incremental interpretation in second language sentence processing. *Bilingualism: Language and Cognition*, *9*(1), 71–88.

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, *7*(7), 701–702.

Wlotko, E. W., & Federmeier, K. D. (2013). Two sides of meaning: The scalp-recorded N400 reflects distinct contributions from the cerebral hemispheres. *Frontiers in Psychology*, *4*, 181.

Wlotko, E. W., & Federmeier, K. D. (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading. *Cortex*, *68*, 20–32.

Xiang, M., & Kuperberg, G. R. (2015). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience*, *30*(6), 648–672.

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *32*(1), 1–14.

8 Appendix

8.1 Study 1: Critical sentences and visual objects

Sentence	**Target**	**Semantic competitor**	**Distractors**
The boy will close/ touch the cabinet.	cabinet	table	teddy bear, yo-yo
The lady will fold/ find the scarf.	scarf	high heels	piano, violin
The boy will catch/ describe the dragonfly.	dragonfly	eagle	shark, whale
The man will fire/ bring the gun.	gun	bomb	watering can, water hose
The man will fly/ check the airplane.	airplane	motorcycle	computer, television
The boy will beat/ choose the drum.	drum	guitar	video game, puzzles
The magician will bend/ move the spoon.	spoon	cup	coin, paper
The housewife will mop/ wash the floor.	floor	carpet	pants, skirt
The teacher will answer/ open the door.	door	window	book, letter
The woman will heat/ fetch the pan.	pan	knife	purse, ring
The child will dress/ borrow the doll.	doll	board game	newspaper, comic book
The woman will climb/ use the stairs.	stairs	elevator	bus, train
The woman will light/ clean the lamp.	lamp	bed	plate, bowl
The woman will iron/ wear the shirt.	shirt	shoes	earrings, necklace
The man will shoot/ need the bow.	bow	sword	spatula, fork
The girl will sharpen/ buy the pencil.	pencil	ruler	slippers, running shoes

8.2 Study 2: Critical sentences and visual objects

Critical sentences and the names of critical objects for each condition (predictable, English competitor, Japanese competitor, unrelated, respectively) in Experiments 3 and 4. The names in square brackets are the Japanese translations written in alphabets. Cloze values collected from L1 and L2 speakers are shown after each sentence in brackets respectively. Items No. 8 and 10 were removed from the analyses in L1 speakers.

No.	**Sentence (L1 cloze; L2 cloze)**	**Object names**
1	In order to have a closer look, the dentist asked the man to open his mouth a little wider. (100; 100)	mouth[kuti]/ mouse[nezumi]/ socks[kutusita]/ bone[hone]
2	It takes about an hour to fly from Edinburgh to London, and about 4 hours by train, usually. (92; 83)	train[densya]/ tray[tore-]/ calculator[dentaku]/ goat[yagi]
3	In an emergency, we cannot use a lift; instead, we need to use the stairs for our safety. (100; 83)	stairs[kaidan]/ stapler[hottikisu]/ seashell[kaigara]/ onion[tamanegi]
4	If the sun comes out during a heavy shower, you can sometimes see a rainbow in the sky. (100; 92)	rainbow[niji]/ radio[rajikase]/ meat[niku]/ barrel[taru]
5	The tourists expected rain when the sun went behind the cloud, but the weather got better later. (100; 92)	cloud[kumo]/ clown[piero]/ bear[kuma]/ globe[tikyuugi]

6	The woman forgot to affix a stamp when posting the letter, and she got it back yesterday. (92; 82)	letter[tegami]/ lettuce[retasu]/ handcuff[tejyou]/ cat[neko]
7	The man didn't know the time because he forgot to wear the watch that he usually wears. (100; 92)	watch[tokei]/ washing machine[sentakuki]/ bird[tori]/ stamp[kitte]
8	The expensive wine is made from a special kind of grape that is grown only in the South of France. (92; 100)	grape[budou]/ grave[haka]/ pig[buta]/ comb[kusi]
9	To make sushi, the chef went to the market to buy some fish early in the morning. (100; 83)	fish[sakana]/ finger[yubi]/ dice[saikoro]/ elephant[zou]
10	To protect against an enemy's bullet or arrows, soldiers used to carry a shield all the time. (92; 67)	shield[tate]/ sheep[hituji]/ bamboo[take]/ giraffe[kurin]
11	The man was gathering honey, when he was stung by a bee and gave a cry. (100; 83)	bee[hati]/ bean[mame]/ flag[hata]/ tiger[tora]
12	The child believed that Santa Claus would come into her house down the chimney at midnight. (100; 83)	chimney[entotu]/ chick[hiyoko]/ pencil[enpitu]/ spoon[supu-n]
13	People can easily go to the island on foot since the government built a bridge last year. (100; 83)	bridge[hasi]/ brick[renga]/ ladder[hasigo]/ key[kagi]
14	The traveller went to the desert because he wanted to ride a camel and go exploring. (100; 100)	camel[rakuda]/ camera[kamera]/ racket[raketto]/ toothbrush[haburasi]
15	The woman found the room was too hot and humid, so to get some fresh air, she opened the window completely. (94; 100)	window[mado]/ windmill[huusya]/ match[matti]/ corn[toumorokosi]
16	The bird cannot fly because it injured its wing when it had a fight with another bird. (100; 83)	wing[hane]/ witch[majyo]/ nose[hana]/ candle[rousoku]

8.3 Study 3 & 4: Critical sentences

All 160 sentences from Experiments 5 – 8 are listed below with critical words for each of the four conditions (predictable word, form-related word, semantically related word, unrelated word, respectively). The mean plausibility ratings from L1 and L2 speakers for the four conditions per sentences are shown in brackets after each critical word. The mean cloze values of the predictable word collected from L1 and L2 speakers are shown in brackets after each sentence.

1. The student is going to the library to borrow a book (*L1 plausibility* 4.7; *L2 plausibility* 4.8)/ hook (1.3; 1.2)/ page (1.9; 2)/ sofa (2.1; 1.6) tomorrow. (*L1 cloze value* 100; *L2 cloze value* 100)

2. Living alone is too expensive, so the students will share a flat (4.7; 4.8)/ flag (1.5; 1.8)/ wall (1.2; 1.4)/ bell (1.6; 1.2) together. (92; 67)

3. The family enjoyed the sunny day, but there will be rain (4; 4.4)/ pain (2.5; 2)/ sky (1.8; 1.6)/ loss (1.8; 1.2) tomorrow. (72; 50)

4. Jack studied medicine in a university and works as a doctor (4.3; 5)/ factor (1.5; 1.2)/ patient (1.2; 1.2)/ tenant (2.1; 2.4) now. (94; 83)

5. Amelia got a driving licence, and will buy her own car (4.4; 5)/ jar (2.1; 1.6)/ tyre (1.7; 2)/ rat (2.4; 1.6) soon. (97; 83)

6. Oliver doesn't have a watch, so he doesn't know the time (4.3; 4.4)/ lime (1.3; 1)/ rest (1.7; 2)/ bean (1.3; 1.2) now. (100; 83)

7. Rachel will go to the cinema to watch a new film (4.6; 4.6)/ firm (1.6; 1.2)/ camera (1.6; 1.4)/ bird (2.4; 1.8) tomorrow. (81; 50)

8. Paul is trying to stand on one leg (4.7; 5)/ lag (1.8; 2)/ hip (1.3; 2.4)/ kid (2.4; 2.2) now. (64; 17)

9. The gambler kept losing, so he doesn't have any money (4.7; 4.8)/ honey (1.7; 1)/ wallet (2.2; 1.8)/ candle (1.3; 1.2) left. (92; 58)

10. Harry intends to propose to Emily and give her the ring (4.6; 5)/ wing (1.6; 1.2)/ finger (1.8; 1.6)/ memo (2.5; 2) tomorrow. (100; 92)

11. John nervously asked the attractive girl out on a date (4.6; 4.8)/ gate (1.7; 2)/ cancel (1.2; 1)/ pin (1.4; 1) yesterday. (100; 67)

12. As a lifetime vegetarian, Olivia doesn't miss eating meat (4; 4.4)/ mean (1.7; 1.4)/ flour (1.5; 2)/ soil (1.7; 1.2) now. (89; 83)

13. Dylan got lost today, so he will use a map (3.8; 4.6)/ cap (1.6; 2)/ globe (1.9; 1.8)/ job (1.4; 1) tomorrow. (92; 58)

14. The comedian was funny, despite a bad joke (4.6; 4.8)/ coke (1.7; 1.6)/ laugh (1.3; 3.8)/ beef (1.4; 1) yesterday. (97; 75)

15. Jacob found he misspelled the word (4.4; 4.6)/ lord (2; 1.4)/ usage (1.8; 3)/ oven (2.1; 1.8) earlier. (86; 67)

16. Oscar opened the postbox, and found a letter (4.6; 4.2)/ litter (2.2; 1.4)/ heading (1.8; 1.6)/ birth (1.3; 1) there. (89; 58)

17. After struggling with the question, Jessica got the answer (4.6; 4.8)/ dancer (1.5; 1.4)/ inquiry (2.1; 3.8)/ pension (2.3; 2.4) finally. (89; 92)

18. At the airport, James checked in for his flight (4.6; 4.8)/ sight (1.7; 2)/ rocket (1.5; 2.4)/ machine (1.6; 1.4) earlier. (86; 50)

19. The lottery gave Emily a car as a prize (4.2; 3.8)/ price (2.3; 4.4)/ medal (1.8; 2.2)/ child (2.3; 2.2) yesterday. (75; 67)

20. Sophie couldn't recall the recent event, and blamed her bad memory (4.6; 4.8)/ melody (1.9; 2)/ storage (2; 2)/ eraser (1.4; 1.2) yesterday. (92; 83)

21. The shoes were small, so Lily asked for the largest size (3.5; 4.8)/ sign (1.8; 1.2)/ height (1.5; 2.6)/ flip (1.5; 1.4) available. (86; 58)

22. To view the 3D image, people wore special glasses (4.6; 4.6)/ classes (1.3; 2.2)/ eyes (1.8; 2)/ markets (1.4; 1.8) yesterday. (97; 92)

23. At the football match, Bob scored a goal (4.6; 4.8)/ coal (1.2; 1)/ team (1.5; 1.4)/ bear (1.3; 1.8) yesterday. (94; 75)

24. Grace put too much dressing on her salad (4.7; 4.8)/ ballad (1.6; 1.4)/ refrigerator (1.9; 2.2)/ movie (1.6; 1.2) yesterday. (86; 83)

25. Chloe couldn't afford the necklace because of its high price (4.7; 4.8)/ pride (1.3; 1.8)/ seller (1.4; 1.2)/ radio (1.5; 1.4) sadly. (92; 58)

26. The family went to the sea to catch some fish (4.5; 4.4)/ wish (1.5; 1.4)/ pond (1.5; 2.2)/ echo (1.3; 1.4) together. (67; 33)

27. Noah missed the final bus, and needed to take a taxi (4.7; 4.8)/ tax (1.8; 1.2)/ fare (1.4; 2.4)/ seed (1.4; 1.2) yesterday. (69; 42)

28. The workers reported the difficult problem to their boss (4.6; 4.8)/ bass (1.2; 1.2)/ job (2.3; 2.6)/ port (1.9; 2.4) yesterday. (42; 58)

29. Freya had a serious car accident and is afraid of driving (4.7; 5)/ thriving (2.1; 1.2)/ licences (2; 1.2)/ finding (1.7; 1.4) now. (67; 17)

30. The man was convicted for murder and is in prison (4.6; 5)/ poison (1.6; 1.4)/ crime (1.3; 1.4)/ image (1.8; 1.4) now. (53; 25)

31. Beth loved cooking, and has become a world famous chef (4.7; 4.4)/ shelf (1.3; 1.2)/ buffet (1.3; 1.6)/ aunt (1.8; 1.4) now. (97; 67)

32. Kyle asked the dentist to pull out the painful tooth (4.5; 5)/ booth (1.3; 1.4)/ brush (1.5; 3)/ grade (1.4; 1) gently. (94; 75)

33. Having high blood pressure, George reduced his intake of salt (4.5; 4.2)/ malt (2.3; 3.8)/ sea (1.3; 2.8)/ bond (1.3; 1.6) considerably. (33; 42)

34. Children were excited to see the first snow this winter (4.7; 4.8)/ printer (1.3; 1)/ summer (2.4; 2.8)/ effect (1.4; 1.4) yesterday. (72; 58)

35. To see the new-born panda, Lucy will go to the zoo (4.6; 4.8)/ loo (1.4; 1.4)/ lion (1.6; 1.2)/ end (2.3; 1.8) tomorrow. (100; 92)

36. The country girl was overwhelmed to see streets full of people (4.6; 4.8)/ purple (2.9; 1.8)/ customs (2.3; 3.4)/ length (1.8; 1.6) yesterday. (64; 83)

37. Daisy is nine months pregnant and will have her first baby (3.9; 4.2)/ bay (2.2; 3.2)/ nappy (1.9; 2.8)/ agent (1.8; 1.8) soon. (75; 50)

38. Students at the train station are rushing to buy a ticket (4.7; 4.8)/ thicket (1.8; 1.4)/ platform (1.5; 1)/ major (1.5; 1.4) now. (92; 83)

39. Feeling stressed at his workplace, Max relaxed at home (4; 4.6)/ dome (1.8; 1)/ laundry (2.7; 3.4)/ beer (1.8; 1.4) completely. (72; 58)

40. Isabella dyed her hair, but she doesn't like the colour (5; 4.2)/ cutter (2.3; 2.2)/ paint (2.5; 2)/ grape (1.5; 2.6) now. (100; 50)

41. For parking illegally, William was charged a fine (4.8; 3.6)/ line (1.4; 1.2)/ court (1.5; 1.8)/ rest (1.2; 1) yesterday. (81; 67)

42. Jim will go swimming and get a suntan at the beach (4.7; 4.4)/ peach (1.4; 1.2)/ coconut (1.8; 1.2)/ drama (1.2; 1.4) tomorrow. (81; 75)

43. The house is haunted by ghosts (4.7; 4.4)/ boasts (1.8; 1.4)/ halloween (2.5; 1.6)/ eagles (2.1; 2.2) now. (89; 67)

44. The juice isn't cold enough, so Alice is adding some ice (5; 4.2)/ dice (1.4; 2)/ cube (1.9; 3.8)/ wine (2.4; 1.8) now. (100; 92)

45. The restaurant is always busy, so Leo will book a table (4.9; 4.6)/ label (1.8; 2.8)/ chair (2.3; 2)/ field (1.8; 2.4) now. (92; 75)

46. The baby is hungry and needs to drink some milk (4.6; 5)/ silk (1.6; 1.4)/ cow (1.3; 1.6)/ debt (1.2; 1.4) now. (86; 83)

47. Before sending the letter, Daniel licked a stamp (4.8; 4.8)/ stump (2; 1.2)/ payment (1.7; 1.6)/ juice (1.8; 1.4) quickly. (78; 42)

48. The waiter got a generous tip because of his good service (4.9; 4.2)/ surface (1.6; 2)/ complaints (1.5; 2.2)/ million (1.1; 1.4) yesterday. (64; 50)

49. For their mother's birthday, the children will hold a party (4.4; 4.6)/ pasty (1.8; 1.8)/ guest (1.4; 1.2)/ scene (1.9; 2.4) tomorrow. (92; 33)

50. Only one cake is left, so Lilly doesn't have a choice (4.8; 3.6)/ voice (2.2; 1.4)/ future (2; 1.6)/ minute (1.6; 1.4) anyway. (33; 17)

51. The men are watching football and drinking beer in the pub (4.8; 4.8)/ rub (1.4; 2.2)/ owner (1.7; 1.2)/ let (1.8; 2.2) together. (64; 33)

52. The plane crash was avoided by the experienced pilot (4.6; 3.4)/ pallet (1.8; 1.4)/ flight (1.9; 2.6)/ mail (1.2; 1.2) yesterday. (97; 92)

53. The bus driver charged Rosie a regular fare (4.4; 5)/ care (1.5; 1.2)/ cash (1.8; 2.2)/ twin (1.3; 1.2) yesterday. (50; 25)

54. Henry was seriously injured but the doctor saved his life (4.8; 5)/ knife (2.7; 1.4)/ death (1.5; 2.2)/ corn (1.3; 1.8) successfully. (81; 75)

55. Joseph used a lighter to make a fire (4.6; 5)/ hire (1.4; 1.2)/ chimney (1.4; 1.4)/ statue (1.3; 1.8) easily. (67; 50)

56. The famous dancer performed on the stage (4.8; 4.8)/ state (2.2; 2.4)/ actor (1.9; 1.4)/ cloud (2; 1.6) yesterday. (92; 67)

57. The girls are going to have cocktails and dance in a club (4.7; 4.4)/ crab (1.3; 1.6)/ sport (1.8; 1.4)/ mist (2.2; 2.2) together. (89; 25)

58. The supporters wished the team good luck (5; 5)/ duck (1.6; 3)/ yell (2; 1.4)/ view (1.2; 2.2) yesterday. (100; 92)

59. The cricket player wants his own glove, ball and bat (5; 2.8)/ rat (2.9; 2.2)/ hit (2.1; 3)/ sneeze (1.4; 2) now. (78; 33)

60. Hannah bought a calendar and hung it on the wall (4.9; 4.4)/ mall (1.7; 2)/ floor (1.5; 2.2)/ fruit (1.7; 1.4) yesterday. (94; 67)

61. The man with the history of self-harm cut his wrist (4.9; 3.8)/ list (1.8; 1.8)/ grip (2.1; 2.6)/ sky (1.2; 1) suddenly. (81; 8)

62. Susan felt tears coming as she sliced the onion (4.6; 5)/ union (2.8; 2.4)/ flavour (1.3; 1.4)/ error (1.1; 1.6) carefully. (81; 50)

63. The woman was hit by a truck when crossing the road (4.8; 4.2)/ load (1.4; 2.4)/ cyclist (1.9; 1.8)/ math (1.2; 1) yesterday. (89; 25)

64. Tyler knows many cocktail recipes because he works in a bar (4.9; 5)/ war (1.8; 1.4)/ pint (1.3; 1.4)/ joy (1.2; 1.8) now. (100; 67)

65. Kate didn't like coffee, so she ordered a cup of tea (5; 5)/ team (1.3; 1.2)/ mug (1.3; 1.4)/ myth (1.2; 1.4) instead. (97; 83)

66. Sam plays guitar in a popular band (5; 4.8)/ sand (1.4; 1.6)/ music (2.3; 1.8)/ hero (1.1; 2) now. (67; 42)

67. As the trousers were loose, Tony tightened up his belt (4.5; 4.8)/ beat (1.5; 1.4)/ helmet (1.9; 1.6)/ tube (1.3; 1.2) yesterday. (86; 75)

68. Meg will go to the park to walk her dog (5; 4.4)/ fog (1.4; 1.4)/ tail (1.6; 1)/ tyre (1.3; 2) tomorrow. (92; 83)

69. The terrorist shot and killed five people using a gun (4.8; 5)/ gum (1.5; 1.4)/ grip (1.8; 1.8)/ rib (1.3; 2) yesterday. (83; 67)

70. The mole was digging a hole (4.9; 3)/ pole (1.5; 1.8)/ drill (1.3; 2)/ mass (1.6; 1) yesterday. (92; 92)

71. The horse went outside the course, and didn't win the race (4.8; 4.8)/ lace (1.9; 2)/ bike (1.4; 1.2)/ snap (1.8; 1.4) understandably. (78; 83)

72. Emma loves a bargain, and everything she bought was on sale (4.5; 3.8)/ safe (1.5; 2.2)/ refunds (2.4; 3.2)/ noon (1.3; 1.2) actually. (86; 58)

73. The king's throne was taken by his first-born son (4.2; 4)/ ton (1.4; 1.2)/ father (1.4; 1.2)/ fox (1.2; 1.4) yesterday. (78; 17)

74. The camp leader taught children how to pitch a tent (4.9; 3.6)/ cent (1.4; 1.4)/ cave (1.3; 2)/ flaw (1.3; 2.8) yesterday. (83; 0)

75. He enjoys hiking in the woods because of the fresh air (4.7; 3.8)/ hair (1.7; 2.4)/ dust (2.8; 2.2)/ ray (1.3; 2.2) outside. (92; 75)

76. All the colleagues have savings accounts at the same bank (4.3; 4.4)/ back (2.1; 1.8)/ receipt (1.2; 1)/ lake (1.3; 1.4) somehow. (75; 75)

77. Ben went to the gym to swim in the pool (4.8; 4)/ tool (1.4; 1.8)/ sink (1.4; 1.2)/ lump (1.3; 1) earlier. (100; 92)

78. Adam keeps different breeds of cows in his farm (4.3; 5)/ harm (1.6; 1.2)/ yields (1.5; 2.2)/ navy (1.3; 1.6) now. (53; 75)

79. For relaxation, Matilda soaked in a hot bath (4.2; 2.8)/ path (1.4; 1.4)/ soap (2.3; 1.4)/ loaf (1.3; 2.4) yesterday. (94; 33)

80. Lewis lost his memory because of the damage to his brain (4.5; 4.8)/ grain (1.4; 1.2)/ surgeries (1.5; 2.2)/ nation (1.2; 1.8) yesterday. (92; 83)

81. After shuffling, the croupier asked the guest to select one card (3.7; 4.4)/ cart (1.6; 3.4)/ swap (1.5; 1.6)/ roll (1.9; 3.2) only. (75; 75)

82. Nobody knows the time as this room has no clock (4.3; 4.8)/ clerk (1.6; 2.2)/ alarm (1.7; 3.6)/ scarf (1.3; 1.4) now. (97; 75)

83. Ryan refused to invest to avoid taking a risk (4.5; 5)/ disc (1.8; 1.4)/ benefit (1.1; 1.6)/ door (1.2; 1.2) yesterday. (53; 58)

84. The client immediately signed the contract because it was a good deal (4.8; 5)/ meal (1.6; 2.4)/ trust (2.4; 3)/ flower (1.5; 1.4) indeed. (78; 75)

85. By closely examining a painting, one can see all the detail (4.7; 3.8)/ retail (2.1; 1.6)/ paragraphs (2; 1.8)/ syrup (2.6; 1.8) clearly. (39; 0)

86. Toby used to walk to school, but now he takes a bus (4.4; 5)/ bug (1.5; 1)/ seat (2.3; 1.8)/ use (1.3; 1.2) often. (69; 75)

87. In the class, Bella whispered in her friend's ear (4.9; 4.2)/ rear (2.3; 1.6)/ sound (1.4; 2)/ kin (1.4; 1.8) quickly. (97; 92)

88. Two dogs injured each other when they had a fight (5; 4.6)/ light (1.8; 1)/ troop (1.6; 1.2)/ piano (1.5; 1) yesterday. (92; 92)

89. To make meringue, she separated the whites from five eggs (4.4; 4.8)/ pegs (1.4; 1.8)/ nests (2; 1.8)/ toes (1.2; 1.6) carefully. (83; 100)

90. To make two groups, the lecturer split the class in half (5; 4.4)/ calf (1.3; 1.2)/ dozen (1.8; 2)/ lake (1.4; 1.6) quickly. (72; 42)

91. Seeing buds on trees heralds the arrival of spring (4.2; 4.2)/ sprint (1.4; 1.6)/ winter (1.3; 3.6)/ power (1.3; 1.4) surely. (94; 75)

92. The store was so busy that the clerk needed help (5; 4.4)/ heap (1.4; 2)/ lifeguards (1.9; 1.8)/ fog (1.1; 1.2) yesterday. (81; 58)

93. The solution didn't work, and Harley lost all hope (4.9; 4.8)/ hose (1.5; 1.2)/ regret (2.5; 2)/ nails (2.3; 2) yesterday. (33; 33)

94. To remember to buy everything she wanted, Gracie made a list (4.8; 5)/ lift (1.6; 1.4)/ volume (1.2; 1.8)/ disc (2; 1.4) quickly. (100; 92)

95. The computer pointer doesn't move though Tommy is moving the mouse (4.8; 4)/ mouth (1.6; 1.2)/ click (2.4; 3.4)/ lemon (2.1; 1.4) now. (94; 92)

96. Katie looks much younger than her actual age (5; 4.2)/ ace (1.4; 1.4)/ birth (1.4; 2.4)/ oak (1.3; 1.2) now. (94; 50)

97. After the meal, Matthew asked a waiter to bring the bill (5; 5)/ pill (1.8; 2)/ tip (2.1; 2.2)/ mess (1.8; 1.4) quickly. (83; 67)

98. Hearing the noise outside the classroom, the lecturer closed the door (4.9; 4.6)/ donor (1.3; 1.6)/ knob (2; 2.4)/ loan (1.3; 1) immediately. (94; 100)

99. Elizabeth doesn't believe Matt since he has told lies (4.8; 4.8)/ pies (1.4; 1.2)/ truth (2; 2.2)/ ways (2.1; 1.4) before. (83; 100)

100. The thief stole the bike easily as it had no lock (5; 4.6)/ look (1.5; 2.2)/ door (2.1; 1.6)/ sum (1.8; 1.4) yesterday. (75; 42)

101. Anna brings a sandwich and a salad for her lunch (4.9; 5)/ punch (1.8; 1.4)/ kitchen (2.3; 2.2)/ issue (1.9; 1.6) usually. (81; 67)

102. People saw the first spaceship that landed on the moon (4.8; 5)/ mood (1.8; 3.8)/ orbit (2.1; 3.6)/ heel (1.4; 1.4) together. (100; 75)

103. Children made a wish when they saw a shooting star (4.8; 4.4)/ scar (2.4; 1.8)/ galaxy (2.1; 2.8)/ devil (1.3; 2.2) yesterday. (97; 92)

104. Maya got a cold and has a runny nose (5; 3)/ pose (2; 1.8)/ chin (1.8; 1.8)/ term (1.5; 1.8) now. (100; 92)

105. The country has no war and people live in peace (4.8; 4.6)/ peak (1.5; 1)/ treaty (2.5; 2.2)/ yacht (2; 1.6) today. (78; 83)

106. To make the garden more green, Amy is growing various plants (4.8; 4.6)/ planets (1.4; 1.2)/ sunlight (1.1; 1.2)/ limits (1.8; 1.6) now. (81; 67)

107. To prevent spreading her cold, Jane is wearing a mask (4.8; 4.2)/ task (1.4; 1.8)/ face (1.6; 1.6)/ pane (1.7; 1.8) now. (72; 25)

108. The leaking rain was due to a hole in the roof (3.8; 4.2)/ root (1.3; 1.8)/ floor (2.6; 3.4)/ pain (1.3; 1.2) yesterday. (86; 58)

109. The fisherman was trying to catch fish in a net (4.4; 4.2)/ jet (1.6; 2.4)/ gross (1.6; 1.6)/ man (1.4; 1.2) earlier. (72; 42)

110. The plant lacks sunlight because it's placed in the shade (4.2; 4.8)/ shape (1.3; 1)/ grass (2.1; 2.4)/ broom (1.5; 1.6) now. (47; 42)

111. Sophia cannot find a suitable lotion because of her delicate skin (4.9; 5)/ spin (1.8; 1.2)/ sweat (2.3; 2.4)/ copy (1.8; 1.2) now. (92; 67)

112. The driver was stopped as he exceeded the specified speed (4.8; 4.8)/ speech (1.5; 1.6)/ jet (1.5; 1.8)/ tablet (1.8; 1.8) yesterday. (64; 67)

113. Conner went down the stairs and sat on the bottom step (5; 3.8)/ stem (1.8; 1.8)/ process (1.1; 1.2)/ jail (1.6; 2.6) slowly. (86; 33)

114. With a high salary, he needs to pay more tax (3.8; 5)/ wax (1.3; 1)/ income (2; 1.4)/ ash (1.5; 1) accordingly. (89; 92)

115. Tilly's sister did the cleaning today, so tomorrow it's her turn (4.5; 4.8)/ turf (1.5; 1.2)/ gear (1.6; 1.4)/ acid (1.3; 1.4) naturally. (61; 50)

116. There was a recycling campaign to reduce the amount of waste (5; 5)/ paste (1.8; 1)/ toilets (1.8; 1.4)/ relief (1.4; 1.4) more. (58; 50)

117. Jamie bravely stopped the robbery without feeling any fear (4.8; 5)/ year (1.4; 1)/ escape (1.5; 1.4)/ dirt (1.8; 1.8) yesterday. (67; 67)

118. The attendees can't miss the meeting without a good excuse (4.9; 5)/ excise (1.5; 1.2)/ doubt (1.9; 1.6)/ style (1.7; 2) tomorrow. (44; 50)

119. Frank will double-check the notification in case there is any change (4.7; 4.8)/ range (2; 1.4)/ same (1.8; 1.4)/ land (2.5; 1.8) tomorrow. (42; 8)

120. Finishing his study abroad, David will return to his own country (4.5; 5)/ counter (2.1; 1.8)/ import (1.1; 1.8)/ puzzle (2.4; 1.6) finally. (72; 58)

121. The businessman left his laptop on his desk (5; 4)/ dusk (1.2; 1.2)/ receptionist (2.8; 2)/ pine (1.9; 1.4) yesterday. (78; 42)

122. Rose couldn't eat noodles using chopsticks, so used a fork (4.3; 4.8)/ fort (1.2; 1.8)/ cup (2.2; 2.4)/ peer (1.6; 1.4) instead. (94; 92)

123. The bird cannot fly because it injured its wing (4.6; 4.2)/ ring (1.3; 1)/ glide (2; 1.6)/ frog (1.4; 1.2) earlier. (94; 92)

124. Eliza worried about her breath, so she took an extra mint (4.5; 5)/ hint (1.3; 1.6)/ herb (2.6; 2.6)/ toll (2.3; 2.4) yesterday. (89; 25)

125. Changing majors required students to fill out a twenty-page long form (4.6; 3.4)/ norm (1.2; 3.6)/ example (2.2; 1.4)/ rival (1.6; 1.6) usually. (61; 50)

126. Violet left the dirty plates and cups in the sink (4.7; 4.2)/ link (1.3; 2)/ towel (2.3; 1.8)/ army (1.8; 1.8) today. (89; 42)

127. To increase her hair volume, the woman wears a wig (4.2; 3.8)/ pig (1.5; 1)/ comb (2.3; 2.8)/ pea (1.6; 1.2) usually. (53; 42)

128. Eleanor covered the old ugly floor with a large rug (4.7; 4.6)/ rum (1.2; 1.4)/ tie (2.3; 1.8)/ ham (1.9; 1.4) completely. (72; 8)

129. The room with bad ventilation got a ceiling fan (4.1; 4)/ fat (1.2; 1.4)/ air (1.4; 2.6)/ kit (2.2; 1.8) finally. (69; 25)

130. For a removal, Cameron packed the TV into its original box (4.3; 4)/ boa (1.2; 1.8)/ lid (1.6; 1.8)/ oil (1.5; 1.2) carefully. (72; 58)

131. Selling drugs is against the law (4.3; 4.6)/ saw (1.5; 1.8)/ jury (1.8; 2)/ fee (1.5; 1.8) today. (100; 83)

132. The waiter wasn't polite, so he didn't receive a good tip (4.8; 4.6)/ lip (1.3; 1.4)/ thumb (1.8; 1.4)/ van (1.8; 1.6) yesterday. (100; 100)

133. For Christmas, the children are hanging bells on the tree (4.6; 4.6)/ treat (1.4; 1.6)/ squirrel (2.3; 2.2)/ inch (1.5; 1.8) happily. (92; 75)

134. To expand their market, the project team made a rough plan (4.7; 3)/ plank (1.3; 2.8)/ future (1.5; 2.2)/ rifle (1.6; 1.2) together. (42; 8)

135. Because of the storm, the ocean has big waves (4.3; 4.8)/ caves (1.3; 1.4)/ surfers (1.8; 2.4)/ heads (1.9; 1.8) now. (92; 75)

136. The sales staff forgot to attach the price tag (4.6; 3.8)/ tug (1.5; 2)/ name (1.6; 1)/ eve (1.6; 1.6) again. (92; 42)

137. To make a pancake easily, Julia used a pancake mix (4.7; 3.6)/ fix (1.4; 2.4)/ digestion (1; 1)/ bid (1.6; 1.8) yesterday. (56; 33)

138. The new variety show appointed the entertainer as a host (4.4; 4.2)/ post (2.4; 2.6)/ meeting (1.1; 1.2)/ drill (1.8; 1.2) yesterday. (31; 8)

139. Nigel's son should inherit the estate according to his will (4.7; 3.6)/ pill (1.3; 2)/ fact (1.8; 1.6)/ site (2.1; 2.4) naturally. (86; 42)

140. It's expected to snow as it will get very cold (4.3; 4)/ old (1.3; 1)/ hot (1.4; 1)/ tall (1.5; 2.2) tomorrow. (94; 92)

141. The explorer in the desert hopes to ride the camel (4.4; 4)/ caramel (2; 1.8)/ oasis (2.1; 3.2)/ user (1.8; 1.6) tomorrow. (89; 83)

142. For the parade, the king's servants will refurbish the entire castle (4.3; 3.8)/ cattle (1.1; 3.6)/ lords (1.6; 2.4)/ depth (1.5; 2.2) perfectly. (33; 17)

143. Making traditional Indian curry requires using several types of spice (4.8; 5)/ space (1.3; 1.4)/ orient (1.6; 2.4)/ excess (1.7; 1.6) together. (81; 8)

144. To finish the cake, Scarlet spread the whipped cream (4.4; 4.4)/ dream (2.2; 3.2)/ pizzas (1.5; 1.4)/ fibre (1.7; 1.8) generously. (94; 75)

145. Students learned how to convert kilometres to miles (4.8; 4.6)/ piles (1.4; 1.2)/ riders (1.8; 1.4)/ veins (1.5; 1.8) yesterday. (58; 83)

146. To compress the air, the machine applies high pressure (4.1; 4.4)/ pleasure (2.2; 2.6)/ relief (1.5; 2)/ session (1.5; 1.4) constantly. (94; 92)

147. Before exchanging money, Rebecca asks the exchange rate (4.6; 4.4)/ fate (1.2; 2.4)/ pace (2.1; 1.4)/ exit (1.4; 1.8) usually. (69; 67)

148. Lydia cannot eat anymore as she is so full (4.4; 4.6)/ dull (1.7; 2.4)/ half (1.4; 1)/ mild (1.7; 1.6) now. (89; 50)

149. Dogs have a good sense of smell (4.5; 4.4)/ shell (1.2; 1)/ nose (1.6; 2.2)/ cash (1.2; 1.4) naturally. (92; 58)

150. Laura will eat the ice cream quickly before it melts (4.7; 4.2)/ meets (1.6; 1.8)/ boils (1.7; 1.6)/ opens (2.3; 2) down. (100; 83)

151. After the main course, Sara checked the dessert menu (4.8; 4.6)/ venue (1.5; 3)/ chef (1.9; 1.6)/ bond (1.2; 2.4) excitedly. (97; 58)

152. Andrew was late because his train had a delay (4.2; 4.4)/ decay (1.2; 1.4)/ time (1.5; 1.8)/ tone (1.5; 1.8) again. (69; 58)

153. The bomb expert pinpointed a switch to make the bomb explode (4.3; 3.2)/ explore (1.8; 2)/ pour (2.1; 2.4)/ stretch (1.7; 1.8) finally. (33; 67)

154. Immigration exposed Lisa to a different culture (4.3; 3.6)/ vulture (1.3; 1.6)/ sociology (2.1; 3.4)/ ginger (1.8; 1.4) naturally. (72; 42)

155. The bridge was washed away by the flood (4.3; 4.6)/ blood (1.7; 1.8)/ soil (2.5; 1.8)/ glove (1.4; 1.4) yesterday. (39; 8)

156. The story was far from logical and didn't make any sense (4.7; 4.8)/ fence (1.2; 1.2)/ taste (1.8; 1.6)/ button (1.4; 1.6) completely. (100; 83)

157. Having no ink or paper, the office workers couldn't print (4.3; 4)/ point (1.3; 2)/ erase (2.1; 2)/ move (2.3; 1.6) anything. (58; 67)

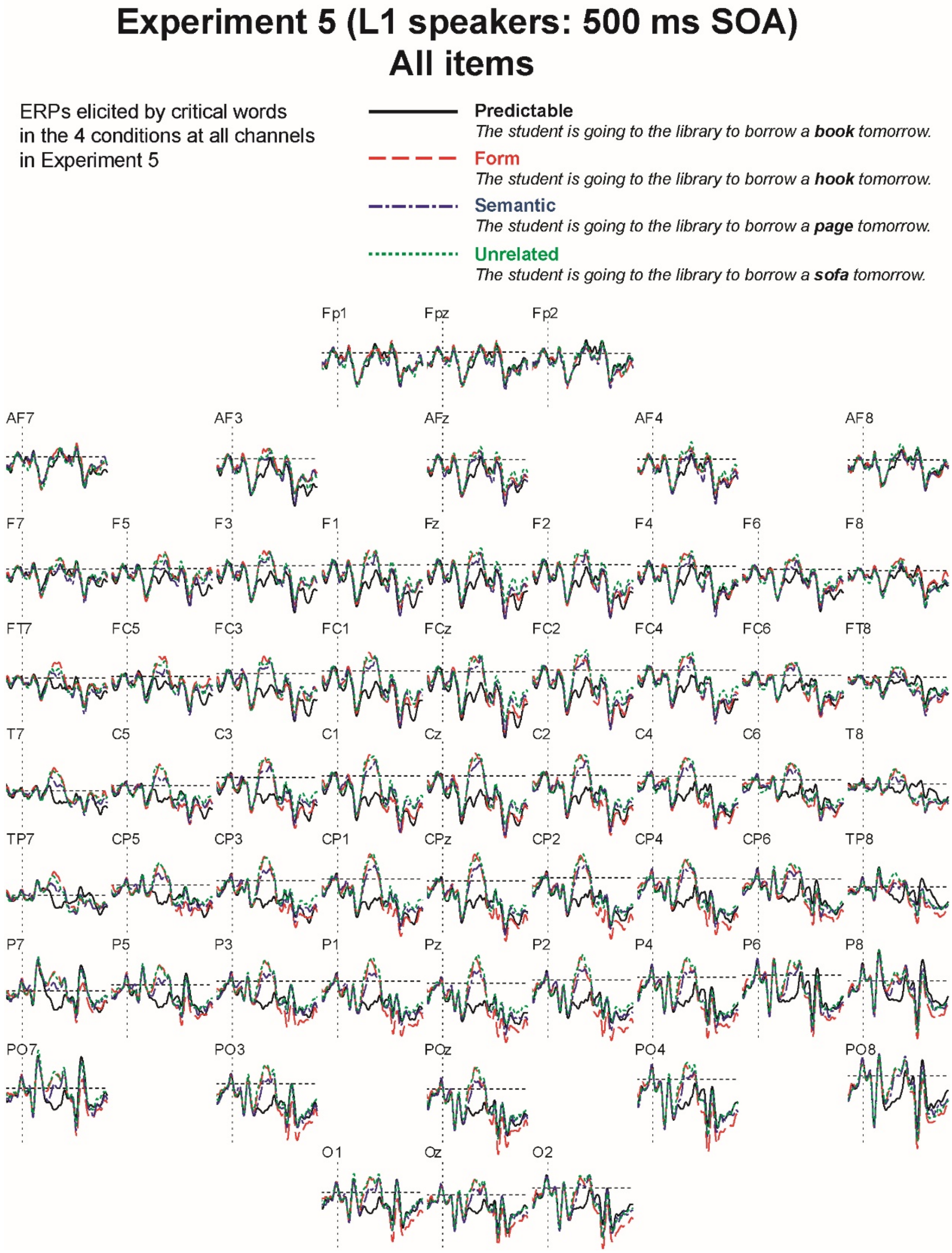
158. In the tennis lesson, Lauren hit the ball with her racket (4.6; 4.4)/ rocket (1.6; 1.8)/ game (1.4; 2)/ area (1.9; 1.6) well. (86; 83)

159. Joe has grown his moustache, but will give it a shave (4.2; 4.2)/ share (1.3; 1.4)/ hair (1.3; 1.2)/ turtle (1.7; 1.2) tomorrow. (44; 8)

160. It's so itchy that Chris can't help scratching the mosquito bite (4.2; 4.2)/ bike (1.2; 1.4)/ chew (1.6; 3)/ tape (2.3; 1.6) constantly. (100; 83)

8.4 Study 3: Supplementary figures

Figures showing results at all channels in Experiments 5 and 6.



Experiment 5 (L1 speakers: 500 ms SOA)
All items

ERPs elicited by critical words in the 4 conditions at all channels in Experiment 5

Experiment 5 (L1 speakers: 500 ms SOA)
High cloze items

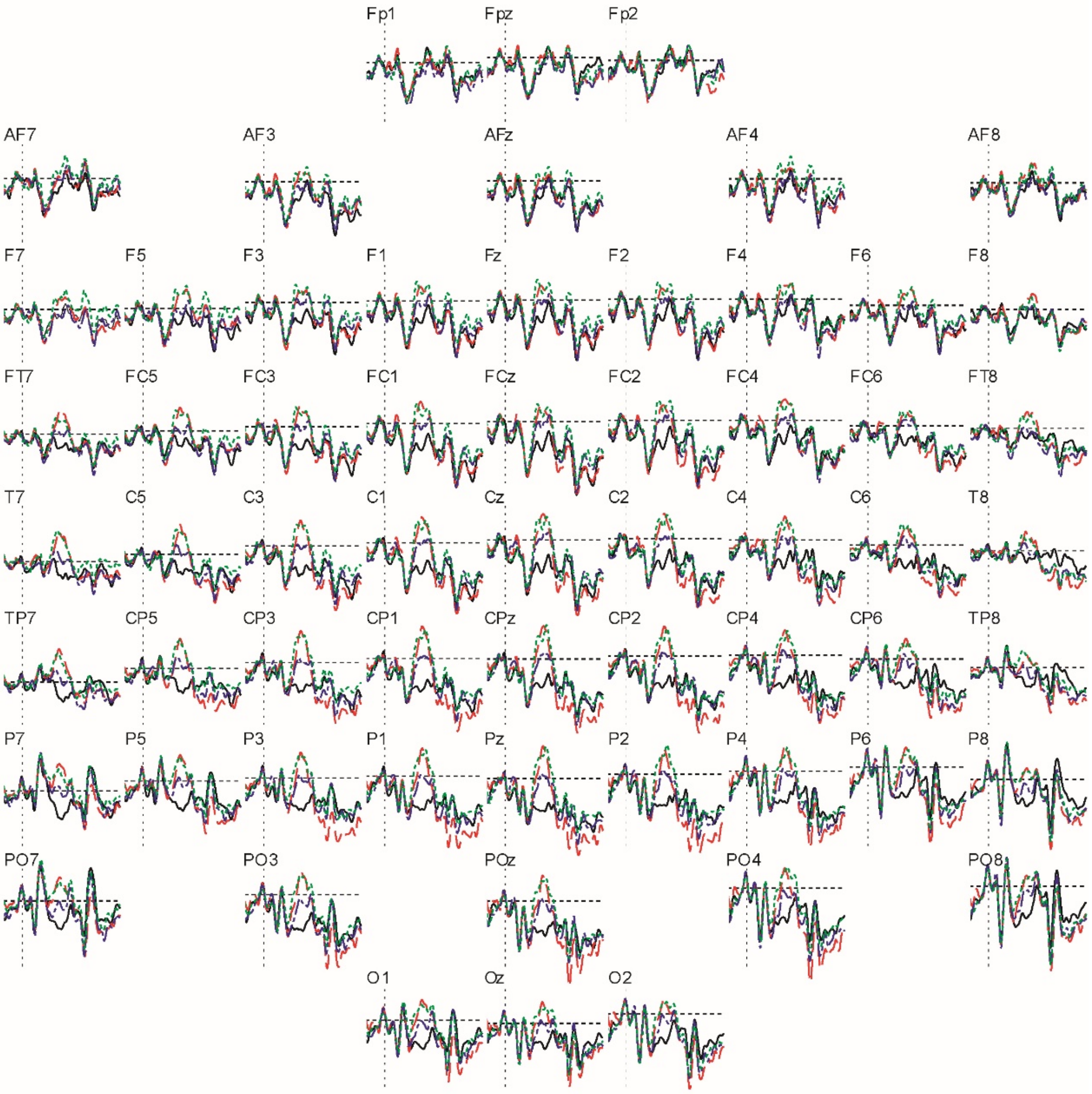
ERPs elicited by critical words
in the 4 conditions at all channels
in Experiment 5

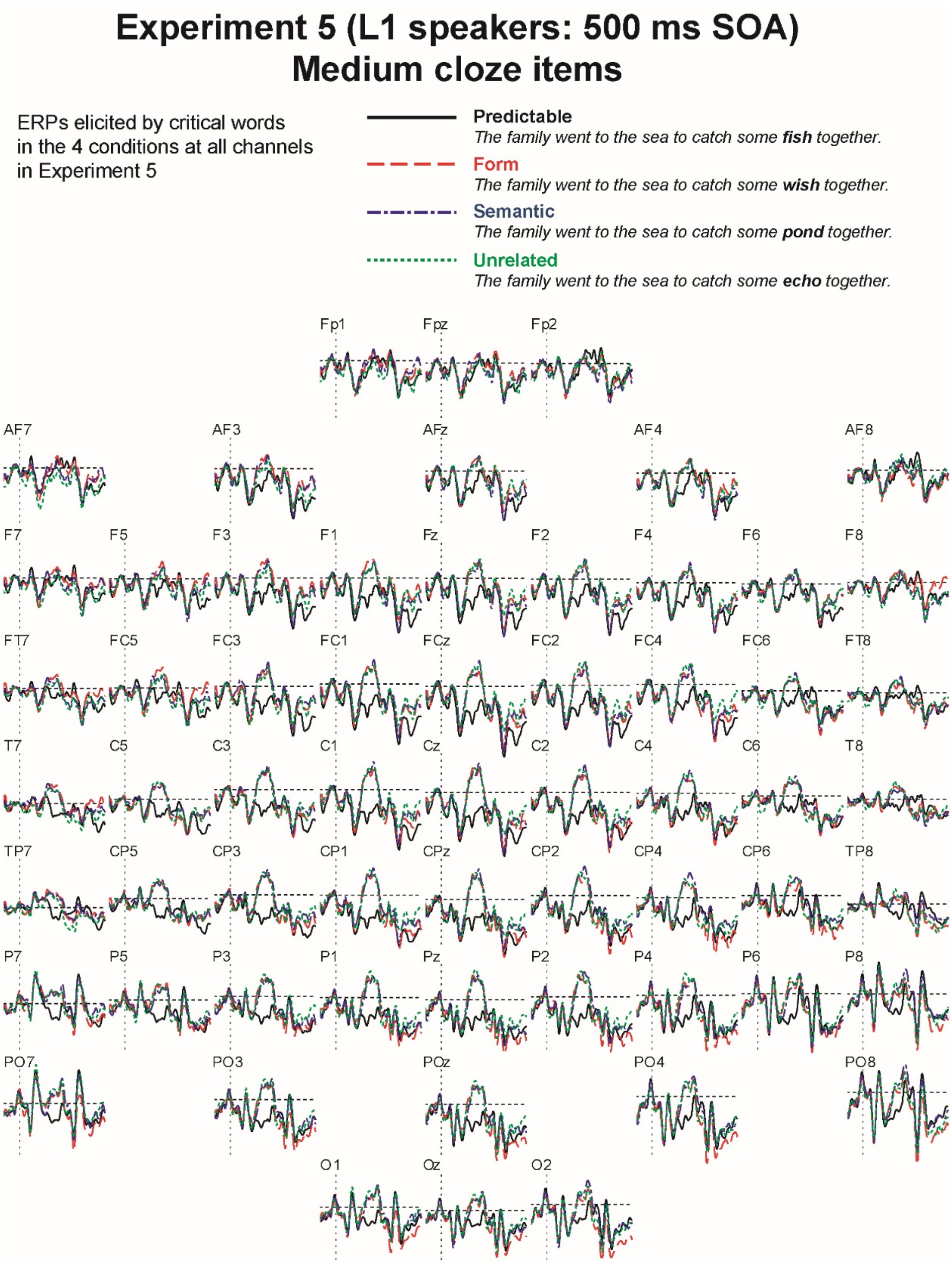
Predictable
The Juice isn't cold enough, so Alice is adding some ***ice*** *now.*

Form
The Juice isn't cold enough, so Alice is adding some ***dice*** *now.*

Semantic
The Juice isn't cold enough, so Alice is adding some ***cube*** *now.*

Unrelated
The Juice isn't cold enough, so Alice is adding some ***wine*** *now.*




Experiment 5 (L1 speakers: 500 ms SOA)
Medium cloze items
ERPs elicited by critical words in the 4 conditions at all channels in Experiment 5
Predictable
The family went to the sea to catch some **fish** together.
Form
The family went to the sea to catch some **wish** together.
Semantic
The family went to the sea to catch some **pond** together.
Unrelated
The family went to the sea to catch some **echo** together.
Fp1
Fpz
Fp2
AF7
AF3
AFz
AF4
AF8
F7
F5
F3
F1
Fz
F2
F4
F6
F8
FT7
FC5
FC3
FC1
FCz
FC2
FC4
FC6
FT8
T7
C5
C3
C1
Cz
C2
C4
C6
T8
TP7
CP5
CP3
CP1
CPz
CP2
CP4
CP6
TP8
P7
P5
P3
P1
Pz
P2
P4
P6
P8
PO7
PO3
POz
PO4
PO8
O1
Oz
O2

Experiment 6 (L1 speakers: 700 ms SOA)
All items

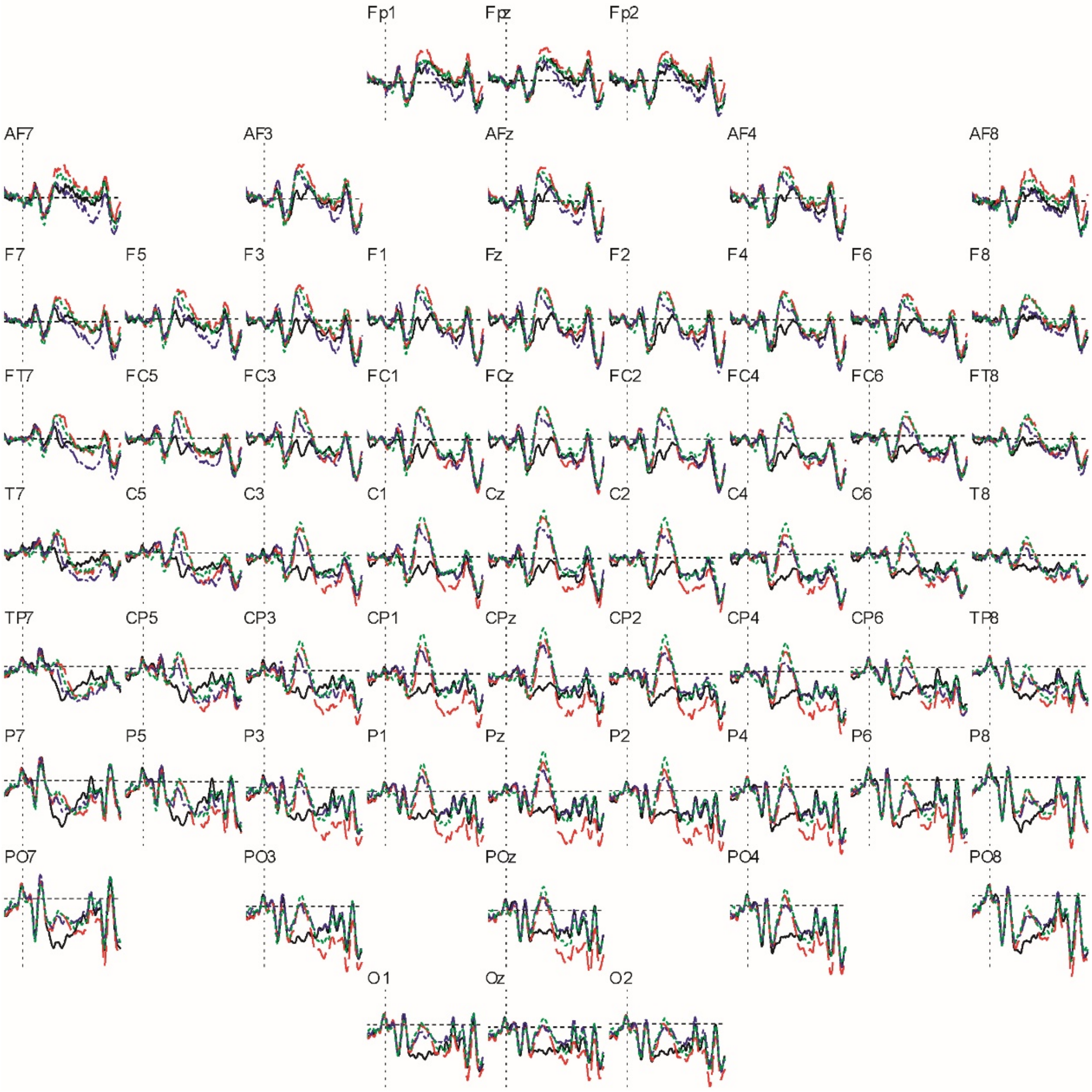
ERPs elicited by critical words
in the 4 conditions at all channels
in Experiment 6

Predictable
*The student is going to the library to borrow a **book** tomorrow.*

Form
*The student is going to the library to borrow a **hook** tomorrow.*

Semantic
*The student is going to the library to borrow a **page** tomorrow.*

Unrelated
*The student is going to the library to borrow a **sofa** tomorrow.*



Experiment 6 (L1 speakers: 700 ms SOA)
High cloze items

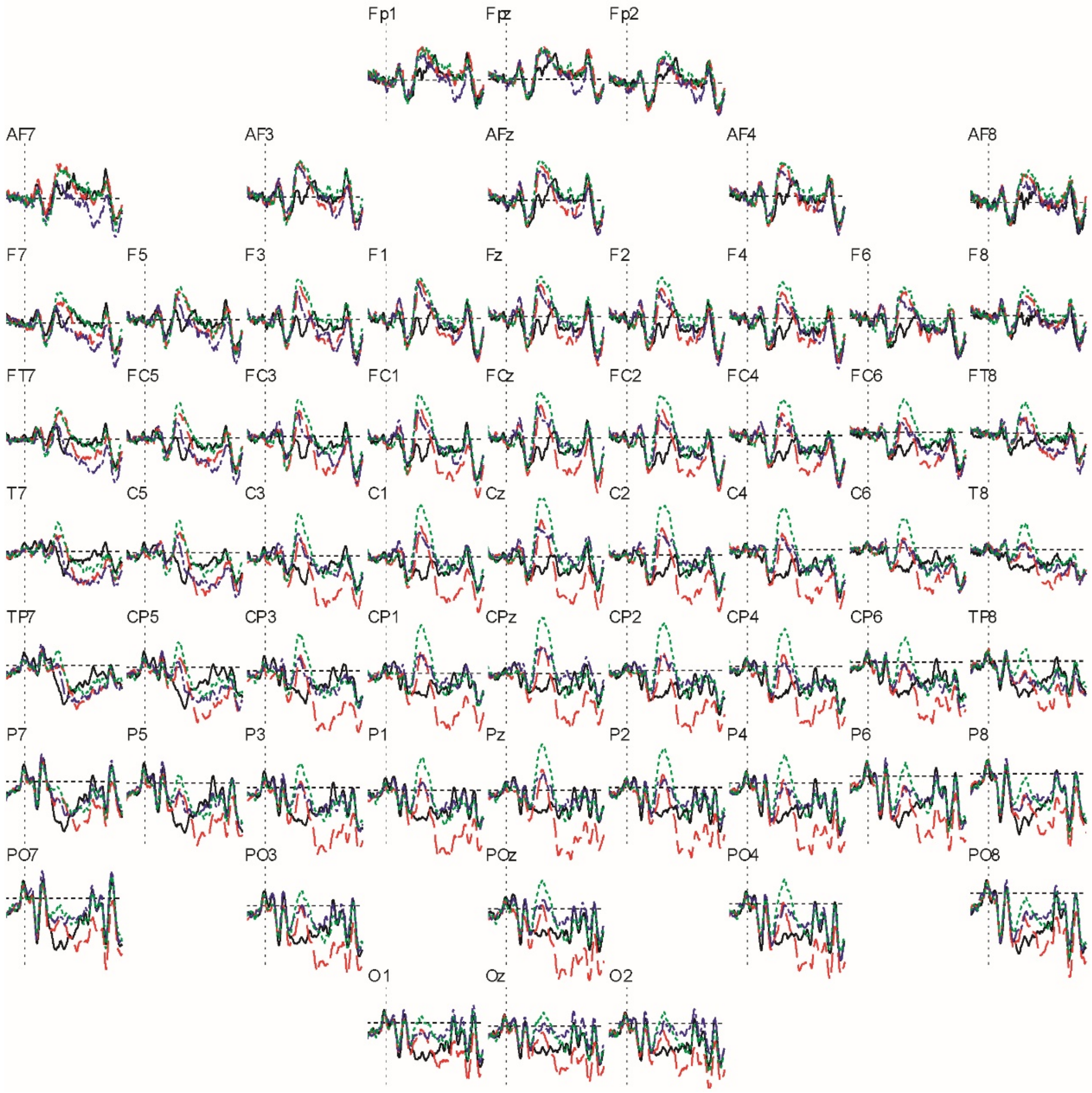
ERPs elicited by critical words in the 4 conditions at all channels in Experiment 6

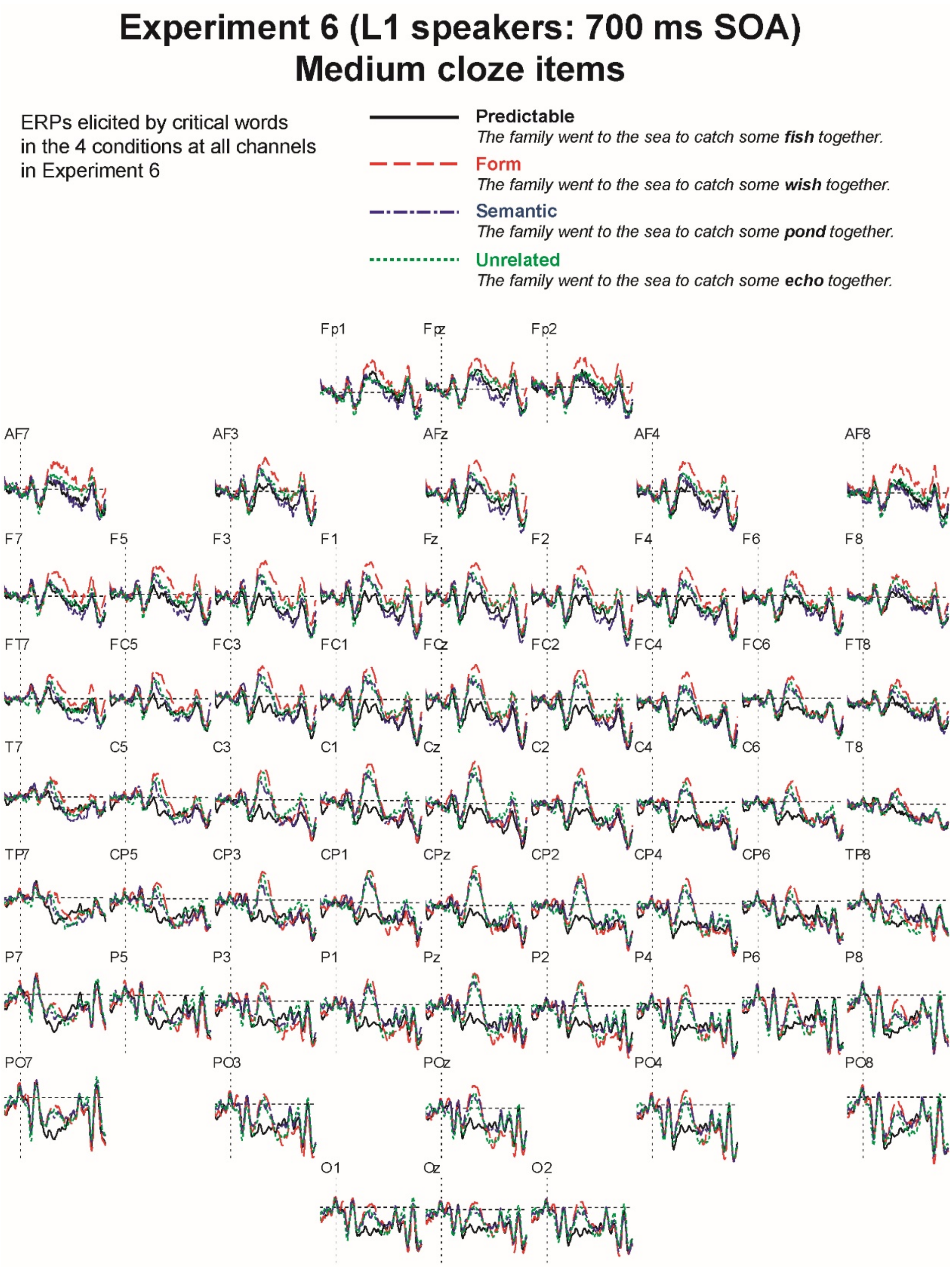
Predictable
The Juice isn't cold enough, so Alice is adding some ***ice*** *now.*

Form
The Juice isn't cold enough, so Alice is adding some ***dice*** *now.*

Semantic
The Juice isn't cold enough, so Alice is adding some ***cube*** *now.*

Unrelated
The Juice isn't cold enough, so Alice is adding some ***wine*** *now.*





Experiment 6 (L1 speakers: 700 ms SOA)
Medium cloze items

ERPs elicited by critical words in the 4 conditions at all channels in Experiment 6

Predictable
The family went to the sea to catch some ***fish*** *together.*

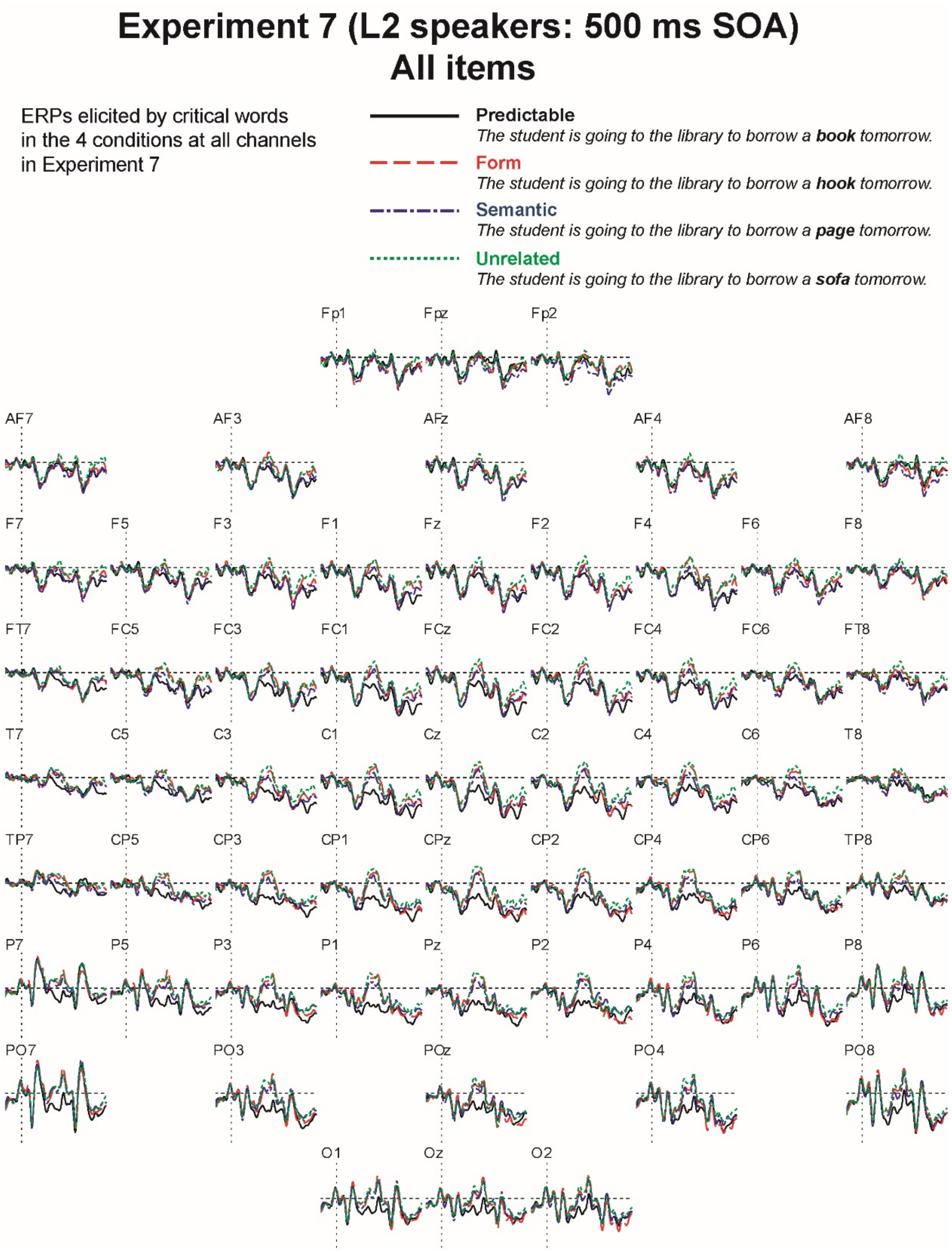
Form
The family went to the sea to catch some ***wish*** *together.*

Semantic
The family went to the sea to catch some ***pond*** *together.*

Unrelated
The family went to the sea to catch some ***echo*** *together.*

8.5 Study 4: Supplementary figures

Figures showing results at all channels in Experiments 7 and 8.



Experiment 8 (L2 speakers: 700 ms SOA)
All items

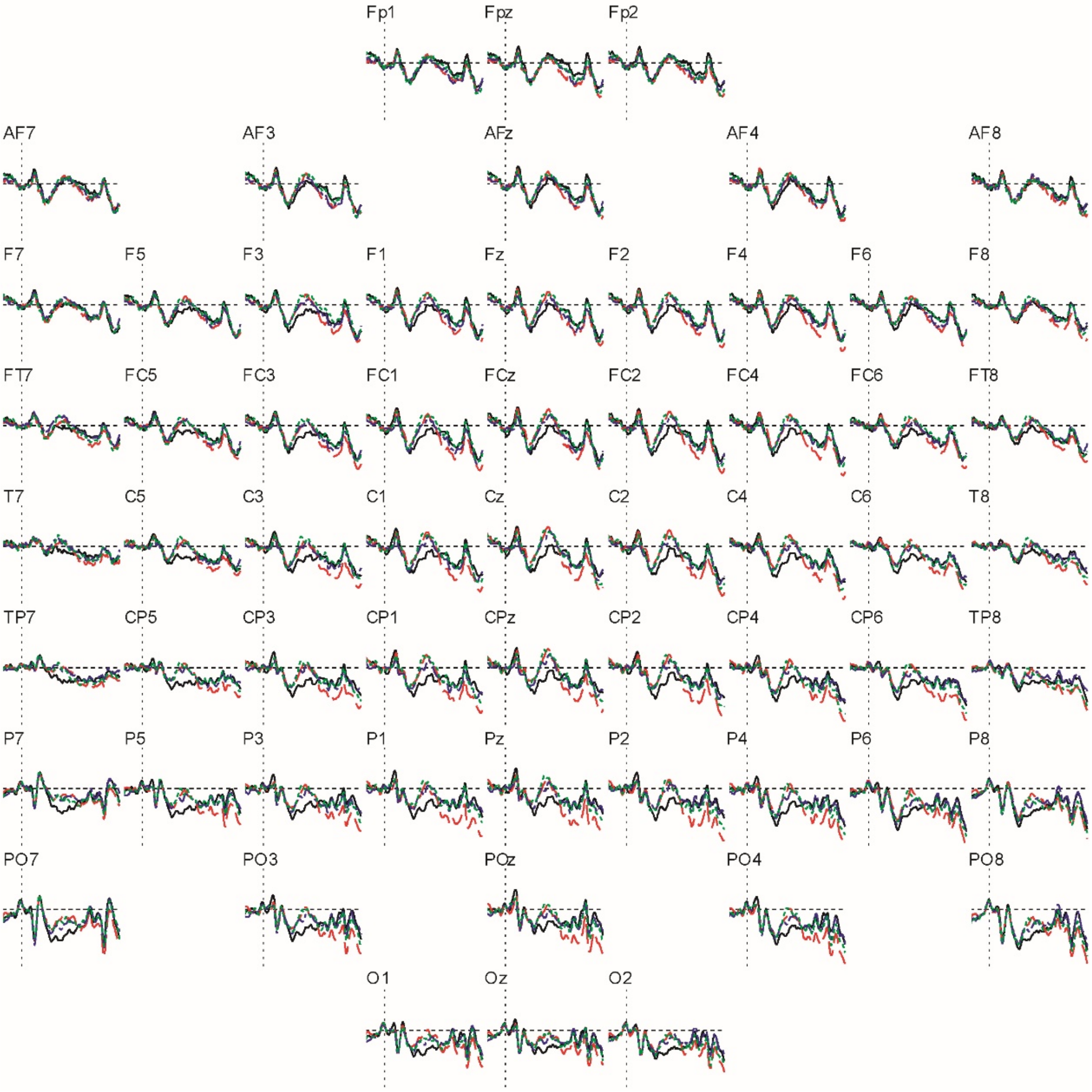
ERPs elicited by critical words
in the 4 conditions at all channels
in Experiment 8

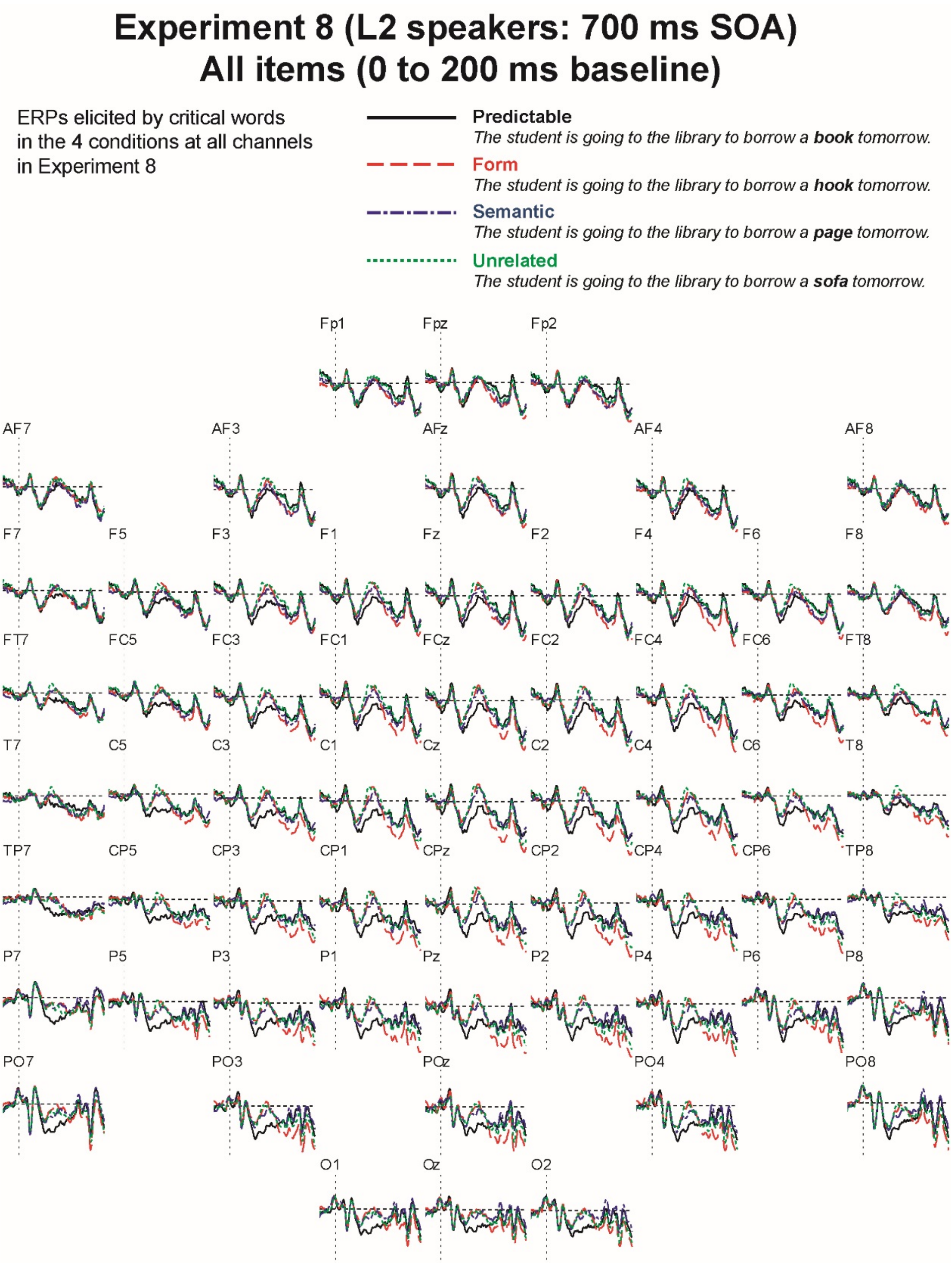
Predictable
*The student is going to the library to borrow a **book** tomorrow.*

Form
*The student is going to the library to borrow a **hook** tomorrow.*

Semantic
*The student is going to the library to borrow a **page** tomorrow.*

Unrelated
*The student is going to the library to borrow a **sofa** tomorrow.*





Experiment 8 (L2 speakers: 700 ms SOA)
All items (0 to 200 ms baseline)

ERPs elicited by critical words in the 4 conditions at all channels in Experiment 8

Predictable
The student is going to the library to borrow a ***book*** *tomorrow.*

Form
The student is going to the library to borrow a ***hook*** *tomorrow.*

Semantic
The student is going to the library to borrow a ***page*** *tomorrow.*

Unrelated
The student is going to the library to borrow a ***sofa*** *tomorrow.*

Experiment 8 (L2 speakers: 700 ms SOA)
High cloze items (0 to 200 ms baseline)

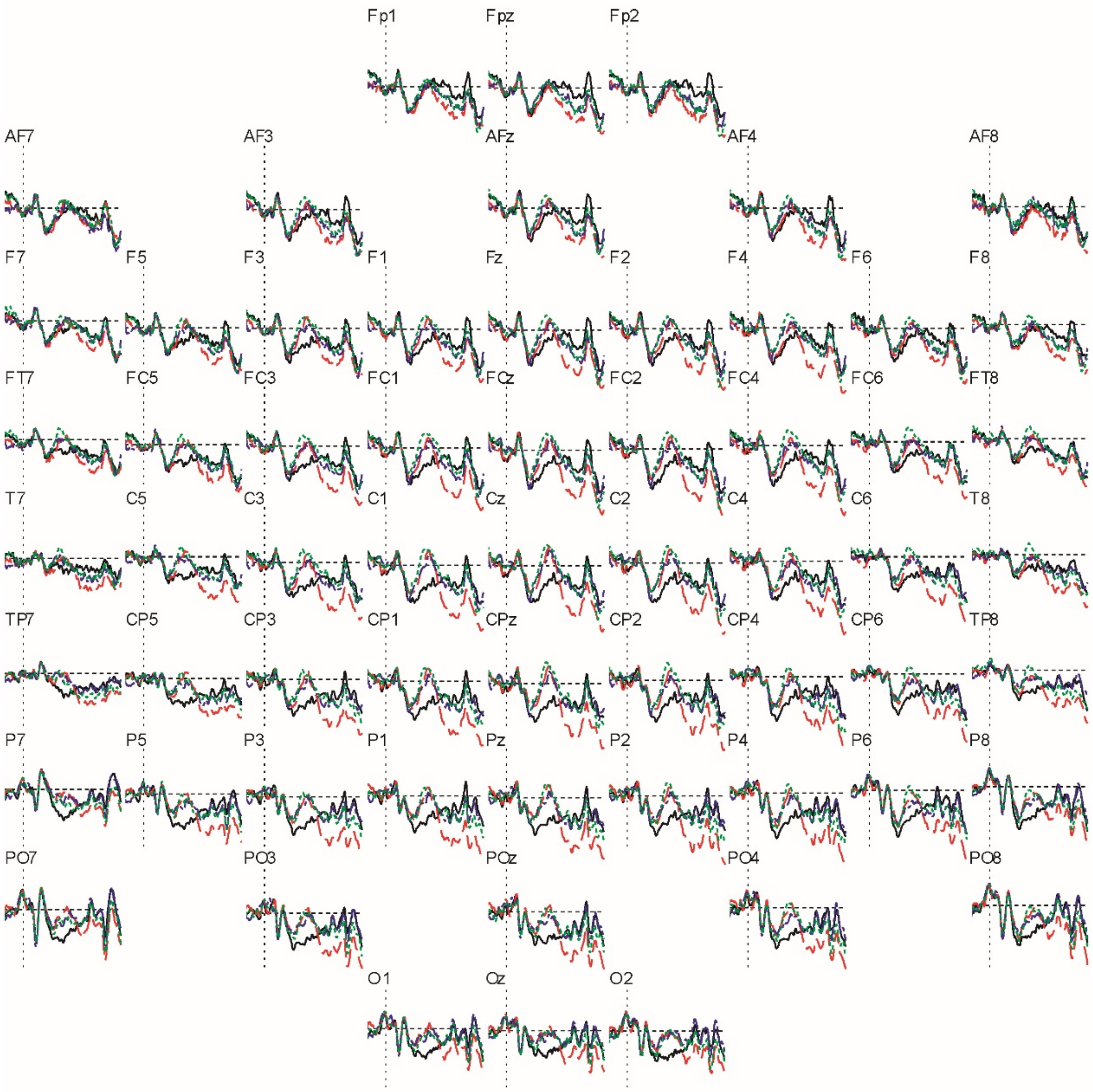
ERPs elicited by critical words
in the 4 conditions at all channels
in Experiment 8

Predictable
*The Juice isn't cold enough, so Alice is adding some **ice** now.*

Form
*The Juice isn't cold enough, so Alice is adding some **dice** now.*

Semantic
*The Juice isn't cold enough, so Alice is adding some **cube** now.*

Unrelated
*The Juice isn't cold enough, so Alice is adding some **wine** now.*



Experiment 8 (L2 speakers: 700 ms SOA)
Medium cloze items (0 to 200 ms baseline)

ERPs elicited by critical words
in the 4 conditions at all channels
in Experiment 8

—— **Predictable**
*The family went to the sea to catch some **fish** together.*

– – – **Form**
*The family went to the sea to catch some **wish** together.*

-·-·- **Semantic**
*The family went to the sea to catch some **pond** together.*

······ **Unrelated**
*The family went to the sea to catch some **echo** together.*