AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology:
Learning, Memory, and Cognition

ISSN: 0278-7393

2024, Vol. 50, No. 12, 1982–2007
https://doi.org/10.1037/xlm0001421

Lexicosemantic Prediction in Native Speakers of English and Swedish-Speaking Learners of English: An Event-Related Potential Study

José Alemán Bañón[1] and Clara D. Martin[2, 3]
[1] Department of Swedish and Multilingualism, Centre for Research on Bilingualism, Stockholm University
[2] Basque Center on Cognition, Brain and Language, Donostia-San Sebastián, Spain
[3] Ikerbasque, Basque Foundation for Science, Bilbao, Spain

The present study uses event-related potentials (ERPs) to investigate lexicosemantic prediction in native speakers (L1) of English and advanced second language (L2) learners of English with Swedish as their L1. The main goal of the study was to examine whether learners recruit predictive mechanisms to the same extent as L1 speakers when a change in the linguistic environment renders prediction a useful strategy to pursue. The study, which uses a relatedness proportion paradigm adapted from Lau et al. (2013), focuses on the N400, an ERP component that is sensitive to the ease of lexical access/retrieval, including lexical prediction. Participants read 800 prime–target pairs, presented word by word and divided into two blocks, while they searched for animal words. Unknown to them, some of the pairs were semantically associated, which is known to reduce the amplitude of the N400 via spreading semantic activation. Most importantly, the proportion of semantically related pairs increased in the second experimental block (via fillers), thereby increasing the reliability of the primes as predictive cues and encouraging prediction. Results from 36 L1-English speakers and 53 L2 learners showed an N400 reduction for related (*remain*-<u>*stay*</u>) relative to unrelated targets (*silver*-<u>*stay*</u>) across blocks. Crucially, this N400 reduction for related targets was significantly larger in the block that encouraged prediction, in both L1 and L2 speakers, consistent with the possibility that both groups recruited similar predictive mechanisms when the context encouraged prediction. These results suggest that, at high levels of proficiency, L2 speakers engage similar predictive strategies to L1 speakers.

Keywords: lexicosemantic prediction, N400 effect, L2 predictive processing, relatedness proportion, individual differences

Supplemental materials: https://doi.org/10.1037/xlm0001421.supp

A central question in second language (L2) acquisition research concerns the extent to which L2 learners recruit similar processing mechanisms to native (L1) speakers while parsing the L2 input (see Hopp, 2022). One such mechanism is prediction, which broadly refers to the use of context to anticipate likely continuations (see Kuperberg & Jaeger, 2016, for a discussion of the different meanings of the term *prediction*). It is well-established that L1 speakers of a language predict at all levels of linguistic representation, although it remains a matter of debate how pervasive prediction is in L1 comprehension (e.g., Huettig, 2015; Huettig & Mani, 2016;

Tessa Warren served as action editor.

José Alemán Bañón https://orcid.org/0000-0003-4210-3174

All experimental materials can be found both in the Supplementary Materials (see Supplementary Materials_S1_Stimuli) and at https://osf.io/r9t3q/ (see Supplementary Materials_S1_Stimuli). All data can be found at https://osf.io/r9t3q/ (see Lexicosemantic prediction_Data.zip), where we also provide the code for all analyses conducted in *R*. The study was not preregistered.

This work was supported by a Riksbankens Jubileumsfond grant to José Alemán Bañón (Grant P18-0756:1). Clara D. Martin acknowledges financial support from the Basque Government (BERC 2022-2025 program), the Spanish State Research Agency (PID2020-113926GB-I00 awarded to Clara D. Martin and Severo Ochoa excellence accreditation CEX2020-001010-S awarded to the Basque Center on Cognition, Brain and Language), and the H2020 European Research Council under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 819093 awarded to Clara D. Martin).

The authors thank Victor Norrman for his help in collecting all the EEG data, Jamie Rinder for his help with participant recruitment and with his native judgments about the related pairs, and all participants for being generous with their time. The authors express their gratitude to Spyridoula Cheimariou, Carrie Jackson, and the associate editor, Tessa Warren, for their constructive comments on the article.

José Alemán Bañón played a lead role in data curation, formal analysis, funding acquisition, investigation, methodology, project administration, writing–original draft, and writing–review and editing and an equal role in conceptualization. Clara D. Martin played a supporting role in funding acquisition, writing–original draft, and writing–review and editing and an equal role in conceptualization.

Correspondence concerning this article should be addressed to José Alemán Bañón, Department of Swedish and Multilingualism, Centre for Research on Bilingualism, Stockholm University, Universitetsvägen 10 (D355), 10691 Stockholm, Sweden. Email: jose.aleman.banon@biling.su.se

Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018). While some researchers argue that prediction is fundamental to language processing (e.g., Clark, 2013) and drives language acquisition in children (e.g., Chang et al., 2006), others claim that prediction is not necessary for successful language comprehension and is, therefore, optional (e.g., Huettig, 2015; Huettig & Mani, 2016; Pickering & Gambi, 2018; see Ryskin & Nieuwland, 2023). Recently, Kuperberg and Jaeger (2016) proposed that whether or not comprehenders predict depends on their goals, their prior knowledge, and their assessment of the reliability of the predictive cues.

With respect to predictive processing in the L2, early reports found that adult L2 learners did not predict to the same extent as native speakers (e.g., Grüter et al., 2012; Kaan et al., 2010; Lew-Williams & Fernald, 2010; Martin et al., 2013), which was interpreted in light of reduced capacity models of L2 processing (e.g., Hopp, 2010; McDonald, 2006). This was, for example, the spirit of the Reduced Ability to Generate Expectations hypothesis, a proposal by Grüter et al. (2012, 2017) according to which L2 learners have a Reduced Ability to Generate Expectations across the board. However, as L2 research on prediction proliferated, evidence accumulated that L2 learners can and sometimes do predict in a native-like manner (e.g., Dijkgraaf et al., 2017; Dussias et al., 2013; Foucart et al., 2014; Hopp, 2013), even though they generally predict to a lesser extent or later than native speakers (see Schlenter, 2023, for a review). This latter evidence echoes a proposal by Kaan (2014) that predictive mechanisms are qualitatively similar in the L1 and the L2 and that the variability observed among L2 learners can be accounted for by linguistic and cognitive factors that also modulate predictive processing in native speakers. Some of these factors include processing speed and working memory (e.g., Huettig & Janse, 2016; Ito, Corley, & Pickering, 2018), the quality of lexical representations (e.g., Hopp, 2013), the type of information used to generate predictions (e.g., Grüter et al., 2020; Hopp, 2015), in addition to factors that are specific to L2 learners, such as L2 proficiency (e.g., Dussias et al., 2013; Hopp, 2013; Hopp & Lemmerth, 2018) and cross-linguistic differences (e.g., Alemán Bañón & Martin, 2021; Dussias et al., 2013; Hopp & Lemmerth, 2018; Şafak & Hopp, 2022; Van Bergen & Flecken, 2017).

More recently, Kaan and Grüter (2021) have adopted Kuperberg and Jaeger's (2016) view that prediction has a utility function, to explain differences in predictive behavior between native speakers and L2 learners (see also Grüter & Rohde, 2021, who provide an updated version of the Reduced Ability to Generate Expectations hypothesis with utility at the center). The idea is that prediction has different levels of utility for L1 and L2 speakers due to differences in the reliability of the predictive cues. For example, some predictive cues are absent or realized differently in the learners' L1, and others carry different weights in the learners' L1 and L2. Alternatively, cue reliability may differ depending on how the relevant information is acquired in native and nonnative language acquisition. This can cause L2 learners to generate predictions that are different from native speakers' and that often go wrong, which eventually makes prediction less useful for L2 learners.

The present study uses event-related potentials (ERPs) to further investigate sources of variability in L2 predictive processing, a line of research advocated by Kaan (2014) and Kaan and Grüter (2021). The study uses a paradigm adapted from Lau et al. (2013) to examine lexicosemantic prediction via semantic associations (e.g., *east-west, slay-kill*) in native speakers of English and Swedish-speaking learners of English. In this paradigm, the proportion of semantically associated pairs increases from 10% to 50% (via fillers) halfway through the experiment, thereby increasing the reliability of the primes as predictive cues. This is assumed to produce a qualitative shift in processing, such that comprehenders start actively using the primes to predict the targets in the second half of the experiment (e.g., Brown et al., 2000; Delaney-Busch et al., 2019; Holcomb, 1988; Lau et al., 2013). Thus, this paradigm can inform us of the extent to which L2 learners engage predictive mechanisms as a function of a change in the linguistic environment. Importantly, a review of the L2 literature on semantic associations reveals that, even offline, learners' associations tend to be less stable, less homogenous, and more erroneous (due to orthographic and phonological interference) than native speakers' (Meara, 2009; van Hell & de Groot, 1998; see Tokowicz, 2015, for a review). Following Kaan's (2014) proposal that the quality of lexical representations modulates predictive abilities (e.g., Mani & Huettig, 2014; Mishra et al., 2012) and that L2 learners have more lower quality lexical representations than native speakers, we reasoned that if L2 learners' semantic networks are smaller, weakly connected, less stable, and more prone to error, learners might predict to a lesser degree than native speakers, since the gain might be reduced (e.g., Kaan & Grüter, 2021).

Another question that we engaged concerns the role of working memory and processing speed in predictive processing among both L1 and L2 speakers. Lau et al. (2013) argued that prediction requires keeping a contextual representation in working memory and incorporating the preactivated representation ahead of the bottom-up input, a process that consumes processing resources and requires time. In fact, previous studies manipulating relatedness proportion that used a fast presentation rate found no evidence that the primes gained predictive strength as a function of proportion (Grossi, 2006). Thus, comprehenders with better working memory and faster processing speed should be better equipped to engage predictive mechanisms. Although Lau et al.'s (2013) study did not address this question, subsequent studies on prediction have provided evidence along these lines with different paradigms (Huettig & Janse, 2016; Ito, Corley, & Pickering, 2018). In addition, previous behavioral studies have reported that generating semantic associations in the L2 takes longer than in the L1 (e.g., Fitzpatrick & Izura, 2011), suggesting that fast processing speed might be necessary for L2 learners to retrieve semantic associates ahead of the bottom-up input. Our study explores this possibility.

Event-Related Potentials

The present study examines lexicosemantic prediction with ERPs, which are brain responses that are time locked to stimuli of interest (e.g., expected words). The study focuses mainly on the N400, a negative-going deflection in the EEG that typically emerges between 300–500 ms poststimulus presentation in central-posterior electrodes of the EEG cap and that reflects, among other things, the ease of lexical access and retrieval (e.g., Kutas & Hillyard, 1980, 1984; Lau et al., 2008). The evidence for this is that the magnitude of the N400 tends to be reduced for words that follow a supportive context. For example, upon reading a sentence preamble such as *I take my coffee with cream and …* , native English speakers show a smaller N400 for the expected word *sugar*, relative to the unexpected word *honey*, even though both continuations are semantically licit (Kutas & Hillyard, 1984). A similar N400 reduction is typically observed in nonsentential contexts when comparing target words that follow related versus unrelated primes.

For example, the word *dog* shows a reduced N400 when it follows the related word *cat*, compared to when it follows an unrelated word, such as *bus* (Holcomb & Neville, 1990; Lau et al., 2013). The latter semantic relatedness effect on the N400 is central to the present study.

There are, at least, two accounts of this so-called "N400 effect" (e.g., see also DeLong et al., 2014; Martin et al., 2013). One is that the preceding context activates a semantic network, which in turn facilitates the processing of words from that network. For example, *cat* activates the semantic network of animals/pets, thereby facilitating the processing of the related word *dog*. Under such an account, the N400 effect reflects facilitation caused by spreading activation or passive resonance but not active prediction.[1] An alternative interpretation is that, when the context is supportive, the parser uses it to actively generate predictions about upcoming words or semantic content. Under this predictive account, the N400 reduction for supported words reflects facilitation from having preactivated the words or some of their semantic features ahead of the bottom-up input (e.g., Federmeier & Kutas, 1999; see Van Petten & Luka, 2012).

Strong evidence that the N400 effect partly reflects prediction-related facilitation comes from studies manipulating prenominal articles whose form depends on the phonological or lexical properties of the expected yet unencountered nouns (e.g., DeLong et al., 2005; Foucart et al., 2014; Martin et al., 2013; Otten & van Berkum, 2009). For example, DeLong et al. (2005) probed high-constraint sentences such as *The day was breezy, so the boy went outside to fly …* and compared expected continuations, such as *a kite*, with unexpected but congruent ones, such as *an airplane*. Crucially, the expected and unexpected nouns were always preceded by different allomorphs of the English indefinite article, *a/an*. DeLong et al.'s (2005) results revealed a reduced N400 for both expected nouns and articles, relative to their unexpected counterparts. While the N400 effect on the article is difficult to accommodate without assuming that the upcoming nouns had been preactivated (but see Nieuwland et al., 2018, who did not replicate this finding across nine different labs), the nature of the N400 effect on the noun is less clear-cut, since it is consistent with both a spreading activation and a prediction account. Lau et al. (2013) provided complementary evidence that the N400 reduction for supported nouns is also driven by lexical prediction, even more so than by spreading activation. Since the design of the present study mirrors Lau et al.'s (2013) and our materials were selected from the same database as theirs, we provide a detailed description of both the rationale of their study and their findings.

In Lau et al.'s (2013) study, 32 native speakers of American English read prime–target word pairs presented word by word. They were instructed to press a button whenever they encountered an animal word and to withhold their responses for all other words (i.e., a go/no-go task). Unbeknownst to them, some of the pairs showed a strong semantic association (e.g., related: *east*-*west*; unrelated: *rye*-*west*) as per the University of South Florida Association Norms (Nelson et al., 2004). In addition, the proportion of related pairs increased from 10% to 50% across the first and second experimental blocks. Lau et al. (2013) reasoned that the low ratio of related pairs in the first block would not encourage active prediction generation and, thus, an N400 reduction for related relative to unrelated targets would only reflect passive spreading activation. In contrast, the high proportion of related pairs in the second block was expected to encourage participants to use the primes predictively, which the authors hypothesized would yield a larger N400 effect relative to the low-proportion block, driven by a larger N400 reduction for related targets.

Lau et al.'s (2013) predictions concerning the N400 were largely borne out. Their results showed an N400 effect, with related targets yielding a reduced N400 relative to unrelated ones in both blocks. Crucially, this effect was significantly larger in the block that supported active predictive processing, driven by a larger N400 reduction for related targets in the high-proportion block. In addition, the N400 effect emerged earlier in the high-proportion relative to the low-proportion block (200–250 ms vs. 400–450 ms) and had a right posterior maximum in the high-proportion block only, which the authors took as further evidence supporting a dissociation between priming and prediction effects on the N400 effect. Unrelated targets also yielded a late negativity (500–800 ms) relative to related targets but only in the high-proportion block. Lau et al. (2013) hypothesized that this post-N400 negativity might reflect the reevaluation of the semantic relation between primes and targets, consistent with a predictive account (see also Steinhauer et al., 2017). In addition, animal targets showed a P300, a component elicited by task-relevant stimuli (e.g., Coulson et al., 1998; see Leckey & Federmeier, 2020), which was larger in the high-proportion block. Lau et al. (2013) argued that the larger P300 for animal words in the high-proportion block was consistent with the possibility that failed predictions had a cost on response selection. Since the animal words were always unrelated to their primes, they always disconfirmed a prediction in the high-proportion block. For example, upon encountering the word *salt* in the high-proportion block, a prediction might have been generated for *pepper* (based on Nelson et al., 2004), thus requiring a "no-go" response. Upon encountering the animal word *giraffe* instead, this might have created a conflict between the planned "no-go" and the required "go" response. This would explain the larger P300 in the high-proportion block (i.e., the predictive block), consistent with previous studies showing larger positivities in the ERP signal for conflict trials (Larson et al., 2009; West, 2003). Finally, an exploratory analysis revealed that the primes (e.g., *east*–*west*) in the high-proportion block yielded a positivity in the P2 and N400 time windows (~200–300 ms and 350–400 ms, respectively) relative to the low-proportion primes. Since the P2 indexes visual feature analysis, Lau et al. (2013) related the larger P2 for high-proportion primes to prediction instantiation.

Thus, the results by Lau et al. (2013) suggest that the classic N400 effect for supported words in the above studies by Kutas and Hillyard (1984) and DeLong et al. (2005) partly reflects facilitation from having preactivated words or some of their semantic features ahead of the bottom-up input. These findings have important implications for the L2 literature, where the lack of N400 differences on prenominal material has been interpreted as evidence that adult L2 learners do not predict words in sentence comprehension (e.g., Ito et al., 2017a; Martin et al., 2013). For example, Martin et al. (2013) did a conceptual replication of DeLong et al.'s (2005) study, and largely replicated DeLong et al.'s (2005) results in another group of native speakers of English: a reduced N400 for both expected articles and nouns, relative to their unexpected counterparts. In contrast, a group of advanced Spanish-speaking learners of

---

[1] Some researchers conceive spreading activation as a form of prediction that is automatic, unconscious, and uncontrollable. However, they differentiate it from active prediction generation (Ito & Pickering, 2021; Pickering & Gambi, 2018; see Kuperberg & Jaeger, 2016).

English only showed a reduced N400 for expected relative to unexpected nouns, an effect that was smaller and later than in the L1-English group. Martin et al. (2013) interpreted the L2 learners' results as evidence that the processing of the noun was facilitated via passive spreading activation but argued that such an effect was unlikely to reflect lexicosemantic prediction (see also Ito et al., 2017a). In light of the results by Lau et al. (2013), however, it is possible that the L2 learners in Martin et al.'s (2013) study did generate lexicosemantic predictions, although they must have predicted to a lesser extent than the native controls, since the effect was smaller. Our study revisits lexicosemantic prediction among L2 learners with a design based on Lau et al.'s (2013). Before introducing the study in detail, we provide a succinct review of the most relevant literature on lexicosemantic prediction in the L2.

Lexicosemantic Prediction in the L2

Prediction at the level of lexical semantics has attracted substantial interest from L2 researchers (e.g., Chambers & Cooke, 2009; Chun & Kaan, 2019; Dijkgraaf et al., 2017, 2019; Foucart et al., 2014; Gambi, 2021; Hopp, 2015; Ito, Corley, & Pickering, 2018; Ito et al., 2017a, 2017b; Ito, Pickering, & Corley, 2018; Martin et al., 2013; Schlenter & Felser, 2021). One generalization that has been drawn is that lexicosemantic prediction in the L2 is unproblematic or intact (e.g., Dijkgraaf et al., 2017; Ito, Corley, & Pickering, 2018), with even intermediate learners being able to generate lexicosemantic predictions successfully (e.g., Chambers & Cooke, 2009; Dijkgraaf et al., 2017; Hopp, 2015). This contrasts with what has been observed in other domains of grammar, such as morphosyntax (e.g., Grüter et al., 2012; Hopp, 2015; Mitsugi, 2017; Mitsugi & MacWhinney, 2016) and, especially, syntax, where even high-proficiency learners often fail to generate predictions or do so in a non-native-like manner (e.g., Covey et al., 2024; Hopp, 2015; Kaan et al., 2016; see Hopp, 2022). Nevertheless, the evidence that L2 learners use lexicosemantic information predictively to the same extent as native speakers comes mainly from studies using the visual world paradigm (VWP) and probing straightforward predictions (e.g., those that can be generated based on the knowledge of which entities can and cannot be folded, eaten). In these studies, participants look at a display of two to four pictures while they listen to sentences that are informative or uninformative with respect to the identity of the upcoming noun, as in *The lady will fold/find the …* with pictures of a scarf, a violin, a piano, and a pair of shoes (example from Ito, Corley, & Pickering, 2018; see also Chambers & Cooke, 2009; Hopp, 2015; Schlenter & Felser, 2021). In such setups, L2 learners might be better able to generate lexicosemantic predictions like native speakers, especially since they are given time to familiarize themselves with the pictures before the onset of the carrier sentence (e.g., Chambers & Cooke, 2009; Dijkgraaf et al., 2017; Ito, Corley, & Pickering, 2018; Schlenter & Felser, 2021; see Huettig, 2015), and sometimes, the pictures are named (e.g., Hopp, 2015), thereby minimizing the burden of lexical retrieval. In fact, with more complex setups, learners' predictions have been found to be reduced or delayed. This was the case in a recent study by Chun and Kaan (2019), where the predictive cue belonged inside a relative clause that was ambiguous with respect to which noun phrase it attached to (e.g., *the friend of the dancer who will <u>get/open</u> the present*, with the verb *open* being predictive of *present*). Likewise, a recent study by Dijkgraaf et al. (2019) where Dutch–English bilinguals only saw the visual display 500 ms before the onset of the target word found that, although participants predicted in both their L1 and their L2, the spread of semantic activation (as measured by looks to a semantic competitor) was reduced and delayed in the L2. Furthermore, Van Bergen and Flecken (2017) provided evidence that, if learners lack the relevant semantic constraint in their L1, they are less likely to use it predictively. In their study, L1-German L2-Dutch learners launched anticipatory looks to target objects based on the object placement information encoded in Dutch verbs (*zetten*: "put.STAND" vs. *leggen*: "put.LIE"), a constraint that German verbs also encode. In contrast, L2 learners who were native speakers of English or French, two languages where placement verbs are underspecified with respect to object position, looked at the target objects later.

Unlike studies using the VWP, studies investigating semantic prediction via ERPs typically measure brain responses to (un)expected words at the point when the words are first encountered, suggesting that comprehenders must incrementally narrow down their entire lexicon until they can retrieve the target word (e.g., Foucart et al., 2014; Ito et al., 2017a, 2017b; Martin et al., 2013). In these studies, it is often the case that L2 learners either do not evince the same effects as native speakers or show effects that are reduced, delayed, or ambiguous. This was, for example, the case in the abovementioned study by Martin et al. (2013), where Spanish-speaking learners of English showed a smaller N400 reduction for expected nouns than native speakers and no N400 reduction for expected articles, unlike the native controls. In addition, a conceptual replication of Martin et al.'s (2013) study by Ito et al. (2017a) found no N400 reduction for expected nouns in another group of L1-Spanish learners of English when sentences were presented with a 500-ms stimulus onset asynchrony (SOA). Ito et al. (2017a) followed up on these results in a second experiment using a 700-ms SOA, the same presentation rate as in Martin et al. (2013), and found that a different group of L1-Spanish L2-English learners did show an N400 effect on the noun, comparable to that of the L1-English controls (although, in this study, neither group yielded an N400 effect on prenominal articles).

In another study, Ito et al. (2016, 2017b) had native and nonnative speakers of English read high-constraint frames, such as *The student is going to the library to borrow a …* , which were followed either by the expected word (*book*) or by an unexpected and implausible word. The unexpected word was either related to the expected word in form (*hook*) or meaning (*page*) or unrelated to it (*sofa*). In the L1-English group (Ito et al., 2016), words that were unexpected but related in form or meaning to the predictable word showed a reduced N400 relative to unrelated words (*hook* and *page*, relative to *sofa*). Crucially, this N400 reduction was accounted for by the cloze probability of the expected word, suggesting that native speakers had preactivated both the form and the meaning of the expected word. In contrast, a group of Spanish-speaking learners of English only showed an N400 reduction for expected relative to all unexpected words (Ito et al., 2017b). Although unexpected but semantically related words showed a reduced N400 compared to unrelated words (*page*, relative to *sofa*), this effect was not explained by the cloze probability of the expected word, which the authors interpreted as evidence that the L2 learners did not preactivate words, but rather showed sensitivity to differences in plausibility between related and unrelated words.

To our knowledge, only Foucart et al. (2014) found native-like brain responses to violations of lexicosemantic predictions in the L2.

The authors examined whether native speakers of Spanish and French-speaking learners of Spanish preactivated the lexical gender of highly predictable nouns, as realized on prenominal articles. Their results revealed that both L1 and L2 speakers of Spanish elicited a comparable N400 reduction for expected articles and nouns, relative to their unexpected counterparts.[2]

It is unclear what explains the different levels of success among L2 learners across the studies by Martin et al. (2013) and Ito et al. (2017a, 2017b); none of which found evidence that L2 learners generated lexicosemantic predictions to the same extent as native speakers, and Foucart et al. (2014), where the L2 learners predicted similarly to the native controls. Two straightforward explanations are differences in L2 proficiency and mean cloze probability. With respect to L2 proficiency, although all learner groups self-rated their proficiency as upper-intermediate, it cannot be ruled out that the learners in Foucart et al. (2014) were more advanced, since only Foucart et al. (2014) tested L2 proficiency independently.[3] Likewise, although mean cloze probability for expected nouns in Martin et al. (2013) was comparable for L1 and L2 speakers (.69 vs. .65, respectively), it was lower than in Foucart et al.'s (2014) study (.81 vs. .82, respectively). In addition, in the studies by Ito et al. (2017a, 2017b), mean cloze probability for expected nouns was not matched in L1 and L2 speakers (Ito et al., 2017a: .65 vs. .57, respectively; Ito et al., 2017b: .80 vs. .60, respectively). These two factors alone or in combination may partly explain why the learners' processing in the studies by Martin et al. (2013) and Ito et al. (2017a, 2017b) was short of native-likeness.

The present study addresses these issues in the following way. First, we pretested the L2 learners for proficiency and selected the most proficient learners that we could find. Second, we used primes for which both L1-English speakers and Swedish-speaking learners of English had the same strongest semantic associate, as determined by a free associations study including 104 high-proficiency Swedish-speaking learners of English. Thus, if we find evidence for reduced prediction among L2 learners, such limitation is unlikely to be due to differences in association strength between L1 and L2 speakers (as might have been the case in the studies by Ito et al., 2017a, 2017b, with respect to cloze probability) or to learners having different semantic associations. Recall, however, that L2 learners often produce semantic associations that are less homogeneous, less stable, and more prone to error than those of native speakers. Thus, learners might still not show the same N400 reduction for predicted targets as native speakers. As Kaan (2014) pointed out, learners may well have the same biases (associations, in this case) as native speakers. However, if they represent this information less consistently and experience difficulty in retrieving it, they might not be able to use it predictively or they might generate predictions that go wrong, which might discourage them from engaging predictive mechanisms. Finally, another strength of the paradigm that we used is that the L2 learners are tested in two environments, one that does not encourage active prediction generation and one that does (i.e., the low- vs. high-proportion blocks), which can shed light on how learners engage predictive mechanisms as a function of a change in the linguistic environment.

The Present Study

The present ERP study probes lexicosemantic prediction via semantic associations in native speakers of English and advanced Swedish-speaking learners of English. The study uses the same relatedness proportion paradigm as Lau et al. (2013) but with a different set of materials suitable for both L1-English and L1-Swedish speakers. This paradigm involves a change in the proportion of semantically associated pairs halfway through the experiment (via fillers), which increases the predictive reliability of the primes (Kaan & Grüter, 2021; Kuperberg & Jaeger, 2016). Given that L2 learners are hypothesized to have more lower quality lexical representations (Kaan, 2014) and that semantic associations are less homogeneous, less consistent, and more prone to error in L2 learners (Meara, 2009), we examine whether the primes in the predictive block will gain similar predictive validity among the L2 learners (e.g., Kaan & Grüter, 2021). Ultimately, the study investigates whether L2 learners generate lexicosemantic predictions at the word level to the same extent as native speakers. This is a timely question in light of recent claims in the L2 literature that efficient lexical processing, which prediction is a part of, is a prerequisite for successful processing at higher levels of linguistic representation, such as syntax (e.g., Hopp's 2018 lexical bottleneck hypothesis; see also Kim & Grüter, 2021).

The study also investigates the extent to which working memory and processing speed modulate lexicosemantic prediction in both L1 and L2 speakers. This question is motivated by views of prediction as a process that involves storing and rapidly updating a contextual representation in working memory. Below, we formulate our research questions and provide specific predictions.

> *Research Question 1:* Do advanced Swedish-speaking learners of English generate lexicosemantic predictions to the same extent as native English speakers?

If so, both groups should display a pattern of results similar to that of Lau et al. (2013). That is, the N400 effect for related versus unrelated targets (e.g., *east-<u>west</u>, rye-<u>west</u>*) should be larger in the block that encourages predictive processing, driven by a larger N400 reduction for related targets in the high-proportion block. It is also possible that the N400 effect will show an earlier onset and a different topographical distribution (right-posterior) in the high-proportion block. Differences in predictive behavior between L1 and L2 speakers should result in qualitative and/or quantitative differences between the two groups in the above pattern of results. Although we focus mainly on the N400 effect, we will also compare how L1 and L2 speakers process the animal targets and the primes in the low- and high-proportion blocks. Animal targets should yield a larger P300 in the high- relative to the low-proportion block, which would reflect the conflict between the expected ("no-go") and the required ("go") behavioral response, consistent with a predictive account. Finally, the primes (e.g., *<u>east</u>-west*) might yield a positivity in the high-proportion block (~200–300 ms, ~350–400 ms), relative to the low-proportion block, which might reflect the process of launching a prediction.

---

[2] Alemán Bañón and Martin (2021) also found native-like brain responses to violations of lexicosemantic predictions in the L2. In this study, however, participants first encountered the target nouns in a context leading to the sentence where brain responses were recorded, which makes it more comparable to VWP studies.

[3] Grüter and Rohde (2021), Grüter et al. (2021), and Kim and Grüter (2021), however, remain skeptical about the relation between proficiency and predictive processing.

> *Research Question 2:* Do individual differences in working memory and processing speed account for variability in predictive processing in native and nonnative speakers?

If so, both working memory and processing speed should make a significant individual contribution toward explaining variability in the N400 relatedness effect in the high-proportion block, in both groups.

Method

All experimental procedures were reviewed and approved by the regional ethics committee in Stockholm (Project Number: 2021-03472). The study was not preregistered.

Participants

The participants include 36 native speakers of English (24 females) and 53 Swedish-speaking learners of English (22 females). An additional eight learners were tested but excluded from all analyses. Five of them either showed excessive drift/noise in the EEG recording or had too few artifact-free epochs after processing (i.e., below 20 per condition). Two learners showed below 65% accuracy in the monitoring task in the high-proportion block, suggesting that they had disengaged from the task. Finally, one learner reported (after the testing took place) having acquired Swedish after Elfdalian, a North Germanic language that is not mutually intelligible with Swedish.

English proficiency was monitored with the Lexical Test for Advanced Learners of English (LexTALE; Lemhöfer & Broersma, 2012), a test of English vocabulary. As Table 1 shows, the learners scored very high in the proficiency test, within the native speaker range. However, they were outperformed by the native speakers, as revealed by a one-way analysis of variance (ANOVA), $F(1, 87) = 10.788$, $p = .001$, $\eta^2 = .110$. A battery of individual differences tasks administered in the participants' L1 revealed that the two groups did not differ with respect to general intelligence, as measured by Raven's Progressive Matrices (Raven & Raven, 2003), $F(1, 87) = 0.202$, $p = .654$, $\eta^2 = .002$. Both groups were also matched with respect to processing speed, as measured by a composite score calculated by averaging across participants' scores in a Letter Comparison Task (Earles & Salthouse, 1995) and a Digit-Symbol Substitution Task (*Wechsler Adult Intelligence Scale* [WAIS], 2004), $F(1, 87) = 0.929$, $p = .338$, $\eta^2 = .011$. With respect to working memory, the L1-English speakers outperformed the L1-Swedish speakers, as measured by a composite score calculated by averaging across participants' scores in a Backwards Digit Span Task (WAIS, 2004) and a Reading Span Task administered in the participants' L1 (adapted from Daneman & Carpenter, 1980), $F(1, 86.517) = 6.319$, $p = .014$, $\eta^2 = .058$ (equal variances not assumed). Both working memory tasks were scored via a partial-credit load scoring procedure (Conway et al., 2005), but similar results were obtained when following a partial-unit procedure.[4]

All L1-English speakers reported having grown up as monolingual speakers of English. Since the testing took place in Stockholm (Sweden), we did not restrict the L1-English group to native speakers of American English, as in Lau et al.'s (2013) study, since this ran the risk of curtailing data collection. Thus, our sample includes 28 native speakers of American English (27 from the United States and one from Canada) and eight native speakers of British English (all from England). Although free association norms can be sensitive to cultural differences (Nelson et al., 2004, p. 406), it is unlikely that differences across these varieties of English and cultures significantly impacted association strength for the related pairs selected for the present study, since all of the pairs were shown to be strongly associated even among nonnative speakers of English from Sweden (see the Free Associations Study section). All but three native speakers reported knowledge of one or more foreign languages (Danish, Dutch, Farsi, French, German, Italian, Lithuanian, Norwegian, Russian, Spanish, Swedish) to different levels of proficiency, and all but one had a college education or higher.

All L1-Swedish learners reported having grown up as monolingual speakers of Swedish and all but one listed English as their L2. The only learner who did not list English as their L2 reported having been exposed to Finnish before English, but this participant also rated their proficiency in Finnish as very low. Thirty-five learners knew at least one more foreign language, also to varying levels of proficiency (Chinese, Farsi, French, German, Hindi, Icelandic, Italian, Japanese, Portuguese, Russian, Spanish). The learners' educational level ranged from secondary education to postdoctoral studies.

All 89 participants had normal or corrected-to-normal vision and no record of neurological impairments. All but two participants (two L1-English speakers) were right-handed, as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). All participants received compensation for their time.

Materials

Experimental Design

Similar to Lau et al. (2013), the study used a 2 × 2 design with Relatedness (related, unrelated) and Proportion (low, high) as the main factors. The experiment comprised two blocks, each including 400 word pairs, presented one word at a time. Each block included 40 semantically related pairs (e.g., *umbrella-rain*) and 40 semantically unrelated pairs (e.g., *low-nut*), which were the critical items that were used to examine the effects of semantic relatedness on the targets. Each block also encompassed 40 semantically unrelated pairs with an animal target (e.g., *phase-cat*), animals being the semantic category that participants were asked to identify in the experiment. Finally, each block also included 280 filler pairs, which were used to manipulate relatedness proportion across the blocks. In the first block, none of the 280 filler pairs were semantically associated, which kept relatedness proportion at 10% (40/400). In the second block, 160/280 filler pairs were semantically associated, which raised relatedness proportion to 50% (200/400). Table 2 provides a schematic of this design. Both the semantic relatedness and the proportion manipulations were concealed to participants, and the low-proportion block (10%) was always presented first.

Free Associations Study

To create the materials for the relatedness manipulation, we first conducted a free associations study with 104 Swedish-speaking

---

[4] These group analyses on the cognitive tasks were conducted on uncorrected data. However, we reran all analyses after correcting the data for outliers, and the pattern of results did not change.

Table 1
Participants' Information

Variable	L1 English (*N* = 36) *M*	*SD*	Range	L1 Swedish (*N* = 53) *M*	*SD*	Range
Age at testing	31	7	21–45	28	6	18–45
AoA of English				8	2	4–12
LexTALE score	95	6	70–100	91	6	75–100
Swedish test score	74	11	53–95	94	7	73–100
Raven score	54	5	38–60	54	4	43–60
Processing speed	1,285	191	860–1,777	1,328	211	937–1,888
Working memory	0.84	0.10	0.55–0.95	0.71	0.15	0.34–1

Note. L1 = native language; AoA of English = age of acquisition of English; LexTALE = Lexical Test for Advanced Learners of English, LexTALE scores are provided as percentages; Processing speed = composite score (ms), calculated by averaging across scores in a Letter Comparison Task and a Digit-Symbol Substitution Task; working memory = composite score (partial-credit load; see Conway et al., 2005), calculated by averaging across scores in a Backwards Digit Span Task and a Reading Span Task.

learners of English of upper-intermediate to advanced proficiency, as measured by the LexTALE (*M* = 87, *SD* = 9, range = 65–100). Here, we only provide a succinct description of the study, which will be reported in detail in Alemán Bañón and Martin. First, we selected 1,244 related pairs from the University of South Florida Association Norms (Nelson et al., 2004) with a forward association strength (FSG) between .40 and .94, suggesting that at least 40% of native speakers of English would provide the target as the first word that would come to mind when presented with the cue. We then collected associations from the L2 learners for all 1,244 cues, and we calculated the proportion of L2 learners who provided the same strongest associate as the L1-English speakers in Nelson et al.'s (2004) study, that is, the FSG for the pair listed in Nelson et al.'s (2004) norms. The two leftmost violin plots in Figure 1 (*Norming Study*) show the range and density of FSG values across the 1,244 pairs among the L1-Swedish learners. The FSG values reported by Nelson et al. (2004) for L1-English speakers are plotted for comparison. As can be seen, the L1-Swedish learners show a more elongated distribution of FSG values than the L1-English speakers in Nelson et al.'s (2004) norms. This suggests that the distribution of FSG values for these materials is more heterogeneous among L2 learners (range = 0–.97). Most importantly, for almost half of the pairs (548/1244), the proportion of learners who provided the same semantic associate as the native English speakers in Nelson et al.'s (2004) norms falls under 40%, the lowest bound value in the L1-English group, and can be as low as 0%. This is consistent with the idea that, even at high levels of proficiency, L2 learners produce a number of semantic associations that differ from those of native speakers.

Table 2
Distribution of Materials Across Conditions

Type of material	Block 1, low-proportion	Block 2, high-proportion
Critical related pairs	40	40
Critical unrelated pairs	40	40
Unrelated pairs, animal target	40	40
Filler pairs, unrelated	280	120
Filler pairs, related	0	160
Total	400	400

Relatedness Manipulation

For the ERP study, only semantically associated pairs with an FSG ≥ .5 both in Nelson et al.'s (2004) norms for L1-English speakers and in our own norms for L1-Swedish learners were selected. In general, we aimed for the pairs with the highest FSG values in both sets of norms, although we applied some restrictions, which are described below.

Related Pairs. The mean FSG for the 160 semantically associated pairs selected for the relatedness manipulation was .67 (*SD* = 0.11, range = .50–.94) in the L1-English norms (Nelson et al., 2004) and .69 (*SD* = 0.11, range = .51–.97) in the L1-Swedish L2-English norms, respectively. The two middle violin plots in Figure 1 (*Related Targets [(ERP Study)]*) show that the distribution of FSG values across the 160 related pairs was comparable for L1-English speakers and L1-Swedish L2-English learners. Importantly, mean FSG in the present study was very similar to Lau et al.'s (2013) study (.65). All pairs had been rated by at least 116 native speakers of English (Nelson et al., 2004) and 102 L1-Swedish learners of English.[5]

We calculated the mean log frequency of the primes and targets with the Corpus of Contemporary American English (Davies, 2008). The primes had a mean log frequency of 1.10 (*SD* = 0.61), and the mean log frequency of the targets was 1.88 (*SD* = 0.51). With respect to length, the mean number of characters of the primes and the targets was 6 (*SD* = 2) and 5 (*SD* = 1), respectively. Based on Brysbaert et al.'s (2014) concreteness ratings, the mean concreteness of the primes was 3.73/5 (*SD* = 1.01), and the mean concreteness of the targets was 3.93/5 (*SD* = 0.91).

English and Swedish being related languages, we did not exclude cognates, since this would have resulted in an unnatural selection of materials.[6] We did, however, exclude associated pairs where the target was an identical English–Swedish cognate (e.g., *glove*-*hand*, which translates as *handske*-*hand* in Swedish), since brain responses

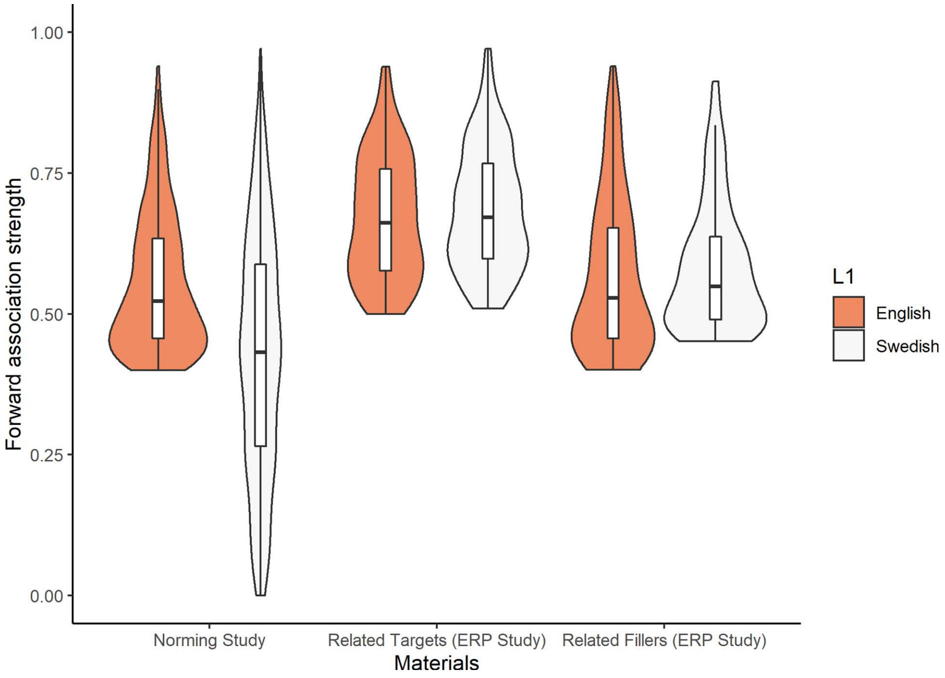
---

[5] Two learners did not complete the second session of the free associations study.

[6] Since most participants in both groups knew other languages, some of which are related to both English and Swedish (e.g., Danish, Dutch, German, Icelandic, Norwegian), we cannot reliably assess the extent to which the participants' English was impacted by coactivation of Swedish and their other languages.

Figure 1

Distribution of Forward Association Strength Values for Three Sets of Related Pairs: Norming Study (1,244 Pairs), Related Targets Selected for the ERP Study (160 Pairs), and Related Fillers Selected for the ERP Study (160 Pairs)



Note. All pairs were selected from "The University of South Florida Free Association, Rhyme, and Word Fragment Norms," by D. L. Nelson, C. L. McEvoy, and T. A. Schreiber, 2004, *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–407 (https://doi.org/10.3758/BF03195588), and normed by 104 Swedish-speaking learners of L2 English. ERP = event-related potential; L1 = native language. See the online article for the color version of this figure.

to such targets cannot unequivocally be taken to reflect L2 lexical processing. Since our main hypotheses concerned the effects reported by Lau et al. (2013) on the targets, we did use pairs where the prime was an identical cognate (e.g., *planet-earth*; Swedish: *planet-jorden*), although our materials include only 12 such primes (out of 160). Targets that only differed in one character between English and Swedish (i.e., near identical cognates), such as *library-book*, which translates as *bibliotek-bok* in Swedish, did not pose the same problem and, thus, they were used.[7]

We dismissed pairs including obvious identical interlingual homographs with Swedish (i.e., words with identical orthography but different meaning), such as *worst-best*, where *best* corresponds to the Swedish word for *beast*. We reasoned that interlingual homographs could activate different conceptual representations in the L2 group. Similar to Lau et al.'s (2013) study, associated pairs where the prime and the target showed obvious morphological overlap (e.g., *implode-explode*) were also excluded, as were pairs including animal words, the semantic category for which participants monitored. Finally, we dismissed all pairs consisting of function words (e.g., *up-down, you-me*).

Unrelated Pairs. The 160 semantically unrelated pairs were created by randomly reshuffling the related primes across the targets and by manually redistributing them to avoid meaningful pairs. Then, for each unrelated pair, we checked that none of the learners in the free associations study had associated the unrelated target to the prime. We also checked that the unrelated targets had not been associated to the primes in the Nelson et al.'s (2004) norms for L1-English speakers (Appendices A and D in Nelson et al.'s 2004 study).

Animal Words. We created an additional set of 80 unrelated pairs with an animal word as the target. As with the related pairs, we excluded identical English–Swedish cognates (e.g., *gorilla*) but used near identical cognates (e.g., *otter*; Swedish: *utter*). Two animal words that are identical interlingual homographs with Swedish in the singular (*dog* "died"; *tiger* "keep quiet") were used in the plural.

Fillers. Two hundred eighty filler pairs were created for each block. In the low-proportion block, none of the filler pairs showed a strong semantic association among L1-English speakers (according to Nelson et al.'s 2004 norms) or L1-Swedish L2-English learners. This kept relatedness proportion at 10% (40/400 trials) in the low-proportion block. In the high-proportion block, 160 of the fillers were semantically associated among both L1-English and L1-Swedish L2-English speakers, which elevated relatedness proportion to 50% (200/400 trials). For the related fillers, we selected pairs with FSG values ≥.40 in both groups, including those that did not meet the criteria for the 160 critical related pairs. The mean FSG for the

---

[7] A translation study with 20 L1-Swedish learners of English of advanced proficiency (LexTALE: $M = 87$, $SD = 6$, range = 75–100) showed that our materials included a comparable number of cognates and noncognates.

160 related fillers in the high-proportion block was .57 (*SD* = 0.13, range = .40–.94), according to Nelson et al.'s (2004) norms for L1-English speakers and .58 (*SD* = 0.11, range = .45–.91) based on the norms for Swedish-speaking learners. The two rightmost violin plots in Figure 1 (*Related Fillers [ERP Study]*) show that the distribution of FSG values across the 160 related fillers was also comparable for L1-English speakers and L1-Swedish L2-English learners.

Lists. The critical pairs were distributed across four experimental lists following a Latin square design, such that each prime and target appeared only once per list. Across lists and participants, each prime and target appeared in all four experimental conditions. Lists 1 and 3 included the same 80 related and 80 unrelated pairs, but in different blocks. Lists 2 and 4 included the remaining 80 related and 80 unrelated pairs, also in different blocks. Table 3 shows how the targets *rain, nut, book*, and *draw* were rotated across the four experimental lists and across the two levels of relatedness and proportion. The animal probes were also distributed across the four lists, such that each pair appeared only once per list. Each animal probe appeared in the low-proportion condition in two of the lists and in the high-proportion condition in the other two lists and always with the same targets. Thus, if the animal pairs had any impact on the critical targets, it must have been the same across the relatedness and proportion conditions (see Table 3 for the probes *bull* and *cat*). The fillers could not be counterbalanced, since the proportion of related and unrelated fillers necessarily differed across the blocks. Thus, the same sets of fillers were used in the low- versus high-proportion blocks in the four experimental lists.

The mean FSG of the critical related items was similar in the low- and high-proportion blocks in Lists 1 and 3, according to both the L1-English norms (low-proportion: $M = 0.67$, $SD = 0.12$; high-proportion: $M = 0.68$, $SD = 0.11$) and the L1-Swedish L2-English norms (low-proportion: $M = 0.69$, $SD = 0.11$; high-proportion: $M = 0.69$, $SD = 0.10$). Mean FSG was also similar in the two blocks in Lists 2 and 4 (L1-English, low-proportion: $M = 0.68$, $SD = 0.10$; high-proportion: $M = 0.66$, $SD = 0.11$; L1-Swedish L2-English, low-proportion: $M = 0.68$, $SD = 0.11$; high-proportion: $M = 0.67$, $SD = 0.11$). The mean log frequency of the targets was also similar in the two blocks in all lists (Lists 1 and 3, low-proportion: $M = 1.84$, $SD = 0.50$; high-proportion: $M = 1.92$, $SD = 0.52$; Lists 2 and 4, low-proportion: $M = 1.84$, $SD = 0.51$; high-proportion: $M = 1.91$, $SD = 0.51$).

In the L1-English group, one of the lists was used 10 times, two of the lists nine times, and one list eight times. In the L1-Swedish L2-English group, three lists were used 13 times and one list 12 times. The complete list of critical related/unrelated pairs, unrelated pairs with animal targets, related filler pairs, and unrelated filler pairs can be found both in the Supplemental Materials (see Supplementary Materials_S1_Stimuli) and at https://osf.io/r9t3q/ (see Supplementary Materials_S1_Stimuli; Alemán Bañón, 2024).

Procedure

The testing involved one 3-hr visit to the Multilingualism Lab at Stockholm University. The learners were pretested (privately on Zoom) for English proficiency, and only those learners who scored 75 or above in the LexTALE were invited to participate in the study. After providing their informed written consent, participants filled in a background questionnaire and the Edinburgh Handedness Inventory (Oldfield, 1971). Once they had been prepared for the EEG recording, they sat on a comfortable chair facing a computer monitor and received instructions that they would read a series of English words, one at a time. The participants' task was to press a button on a response pad as fast and accurately as possible whenever they encountered an animal word, for which they could only use the middle finger of their left hand. Participants were asked to avoid blinks and body movements while reading the words and to rest their eyes during the blank screens.

Participants also had an opportunity to practice with 13 prime–target pairs, and they received feedback for three of them. In total, the practice encompassed five animal words (i.e., five "go" trials), all of which were identical English–Swedish cognates that could not be used in the main experiment (*hyena, gorilla, alligator, hamster, flamingo*). Since two animal words in the main experiment were used in the plural, one of the animal words in the practice (*flamingos*) was also used in the plural.

The experiment began right after the practice. Within each of the two blocks, participants were given a short break after every 100 pairs (i.e., a total of three breaks per block). Within each block, the critical related and unrelated pairs were intermixed and randomized. Words were displayed in black text (Courier New font) against a gray background. The experiment was run on *PsychoPy* (Peirce, 2007; Peirce & MacAskill, 2018).

Trial Structure

The trial structure in the present study closely follows Lau et al.'s (2013). Each trial began with a central fixation cross, which was presented for 700 ms and followed by a 100-ms pause. Then, the prime was presented for 500 ms, followed by another 100-ms pause. The target was then presented for 900 ms, followed by a 100-ms pause. We added an interval ranging from 500 to 1,000 ms at the beginning of each pair, pseudorandomly varied at 50-ms increments.

Table 3

Sample Distribution of the Critical Pairs Across Lists and Conditions

Condition	List 1	List 2	List 3	List 4
Related, low	umbrella-*rain*	cashew-*nut*	library-*book*	sketch-*draw*
Unrelated, low	low-*nut*	cash-*rain*	salt-*draw*	starving-*book*
Unrelated animal, low	phase-*cat*	phase-*cat*	paddle-*bull*	paddle-*bull*
Related, high	library-*book*	sketch-*draw*	umbrella-*rain*	cashew-*nut*
Unrelated, high	salt-*draw*	starving-*book*	low-*nut*	cash-*rain*
Unrelated animal, high	paddle-*bull*	paddle-*bull*	phase-*cat*	phase-*cat*

Note. Words in italics are targets.

Individual Differences Measures and Proficiency Tests

After the EEG recording, participants completed the Backwards Digit Span Task (WAIS, 2004) and the Letter Comparison Task (Earles & Salthouse, 1995). Then, after a short break, they were administered the Reading Span Task (Daneman & Carpenter, 1980) and the Digit-Symbol Substitution Task (WAIS, 2004). Then, they took Raven's Progressive Matrices (Raven & Raven, 2003). Finally, the L1-English participants took the LexTALE (Lemhöfer & Broersma, 2012), whereas the L2 learners had taken it during recruitment, and all participants took a short (nonstandardized) test of Swedish proficiency, modeled after the LexTALE (Borg, 2021).

EEG Recording

The EEG was recorded continuously from 32 sintered Ag/AgCl active electrodes mounted in an elastic cap (Biosemi, Amsterdam, the Netherlands).[8] The electrodes were placed according to the International 10-20 System (midline: Fz, Cz, Pz, Oz; lateral: FP1/2, AF3/4, F3/4, F7/8, FC1/2, FC5/6, C3/4, T7/8, CP1/2, CP5/6, P3/4, P7/8, PO3/4, O1/2). Electrodes Common Mode Sense (between C3 and Cz) and Driven Right Leg (between Cz and C4) served as online reference and ground, respectively. The electrooculogram was recorded with four flat electrodes: two placed above and below the left eye (to detect blinks) and two placed on the left and right outer canthi (to detect horizontal eye movements). Activity at the mastoids was recorded with two additional flat electrodes. The use of active electrodes kept impedances very low. The signal was amplified with an ActiveTwo amplifier (Biosemi, Amsterdam, the Netherlands) and digitized continuously with a sampling rate of 2,048 Hz. The recordings were decimated offline to 1,024 Hz, with an anti-aliasing filter.

We used the Brain Vision Analyzer 2.1 software (Brain Products, GmbH, Germany) to preprocess the EEG data. We started by applying a 0.1-Hz high-pass filter to remove drift. We then rereferenced the recordings to averaged left and right mastoids. After removing electromyogram artifacts, we ran independent component analysis to remove blinks and eye movements. Bad electrodes (based on visual inspection) were interpolated via spherical spline interpolation. The EEG was then segmented into epochs in the interval from −100 to +1,000 ms relative to the onset of the critical targets. Remaining artifacts exceeding ±75 µV were automatically rejected. Trials associated with incorrect behavioral responses (i.e., failing to press the button for animal probes or pressing the button for nonanimal words) were removed from analysis. The mean number of trials per condition was comparable across the two levels of Relatedness and Proportion and across the two groups (L1-English: low-proportion related, $M = 38$, range = 32–40; low-proportion unrelated, $M = 38$, range = 31–40; high-proportion related, $M = 39$, range = 35–40; high-proportion unrelated, $M = 39$, range = 33–40; L1-Swedish: low-proportion related, $M = 38$, range = 31–40; low-proportion unrelated, $M = 38$, range = 28–40; high-proportion related, $M = 38$, range = 30–40; high-proportion unrelated, $M = 38$, range = 29–40). The same was true for the animal probes (L1-English: low-proportion, $M = 37$, range = 29–40; high-proportion, $M = 37$, range = 30–40; L1-Swedish: low-proportion, $M = 36$, range = 27–40; high-proportion, $M = 36$, range = 23–40). As for the primes, we used all of the primes that were counterbalanced, that is, those for related targets, unrelated targets, and unrelated animal words (120 trials per block). Importantly, the mean number of trials was also similar across the low- and high-proportion conditions and across the two groups (L1-English: low-proportion, $M = 114$, range = 96–120; high-proportion, $M = 114$, range = 96–120; L1-Swedish: low-proportion, $M = 113$, range = 86–119; high-proportion, $M = 113$, range = 79–120). After data trimming, we baseline-corrected the epochs relative to the 100-ms prestimulus interval. Epochs were then averaged per condition and per subject. Finally, the averaged waveforms were filtered with a phase-shift free infinite impulse response Butterworth filter, with a high cutoff of 30 Hz and a 12-dB/octave roll-off.

Results

Here, we only report effects significant at $p < .05$ and relevant nonsignificant effects. In the Supplemental Materials (see Supplementary Materials_S2_OmnibusTests), we provide tables reporting all effects from the omnibus tests. All data can be found at https://osf.io/r9t3q/ (see Lexicosemantic prediction_Data.zip), where we also provide the code for all analyses conducted in *R* (Alemán Bañón, 2024). We used a false discovery rate correction (Benjamini & Hochberg, 1995) for follow-up tests.

Behavioral Results

Mean accuracy in detecting animal words in the low-proportion block was 97% ($SD = 4\%$; range = 80%–100%) in the L1-English group and 95% ($SD = 4\%$; range = 85%–100%) in the L1-Swedish group. In the high-proportion block, mean accuracy was 95% ($SD = 5\%$; range = 78%–100%) in the L1 English group and 93% ($SD = 6\%$; range = 73%–100%) in the L1-Swedish group. We analyzed the accuracy data using generalized linear mixed-effects models via the *glmer* function from the *lme4* package (Version 1.1–35.1) in *R* (Version 4.3.2). The model included contrast-coded fixed effects for Proportion (−.5 = low, .5 = high) and Group (−.5 = L1-English, .5 = L1-Swedish), their interaction, and the maximum random effects structure that converged: random intercepts for participants and items and by-participant random slopes for proportion.[9] The results revealed a significant main effect of Proportion (estimate = −.434, $SE = .130$, $t = -3.331$, $p < .001$), with lower accuracy in the high- ($M = 94\%$, $SD = 6\%$) relative to the low-proportion block ($M = 96\%$, $SD = 4\%$), and a significant main effect of Group (estimate = −.374, $SE = .176$, $t = -2.117$, $p = .034$), with L1-English speakers showing higher accuracy overall ($M = 96\%$, $SD = 6\%$) than L2 learners ($M = 94\%$, $SD = 5\%$). The Proportion × Group interaction was not significant.

Mean response time (RT) in detecting the animal probes in the low-proportion block was 552 ms ($SD = 37$ ms) in the L1-English group and 585 ms ($SD = 41$ ms) in the L1-Swedish group, respectively. In the high-proportion block, mean RT was 574 ms ($SD = 45$ ms) in the L1-English group and 605 ms ($SD = 42$ ms) in the L1-Swedish group, respectively. We used mixed-effects models to analyze the log-transformed RT data. The model included contrast-coded fixed effects for Proportion (−.5 = low, .5 = high) and Group

---

[8] Sintering involves compressing and heating the metal particles, thus providing very low-noise measurements.

[9] The structure of the model was glmer(target.accuracy ~ proportion*group + (1 | item) + (proportion | participant)).

($-.5$ = L1-English, .5 = L1-Swedish) and the maximum random effects structure that converged: random intercepts for participants and items, by-item random slopes for group, and by-participant random slopes for proportion.[10] Analyses revealed a significant main effect of Proportion (estimate = .039, $SE = .005$, $t = 8.103$, $p < .001$), driven by the fact that both groups were slower in detecting animal words in the high- (M = 590 ms, SD = 44 ms) than in the low-proportion block (M = 569 ms, SD = 40 ms), and a significant main effect of Group (estimate = .056, $SD = 0.014$, $t = 3.892$, $p < .001$), with L1-English speakers showing faster responses (M = 563 ms, SD = 57 ms) than L2 learners (M = 595, SD = 47 ms) overall. The Proportion × Group interaction was not significant.

Thus, relatedness proportion impacted accuracy for both groups and, similar to the native English speakers in Lau et al.'s (2013) study, proportion also impacted RTs in both groups. Finally, the L1 speakers were faster and more accurate detecting the probes than the learners.

EEG Results

Our analyses are largely based on those by Lau et al. (2013), since our study is a conceptual replication of their study with native speakers of English, and we address similar questions with respect to L2 learners' ability to generate lexicosemantic predictions in the L2. Below, we explain and motivate each analysis, and we immediately report its results.

Effects of Relatedness Proportion on the Magnitude of the N400 Effect

ERPs for related/unrelated targets were quantified via mean amplitudes in the time window between 300 and 500 ms relative to the onset of the targets. This time window is standard in both theoretical and experimental reports on the N400 (e.g., Friederici, 2002; Kuperberg, 2007; Kutas & Hillyard, 1984; Lau et al., 2008) and corresponds to the time window where relatedness and proportion impacted N400 amplitude for related targets in Lau et al.'s (2013) study. Mean amplitudes across all electrodes were submitted to a mixed-effects ANOVA with Proportion (high, low) and Relatedness (related, unrelated) as the within-subjects factors and Group (L1-English, L1-Swedish) as the between-subjects factor.

Figures 2 and 3 show the results of the relatedness and proportion manipulations for the L1-English and L1-Swedish groups, respectively. The results of the omnibus ANOVA showed a large main effect of Relatedness, $F(1, 87) = 49.251$, $p < .001$, $\eta^2 = .361$, driven by the fact that related targets overall yielded less negative waveforms than unrelated ones (related: M = 1.080 μV, SD = 1.95; unrelated: M = 0.080 μV, SD = 1.68), suggestive of a priming effect on the N400. Importantly, the main effect of Relatedness was qualified by an interaction with Proportion, $F(1, 87) = 5.126$, $p = .026$, $\eta^2 = .056$, an effect of medium size. Follow-ups to this interaction at each level of Proportion revealed that, although the main effect of Relatedness was significant in both the low-proportion, $F(1, 87) = 16.612$, $p < .001$ ($q^* = .025$), $\eta^2 = .160$, and the high-proportion blocks, $F(1, 87) = 44.684$, $p < .001$ ($q^* = .012$), $\eta^2 = .339$, the effect was larger in the latter (mean amplitude for unrelated minus related, low-proportion: M = -0.733 μV, SD = 1.697; high-proportion: M = -1.266 μV, SD = 1.786). As in Lau et al. (2013), the larger N400 effect in the high-proportion block was driven by related targets becoming less negative in the high-proportion block (M = 1.303 μV, SD = 2.040), compared to the low-proportion block (M = 0.856 μV, SD = 2.254), $F(1, 87) = 5.435$, $p = .022$ ($q^* = .037$), $\eta^2 = .059$. In contrast, the waveforms for unrelated targets did not significantly differ across the high- (M = 0.037 μV, SD = 1.906) and low-proportion blocks (M = 0.123 μV, SD = 1.831).

The main effect of Group was not significant and did not interact with Proportion, Relatedness, or with the Proportion × Relatedness interaction. Thus, these results reveal a clear picture. Across L1 and L2 speakers of English, semantic relatedness yielded an N400 effect in both the low- and the high-proportion blocks. This priming effect was, however, larger in the high-proportion block, with related targets becoming less negative in the high-proportion block, relative to the same related targets in the low-proportion block. This pattern of results for L1 and L2 speakers of English is similar to that in Lau et al.'s (2013) study for L1-English speakers.

Effects of Relatedness Proportion on the Topography of the N400 Effect

In order to explore the topographical distribution of the relatedness and proportion effects on the N400, a quadrant analysis was carried out on N400 effect magnitude in the 300–500 ms time window. N400 effect magnitude was calculated by subtracting the related from the unrelated condition, separately for the low- and high-proportion blocks. We created four regions of interest representing the four quadrants of the EEG cap: Left Anterior (FP1, F7, F3, FC5, and FC1), Right Anterior (FP2, F8, F4, FC6, and FC2), Left Posterior (CP5, CP1, PO3, P3, and O1), and Right Posterior (CP6, CP2, PO4, P4, and O2). These regions are largely similar to those in Lau et al.'s (2013) study, the only difference being that we used PO3/4 instead of T5/6 in the posterior regions, since we used a different EEG system with a different array of electrodes. For this analysis, N400 magnitudes were submitted to a mixed-effects ANOVA with Proportion (high, low), Anterior–Posterior (anterior, posterior), and Hemisphere (left, right) as the within-subjects factors and Group (L1-English, L1-Swedish) as the between-subjects factor.

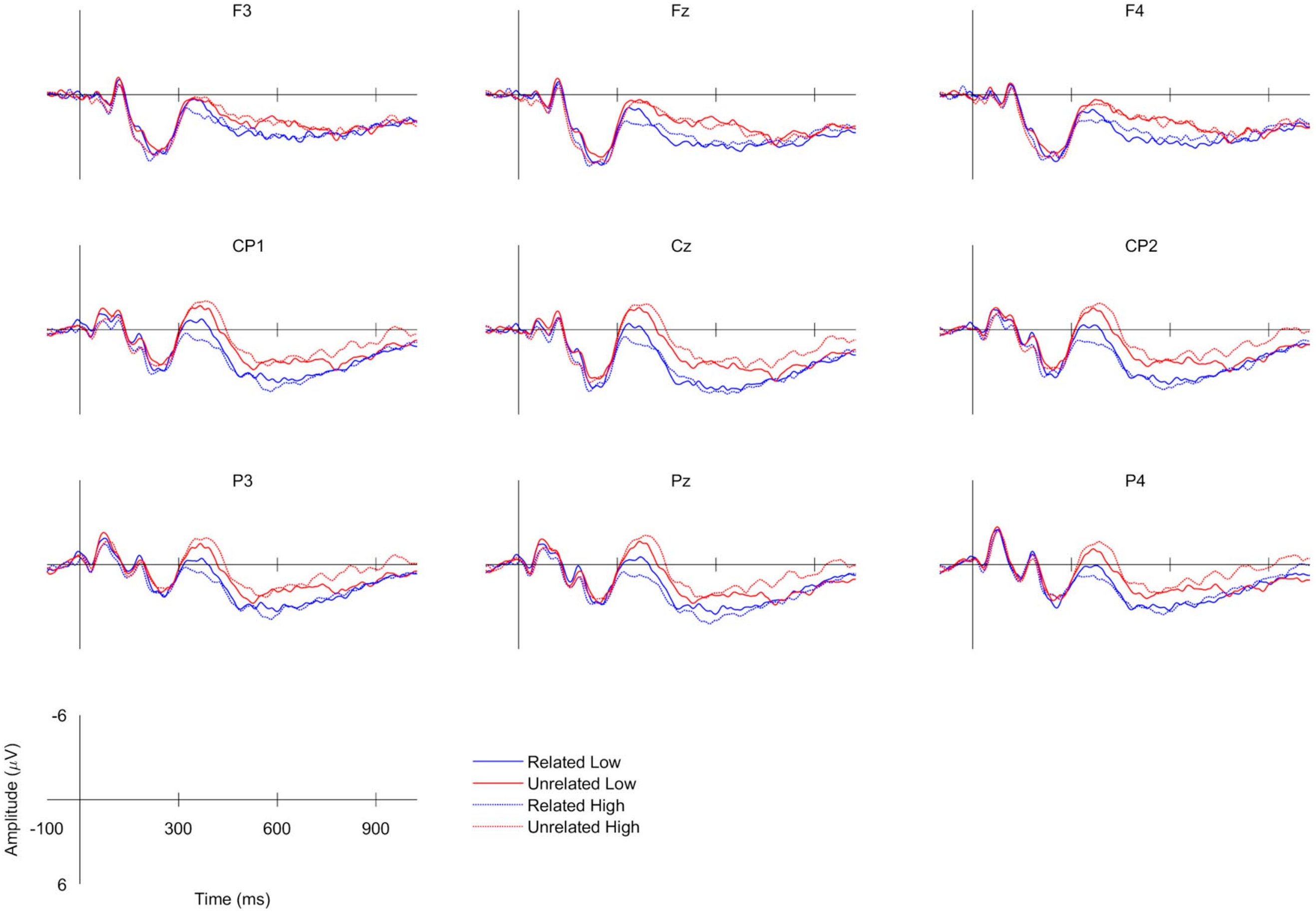
Figure 4 provides topographic plots of the effects of relatedness proportion for both groups. The results of the quadrant analysis on N400 magnitudes revealed a main effect of Proportion of medium size, $F(1, 87) = 4.898$, $p = .030$, $\eta^2 = .053$, driven by the fact that, across regions, the N400 effect was larger in the high- (M = -1.281 μV, SD = 1.800) relative to the low-proportion block (M = $-.752$ μV, SD = 1.715). This is consistent with the significant Relatedness × Proportion interaction across all electrode sites (see the Effects of Relatedness Proportion on the Magnitude of the N400 Effect section). In addition, the main effect of Anterior–Posterior was significant, $F(1, 87) = 14.125$, $p = .00031$, $\eta^2 = .17$, driven by the fact that N400 effects overall (i.e., across blocks) were larger in the posterior (M = -1.232 μV, SD = 1.506) relative to the anterior regions (M = -0.801 μV, SD = 1.400), consistent with the topography of the N400. The main effect of Hemisphere was also significant, $F(1, 87) = 6.805$, $p = .011$, $\eta^2 = .073$, with N400 effects overall being larger in the right hemisphere (M = -1.097 μV, SD = 1.351), compared to the

---

[10] The structure of the model was lmer(log(target.rt) ~ proportion*group + (group | item) + (proportion | participant)).

Figure 2

Grand Average ERPs for Related (Blue) and Unrelated (Red) Targets in the Low-Proportion (Solid) and High-Proportion (Dotted) Blocks in the L1-English Group



Note. ERPs are plotted for six electrodes within each region of interest, in addition to three midline electrodes. ERP = event-related potential; L1 = native language. See the online article for the color version of this figure.

left one ($M = -0.937$ μV, $SD = 1.409$), also consistent with the right-hemisphere bias that is characteristic of the N400. In line with this, the Anterior–Posterior × Hemisphere interaction was also significant, $F(1, 87) = 4.251$, $p = .042$, $\eta^2 = .047$. This interaction was driven by the fact that N400 magnitudes were larger in Right Anterior relative to Left Anterior, $F(1, 87) = 9.551$, $p = .003$ ($q^* = .025$), $\eta^2 = .099$ (Right Anterior: $M = -.919$ μV, $SD = 1.357$; Left Anterior: $M = -0.683$ μV, $SD = 1.529$), but there was no difference between the two posterior regions.

Importantly, Proportion interacted with Anterior–Posterior, $F(1, 87) = 11.596$, $p = .001$, $\eta^2 = .118$. Follow-up analyses at each level of Proportion revealed that the main effect of Anterior–Posterior was only significant in the high-proportion block, $F(1, 87) = 29.881$, $p < .001$ ($q^* = .012$), $\eta^2 = .256$. Examination of the condition means suggests that the N400 priming effect in the high-proportion block had a more posterior focus (posterior: $M = -1.655$ μV, $SD = 2.075$; anterior: $M = -0.907$ μV, $SD = 1.734$).

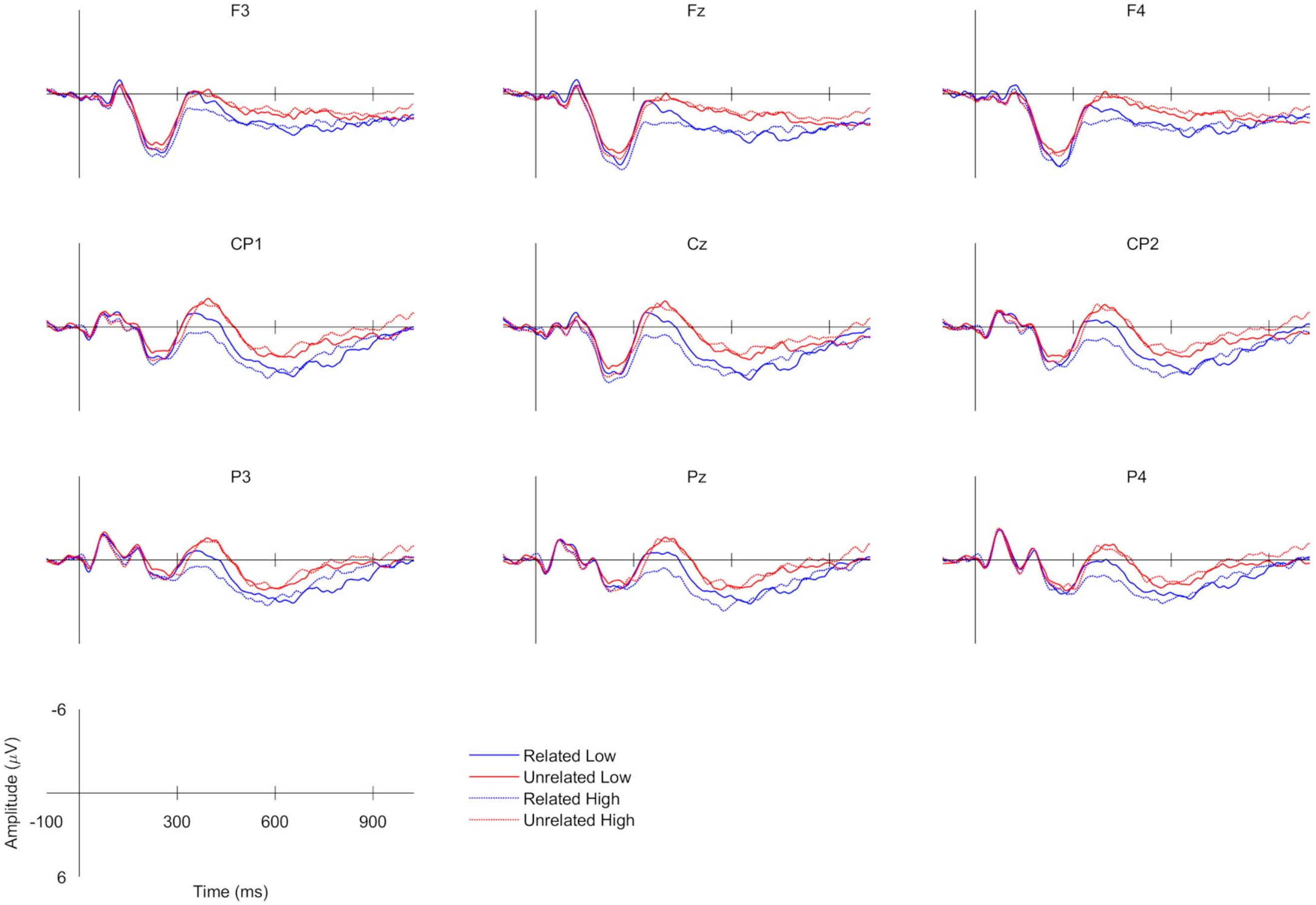
As with the analyses on the effects of relatedness proportion on the size of the N400 priming effect, the main effect of Group was not significant and did not interact with Proportion, with the topographical factors, or with their interaction terms. Thus, these results suggest that, across native and nonnative speakers, the N400 effect only had a posterior focus in the high-proportion block. These results are similar to those in Lau et al.'s (2013) in that relatedness proportion impacted the topographical distribution of the N400 priming effect, although the specific pattern of results differs slightly across the two studies, since the N400 effect in the high-proportion block in our study did not show a right-hemisphere focus, only a posterior one.

Effects of Relatedness Proportion on the Onset of the N400 Effect

In order to examine the effects of the proportion manipulation on the onset of the N400 priming effect, mean amplitudes for related/unrelated targets were quantified in the time window between 200 and 250 ms relative to the onset of the targets. This is the time window where Lau et al. (2013) reported an earlier onset for the N400 priming effect in the high- versus the low-proportion block, based on the results of a permutation test on electrode Cz.

Figure 3

Grand Average ERPs for Related (Blue) and Unrelated (Red) Targets in the Low-Proportion (Solid) and High-Proportion (Dotted) Blocks in the L1-Swedish Group



Note. ERPs are plotted for six electrodes within each region of interest, in addition to three midline electrodes. ERP = event-related potential; L1 = native language. See the online article for the color version of this figure.

For these analyses, mean amplitudes across all electrodes were submitted to a mixed-effects ANOVA with Proportion (high, low) and Relatedness (related, unrelated) as the within-subjects factors and Group (L1-English, L1-Swedish) as the between-subjects factor.

The results only revealed a main effect of Relatedness of medium size, $F(1, 87) = 6.013$, $p = .016$, $\eta^2 = .065$, driven by the fact that related targets yielded less negative waveforms than unrelated ones overall (related: $M = 2.390$ μV, $SD = 1.930$; unrelated: $M = 2.154$ μV, $SD = 2.049$). A similar pattern of results emerged when only electrode Cz was analyzed (as in Lau et al., 2013), $F(1, 87) = 8.228$, $p = .005$, $\eta^2 = .086$. As a follow-up, we also explored the 150–200 ms time window (Brothers et al., 2019), but here the N400 effect was yet to emerge. Thus, unlike in Lau et al.'s (2013) study, our results provide no evidence that the N400 effect began earlier in the high-proportion block. Between 200 and 250 ms, the N400 effect had already emerged in both blocks (see Figures 2 and 3). Once again, the main effect of Group was not significant and did not interact with Relatedness, Proportion, or with the Relatedness × Proportion interaction.

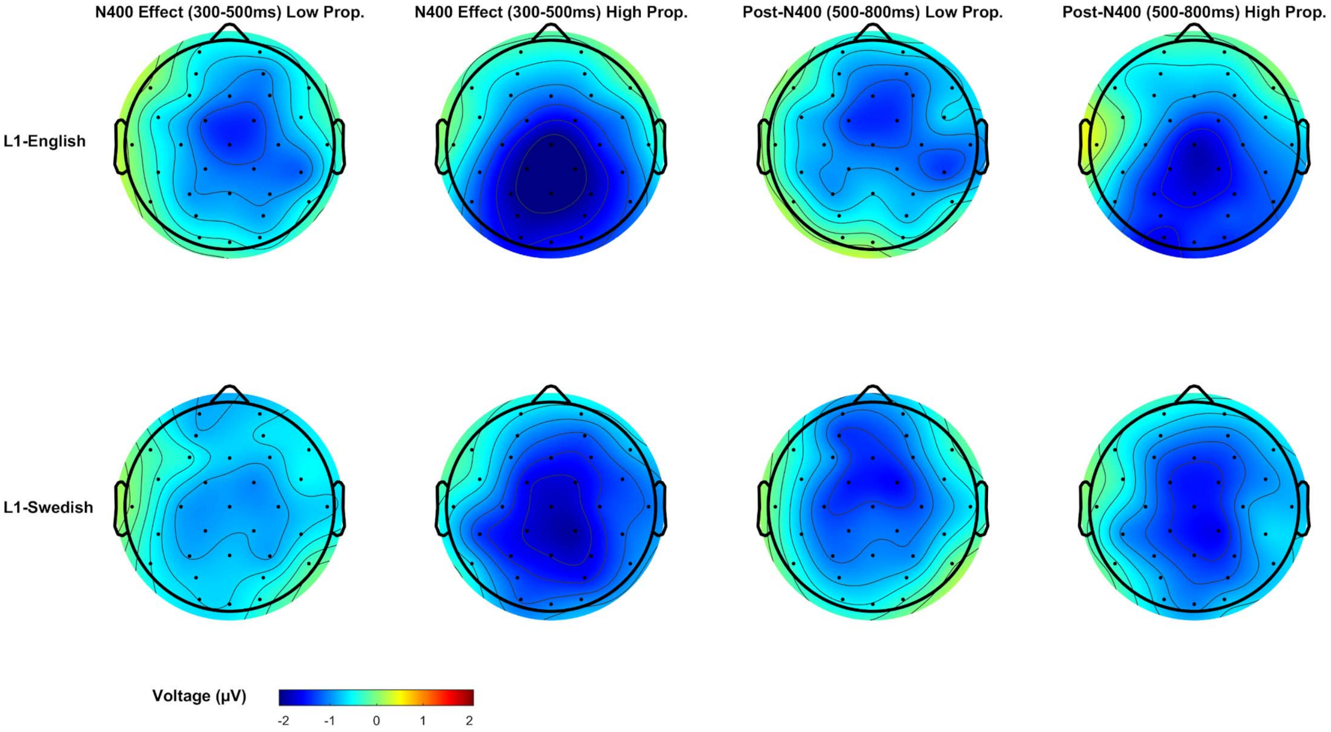
Effects of Unmet Predictions on Targets

In order to examine the cost of failed predictions, ERPs for related/unrelated targets were quantified via mean amplitudes in the time window between 500 and 800 ms relative to the onset of the targets, corresponding to the time window where Lau et al. (2013) found a larger negativity for unrelated relative to related targets in the high-proportion block, relative to the low-proportion block. Mean amplitudes across all electrodes were submitted to a mixed-effects ANOVA with Proportion (high, low) and Relatedness (related, unrelated) as the within-subjects factors and Group (L1-English, L1-Swedish) as the between-subjects factor.

These analyses only revealed a significant main effect of Relatedness, $F(1, 87) = 25.936$, $p < .001$, $\eta^2 = .023$, driven by the

Figure 4

Topographic Plots for the N400 Effect and the Effects in the Post-N400 Window in the Low-Proportion and High-Proportion Blocks for L1-English Speakers (Upper Row) and L1-Swedish L2-English Learners (Lower Row)



Note. Plots were computed by subtracting the related from the unrelated condition between 300–500 ms and 500–800 ms. Prop. = proportion; L1 = native language; L2 = second language. See the online article for the color version of this figure.

fact that unrelated targets yielded more negative brain responses than related ones across both levels of Proportion (unrelated: $M = 1.312$ µV, $SD = 1.602$; related: $M = 2.189$ µV, $SD = 2.055$). Thus, similar to Lau et al. (2013), we found a negativity for unrelated targets in the 500–800 ms time window. However, unlike Lau et al. (2013), we did not find that this later negativity for unrelated targets was restricted to the block that encouraged predictive processing. This can be seen in Figures 2–4.

We hypothesized, however, that the effects of relatedness proportion on the negativity for unrelated targets might be topographically restricted and, therefore, we ran an additional ANOVA on difference waves (mean amplitudes for unrelated minus related targets) with Proportion (low, high), Anterior–Posterior (anterior, posterior), and Hemisphere (left, right) as the within-subjects factors and Group (L1-English, L1-Swedish) as the between-subjects factor. Results showed a significant main effect of Anterior–Posterior, $F(1, 87) = 4.612$, $p = .035$, $\eta^2 = .050$, which was qualified by an interaction with Proportion, $F(1, 87) = 7.243$, $p = .009$, $\eta^2 = .077$. We followed up on the significant Proportion × Anterior–Posterior interaction by examining the effects of Proportion separately in the anterior and posterior regions, but these analyses yielded no significant effects of Proportion, consistent with the analyses across all electrodes. We also examined the effects of Anterior–Posterior (i.e., the topographical distribution of the negativity) at each level of Proportion and found a main effect of Anterior–Posterior in the high-proportion block only, $F(1, 87) = 15.520$, $p < .00016$ ($q^* = .012$), $\eta^2 = .151$, driven by the fact that the negativity was larger in the posterior ($M = -1.226$ µV, $SD = 2.327$) relative to the anterior regions ($M = -.702$ µV, $SD = 1.999$). To sum up, these analyses showed a late negativity for unrelated relative to related targets across both the low- and the high-proportion blocks, in both L1 and L2 speakers. Although the size of the negativity was not impacted by the relatedness proportion manipulation, it showed a posterior maximum only in the high-proportion block.

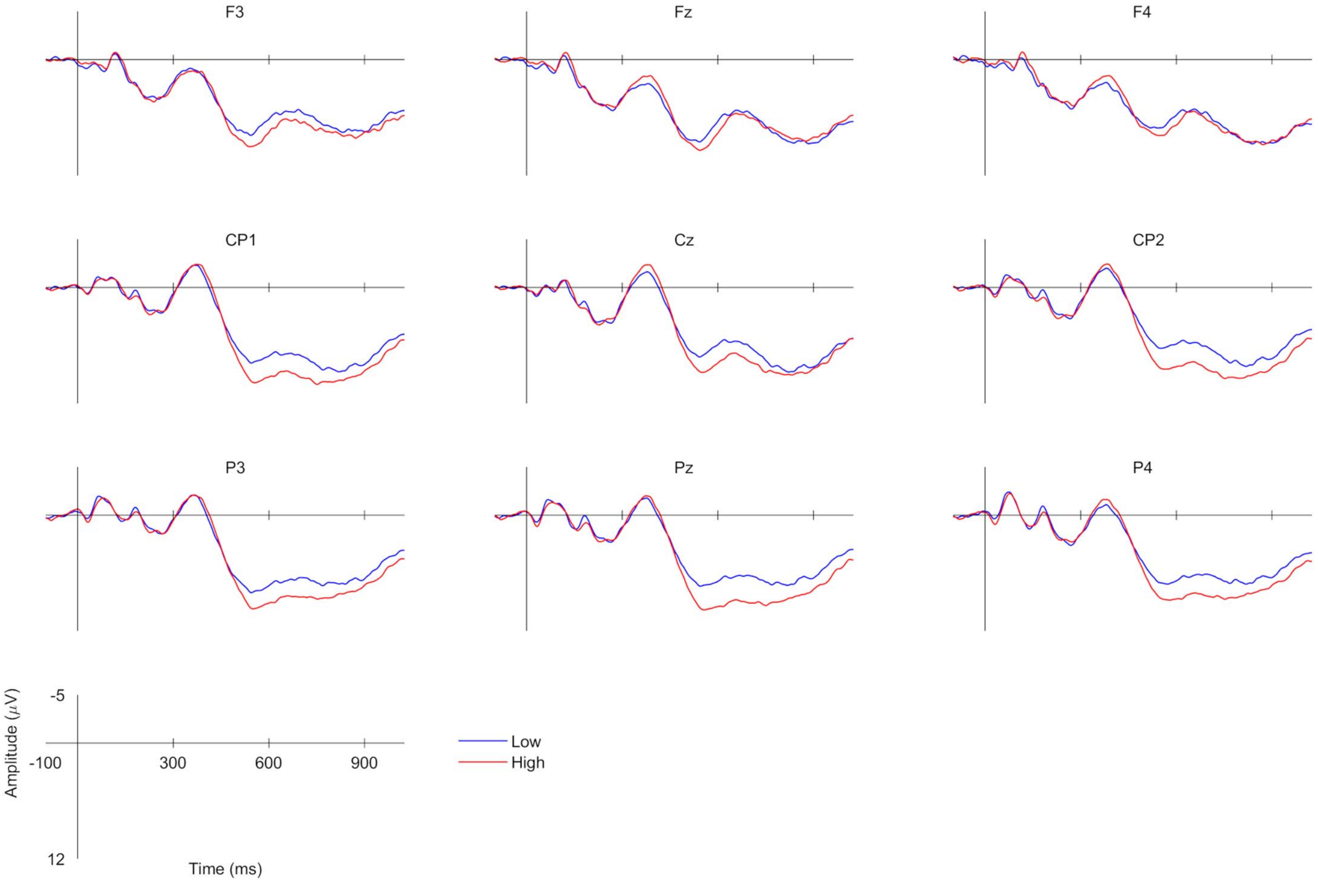
Effects of Relatedness Proportion on Animal Probes

For this analysis, ERPs were quantified via mean amplitudes between 500 and 800 ms relative to the onset of the animal probes. This time window was selected based on visual inspection of the waveforms from both L1 and L2 speakers, and it comprises the 600–700 ms window where Lau et al. (2013) found a larger P300 for animal probes in the high- versus the low-proportion conditions via a permutation test (see also Folstein & Van Petten, 2011, on P300 latency). We opted for a quadrant analysis, since the P300 comprises (at least) two subcomponents that are sensitive to different factors (P3a: novelty of the stimuli; P3b: stimulus relevance) and exhibit different topographical maxima (frontal vs. posterior). Thus, mean amplitudes were submitted to a mixed-effects ANOVA with Proportion (high, low), Anterior–Posterior (anterior, posterior), and Hemisphere (left, right) as the within-subjects factors and Group (L1-English, L1-Swedish) as the between-subjects factor.

Figures 5 and 6 show the effects of Proportion on the animal words in the L1-English and L1-Swedish groups, respectively. Similar to Lau et al. (2013), both L1 and L2 speakers in our study

Figure 5

Grand Average ERPs for Animal Probes in the Low-Proportion (Blue) and High-Proportion (Red) Blocks in the L1-English Group



Note. ERPs are plotted for six electrodes within each region of interest, in addition to three midline electrodes. ERP = event-related potential; L1 = native language. See the online article for the color version of this figure.

elicited a P300 component for animal words in the two blocks, as expected for go/no-go behavioral tasks.

The results of the omnibus ANOVA in the 500–800 ms time window revealed a significant main effect of Proportion, $F(1, 87) = 19.572$, $p < .001$, $\eta^2 = .184$, which was qualified by an interaction with Anterior–Posterior, $F(1, 87) = 16.830$, $p < .001$, $\eta^2 = .162$, and by an interaction with Anterior–Posterior and Hemisphere, $F(1, 87) = 4.635$, $p = .034$, $\eta^2 = .051$. We evaluated the three-way interaction by examining the effects of Proportion within each of the four regions and found that, after controlling for Type I error, the larger P300 for animal probes in the high-proportion block was significant in all four regions: Left Anterior, $F(1, 87) = 9.648$, $p = .003$ ($q^* = .037$), $\eta^2 = .100$ (high-proportion: $M = 5.741$ μV, $SD = 4.193$; low-proportion: $M = 4.908$ μV, $SD = 3.748$); Right Anterior, $F(1, 87) = 6.254$, $p = .014$ ($q^* = .05$), $\eta^2 = .067$ (high-proportion: $M = 5.407$ μV, $SD = 4.377$; low-proportion: $M = 4.685$ μV, $SD = 3.761$); Left Posterior, $F(1, 87) = 25.283$, $p < .001$ ($q^* = .025$), $\eta^2 = .225$ (high-proportion: $M = 7.903$ μV, $SD = 5.215$; low-proportion: $M = 6.370$ μV, $SD = 4.803$), and Right Posterior, $F(1, 87) = 28.699$, $p < .001$ ($q^* = .012$), $\eta^2 = .248$ (high-proportion: $M = 7.474$ μV, $SD = 5.009$; low-proportion: $M = 5.746$ μV, $SD = 4.758$). Evaluation of the condition means suggests that the three-way interaction was driven by the fact that the larger P300 for animal probes in the high-proportion block had a posterior maximum, which was more pronounced in the right hemisphere.

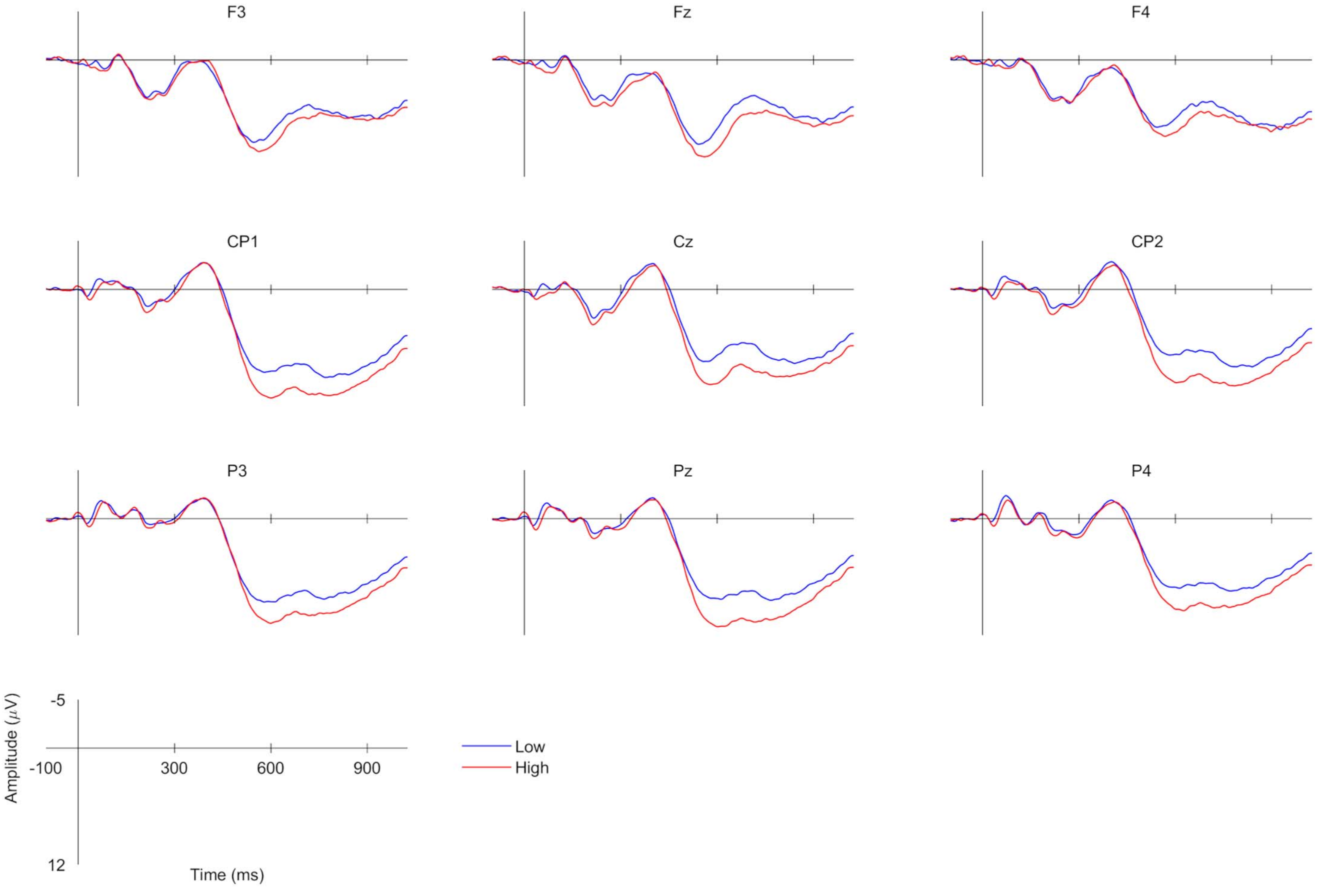
Once again, the main effect of Group was not significant and did not interact with Proportion, with the topographical factors, or with their interaction terms. To summarize, similar to Lau et al.'s (2013) study, these results show a larger P300 for animal probes in the block that encouraged predictive processing, mainly in posterior regions, across both native and nonnative speakers.

Effects of Relatedness Proportion on the Primes

For this analysis, we quantified mean amplitudes in the time windows between 200–300 ms and 350–400 ms relative to the onset of the primes. These are the time windows where Lau et al. (2013) found a frontocentral positivity for the primes in the high-proportion block, relative to their low-proportion counterparts, which might reflect prediction instantiation. Mean amplitudes were submitted to a mixed-effects ANOVA with Proportion (high, low), Anterior–Posterior (anterior, posterior), and Hemisphere (left, right) as the within-subjects factors and Group (L1-English, L1-Swedish) as the between-subjects factor.

Figure 6

Grand Average ERPs for Animal Probes in the Low-Proportion (Blue) and High-Proportion (Red) Blocks in the L1-Swedish Group



Note. ERPs are plotted for six electrodes within each region of interest, in addition to three midline electrodes. ERP = event-related potential; L1 = native language. See the online article for the color version of this figure.

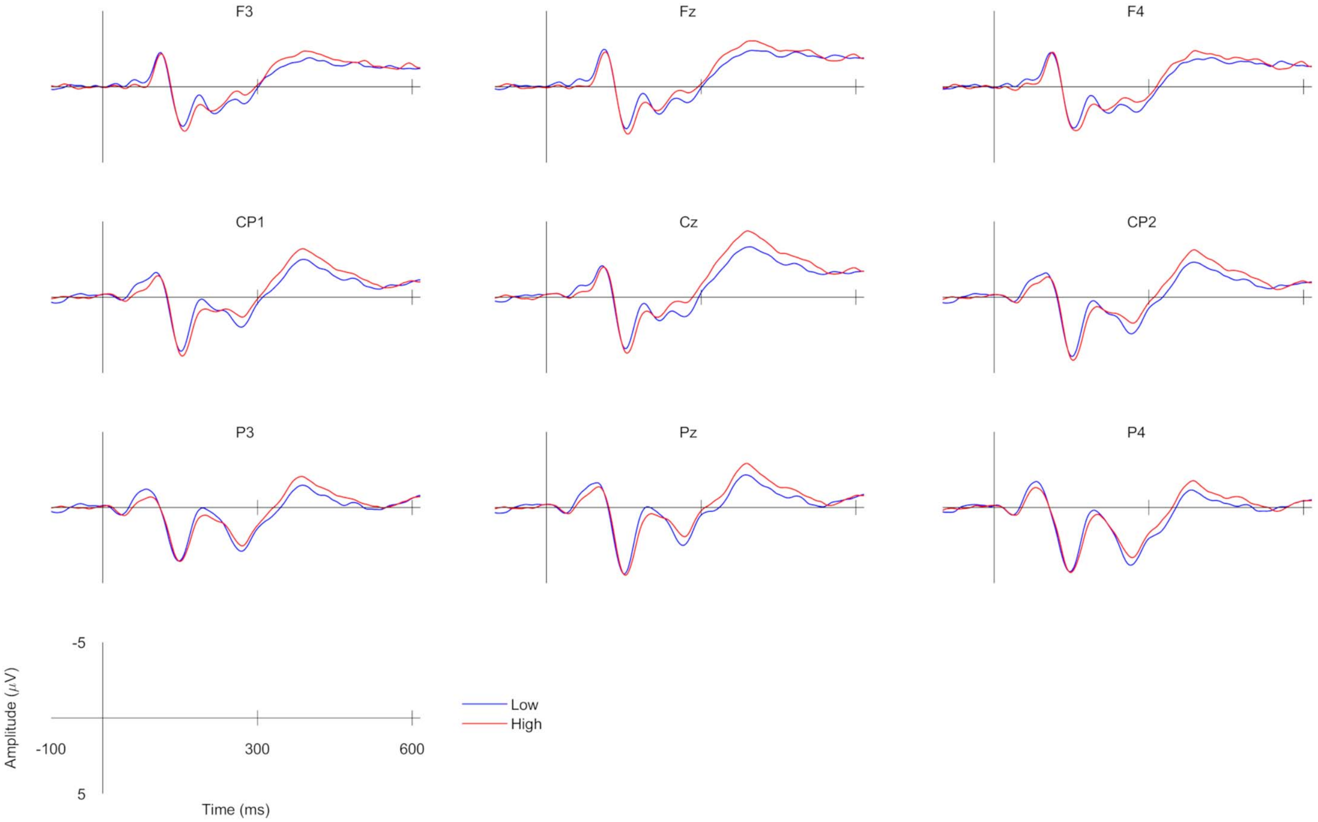
Figures 7 and 8 show the effects of relatedness proportion on the primes in the L1-English and L1-Swedish groups, respectively. In both the 200–300 ms and the 350–400 ms time windows, the omnibus ANOVA revealed a significant Proportion × Group interaction, 200–300 ms: $F(1, 87) = 5.977$, $p = .017$, $\eta^2 = .064$; 350–400 ms: $F(1, 87) = 9.140$, $p = .003$, $\eta^2 = .095$. In both cases, the L1-English speakers yielded more negative waveforms for high- versus low-proportion primes (unlike the L1-English speakers in Lau et al.'s 2013 study), while the L1-Swedish speakers showed the reverse pattern (in line with the L1-English speakers in Lau et al.'s 2013 study). In the 200–300 ms window, follow-up analyses by Group revealed that the negativity was not significant in the L1-English group. In the L1-Swedish group, the positivity was only marginal, even before correcting for Type I error, $F(1, 52) = 3.575$, $p = .064$, $\eta^2 = .064$. In the 350–400 ms window, follow-up analyses by Group only revealed a significant negativity in the L1-English group, $F(1, 35) = 8.382$, $p = .004$ ($q^* = .025$), $\eta^2 = .211$ (again, unlike the L1-English speakers in Lau et al.'s 2013 study).

We also conducted an exploratory analysis which consisted of point-by-point paired-samples *t* tests comparing every sampling point between 100 ms postonset of the prime and 600 ms (corresponding to the end of the pause between the prime and the target) in the low- and high-proportion blocks, in all 32 electrodes. To control for Type I error, only cases where the difference between the two conditions was significant at $p < .005$ over at least 15 consecutive samples in at least five adjacent electrodes simultaneously were considered significant (see Szűcs & Soltész, 2010). This analysis was conducted separately for the L1-English and the L1-Swedish speakers.

The results of this conservative point-by-point analysis revealed that the L1-English group elicited a positivity for the primes in the high- relative to the low-proportion block between 161 and 186 ms mainly in frontocentral electrodes (FP1, FP2, AF3, AF4, F7, F3, Fz, F4, FC5, FC1, FC2, C3, Cz, C4, CP1, CP2, CP6). In addition, consistent with the above analyses in our predetermined time windows, the L1-English speakers showed a negativity for the primes in the high-proportion block, relative to their low-proportion counterparts, between 370 and 386 ms in central electrodes (FC1, FC2, Cz, C4, CP2, CP6). In the L1-Swedish group, this conservative analysis only revealed a positivity for the high-proportion primes between 582 and 600 ms, corresponding to the pause between the prime and the target, in frontal electrodes (AF3, AF4, F3, Fz, F4).

Figure 7

Grand Average ERPs for Prime Words in the Low-Proportion (Blue) and High-Proportion (Red) Blocks in the L1-English Group



Note. ERPs are plotted for six electrodes within each region of interest, in addition to three midline electrodes. ERP = event-related potential; L1 = native language. See the online article for the color version of this figure.

Individual Differences in Working Memory and Processing Speed

In order to investigate the extent to which individual differences in working memory and processing speed explained variability in L1 and L2 speakers' predictive processing (Research Question 2), we modeled N400 amplitude in the high-proportion block (i.e., the predictive block) using mixed-effects models via the *lmer* function from the *lme4* package (Version 1.1-35.1) in *R* (Version 4.3.2). The model included contrast-coded fixed effects for Relatedness (−.5 = unrelated, .5 = related) and Group (−.5 = L1-English, .5 = L1-Swedish), in addition to three continuous (centered) predictors: Working_Memory, Processing_Speed, and Raven's_Score. The model also included four two-way interactions: Relatedness × Group, Relatedness × Working_Memory, Relatedness × Processing_Speed, and Relatedness × Raven's_Score. Finally, the model allowed by-item and by-participant random intercepts and slopes for Relatedness.[11]

Working_Memory and Processing_Speed were operationalized as composite scores calculated by averaging across two separate scores for each factor (see the Participants section). We included Raven's_Score in the model, in order to determine the individual contribution of working memory and processing speed over and above the effects of general intelligence. Likewise, although the above ANOVAs (see the Effects of Relatedness Proportion on the Magnitude of the N400 Effect section) revealed comparable effects of relatedness proportion on N400 magnitude for both L1 and L2 speakers, we included Group as a predictor, in order to evaluate the individual contributions of Working_Memory and Group, given that the L1-English group outperformed the L1-Swedish speakers in the working memory tasks (see the Participants section).

After running the model, we checked for multicollinearity with the *check_collinearity* function from the *performance* package (Version 0.11.0) in *R*, which revealed that multicollinearity was not a concern. As expected, the results of these analyses, which are summarized in Table 4, revealed a significant main effect of Relatedness, consistent with the above ANOVAs. However, Relatedness did not significantly interact with any of the other factors, suggesting that the prediction effect in our study was not modulated by working memory, processing speed, general intelligence, or native status. Raven's_Score did significantly predict overall N400 amplitude, but it did not interact with Relatedness.

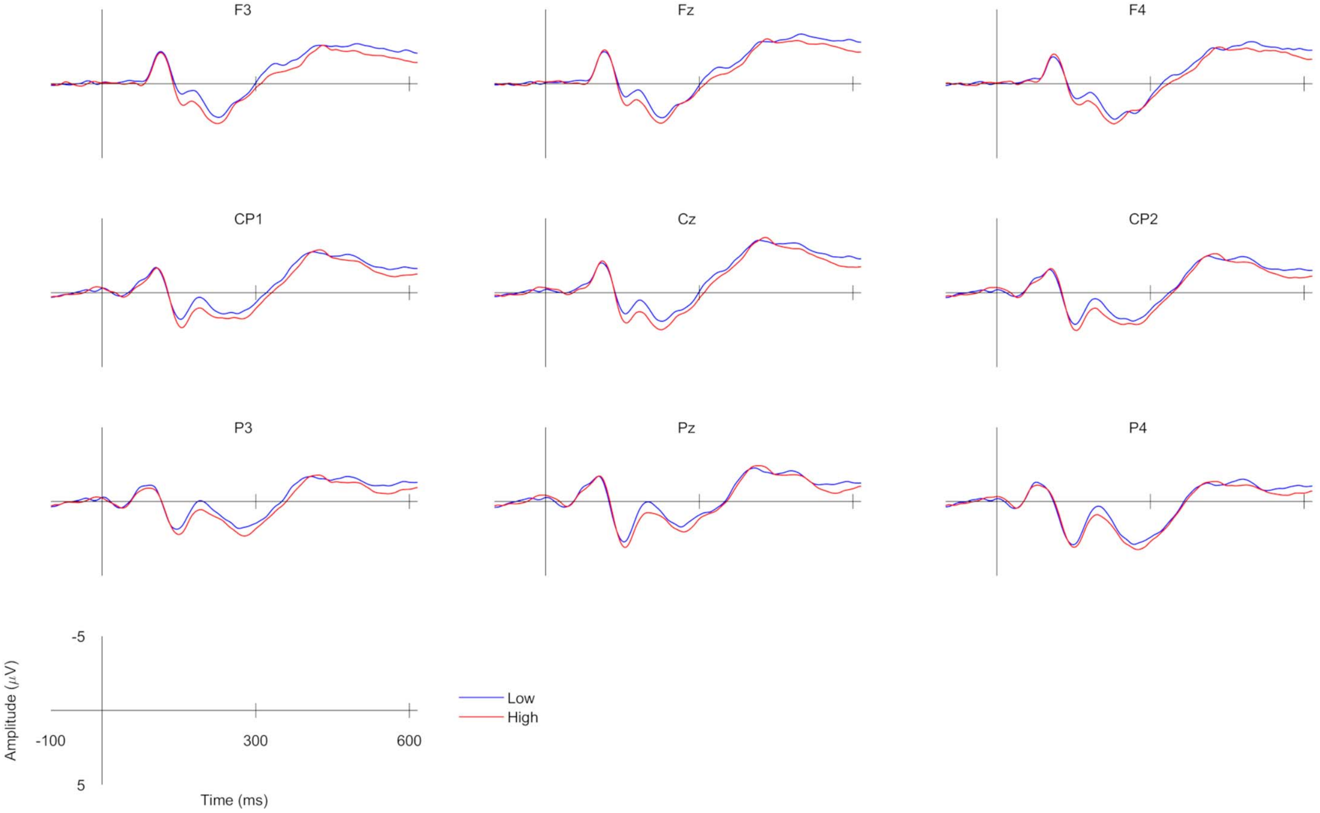
We then explored the possibility that the lack of significant relationships between N400 effect size in the high-proportion block (i.e., the prediction effect) and the cognitive factors that we tested

---

[11] The structure of the model was as follows: lmer(N400 ~ relatedness*group + relatedness*wm + relatedness*speed + relatedness*raven + (relatedness | item) + (relatedness | participant)).

Figure 8

Grand Average ERPs for Prime Words in the Low-Proportion (Blue) and High-Proportion (Red) Blocks in the L1-Swedish Group



Note. ERPs are plotted for six electrodes within each region of interest, in addition to three midline electrodes. ERP = event-related potential; L1 = native language. See the online article for the color version of this figure.

was due to the individual differences in N400 effect size having low reliability. It has recently been pointed out that experimental effects that are large and replicable at the group level, as is the case with the N400 effect in the predictive block ($\eta^2 = .34$), are less likely to show consistent variability across participants, which makes such effects less suitable for investigations of individual differences, since low consistency impacts statistical power (e.g., Hedge et al., 2018; see also Cunnings & Fujita, 2021; James et al., 2018; Staub, 2021). To address this issue, we first verified that our study had indeed captured sufficient variability in N400 effect size across all 89 participants. We did so by comparing two nested linear mixed-effects models analyzing N400 amplitude as a function of a contrast-coded fixed effect of Relatedness (−.5 = unrelated, .5 = related): a base model that did not include by-participant random slopes for Relatedness and a larger model that did. The two models were otherwise identical and both included by-item random intercepts and

Table 4

Results of the Mixed-Models Analysis on N400 Amplitude in the High-Proportion Block

Fixed effect	Estimate	*SE*	*df*	*t*	*p*
Intercept	0.6337	0.2145	0.9141	2.954	<.01
Relatedness	1.6840	0.2399	0.8182	7.017	**<.00001**
Group	−0.0630	0.4311	0.8819	−0.146	.88419
Working memory	0.6998	1.4800	0.0107	0.473	.63719
Processing speed	−0.0020	0.0011	0.8892	−1.821	.07194
Raven score	0.1127	0.0492	0.8843	2.289	**<.05**
Relatedness × Group	0.3142	0.4567	0.8585	0.688	.49330
Relatedness × Working Memory	0.3881	1.6670	0.9178	0.233	.81645
Relatedness × Processing Speed	0.0008	0.0012	0.8608	0.711	.47885
Relatedness × Raven Score	0.0588	0.0522	0.8712	1.126	.26329

Note. N400 amplitude was modeled for the 300–500 ms time window, across all five electrodes of the Right Posterior region. Values in bold indicate effects that are significant at $p < .05$. *SE* = standard error.

slopes for Relatedness and by-participant random intercepts for Relatedness.[12] We omitted Group as a fixed effect since it did not interact with Relatedness in the original model reported in Table 4. This model comparison, which was conducted via the *anova* function in *R*, revealed that the difference between the two models was significant, $\chi^2 = 8.381$, $df = 2$, $p = .015$, suggesting that the inclusion of by-participant random slopes for Relatedness improved model fit. This suggests that the N400 effect in our study did indeed differ across participants.

We then assessed the split-half reliability of those individual differences in N400 effect size, which involved splitting the data into two halves, modeling the N400 effect from each half, and calculating the correlation between the participants' estimates for the N400 effect (i.e., the by-participant slopes for Relatedness) from each half. Although there is no agreed-upon magnitude threshold for assuming reliable split-half correlations, a correlation coefficient of at least .7 has been taken as an indication that the observed individual differences are systematic (Cunnings & Fujita, 2021; Hedge et al., 2018; Staub, 2021). The data were split such that the first half included a random selection of 20 related and 20 unrelated items per participant, and the second half included the remaining 20 related and 20 unrelated trials (before artifact rejection). We then used the *brm* function from the *brms* package (Version 2.21.0; Bürkner, 2017) in *R* (Version 4.3.2) to fit a Bayesian linear mixed-effects model to estimate N400 amplitude as a function of two fixed effects: the relatedness effect in the first half of the data and the relatedness effect in the second half. For each effect, unrelated was coded as −.5, related as .5, and trials from the other half were coded as 0 (e.g., Staub, 2021). The model included by-participant and by-item random intercepts and slopes for both fixed effects.[13] We assumed default priors for all parameters. The function generated four chains, each with 4,000 samples, 1,000 of which were warmup. This process was repeated 100 times, so that split-half reliability would not be determined based on a single data split.

In all cases, the R-hat values for all parameter estimates had a value of 1, which indicates good convergence. The critical parameter for the purposes of assessing split-half reliability is the parameter of the correlation between the by-participant slopes for Relatedness from the two data halves. Importantly, the posterior mean of this correlation parameter (averaged across the 100 permutations) was 0.33 (range = .10–.47), which is well below the accepted .7 value. In addition, the highest density interval was very wide and zero fell under it in all cases, suggesting that the models were highly uncertain about the true value of the parameter.[14]

Discussion

The present study investigated lexicosemantic prediction in 36 native speakers of English and 53 high-proficiency L2 learners of English with Swedish as their L1. The main goal of the study was to examine whether adult L2 learners engage lexicosemantic prediction to the same extent as L1 speakers, in light of recent claims that prediction might be less pervasive among L2 learners due to its reduced utility (Grüter & Rohde, 2021; see also Kaan & Grüter, 2021). More specifically, it has been argued that predictive cues might not have the same reliability for L1 and L2 speakers, which might discourage learners from engaging predictive strategies. In addition, the study investigated the extent to which individual differences in working memory and processing speed modulated predictive processing in both L1 and L2 speakers, as suggested by previous studies (Huettig & Janse, 2016; Ito, Corley, & Pickering, 2018) and theoretical proposals (Lau et al., 2013).

Participants read prime–target pairs in English, one word at a time, and monitored for animal words while their brain activity was recorded with EEG. We focused on the N400, an ERP component that is sensitive to the ease of lexical access and retrieval, including lexical prediction. The study involved two manipulations to which participants were naïve. First, some prime–target pairs showed a strong semantic association, which is known to facilitate the processing of the target, as reflected by a reduction in the amplitude of the N400 for related (*table*-*chair*) relative to unrelated targets (*uncooked*-*chair*). The second (and most crucial) manipulation consisted of increasing the proportion of related pairs from 10% to 50% halfway through the experiment (via fillers), with the aim of increasing the reliability of the primes as predictive cues and encouraging predictive processing. Although the primes had the same strongest associate for both the L1 and the L2 speakers, and mean association strength across all related pairs was similar for both groups, L2 learners' semantic networks are generally less homogenous, less stable, and more erroneous than those of L1 speakers (e.g., Meara, 2009), and learners are slower than L1 speakers when producing semantic associations (e.g., Fitzpatrick & Izura, 2011). Thus, we reasoned that L2 learners would predict to some extent (e.g., Dijkgraaf et al., 2017, 2019; Foucart et al., 2014) but to a lesser extent than L1 speakers. Below, we summarize and discuss our findings.

Effects of Proportion on the Magnitude, Onset, and Scalp Distribution of the N400 Effect

Our results revealed an N400 reduction for related relative to unrelated targets in the 300–500 ms time window, in both blocks and across both groups. Crucially, this N400 reduction for related targets was larger in the high-proportion block across both groups, as indicated by a significant interaction between Relatedness and Proportion, not qualified by Group. These results, which replicate those by Lau et al. (2013) for L1-English speakers, are consistent with the possibility that, when the reliability of the primes as predictive cues increased in the second half of the experiment, both L1 and L2 speakers started using them to predict the targets or, at least, some of their semantic features. In turn, this facilitated the processing of the targets to a larger extent than in the block that did not encourage active prediction generation. It is also interesting that the L2 learners showed a native-like N400 effect in the low-proportion block, which suggests that, at high levels of proficiency, the spread of semantic activation can be as strong in the L2 as in the

---

[12] The structure of the base model was as follows: lmer(N400 ~ relatedness + (1 | participant) + (relatedness | item)), and the structure of the larger model was as follows: lmer(N400 ~ relatedness + (relatedness | participant) + (relatedness | item)).

[13] The structure of the model was as follows: brm(N400 ~ half1 + half2 + (half1 + half2 | participant) + (half1 + half2 | item)).

[14] One caveat with this approach is that, in our study, within-subject effects of relatedness were not expected to remain stable within the block. Rather, they were expected to increase as the cumulative proportion of related pairs increased within the block. By using 100 permutations, this issue must have been ameliorated to some extent, but these results must be interpreted with caution.

L1 (cf. Dijkgraaf et al., 2019). This is important since some researchers conceive passive spreading activation as some form of predictive mechanism that is automatic, uncontrollable, unconscious, and highly inaccurate (Ito & Pickering, 2021; Pickering & Gambi, 2018; see also Huettig, 2015). If on the right track, our results suggest that such mechanism is fully operative among advanced L2 learners.

Relatedness proportion also impacted the scalp distribution of the N400 effect. Our analysis on N400 effect size revealed that, across both L1 and L2 speakers, the N400 effect was broadly distributed in the low-proportion block but had a posterior maximum in the high-proportion block, as corroborated by a significant Proportion × Anterior–Posterior interaction, not qualified by Group. These distributional differences suggest that partly different neural generators were responsible for the two N400 effects. At a minimum, this suggests that the processing mechanisms engaged in the low- and high-proportion blocks were qualitatively different, consistent with the possibility that both L1 and L2 speakers recruited predictive mechanisms in the high-proportion block. Lau et al. (2013) also reported effects of relatedness proportion on the topography of the N400 effect, although in their study both N400 effects showed a posterior distribution, and the effect in the predictive block also had a right posterior bias. It is unclear what explains the topographical differences between our study and Lau et al.'s (2013), especially since the effects of relatedness proportion on the magnitude of the N400 effect were similar across the two studies. For comparability with Lau et al. (2013), we repeated this analysis with only the 36 L1-English speakers, and we found a similar pattern of results to the analysis that also included the 53 L2 learners. Thus, we can rule out the L2 data as the cause for the differential distributional patterns across the two studies. Crucially, both studies provide evidence that an increase in the predictability of the targets impacts the topography of the N400 effect, consistent with the possibility that different neural generators were engaged in the two blocks. As Lau et al. (2013) pointed out, this constitutes support against an alternative interpretation of the larger N400 effect in the high-proportion block as merely reflecting a boost in spreading activation, rather than a qualitative change in processing mechanisms.

Relatedness proportion did not impact the onset of the N400 effect, unlike in Lau et al.'s (2013) study. The two studies seem to differ the most with respect to the low- as opposed to the high-proportion block. In Lau et al.'s (2013) study, the N400 effect emerged between 200 and 250 ms in the high-proportion block and between 400 and 450 ms in the low-proportion block. In the present study, the N400 effect had already emerged between 200 and 250 ms in both blocks. That the N400 effect in the low-proportion block began earlier in our study than in Lau et al.'s (2013) is important because it suggests that there was no general delay in the onset of the N400 effect in the present study due to two characteristics of our participants: their bilingualism and their age. With respect to the former, bilinguals tend to have slower lexical access than monolinguals, even in their dominant language (e.g., Bialystok et al., 2008; Gollan et al., 2005, 2011). Although all of the participants in the present study grew up as monolingual speakers of English and Swedish, respectively, most of them were bilingual (or even multilingual) at the time of testing. Regarding our participants' age, work by Federmeier et al. (2002, 2003, 2005, 2010) has shown that ageing causes a delay or decline in predictive abilities. In the present study, the age range in both the L1 and L2 groups (L1-English: 21–45, L1-Swedish: 18–45) was considerably wider than in Lau et al.'s (2013) study (L1-English: 19–24), although our participants were also much younger than those in Federmeier et al.'s (2002, 2003, 2005, 2010) studies (>60 years). That the onset of the N400 effect across the low- and high-proportion blocks in our study was more similar to Lau et al.'s (2013) high-proportion block suggests that neither bilingualism nor age caused a delay in lexical access or prediction among our participants.

Finally, our results revealed a post-N400 negativity (500–800 ms) for unrelated relative to related targets across the two blocks, in both L1 and L2 speakers. This negativity was not modulated by relatedness proportion. Lau et al. (2013) found a similar post-N400 negativity for unrelated targets, but in their study, the negativity was larger in the high-proportion block. Nevertheless, Figures 2 and 3 show that, in our study, the negativity is larger in the high-proportion block from ~750 ms until the end of the epoch, in central-posterior electrodes, especially in the L1-English group. We ran an exploratory analysis in the 750–950 ms time window, in the posterior regions, which revealed that the Proportion × Relatedness interaction was marginal, $F(1, 87) = 3.873$, $p = .052$, $\eta^2 = .043$, and was not qualified by Group. Thus, if this effect is linked to the reinterpretation of the semantic relation between the prime and the target (e.g., Otten & van Berkum, 2009), our results suggest that both L1 speakers and L2 learners engaged in such reevaluation processes. This is, however, highly speculative given the exploratory nature of these analyses and the latency differences between the negativity in our study and Lau et al.'s (2013).

Having considered this evidence, we answer our first research question *Do advanced Swedish-speaking learners of English generate lexicosemantic predictions to the same extent as native English speakers?* in the positive. When a change in the linguistic environment made prediction a reliable strategy to pursue, the L2 learners engaged mechanisms that were quantitatively and qualitatively similar to those of the native speaker controls (in N400 effect size and topography). If Lau et al. (2013) are correct that those mechanisms involve updating a conceptual representation held in working memory ahead of the bottom-up input, our results suggest that the same mechanisms support lexicosemantic prediction in the L2. Overall, these results support the currently most prevalent position in the literature that L2 learners do not have a reduced ability to generate predictions in the L2 across the board and that L2 predictive processing can be qualitatively and quantitatively native-like when the conditions are favorable, that is, when learners are highly proficient in the L2, have similar biases to L1 speakers, and have enough time for prediction generation (Kaan, 2014; Kaan & Grüter, 2021).

Importantly, these results advance our understanding of lexicosemantic prediction in the L2 in a number of ways. First, our study is unique in that it investigated L2 learners' ability to predict via semantic associations. This allowed us to explore whether L2 learners predict to a lesser extent than native speakers due to their having less integrated semantic networks (Meara, 2009) and more lower quality lexical representations than native speakers, as hypothesized by Kaan (2014). In addition, we used a paradigm that circumvented some of the limitations of previous L2 studies on prediction using eye tracking and the VWP, such as previewing the images, which activates the nouns in the display before the presentation of the carrier sentence (e.g., Spivey & Marian, 1999; see also Chabal & Marian, 2015), or restricting the lexical search to no more than four words. In our study, participants' brain responses to the targets were recorded at the point when they first encountered them, and the lexical search was unrestricted by visual input. This shows that L2 learners can engage lexicosemantic prediction to native-like levels in setups where the lexical search cannot be delimited to a reduced set of already activated words (similar to the reading study by Foucart et al., 2014, who examined lexical prediction in sentence contexts). Moreover, our experimental paradigm allowed us to dissociate the effects of spreading activation from the effects of prediction on the target words in a way that previous L2 ERP studies manipulating word predictability in sentences could not (e.g., Foucart et al., 2014; Martin et al., 2013). For example, upon reading the preamble *If you put a flame in front of a gas pipe you will cause …* , most L1 and L2 speakers of English expect the noun *explosion* as opposed to *fire*, based on cloze probability ratings (Martin et al., 2013). This predicts a reduced N400 for the expected noun *explosion* relative to the unexpected but plausible noun *fire*, which is the pattern of results reported for both L1 and L2 speakers in the studies by Martin et al. (2013) and Foucart et al. (2014). However, it is unclear how the activation of the words in the preamble (e.g., *flame, gas*) would spread to the related words *explosion* and *fire*, and how this modulated the amplitude of the N400 component in previous studies. In our study, we used the same prime–target pairs in the low-proportion and high-proportion blocks across participants, and thus, the lexicosemantic relation between the primes and the targets was identical in the two blocks (Lau et al., 2013). This strengthens our claim that the larger N400 reduction for related versus unrelated targets in the predictive block for both L1 and L2 speakers reflects a mechanism different from passive spreading activation.

To our knowledge, our study is also the first to examine how a change in the input (i.e., a change in the proportion of related pairs in the second block) encourages L2 learners to invoke predictive mechanisms. The fact that our L2 learners predicted to native-like levels when the primes acquired predictive validity demonstrates that their online assessment of cue reliability in the predictive block was as efficient as that of native speakers. This suggests that, when tested on an equal footing with native speakers, that is, when the predictive cues have comparable reliability for L1 and L2 speakers, prediction has similar utility for L1 and L2 speakers. These findings are interesting given that L2 studies on semantic associations have shown that, even under no time pressure, learners' semantic networks tend to be more heterogeneous, unstable, and prone to error than those of native speakers, and our own free associations study showed substantial heterogeneity among learners' associations compared to L1-English speakers (Nelson et al., 2004; see Figure 1 [*Norming Study*]). Overall, these findings lend support to views of prediction which assume that comprehenders adjust their predictive behavior as a function of the reliability of the input (Kuperberg & Jaeger, 2016). In addition, the fact that our L2 learners adjusted their behavior similarly to the native speakers in the predictive block suggests that predictive processing can be as malleable in the L2 as in the L1 as a function of the input, and provides further support for proposals which posit that prediction is qualitatively similar in the L1 and the L2 (Kaan, 2014). At the same time, claims that prediction has reduced utility in the L2 are tightly connected to cross-linguistic differences (e.g., Grüter & Rohde, 2021; Kaan & Grüter, 2021; Schlenter, 2023), which we did not manipulate. In fact, we specifically excluded interlingual homographs from our selection of related pairs, to avoid L1 interference. In addition, we do not know whether the English associations that we probed are similar or different in the learners' L1 Swedish, since our free associations study only collected associations from the learners' L2. Thus, we do not know to what extent the learners' predictions were facilitated or impacted by their L1. For that reason, it remains an open question whether prediction has similar utility in the L2 in cases where the L1 and the L2 are in conflict or weigh predictive cues differently.

An interesting open question is whether our L2 learners would also engage predictive mechanisms with native-like efficiency in sentence comprehension (as in Foucart et al., 2014). In the present study, the predictive cues were individual words, so the learners need not integrate multiple sources of information to generate predictions. In contrast, prediction in sentential contexts involves more complex predictive cues (i.e., sentence preambles consisting of several words) and requires integrating linguistic and nonlinguistic information (see Delaney-Busch et al., 2019), which might consume more of the learners' processing resources and potentially cause them to rely on top-down processing to a lesser extent (e.g., Hopp, 2018, 2022; Martin et al., 2013). A recent study by Brothers et al. (2019) provides indirect evidence that L1 speakers adapt their predictive behavior similarly in sentential contexts and single-word contexts. Their study investigated whether the proportion of reliable predictive cues in sentences impacted lexicosemantic prediction among L1-English speakers. The participants listened to critical sentences intermixed with fillers, and cue reliability was manipulated by using two different speakers, one that always provided the expected word in high-constraint filler sentences, and one that always supplied plausible but unexpected words in the same high-constraint fillers. In the critical sentences, the authors found that the same expected words yielded a larger N400 reduction when the speaker was reliable, relative to the unreliable speaker. These results are similar to those reported by Lau et al. (2013) for single-word contexts, which we largely replicated here in both L1 and L2 speakers. Future studies should examine whether L2 learners, too, can adapt their predictive strategies in sentential contexts with a design similar to Brothers et al.'s (2019). We summarize this part of the discussion with the conclusion that, when immersed in a predictive environment, L2 learners can recruit predictive mechanisms and generate lexicosemantic predictions similarly and to the same extent as native speakers, at least in single-word contexts.

Effects of Proportion on the Animal Targets

Similar to Lau et al.'s (2013) study, animal probes yielded a larger P300 in the high- relative to the low-proportion block between ~500 and 800 ms, across both groups. This effect was broadly distributed, but largest in right posterior electrodes (Osterhout et al., 1996), as revealed by a three-way interaction between Proportion, Anterior–Posterior, and Hemisphere (but not Group). Under the assumption that both L1 and L2 speakers started using the primes predictively in the high-proportion block, animal words must have created a conflict between the planned "no-go" and the required "go" response, since the primes were always unrelated to the animal words (Lau et al., 2013). No such conflict should arise in the low-proportion block, if the participants were not using the primes predictively and planning their behavioral responses ahead of the bottom-up input. This explanation for the larger P300 in the predictive block based on the cost of failed predictions is also consistent with the behavioral data, which showed that both L1 and L2 speakers became slower and less accurate detecting animal words in the high- relative to the low-proportion block. Although the behavioral data could also be accounted for by the participants' becoming fatigued or disengaged over the course of the experiment, the larger P300 in the predictive block is harder to explain on the basis of such arguments.

Finally, the larger P300 in the predictive block also speaks against an interpretation that the larger N400 effect in the high-proportion block *only* reflects an increase in spreading activation. This is because, even if the P300 were modulated by semantic relatedness (Hill et al., 2006), activation of primes like *pickles/insurance/heroine* is highly unlikely to spread to targets like *horse/fox/cow* with or without a boost in spreading activation. Thus, although the strongest evidence that the L1 and L2 speakers in the present study engaged predictive mechanisms comes from the attenuated N400 for related targets in the high-proportion relative to the low-proportion block, we take the larger P300 for animal words in the high-proportion block as indirect evidence that both L1 and L2 speakers engaged different processing mechanisms in the second block.

Effects of Proportion on the Primes

In the L1-English group, our conservative point-by-point analysis ($p < .005$) showed that the primes elicited more positive responses in the high- versus the low-proportion block in the P2 time window (~161–186 ms) in frontocentral electrodes (Federmeier et al., 2005; Kaan & Carlisle, 2014). Lau et al. (2013) reported a similar effect, although in their study the effect was larger and emerged later (200–250 ms). In the L2 group, this positivity was only marginal in our predetermined time window (200–250 ms). However, visual inspection of the waveforms (see Figure 7) suggests that the positivity is largest between ~170 and 195 ms, similar to the native speakers. In fact, Figure 7 shows a sustained positivity from ~170 ms, which became significant between 582 and 600 ms. Our conservative analysis was probably too stringent for this effect to consolidate in the P2 time window in the L2 group. In fact, the positivity is significant at $p < .009$ across at least 15 data points between ~175 and 200 ms in a number of frontocentral electrodes.

Since the P2 has been related to complex visual search (e.g., Evans & Federmeier, 2009; Federmeier et al., 2005), this effect is consistent with the possibility that, in the predictive block, the L1-English controls visually scrutinized the primes to a larger extent for the purposes of prediction. The evidence that the L2 learners also recruited this strategy is weaker. We also point out that, in the L1-English group, the positivity might be smaller than in Lau et al.'s (2013) study due to the presence of a small negativity for high-proportion primes in the 350–400 ms time window, corresponding to the N400 time window. This larger N400, which did not emerge in Lau et al.'s (2013) study, might reflect deeper processing of the primes. Although this is consistent with the possibility that the primes spread more activation to the targets in the high-proportion block (Lau et al., 2013, p. 496), it seems to us that deeper processing of the predictive cues is also consistent with a predictive account. The fact that the L2 learners did not show this larger N400 for high-proportion primes while they still showed a larger N400 effect for related versus unrelated targets in the high-proportion block is overall more compatible with a predictive account.

Individual Differences

Under the assumption that predictive processing requires storing contextual information in working memory and rapidly updating it ahead of the bottom-up input, we hypothesized that both working memory and processing speed might modulate the size of the N400 effect for relatedness in the predictive block, especially among L2 learners, who have been argued to have limited processing capacity in the L2 (e.g., Hopp, 2010; Martin et al., 2013; McDonald, 2006). However, the results of a linear mixed-effects model analysis revealed that none of the factors assessed, that is, working memory, processing speed, general intelligence, and native status, influenced the magnitude of the N400 effect. Although it is possible that none of these constructs modulates the type of predictive processing that we investigated here, another likely explanation for the lack of significant relationships is that N400 effect magnitude in our study was a poor measure of individual differences. Although our study did capture individual differences in N400 effect size, as indicated by the fact that modeling them improved model fit, those individual differences showed low within-subject consistency, suggesting that they were not reliable. Since low reliability has detrimental consequences for statistical power, this might have made it difficult for relationships with the cognitive factors that we tested to emerge (e.g., Hedge et al., 2018; see also Cunnings & Fujita, 2021; Staub, 2021). A similar issue was reported by Staub (2021), who assessed the reliability of individual differences in word predictability on several eye-tracking measures among L1-English speakers, and found low reliability for all measures.

The question arises why the individual differences in prediction-related N400 effect size showed low reliability in our study. In their influential article, Hedge et al. (2018) pointed out that experimental effects that are robust and replicable at the group level tend to be unsuitable as individual differences measures (i.e., they show low split-half reliability). The reason for this "reliability paradox," as the authors named it, is that the lack of consistent between-subject variation is precisely what makes an experimental effect robust and replicable. In the present study, the N400 effect in the high-proportion block replicated findings from comparable samples (e.g., Holcomb, 1988; Lau et al., 2013). In addition, the effect was large, as indicated by a partial η^2 value of .34, which is substantially above the .14 threshold to consider an effect as large. This might explain, at least to some extent, why the effect did not show consistent individual differences. Although this makes our examination of individual differences inconclusive, we highlight that this is not a flaw of our experimental design. That the prediction-related N400 effect in our study was large and robust probably results from our using strong and reliable lexicosemantic associations for both L1-English and L1-Swedish speakers and, in the case of the L2 group, recruiting the most proficient learners that we could find.

Thus, the answer to our second research question *Do individual differences in working memory and processing speed account for variability in predictive processing in native and nonnative speakers?* is inconclusive. We believe that both factors probably modulate lexicosemantic predictions, although other paradigms and metrics might be better suited to address this question. We also echo previous researchers' recommendations to check and report the reliability of measures of individual differences in psycholinguistic effects (Cunnings & Fujita, 2021; James et al., 2018; Staub, 2021), both in cases where significant relations do emerge and inferences are drawn based on those relations, and in cases where significant relations are absent. For example, in their recent review, Kaan and Grüter (2021) pointed out that most studies that set out to investigate the contribution of proficiency to predictive processing found no significant correlations (see also Grüter et al., 2021; Grüter & Rohde, 2021; Kim & Grüter, 2021; Schlenter, 2023). Although the authors are cautious about the lack of proficiency effects, they express skepticism about the relation between global L2 proficiency and native-like predictive processing. We agree with Kaan and Grüter (2021) that high proficiency does not automatically translate into L1-like prediction, but we point out that none of those studies assessed (or at least reported) the reliability of individual differences in the relevant prediction effects. In light of our own findings, the lack of correlations between proficiency and prediction should be interpreted with even more caution.

Conclusion

The present study examined lexicosemantic prediction in prime–target contexts among native speakers of English and Swedish-speaking learners of English at a high level of L2 proficiency. The results show that, after both groups detected an increase in the reliability of the primes as predictive cues, both groups recruited predictive mechanisms that were qualitatively and quantitatively similar. These results inform our understanding of lexicosemantic prediction in a number of ways. First, they show that comprehenders adjust their predictive behavior as a function of their assessment of the reliability of the input, in line with proposals which assume that prediction has a utility function (Kaan & Grüter, 2021; Kuperberg & Jaeger, 2016). With respect to L2 processing, these results indicate that L2 learners do not have a reduced capacity to generate lexicosemantic predictions and that they are able to track the reliability of predictive cues as efficiently as native speakers, at least in single-word contexts. Overall, our findings support Kaan's (2014) proposal that prediction can in principle be quantitatively and qualitatively similar in the L1 and the L2. They also suggest that, at least when the predictive cues have similar reliability for L1 and L2 speakers, prediction has similar utility in the L1 and the L2.

Related to the latter point, the present study highlights the importance of using stimuli that are matched on predictive strength (e.g., semantic association strength here) for L1 and L2 speakers, especially when probing the possibilities and limitations of L2 predictive processing. To our knowledge, our study and the study by Foucart et al. (2014), both of which used materials with similar predictive strength for L1 and L2 speakers, are also the only ERP studies to have found that L2 learners generate lexicosemantic predictions to the same extent as native speakers. Thus, L2 learners are more likely to show native-like predictive processing if they are tested on an equal footing with native speakers, that is, if the predictive cues have comparable strength and reliability for both populations (Kaan, 2014; see also Schlenter, 2023).

References

Alemán Bañón, J. (2024, June 25). *Lexicosemantic prediction in native speakers of English and Swedish-speaking learners of English*. https://osf.io/r9t3q/

Alemán Bañón, J., & Martin, C. (2021). The role of crosslinguistic differences in second language anticipatory processing: An event-related potentials study. *Neuropsychologia*, *155*, Article 107797. https://doi.org/10.1016/j.neuropsychologia.2021.107797

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B, Statistical Methodology*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bialystok, E., Craik, F., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 859–873. https://doi.org/10.1037/0278-7393.34.4.859

Borg, R. (2021). *Facilitative online processing of gender in Swedish as a second language* [Master's thesis, Stockholm University]. DiVA. https://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-199824

Brothers, T., Dave, S., Hoversten, L. J., Traxler, M. J., & Swaab, T. Y. (2019). Flexible predictions during listening comprehension: Speaker reliability affects anticipatory processes. *Neuropsychologia*, *135*, Article 107225. https://doi.org/10.1016/j.neuropsychologia.2019.107225

Brown, C. M., Hagoort, P., & Chwilla, D. J. (2000). An event-related brain potential analysis of visual word priming effects. *Brain and Language*, *72*(2), 158–190. https://doi.org/10.1006/brln.1999.2284

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Chabal, S., & Marian, V. (2015). Speakers of different languages process the visual world differently. *Journal of Experimental Psychology: General*, *144*(3), 539–550. https://doi.org/10.1037/xge0000075

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. https://doi.org/10.1037/a0015901

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272. https://doi.org/10.1037/0033-295X.113.2.234

Chun, E., & Kaan, E. (2019). L2 prediction during complex sentence processing. *Journal of Cultural Cognitive Science*, *3*(2), 203–216. https://doi.org/10.1007/s41809-019-00038-0

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786. https://doi.org/10.3758/BF03196772

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, *13*(1), 21–58. https://doi.org/10.1080/016909698386582

Covey, L., Fiorentino, R., & Gabriele, A. (2024). Island sensitivity in L2 learners: Evidence from acceptability judgments and event-related potentials. *Second Language Research*, *40*(1), 19–50. https://doi.org/10.1177/02676583221116039

Cunnings, I., & Fujita, H. (2021). Quantifying individual differences in native and nonnative sentence processing. *Applied Psycholinguistics*, *42*(3), 579–599. https://doi.org/10.1017/S0142716420000648

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

Davies, M. (2008). *The corpus of contemporary American English (COCA)*. https://www.english-corpora.org/coca/

Delaney-Busch, N., Morgan, E., Lau, E., & Kuperberg, G. R. (2019). Neural evidence for Bayesian trial-by-trial adaptation on the N400 during semantic priming. *Cognition*, *187*, 10–20. https://doi.org/10.1016/j.cognition.2019.01.001

DeLong, K. A., Troyer, M., & Kutas, M. (2014). Pre-processing in sentence comprehension: Sensitivity to likely upcoming meaning and structure. *Language and Linguistics Compass*, *8*(12), 631–645. https://doi.org/10.1111/lnc3.12093

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

Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2017). Predicting upcoming information in native-language and non-native-language auditory word recognition. *Bilingualism: Language and Cognition*, *20*(5), 917–930. https://doi.org/10.1017/S1366728916000547

Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2019). Prediction and integration of semantics during L2 and L1 listening. *Language, Cognition and Neuroscience*, *34*(7), 881–900. https://doi.org/10.1080/23273798.2019.1591469

Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition*, *35*(2), 353–387. https://doi.org/10.1017/S0272263112000915

Earles, J. L., & Salthouse, T. A. (1995). Interrelations of age, health, and speed. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences*, *50B*(1), P33–P41. https://doi.org/10.1093/geronb/50B.1.P33

Evans, K. M., & Federmeier, K. D. (2009). Left and right memory revisited: Electrophysiological investigations of hemispheric asymmetries at retrieval. *Neuropsychologia*, *47*(2), 303–313. https://doi.org/10.1016/j.neuropsychologia.2008.08.027

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. https://doi.org/10.1006/jmla.1999.2660

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. https://doi.org/10.1016/j.bandl.2010.07.006

Federmeier, K. D., Mai, H., & Kutas, M. (2005). Both sides get the point: Hemispheric sensitivities to sentential constraint. *Memory & Cognition*, *33*(5), 871–886. https://doi.org/10.3758/BF03193082

Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*(2), 133–146. https://doi.org/10.1111/1469-8986.3920133

Federmeier, K. D., Van Petten, C., Schwartz, T. J., & Kutas, M. (2003). Sounds, words, sentences: Age-related changes across levels of language processing. *Psychology and Aging*, *18*(4), 858–872. https://doi.org/10.1037/0882-7974.18.4.858

Fitzpatrick, T., & Izura, C. (2011). Word association in L1 and L2. *Studies in Second Language Acquisition*, *33*(3), 373–398. https://doi.org/10.1017/S0272263111000027

Folstein, J. R., & Van Petten, C. (2011). After the P3: Late executive processes in stimulus categorization. *Psychophysiology*, *48*(6), 825–841. https://doi.org/10.1111/j.1469-8986.2010.01146.x

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. https://doi.org/10.1037/a0036756

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*(2), 78–84. https://doi.org/10.1016/S1364-6613(00)01839-8

Gambi, C. (2021). The role of prediction in second language vocabulary learning. In E. Kaan & T. Grüter (Eds.), *Prediction in second language processing and learning* (1st ed., pp. 188–206). John Benjamins Publishing. https://doi.org/10.1075/bpa.12.09gam

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition*, *33*(7), 1220–1234. https://doi.org/10.3758/BF03193224

Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology: General*, *140*(2), 186–209. https://doi.org/10.1037/a0022256

Grossi, G. (2006). Relatedness proportion effects on masked associative priming: An ERP study. *Psychophysiology*, *43*(1), 21–30. https://doi.org/10.1111/j.1469-8986.2006.00383.x

Grüter, T., Lau, E., & Ling, W. (2020). How classifiers facilitate predictive processing in L1 and L2 Chinese: The role of semantic and grammatical cues. *Language, Cognition and Neuroscience*, *35*(2), 221–234. https://doi.org/10.1080/23273798.2019.1648840

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. https://doi.org/10.1177/0267658312437990

Grüter, T., & Rohde, H. (2021). Limits on expectation-based processing: Use of grammatical aspect for co-reference in L2. *Applied Psycholinguistics*, *42*(1), 51–75. https://doi.org/10.1017/S0142716420000582

Grüter, T., Rohde, H., & Schafer, A. J. (2017). Coreference and discourse coherence in L2: The roles of grammatical aspect and referential form. *Linguistic Approaches to Bilingualism*, *7*(2), 199–229. https://doi.org/10.1075/lab.15011.gru

Grüter, T., Zhu, Y. A., & Jackson, C. N. (2021). Forcing prediction increases priming and adaption in second language production. In E. Kaan & T. Grüter (Eds.), *Prediction in second language processing and learning* (1st ed., pp. 208–232). John Benjamins Publishing. https://doi.org/10.1075/bpa.12.10gru

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*, 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

Hill, H., Ott, F., Herbert, C., & Weisbrod, M. (2006). Response execution in lexical decision tasks obscures sex-specific lateralization effects in language processing: Evidence from event-related potential measures during word reading. *Cerebral Cortex*, *16*(7), 978–989. https://doi.org/10.1093/cercor/bhj040

Holcomb, P. J. (1988). Automatic and attentional processing: An event-related brain potential analysis of semantic priming. *Brain and Language*, *35*(1), 66–85. https://doi.org/10.1016/0093-934X(88)90101-0

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes*, *5*(4), 281–312. https://doi.org/10.1080/01690969008407065

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, *120*(4), 901–931. https://doi.org/10.1016/j.lingua.2009.06.004

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*(1), 33–56. https://doi.org/10.1177/0267658312461803

Hopp, H. (2015). Semantics and morphosyntax in predictive L2 sentence processing. *International Review of Applied Linguistics in Language Teaching*, *53*(3), 277–306. https://doi.org/10.1515/iral-2015-0014

Hopp, H. (2018). The bilingual mental lexicon in L2 sentence processing. *Second Language*, *17*, 5–17. https://www.jstage.jst.go.jp/article/secondlanguage/17/0/17_5/_pdf/-char/ja

Hopp, H. (2022). Second language sentence processing. *Annual Review of Linguistics*, *8*(1), 235–256. https://doi.org/10.1146/annurev-linguistics-030821-054113

Hopp, H., & Lemmerth, N. (2018). Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition*, *40*(1), 171–199. https://doi.org/10.1017/S0272263116000437

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–135. https://doi.org/10.1016/j.brainres.2015.02.014

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. https://doi.org/10.1080/23273798.2015.1047459

Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience*, *31*(1), 19–31. https://doi.org/10.1080/23273798.2015.1072223

Ito, A., Corley, M., & Pickering, M. J. (2018). A cognitive load delays predictive eye movements similarly during L1 and L2 comprehension. *Bilingualism: Language and Cognition*, *21*(2), 251–264. https://doi.org/10.1017/S1366728917000050

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. https://doi.org/10.1016/j.jml.2015.10.007

Ito, A., Martin, A. E., & Nieuwland, M. S. (2017a). How robust are prediction effects in language comprehension? Failure to replicate article-elicited N400 effects. *Language, Cognition and Neuroscience*, *32*(8), 954–965. https://doi.org/10.1080/23273798.2016.1242761

Ito, A., Martin, A. E., & Nieuwland, M. S. (2017b). On predicting form and meaning in a second language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 635–652. https://doi.org/10.1037/xlm0000315

Ito, A., & Pickering, M. (2021). Automaticity and prediction in non-native language comprehension. In E. Kaan & T. Grüter (Eds.), *Prediction in second language processing and learning* (1st ed., pp. 26–46). John Benjamins Publishing. https://doi.org/10.1075/bpa.12.02ito

Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language*, *98*, 1–11. https://doi.org/10.1016/j.jml.2017.09.002

James, A. N., Fraundorf, S. H., Lee, E. K., & Watson, D. G. (2018). Individual differences in syntactic processing: Is there evidence for reader-text interactions? *Journal of Memory and Language*, *102*, 155–181. https://doi.org/10.1016/j.jml.2018.05.006

Kaan, E. (2014). Predictive sentence processing in L1 and L2. What's different? *Linguistic Approaches to Bilingualism*, *4*(2), 257–282. https://doi.org/10.1075/lab.4.2.05kaa

Kaan, E., & Carlisle, E. (2014). ERP indices of stimulus prediction in letter sequences. *Brain Sciences*, *4*(4), 509–531. https://doi.org/10.3390/brainsci4040509

Kaan, E., Dallas, A., & Wijnen, F. (2010). Syntactic predictions in second-language sentence processing. In C. J. Zwart, M. D. Vries, & J. Koster (Eds.), *Structure preserved: Studies in syntax for Jan Koster* (pp. 207–213). John Benjamins. https://doi.org/10.1075/la.164.23kaa

Kaan, E., & Grüter, T. (2021). Prediction in second language processing and learning: Advances and directions. In E. Kaan & T. Grüter (Eds.), *Prediction in second language processing and learning* (1st ed., pp. 2–24). John Benjamins Publishing. https://doi.org/10.1075/bpa.12.01kaa

Kaan, E., Kirkham, J., & Wijnen, F. (2016). Prediction and integration in native and second-language processing of elliptical structures. *Bilingualism: Language and Cognition*, *19*(1), 1–18. https://doi.org/10.1017/S1366728914000844

Kim, H., & Grüter, T. (2021). Predictive processing of implicit causality in a second language. *Studies in Second Language Acquisition*, *43*(1), 133–154. https://doi.org/10.1017/S0272263120000443

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 23–49. https://doi.org/10.1016/j.brainres.2006.12.063

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163. https://doi.org/10.1038/307161a0

Larson, M. J., Kaufman, D. A. S., & Perlstein, W. M. (2009). Neural time course of conflict adaptation effects on the Stroop task. *Neuropsychologia*, *47*(3), 663–670. https://doi.org/10.1016/j.neuropsychologia.2008.11.013

Lau, E. F., 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. https://doi.org/10.1162/jocn_a_00328

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920–933. https://doi.org/10.1038/nrn2532

Leckey, M., & Federmeier, K. D. (2020). The P3b and P600(s): Positive contributions to language comprehension. *Psychophysiology*, *57*(7), Article e13351. https://doi.org/10.1111/psyp.13351

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

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. https://doi.org/10.1016/j.jml.2010.07.003

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. https://doi.org/10.1016/j.jecp.2014.05.004

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. https://doi.org/10.1016/j.jml.2013.08.001

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, *55*(3), 381–401. https://doi.org/10.1016/j.jml.2006.06.006

Meara, P. (2009). *Connected words: Word associations and second language vocabulary acquisition*. John Benjamins Publishing. https://doi.org/10.1075/lllt.24

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), 1–10. https://doi.org/10.16910/jemr.5.1.3

Mitsugi, S. (2017). Syntactic prediction in L2 comprehension: Evidence from Japanese adverbials. In M. LaMendola & J. Scott (Eds.), *Proceedings of the 41st annual Boston University conference on language development* (pp. 509–521). Cascadilla Press.

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. https://doi.org/10.1017/S1366728914000881

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms.

Behavior Research Methods, Instruments, & Computers, *36*(3), 402–407. https://doi.org/10.3758/BF03195588

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., Von Grebmer Zu Wolfsthurn, S., Bartolozzi, F., Kogan, V., Ito, A., Mézière, D., Barr, D. J., Rousselet, G. A., Ferguson, H. J., Busch-Moreno, S., Fu, X., Tuomainen, J., Kulakova, E., Husband, E. M., … Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife*, *7*, Article e33468. https://doi.org/10.7554/eLife.33468

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Osterhout, L., McKinnon, R., Bersick, M., & Corey, V. (1996). On the language specificity of the brain response to syntactic anomalies: Is the syntactic positive shift a member of the p300 family? *Journal of Cognitive Neuroscience*, *8*(6), 507–526. https://doi.org/10.1162/jocn.1996.8.6.507

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. https://doi.org/10.1016/j.brainres.2009.07.042

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Peirce, J. W., & MacAskill, M. (2018). *Building experiments in PsychoPy* (1st ed.). SAGE Publications.

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002–1044. https://doi.org/10.1037/bul0000158

Raven, J., & Raven, J. (2003). Raven progressive matrices. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment* (pp. 223–237). Springer. https://doi.org/10.1007/978-1-4615-0153-4_11

Ryskin, R., & Nieuwland, M. S. (2023). Prediction during language comprehension: What is next? *Trends in Cognitive Sciences*, *27*(11), 1032–1052. https://doi.org/10.1016/j.tics.2023.08.003

Şafak, D. F., & Hopp, H. (2022). Verb bias and semantic persistence effects in L2 ambiguity resolution. *Second Language Research*, *38*(4), 705–736. https://doi.org/10.1177/0267658321997904

Schlenter, J. (2023). Prediction in bilingual sentence processing: How prediction differs in a later learned language from a first language. *Bilingualism: Language and Cognition*, *26*(2), 253–267. https://doi.org/10.1017/S1366728922000736

Schlenter, J., & Felser, C. (2021). Second language prediction ability across different linguistic domains. In E. Kaan & T. Grüter (Eds.), *Prediction in second language processing and learning* (1st ed., pp. 48–68). John Benjamins Publishing. https://doi.org/10.1075/bpa.12.03sch

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. https://doi.org/10.1111/1467-9280.00151

Staub, A. (2021). How reliable are individual differences in eye movements in reading? *Journal of Memory and Language*, *116*, Article 104190. https://doi.org/10.1016/j.jml.2020.104190

Steinhauer, K., Royle, P., Drury, J. E., & Fromont, L. A. (2017). The priming of priming: Evidence that the N400 reflects context-dependent post-retrieval word integration in working memory. *Neuroscience Letters*, *651*, 192–197. https://doi.org/10.1016/j.neulet.2017.05.007

Szűcs, D., & Soltész, F. (2010). Event-related brain potentials to violations of arithmetic syntax represented by place value structure. *Biological Psychology*, *84*(2), 354–367. https://doi.org/10.1016/j.biopsycho.2010.04.002

Tokowicz, N. (2015). *Lexical processing and second language acquisition*. Routledge.

Van Bergen, G., & Flecken, M. (2017). Putting things in new places: Linguistic experience modulates the predictive power of placement verb semantics. *Journal of Memory and Language*, *92*, 26–42. https://doi.org/10.1016/j.jml.2016.05.003

van Hell, J. G., & de Groot, A. M. B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition*, *1*(3), 193–211. https://doi.org/10.1017/S1366728998000352

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015

Wechsler Adult Intelligence Scale. (4th ed.). (2004). Pearson.

West, R. (2003). Neural correlates of cognitive control and conflict detection in the Stroop and digit-location tasks. *Neuropsychologia*, *41*(8), 1122–1135. https://doi.org/10.1016/S0028-3932(02)00297-X

Received June 27, 2023
Revision received June 26, 2024
Accepted September 3, 2024 ■