



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/177530/>

Version: Accepted Version

Article:

Bovolenta, Giulia and Marsden, Emma (2022) Prediction and error-based learning in L2 processing and acquisition: a conceptual review. *Studies in Second Language Acquisition*. pp. 1384-1409. ISSN: 1470-1545

<https://doi.org/10.1017/S0272263121000723>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Abstract

There is currently much interest in the role of prediction in language processing, both in L1 and L2. For language acquisition researchers, this has prompted debate on the role that predictive processing may play in both L1 and L2 language learning, if any. In this conceptual review, we explore the role of prediction and prediction error as a potential learning aid. We examine different proposed prediction mechanisms and the empirical evidence for them, alongside the factors constraining prediction for both L1 and L2 speakers. We then review the evidence on the role of prediction in learning languages. We report computational modelling which underpins a number of proposals on the role of prediction in L1 and L2 learning, then lay out the empirical evidence supporting the predictions made by modelling, from research into priming and adaptation. Finally, we point out the limitations of these mechanisms in both L1 and L2 speakers.

PREDICTION AND ERROR-BASED LEARNING IN L2 PROCESSING AND ACQUISITION: A CONCEPTUAL REVIEW

There is currently great interest in whether the same mechanisms that underpin language processing also drive language learning. This interest stems, in part, from a desire to account for language learning in the absence of any kind of pre-determined grammatical hard-wiring. That is, accounting for both learning phenomena and processing phenomena within the same model would achieve a desirable theoretical parsimony (O’Grady, 2005). In addition, investigating the nature and role of processing in second language (L2) acquisition potentially offers a way to shed light on the ways in which L2 acquisition may differ from first language (L1) acquisition¹.

In the field of second language acquisition (SLA), it has been suggested that prediction has a role in language learning and, specifically, for acquiring complex contingencies, thought to be among the hardest phenomena to learn. For example, prediction during sentence comprehension may be useful for L2 learners by allowing for *hypothesis testing*, which could help them retreat from overgeneralisation when their predictions (their hypotheses) are disconfirmed (Phillips & Ehrenhofer, 2015). The idea that prediction may serve as a learning tool is supported by computational modelling showing that data from L1 acquisition and processing can be reproduced by recurrent neural networks that use prediction and error-based learning (Chang et al., 2006). However, there is evidence that L2 speakers often lag behind L1 speakers in their ability to predict upcoming input, as shown by data from both eye-tracking and EEG (electroencephalography) studies. Indeed, it has been suggested that L2 speakers, in particular, may be affected by a reduced ability to generate expectations (Grüter et al., 2014,

¹ We limit our review to L2, though it has relevance to additional languages (Lx) too. Lx involves other issues that would merit further review and as such are beyond the scope of our review.

2017). This possibility sparked a concern that limitations in L2 learners' ability to predict, relative to L1 speakers, may prevent them from using a prediction-based learning mechanism (e.g., Kaan et al., 2019; Phillips & Ehrenhofer, 2015)². Therefore, understanding exactly what is meant by prediction, how L2 learners may differ from native speakers in their ability to predict, and what can actually be learned through prediction, will be necessary to address these concerns. The aim of this article is to provide an overview of research into prediction in L1 and L2 processing and learning, in order to provide a frame of reference for those interested in the role prediction may play in L2 acquisition in particular.

Our review is structured as follows: In the first section, we define prediction and describe different prediction mechanisms which have been identified in the literature, showing that it can be conceived of as a continuum going from simpler to more complex instances of prediction. We present the empirical evidence for the different types of prediction in L1 and L2 speakers, and highlight the factors which can constrain prediction in both groups. In the second section, we introduce the theoretical debate on the role of prediction in language learning – both L1 and L2. We introduce computational models of L1 processing which show that language acquisition and priming phenomena can be explained by error-based learning. We then review the available empirical evidence for a learning mechanism based on prediction error, from studies on priming and adaptation in both L1 and L2, and conclude by highlighting potential limitations of this mechanism.

² For our purposes, L2 learning will be defined here as a change in a speaker's L2 representations resulting from exposure to the language, in any form.

Evidence for prediction in L1 and L2 speakers

Defining ‘prediction’

We need to make here a preliminary distinction between different conceptions of prediction: prediction as the formulation of expectations during sentence comprehension (as in ‘pre-processing’, DeLong et al., 2014), which is the focus of this review, and a more general sense of prediction as inference generation (for instance, using contextual cues to assign referents to ambiguous pronouns). Grüter et al. (2017) note that prediction in the narrower sense of ‘pre-processing’ is still not often investigated by SLA researchers. They add that:

The term ‘prediction’ has been used in the SLA literature, primarily in the context of L2 reading, to refer to inference generation, or guessing (e.g., in fill-the-gap tasks), more generally (e.g., McLaughlin, 1987). This usage does not specify the temporal aspect of this process, i.e., when such inference generation takes place during the incremental construction of meaning as we read/listen. As such, [this usage of the term prediction] is compatible with both (retroactive) information integration and prediction in terms of (proactive) linguistic pre-processing. (Grüter et al., 2017, footnote to p. 224).

One example of this generic, temporally non-specific usage of ‘prediction’ in SLA is the literature on *statistical preemption* (Ambridge & Brandt, 2013; Boyd & Goldberg, 2011; Foraker et al., 2009; Robenalt & Goldberg, 2016), a proposed learning mechanism which is driven by associative learning: Every time an expected outcome is not encountered after a given cue, the strength of its association with that cue diminishes (Rescorla & Wagner, 1972). This line of research has to date examined inference generation through offline tasks such as acceptability judgments (e.g., Robenalt & Goldberg, 2016) to determine to what extent learners take into account potential alternatives to structures they encounter. While it cannot be ruled out that prediction *during* processing also plays a role in determining acceptability, these kinds

of acceptability tasks also capture the result of processes (such as retroactive information integration) which are not part of prediction in the narrower sense (i.e., linguistic pre-processing), and thus this line of research is not part of the scope of the current review.

Another field of research on prediction which is temporally non-specific investigates the effect of expectation violation on the formation of new declarative memories. This line of research does not originate in SLA research, but it is relevant to language learning, as well as learning, more generally: Findings have shown that novel associations which violate established patterns are remembered better than those which do not (Brod et al., 2018; De Loof et al., 2018; Greve et al., 2017; Greve et al., 2019) and that generating incorrect guesses followed by corrective feedback can, under some circumstances, lead to better learning than simply being exposed to the correct answer (Potts, Davies, & Shanks, 2019; Potts & Shanks, 2014). While these effects have been observed for a variety of stimuli, such as conceptual knowledge (Brod et al., 2018) and arbitrary picture-word mappings (Greve et al., 2017), it is its role in the acquisition of vocabulary which may be most relevant to SLA. Expectation violation has been shown to aid the acquisition of L1 vocabulary in young children (Stahl & Feigenson, 2017), as well as Dutch-Swahili translation word pairs in adult Dutch L1 speakers (De Loof et al., 2018; see also Gambi (cited in Kaan & Grüter, forthcoming) for more recent work on L2 vocabulary learning). While it is of relevance to SLA, however, this particular conceptualisation of prediction is, too, outside the scope of our review, which is on prediction as linguistic ‘pre-processing’, or the *incremental formulation of expectations during sentence comprehension*. We will now turn to prediction as linguistic ‘pre-processing’ and the ways in which it can be conceptualised.

In the literature on prediction during language comprehension, there is a great amount of variation in terms of approaches and terminology used, with different authors focusing on

different aspects of the phenomenon (see Kuperberg & Jaeger, 2016 for a review). Pickering and colleagues (Pickering & Gambi, 2018; Pickering & Garrod, 2013), distinguish between two types of prediction, *prediction-by-association* and *prediction-by-production*. The first mechanism, *prediction-by-association*, is driven by basic associative mechanisms like spreading activation, and may constitute the stage prior to *prediction-by-production*. A feature specific to this account is that in the more complex route (*prediction-by-production*) preactivation involves forward speech planning: comprehenders use the language production system to formulate predictions about upcoming input. The *prediction-by-production* route is considered to be very accurate and, therefore, to aid processing, but it is thought to be an optional mechanism; for example, it is not always available, especially in L2, or populations with cognitive limitations (Pickering & Gambi, 2018). By contrast, *prediction-by-association* is less precise and less effective than *prediction-by-production*; however, it is an integral part of comprehension and, being automatic, it does not take up cognitive resources, which is why it should remain unimpaired even in comprehenders with limited resources (Pickering & Gambi, 2018, p. 1030).

Similarly to Pickering and colleagues, other authors also distinguish between two broad types of ‘pre-processing’ prediction: a simple, automatic kind, generally limited to the semantic domain, and a more complex, resource intensive type, involving prediction of specific linguistic features. Kuperberg & Jaeger (2016) contrast a basic sense of prediction, as expectations based on discourse context, with ‘predictive activation’ of low-level (e.g., phonological, morphological) features. In this latter case, comprehenders can ‘predictively pre-activate’ low-level representations (e.g., phonological form) based on high-level inferences, before encountering them in the input, rather than just making a high-level event hypothesis, as happens in the more basic, simple case of prediction. Another potential dual-route account of prediction is also offered by Huettig (2015), modelled on Kahneman’s (2011) dual-system

model of reasoning: a ‘dumb’ route (System 1), based on simple associative mechanisms, which is contrasted with a “smart” route (System 2), linked to more effortful active reasoning.

The conceptualisations of prediction we have just seen could all be seen as, essentially, dichotomous distinctions; however, the empirical evidence (which will be reviewed below) suggests a more graded process, which can vary in complexity and specificity depending on a variety of factors including context, language proficiency, and the nature of the task. In light of this complexity, Huettig (2015) proposes a multiple-mechanisms account of prediction, called PACS (production-, association-, combinatorial-, simulation-based prediction). According to this account, prediction can be driven by diverse mechanisms. One is basic association, which is often for semantic information, but may also involve other types of representation (e.g., phonological); another is production, where prediction happens through covert speech production. There is also a combinatorial route – where meaning is built by drawing on multiple linguistic constraints, and an event simulation route, where mental imagery may be used to pre-activate linguistic representations. Crucially, these four mechanisms interact with each other: For instance, basic association may provide input which then feeds into the combinatorial route (Huettig, 2015).

To further illustrate the graded nature of prediction processes and its context-, task-, and individual-specificity, we now review evidence for prediction in L1 and L2 speakers in growing order of complexity: from basic sensitivity towards word predictability based on semantic context to preactivation of specific morphological and phonological features, but all aligning with our broad working definition of prediction for the purposes of this review as *incremental formulation of expectations during sentence comprehension*. In the subsequent section, we then move on to examine the factors which can constrain the extent of predictive processing in both L1 and L2 speakers.

Types of prediction: from basic expectations to preactivation of specific features

A simple type of prediction: sensitivity to word predictability

The predictability of a word from context is known to affect the way it is processed during comprehension. Words that are predictable from their semantic context are easier to process: L1 speakers spend less time fixating on them during reading (Balota et al., 1985; Demberg & Keller, 2008; Ehrlich & Rayner, 1981; McDonald & Shillcock, 2003), and are quicker to react to them in behavioural tasks such as lexical decisions (Schwanenflugel & LaCount, 1988; Schwanenflugel & White, 1991; Stanovich & West, 1983) and naming tasks (Forster, 1981; Stanovich & West, 1981, 1983; Traxler & Foss, 2000). Using EEG, words that are highly predictable from context elicit a reduced N400³ relative to unexpected words (Kutas & Hillyard, 1980, 1984), a finding which has been widely replicated (see Kutas & Federmeier, 2011 for a review). The size of the N400 elicited by an unexpected sentence-final word is inversely proportional to the cloze probability of the word itself, i.e., how likely the word is to occur at the end of that sentence (DeLong, Groppe, Urbach, & Kutas, 2012; DeLong, Urbach, & Kutas, 2005; Luke & Christianson, 2016), although it is not affected by the number of potential alternative completions (Kuperberg et al., 2020; Kutas & Hillyard, 1984). The N400 reduction to predictable words is found in L2 speakers, too (Martin et al., 2013). Finally, both L1 and L2 speakers can also exhibit sensitivity to word predictability based on structural cues, not just semantic ones, evidenced by data from EEG (Kaan et al., 2016) and self-paced reading (Leal et al., 2017).

³ The N400 is a negative shift in potential, detected in the centroparietal region of the scalp, which peaks approximately 400ms after encountering a new stimulus. In language studies, the stimulus will normally be a word (presented visually or aurally), but this component is also observed in response to other kinds of stimuli, such as pictures.

In the L1 processing literature, there was initial resistance to accept evidence of sensitivity to word predictability as evidence of ‘prediction’, on the grounds that it could also simply be interpreted as an effect of ‘integration’, that is, the ease with which the word’s meaning could be accessed or combined with that of preceding words (see Kutas & Federmeier, 2011 and Van Petten & Luka, 2012, for reviews). Indeed, it is very difficult to distinguish between these two accounts (*prediction* and *integration*) experimentally. For example, to explain the observation that the N400 in response to predictable words is smaller than that to less predictable words, one could argue that it is because comprehenders were expecting to encounter the specific, highly predictable word (thus, this observation is usable as evidence of *prediction*). But it is also possible that comprehenders were not expecting anything in particular, and that upon hearing a highly predictable word, it was simply easier for them to process due to its closer semantic fit with the preceding context (thus, this observation could be usable as evidence of *integration*).

In the body of evidence we have seen so far, a clear-cut distinction between *prediction* and *integration* may not always be found; it is now generally accepted that the N400 indexes a cascade of processes that happen both before, during and after word recognition (Nieuwland et al., 2020). However, it is clear that basic sensitivity to word predictability, based on frequency information and often drawing on associative mechanisms, appears to be a robust feature of language processing, both in L1 and L2. Comprehenders use sentential context, whether highly constraining or not, to update their expectations about the likelihood of potential continuations, in a probabilistic fashion (i.e., where multiple possibilities have varying likelihoods). These expectations then affect processing of upcoming input, depending on how expected each was. However, in the next section we will see evidence for how, when context allows it, comprehenders can also make use of these expectations ahead of

encountering input in order to narrow down the range of possible continuations, thus constituting a more complex type of prediction.

Anticipating content: Integrating cues with context

In experimental settings, at least, it has been shown that comprehenders can combine their expectations with context to identify the most likely referent of an upcoming word from a limited set of candidates. Studies using the visual world eye-tracking paradigm have shown that comprehenders use cues such as verb selectional restrictions to form expectations for upcoming content as the sentence unfolds, and identify likely referents from a set of options based on how well they fit these expectations. For instance, when hearing a verb such as *eat* in “The boy will eat...”, L1 speakers already restrict the range of potential expected completions to items that can be the object of *eat*; if the visual scene only contains one item that fits that category (e.g., a cake), they will automatically look at the picture of the cake even before hearing the word *cake* (Altmann & Kamide, 1999). This shows a more active kind of anticipation, that goes beyond simple sensitivity to word likelihood: rather than responding to a word based on how likely it was, comprehenders used their expectations to narrow down the range of potential referents for an upcoming word, ahead of encountering the word itself, by picking out the most likely candidate from the ones available.

Again, these effects have been observed in L2 speakers, too, although not to the same extent as in L1 speakers. High-proficiency L2 English speakers behaved similarly to L1 speakers when the object of a sentence could be predicted based on the verb’s meaning (Ito, Corley, et al., 2018) or from the situational context more generally (Ito, Pickering, et al., 2018), giving anticipatory looks to suitable objects (from a constrained set) to the same extent as L1 speakers. However, there is also evidence that these effects are slower and weaker in L2

speakers (Dijkgraaf et al., 2019). They are also modulated by proficiency: Lower-skilled L2 speakers are more likely to fixate (give a prolonged gaze on) less relevant themes in a visual world paradigm (e.g., “cat” when listening to the sentence “The pirate will chase... (the ship)”) compared to higher-skilled bilinguals (Peters et al., 2018).

The same visual world paradigm has been used in different languages to investigate features such as gender marking (Lew-Williams & Fernald, 2007) and case marking (Kamide et al., 2003), showing that L1 speakers can also use morphological cues to select possible referents from the items in a visual scene. L2 learners have often failed to show the same ability to anticipate content, whether on the basis of gender marking (Grüter et al., 2012; Lew-Williams & Fernald, 2010), morphosyntactic (Andringa & Curcic, 2015; Hopp, 2015), or morphological information (Mitsugi & MacWhinney, 2016). However, there are also instances of L2 speakers performing similarly to L1 speakers in studies using the visual world eye-tracking paradigm with morphological cues (Dussias et al., 2013; Hopp & Lemmerth, 2018). High-proficiency L1 Russian-L2 German speakers, too, showed native speaker-like prediction using determiner gender marking, even though Russian does not have gender-marked prenominal articles (Hopp & Lemmerth, 2018). Dussias et al. (2013) used the paradigm employed by Lew-Williams & Fernald (2007) to investigate predictive processing of gender marking, extending it to L2 speakers. They showed that highly proficient L1 English (a -gender language) and L1 Italian (a +gender language) speakers of L2 Spanish could use gender agreement in prenominal determiners, as done by the L1 Spanish speakers in Lew-Williams & Fernald’s (2007) study, with all groups of participants giving anticipatory looks to appropriate objects in the visual scene. By contrast, low-proficiency L1 English speakers did not show nativelike prediction (however, it should be noted that the L1 Italian group, who had low proficiency, only showed anticipatory looking for feminine determiners).

The evidence we considered earlier shows that both L1 and L2 speakers are sensitive to word predictability, as evidenced by their processing of more or less predictable words. We have now seen that L1 speakers (and sometimes, highly proficient L2 speakers too) can also use their expectations for upcoming content, based on cues in the input (which may be semantic, such as verb selectional restrictions, or grammatical, such as morphological gender marking) to select suitable referents from those made available by context. However, even preferential looking to suitable targets in a visual world eye-tracking study does not necessarily imply preactivation of a specific lexical item or feature: The visual world paradigm provides the item to the participants, who identify it from a set of options as that which most closely matches their expectations. A desire to establish conclusive evidence of prediction in the strictest sense (rather than integration) has informed more complex experimental work using EEG, aimed at showing that preactivation of specific features is possible, in the appropriate circumstances.

Preactivation of specific features: Evidence from EEG

A series of EEG studies has examined prediction by manipulating the morphological and phonological dependencies between highly predictable words and prior elements in the sentence, such as adjectives and determiners (DeLong, 2009; DeLong et al., 2005; Otten & Van Berkum, 2008; Szewczyk & Schriefers, 2013; Van Berkum et al., 2005; Wicha et al., 2004). For example, DeLong and colleagues (DeLong, 2009; DeLong et al., 2005) investigated whether specific words were being predicted by their participants by manipulating the phonological alternation of the English singular indefinite article (*a/an*). Participants read sentences such as “The day was breezy so the boy went to fly...”, which is highly constraining for the completion (*a*) *kite*. At this point, encountering the *an* form of the determiner (potentially leading to a less expected noun e.g., *an airplane*) elicited a significantly larger

N400 compared to the form *a*, compatible with the more likely *kite*. This suggests that the expectation for *kite* was being used to preactivate a specific linguistic representation including its phonological form (the initial consonant), in turn generating an expectation for *a* instead of *an*. (However, see Nieuwland et al., 2017, 2018 for a failure to replicate this effect in L1 speakers.) The size of the N400 effect on the article was graded based on the cloze probability of the target noun (i.e., how likely subjects were to expect it as the next word, based on an offline sentence completion task done by native speakers), suggesting that participants were making probabilistic predictions of specific words. Similar results using EEG have been obtained by manipulating gender agreement between nouns and determiners in Spanish (Wicha et al., 2004) and Dutch (Van Berkum et al., 2005), as well as animacy marking agreement between nouns and adjectives in Polish (Szewczyk & Schriefers, 2013).

Compared to the simpler types of prediction seen previously, there seems to be a greater gap between L1 and L2 speakers when it comes to preactivating specific features. Martin et al. (2013) used the EEG paradigm from DeLong et al. (2005), which required participants to preactivate phonological forms (*a / an kite*), but they failed to replicate in L2 English speakers the effect observed by DeLong et al. for L1 speakers on the determiner. However, Martin et al. still found a basic effect of noun predictability on the noun itself: Replacing a highly predictable noun with a less predictable one elicited an increased N400, as it did in L1 speakers. This means Martin et al.'s participants did have probabilistic expectations about possible upcoming nouns, even though they were not building on these to predict the appropriate determiner, as did L1 speakers in DeLong et al.'s study. Furthermore, subsequent research suggests that L2 speakers can preactivate specific features in a manner similar to native speakers, at least if those features exist in their L1. Foucart, Martin, Moreno, & Costa (2014) exposed native Spanish speakers and two groups of L2 Romance bilinguals (French-Spanish late bilinguals, and Spanish-Catalan early bilinguals) to Spanish sentences with highly predictable final nouns,

manipulating the gender of the preceding determiner following Wicha et al.'s (2004) design. Unlike the English *a/an* alternation, which the L2 participants in the Martin et al. (2013) study did not have in their L1 (Spanish), gender agreement between determiners and nouns is a common feature of Romance languages, meaning that both the bilingual groups in Foucart et al. (2014) would be familiar with this feature from their L1. When morphological gender marking on the determiner was incongruent with the gender of the expected noun, all three groups – Spanish monolinguals, and the two bilingual groups - exhibited an increased N400 response, suggesting that L2 speakers were preactivating gender features in a way similar to L1 speakers. This kind of study arguably provides stronger evidence for the preactivation of specific forms than visual world eye-tracking studies because participants were not provided with the possible completions.

The studies we have reviewed here showed that L1 comprehenders (and sometimes, L2 comprehenders too) do not just form expectations based on context, and integrate them with other information, but can also preactivate the phonological and morphological features of the most likely completions, and use those to form further expectations about other elements in the sentence. However, we have also seen that there is variation due to factors (such as L1-L2 similarity for L2 speakers) which can constrain the extent to which comprehenders engage in prediction. In the next section, we review in more detail the factors which constitute the main limitations to predictive mechanisms, in both L1 and L2 speakers.

Limitations to prediction during sentence comprehension in both L1 and L2 speakers

It has been suggested that L2 speakers may suffer from a Reduced Ability to Generate Expectations, or RAGE (Grüter et al., 2014, 2017). Grüter et al. (2014, 2017) showed that L2 English speakers (L1 Japanese and L1 Korean) do not take verb aspect information into consideration when formulating expectations about discourse context, as L1 English speakers

do, despite the fact that, in both Japanese and Korean, verb aspect has the same discourse implications as it does in English. Grüter et al. argue that processing limitations make it difficult for L2 speakers to integrate cues to formulate predictions.

However, as the authors themselves point out, the distinction between prediction in the L1 and L2 is far from a monolithic one. L2 speakers' prediction abilities vary depending on proficiency, and L1 speakers can also be limited in their prediction abilities, prompting several authors to conclude that the difference between L1 and L2 in prediction is probably a quantitative, rather than a qualitative, one (Grüter et al., 2017; Kaan, 2014; Phillips & Ehrenhofer, 2015). Therefore, rather than asking whether or not L2 speakers can predict, we can look at factors which affect prediction in both L1 and L2 speakers, and which can tend to affect L2 speakers in a specific way. The prediction of specific linguistic input based on expectations is constrained by a number of factors, both relating to the input (linguistic constraints and the context) and to the comprehender (cognitive abilities, processing speed, and proficiency). Here, we highlight just three of these constraints on prediction, which are not unique to L2 speakers, but can affect processing in both L1 and L2: cognitive abilities, proficiency, and task design. This is not meant to be a comprehensive list of the factors affecting prediction in L2 speakers (see Kaan, 2014 for a review). Rather, it shows how these factors can vary from being intrinsic to speakers (cognitive abilities) to completely extrinsic (task design), highlighting the complexity of mechanisms involved.

First, predictive mechanisms, other than the most basic sensitivity to word cloze probability, are cognitively demanding and are not consistently observed, even in L1 speakers: They are impaired in elderly L1 speakers (DeLong et al., 2012; see also Huettig, 2015) and low-literacy populations (Mishra et al., 2012). Predictive looks in visual world paradigm eye-tracking studies correlate with working memory capacity and processing speed (Huettig &

Janse, 2016) and are delayed under memory load (Ito, Corley, et al., 2018). In a visual world eye-tracking study using Russian, Sekerina (2015) showed a gradient in the speed with which different Russian-speaking populations (L1 adult, heritage speaker adult, L1 child) showed preferential looking towards the upcoming noun based on gender information from the preceding adjective. Prediction in L1 speakers can also vary in speed depending on the cues used to formulate expectations (Chow et al., 2016, 2018). L2 speakers may be particularly affected by time constraints as they tend to be slower in their processing compared to L1 speakers (Frenck-Mestre, 2002; Frenck-Mestre et al., 2014; Hahne, 2001), and so predictive behaviour may not be observable.

Second, anticipation of linguistic material is heavily dependent on proficiency, both in the sense of correct knowledge representations, and in the procedural sense: Using morphological dependencies to generate expectations (as in visual world paradigms) or to probe for them (EEG studies) relies on participants both having a knowledge of these dependencies, and being able to deploy it rapidly during processing. Such automatized knowledge may not be available to all, and perhaps only to the most advanced L2 learners for some linguistic dependencies⁴. In fact, as reviewed above, while L2 speakers often do not predict to the same extent as L1 speakers, several studies have replicated prediction findings from L1 using high-proficiency L2 speakers, both using eye-tracking and EEG (Dussias et al., 2013; Foucart et al., 2014; Ito, Corley, et al., 2018; Ito, Pickering, et al., 2018). Effects of proficiency have been observed in L1, too, both with regards to proficiency in the sense of knowledge representations (e.g., vocabulary size) and in the procedural sense (e.g., verbal fluency). Speed of anticipatory looking in a visual world eye-tracking paradigm correlated

⁴ However, there is also evidence of predictive processing of determiner-noun agreement emerging after relatively brief exposure to a miniature language (Curcic et al., 2019), but only in subjects who became aware of the agreement rule.

positively with vocabulary size in both adults and children (Borovsky et al., 2012), and with word reading skills in children (Mani & Huettig, 2014). In adults, anticipatory looking based on semantic cues was found to correlate with verbal fluency (Hintz et al., 2014), which is compatible with the idea of a *prediction-by-production* route (Pickering & Gambi, 2018; Pickering & Garrod, 2013). Following this account, reduced production skills may explain differences in prediction performance between L2 and L1 speakers, too. Grüter et al. (2012) found that L2 speakers who were unable to use gender cues to anticipate nouns also made errors in gender assignment on determiners in elicited production. Similarly, Hopp (2013) observed that English learners of German showed nativelike anticipatory use of gender information in a visual world paradigm only if they were able to accurately and consistently produce the right gender assignment for those nouns. On the other hand, high proficiency in L2 speakers does not necessarily lead to native speaker-like prediction: even highly proficient L2 speakers may fail to display fully native speaker-like prediction (Dijkgraf et al., 2019; Kaan et al., 2016) and studies investigating the relation between prediction and L2 proficiency have not found a direct correlation between the two (e.g., Ito, Corley, & Pickering, 2018; Kim & Grüter, 2021; see Kaan & Grüter (forthcoming) for a discussion).

Third, the nature of the task used can have a significant effect on the emergence and experimental detection of predictive processing, in multiple ways. On one hand, prediction studies demonstrating prediction during language processing generally employ highly constraining contexts, which are rare in natural language use (as noted by Luke & Christianson, 2016). In fact, due to the rarity of highly constraining contexts in everyday language use, the relevance of predictive processes has been questioned, both in relation to language comprehension (Huettig & Mani, 2016) and L1 acquisition (Rabagliati et al., 2016). On the other hand, even given a context which encourages predictive processing, prediction may not be detected if not enough time is available. Huettig and Guerra (2015) found that anticipatory

looking by L1 Dutch speakers based on gendered determiner cues was observed if the visual targets appeared on screen four seconds before the spoken sentence, but not if they only appeared one second before. Trenkic, Mirković, & Altmann (2014) found that L2 English speakers performed similarly to (though more slowly than) native speakers when processing English determiners in a visual world eye-tracking paradigm, even though they didn't have the equivalent feature in their L1 (Mandarin). In fact, neither native speakers nor the L2 speakers showed evidence of prediction, as preferential looking emerged *after* the onset of the noun following the determiner in both groups (rather than prior to the noun); however, even this effect was slower to emerge in the L2 group, relative to native speakers. These findings illustrate the critical role of timing in detecting prediction. The reason why the data from Trenkic et al. (2014) did not count as evidence of 'prediction' is that participants did not begin looking at potential referents before the onset of the noun; however, as we have seen, the speed with which preferential looking emerges is affected by several factors such as task timing and memory load, even in L1 speakers. Therefore, it is possible that, if participants had more time, preferential looking would have been observable even without needing to hear the noun first. The same applies more generally to L2 speakers, when they fail to show predictive behaviour in eye-tracking experiments. When preferential looking emerges ahead of the onset of the target for L1 speakers but not for L2 speakers it may simply reflect slower processing in L2 speakers (in a context that did not allow for detection at longer time intervals), rather than a qualitative difference between the groups.

In reality, all the factors described above - cognitive abilities, proficiency, and task design - are likely to interact with each other. For instance, whether a task will show evidence of prediction depends, among other things, on whether it allows enough time for prediction to emerge and be observed in the particular experimental paradigm being used; in turn, what constitutes "enough time" will be affected by individual differences such as proficiency, verbal

abilities and working memory, for both L1 and L2 speakers. These limitations are relevant to our core question about the extent to which language learning, and L2 learning in particular, may draw on prediction as a learning mechanism, to which we now turn our attention.

Learning from error: prediction as a learning mechanism

Having examined the extent to which language processing during sentence comprehension involves prediction, and the factors constraining it in both L1 and L2 speakers, we now turn to the question of whether prediction can serve as a learning mechanism. First, we lay out different accounts of the potential role of prediction in SLA, and in language acquisition more generally. We then examine the evidence for error-based learning, starting from the computational modelling which has inspired proposals on the role of prediction in SLA, and also covering empirical evidence from priming and adaptation in both L1 and L2.

What role may prediction play in L1 and L2 acquisition?

While there is abundant evidence that predictive mechanisms operate in language comprehension, the extent to which they may also contribute to language acquisition is debated. While some argue that prediction drives L1 acquisition (Chang et al., 2013; Rowland et al., 2012), there is scepticism on the importance of prediction in this respect (Huettig, 2015; Huettig & Mani, 2016; Rabagliati et al., 2016). Enabling learning is one of the main functions that have been proposed for predictive processing (see Huettig, 2015 for a discussion). Specifically, it has been suggested that prediction and error-based learning are necessary for L1 acquisition, partly due to the score of studies on statistical learning showing that children use forward transitional probabilities (the likelihood of an element being followed by another) to acquire language (Saffran, 1996). However, tracking these probabilities does not necessarily involve predictive processing; in fact, *backwards* probabilities (the likelihood of an element

being preceded by another) are also used by children (Pelucchi et al., 2009). The fact that learning can occur without prediction, then, casts doubt on claims that prediction is absolutely necessary for language learning (Huettig, 2015). Overall, the empirical evidence on whether children actually use prediction for learning their L1 is mixed (Rabagliati et al., 2016). While it is not clear whether prediction during processing is a necessary or pervasive element of L1 acquisition, there is, however, certainly evidence that prediction *can* be a source of learning. Computational models using error-based learning, which rely on prediction, can model data from L1 syntactic acquisition (Chang et al., 2006) and from priming studies in L1 and L2, supporting claims that error-based learning may be the mechanism underpinning these phenomena (Bock et al., 2007). This evidence is reviewed more fully in the following section.

Against this backdrop, it has been suggested that prediction may serve as a learning mechanism for certain aspects of SLA (Phillips & Ehrenhofer, 2015). Specifically, this proposal aims to address the problem of *overgeneralisation*, traditionally stressed by generativist approaches to language acquisition: How learners can learn to use rules productively yet avoid producing ungrammatical forms (e.g., “I goed” instead of “I went”), even though they have no direct evidence that such forms are not allowed in the language. In L1 acquisition, children often overgeneralise rules, but eventually converge on the target variety of their language (Pinker, 2009). According to Phillips & Ehrenhofer’s proposal, prediction may offer a way out of overgeneralisation for L2 learners, especially when learning complex phenomena (e.g., those that require learners to integrate information from syntax and semantics), by providing the opportunity for *hypothesis testing*: The ability to make sophisticated predictions about upcoming input, using multiple cues, may allow learners to acquire complex contingencies and, crucially, for retreat from overgeneralisation when these hypotheses are not confirmed (Phillips & Ehrenhofer, 2015).

In order to exploit this mechanism, learners would need to rapidly integrate multiple cues as they process speech; thus, the fact that the proposed hypothesis-testing mechanism relies on processing speed and proficiency makes it an unlikely candidate for early L1 acquisition, as Phillips & Ehrenhofer (2015) acknowledge. In L2 learners, however, it may serve a useful function, if they can formulate the relevant prediction quickly enough and track the source of their predictions so that they may readjust their prediction based on that cue for the next time they experience it. To our knowledge, this proposal has not yet been investigated empirically. We will, however, look at the existing evidence for error-based learning. In the following sections, we first review computational models showing that data from L1 processing and acquisition is consistent with a learning mechanism driven by error-based learning which, in turn, requires prediction to occur. We then review empirical evidence from human language processing, which is compatible with the predictions made by these models, and which shows evidence of error-based learning during both L1 and L2 processing.

Insights from computational modelling of L1 processing and acquisition

Computational modelling has shown that certain aspects of language can be acquired through the same mechanisms that are used to process it (Elman, 1990). The models in question use so-called neural networks, a particular type of computational model loosely inspired by brain architecture, which consists of units connected to each other in a network. As each unit is activated, it transmits a signal to the units it is connected to. The connections between units are weighted, meaning that the extent to which one unit affects the next can be adjusted. Neural network models can be used for a variety of tasks, such as classifying data (for instance, determining whether an image is a picture of a bee). A model can learn to perform the required task through supervised training: It is given a ‘training set’ consisting of input (e.g., a set of images) and a desired, or target, output (e.g., a set of labels, either “bee” or “no bee”) for that

input. As the model works its way through the input, it produces its own output (i.e., a label for each image). At each step, the model compares its own output to the target output (i.e., the desired label). The difference between the model's output and the target output is known as the *prediction error*, and is used by the model to adjust its connection weights, so that the next time it encounters that input, the output it produces will be closer to the target output. In this manner, by gradually adjusting its connection weights, the model learns to perform the required task.

Sentence processing, too, can be modelled with neural networks. It is often modelled using a particular type of neural network called a recurrent neural network, or RNN (Elman, 1990). In an RNN, an additional series of connections allows the model to keep track of its previous states (akin to keeping track of words experienced in a sentence), which allows it to process input unfolding over time. Elman (1990) first used this architecture to train a model on next-word prediction in a miniature language. As it encountered each word in the sentence, the model's output was a pattern of activation reflecting the probabilities of possible continuations. Any difference between its output and the actual next word (*prediction error*) was then used to adjust its connections. As the model learned the word order patterns in the language, words belonging to the same syntactic categories began to produce similar patterns of activation, even though the model had no initial notion of word category. This suggested that it is possible to acquire syntactic structure simply through processing language, by estimating the likelihood of possible continuations, and adjusting it based on experience (Elman, 1990, 1993; see Mikolov et al. 2013, for similar results obtained with a natural language corpus).

The potential relevance of these models to prediction during human language processing, and, in turn, language learning, is demonstrated by research showing that the magnitude of prediction error the models encounter positively correlates with sensitivity to

word predictability in humans. RNNs trained on next-word prediction were first trained on natural language corpora and then applied to the same materials that were given to human participants in experimental studies, making it possible to compare model performance with human processing. Word-by-word prediction error from these models has been shown to reflect reading times (Frank, 2013; Frank & Hoeks, 2019; Goodkind & Bicknell, 2018; Monsalve et al., 2012; Van Schijndel & Linzen, 2018), N400 amplitudes during EEG (Frank et al., 2013, 2015), and MEG responses (Wehbe et al., 2014). In other words, the ‘error signal’ used by neural network models to do error-based learning positively correlates with language users’ expectations about upcoming input, which suggests that these expectations may be what supports error-based learning in humans too.

More support for a potential role of prediction in learning comes from the Dual-Path model (Chang, 2002; Chang et al., 2006). This is a specific instance of RNN model that is particularly relevant to current debate on prediction in SLA, because it has been cited as the theoretical underpinning for error-based learning in L2 and for the potential role of prediction in such learning (e.g., Jackson & Hopp, 2020; Leal et al., 2017; Phillips & Ehrenhofer, 2015). Originally developed as a model of language production (Chang, 2002), the Dual-Path model was adapted to next-word prediction by harnessing its production output to formulate predictions for upcoming words (Chang et al., 2006); that is, it simulates the *prediction-by-production* route in humans (Huettig, 2015; Pickering & Garrod, 2013). The model was evaluated against data from L1 acquisition, showing that it could simulate findings from preferential looking studies on the acquisition of transitive structures (i.e., Hirsh-Pasek & Golinkoff, 1996; Naigles, 1990). It could also reproduce data on structural priming (Chang et al., 2006) and the acquisition of word order biases in English and Japanese (Chang, 2009). Priming itself has been suggested to be a case of implicit error-based learning, and we review evidence for this claim in the next section.

Evidence from priming and adaptation effects in L1 and L2 speakers

The main source of experimental behavioural evidence for error-based learning, in both L1 and L2 speakers, comes from studies on structural priming and adaptation. *Structural priming* refers to the fact that when language users encounter a particular syntactic construction, they are more likely to expect it again, or to use it in production, than they were before encountering it (Arai et al., 2007; Bock, 1986; Ferreira & Bock, 2006; Ledoux et al., 2007). Structural priming effects begin early: They have been observed in children as young as 3 years of age, with priming effects lasting across learning sessions (Branigan & Messenger, 2016; Rowland et al., 2012) and during the earliest stages of L2 learning (Weber et al., 2019). It has been suggested that structural priming should be regarded as a case of implicit error-based learning, which modifies a comprehender's language system, rather than simply inducing a temporary activation of representations (Bock et al., 2007; Bock & Griffin, 2000; Chang et al., 2012).

When the priming effect is persistent, it is often called *adaptation*. Kaan & Chun (2018a) define syntactic adaptation as 'persistent' or 'cumulative' priming, where "...comprehension or production is not or not only affected by the most recently encountered structure, but by the cumulative prior exposure to structures of the same type." (p. 87). In computational modelling terms, the updating of expectations seen in adaptation would be akin to 'adjusting one's weights.' Adaptation can be measured by tracking the increase in priming effect following repeated exposure to a structure over time, which may manifest itself as increased likelihood to use it in production (Kaan & Chun, 2018b) or reduced response times when encountering it in comprehension (Fine & Jaeger, 2016). Another method, more familiar to SLA research, is to use a pre-test / post-test design (Jackson & Ruf, 2017; Jackson & Hopp, 2020). Adaptation to syntactic structure alternations (such as that between prepositional object and double object dative constructions in English) has been observed in L1 production (Jaeger

& Snider, 2013; Kaan & Chun, 2018b; Kaschak, 2007; Kaschak et al., 2006, 2011; Kaschak & Borreggine, 2008), and in L1 comprehension (Farmer et al., 2014; Fine et al., 2013; Fine & Jaeger, 2016). Adaptation effects have also frequently been observed in L2 speakers (Hopp, 2020; Jackson & Ruf, 2017; Kaan & Chun, 2018b; McDonough & Trofimovich, 2015; Montero-Melis & Jaeger, 2020; Shin & Christianson, 2012; see Jackson, 2018 for a review; and see Jackson & Hopp, 2020 for an instance of priming without adaptation).

The permanence of priming effects observed in adaptation is compatible with accounts of priming as an instance of learning; however, it does not specifically implicate a role for prediction error as the driving learning mechanism. Additional support for the claim that priming (and adaptation) is a learning mechanism, and specifically an error-based learning mechanism, comes from the observation of *inverse frequency effects*, which are predicted by an error-based learning model. In the Dual-Path model, low-frequency words would generate greater prediction error, causing a larger adjustment in the weights and therefore a larger learning effect (Chang et al., 2006). Inverse frequency effects are also observed at the level of structure, not just words: In both L1 and L2, structures that have lower frequency in the input elicit greater priming effects (Hartsuiker et al., 1999; Hartsuiker & Westenberg, 2000; Jaeger & Snider, 2013; Kaan & Chun, 2018b; Kaschak, 2007; Kaschak et al., 2006; Montero-Melis & Jaeger, 2020). In L2 learners, frequency effects appear to be based on the statistics of the L1 at lower proficiency levels, moving to more native speaker-like expectations as proficiency increases (Jackson & Ruf, 2017; Montero-Melis & Jaeger, 2020; see Jackson's 2018 review). Finally, at least in L1, frequency effects also extend to adaptation, with greater adaptation observed for dative structures that are encountered in unexpected contexts (Fazekas et al., 2020). More recent research has begun to directly investigate the link between adaptation and language acquisition, showing that children can adapt to different syntactic structures and use their adapted predictions to infer the meaning of new words and interpret ambiguous words

(Havron et al., 2019; Havron et al., 2021). All these findings suggest that structural priming and adaptation, driven by a prediction-based mechanism, could potentially play a role in both L1 and L2 acquisition.

Limitations of error-based learning during processing

Even adaptation, however, is subject to variability. For instance, adaptation to structural alternation may depend on which specific semantically constrained constructions are used. In Experiment 1 in Jackson & Ruf (2017), intermediate English-German L2 learners were exposed to fronted temporal adverbial phrases such as *Im Winter trägt Paul eine Jacke* (“In winter Paul wears a jacket”), which are marked in both English and German, but more frequent in German. They showed both immediate priming and adaptation to these structures, as measured in a post-test. However, in Experiment 2, exposure to fronted adverbial phrases using locative instead of temporal expressions (e.g., *Auf dem Berg trägt der Schüler eine Jacke*, “On the mountain the pupil wears a jacket”) led to short-term priming, but no adaptation. Similarly, Jackson & Hopp (2020) found L2 English speakers exposed to fronted adverbials in English exhibited immediate priming but no evidence of adaptation in the post-test, unlike L1 English speakers.

In particular, adaptation to garden-path sentences appears to be less robust than adaptation to simple structural alternations, in both L1 and L2 speakers. While some studies find an adaptation effect to garden-path sentences in L1 (Farmer et al., 2014; Fine et al., 2013; Kaan et al., 2019), others find no such evidence (Dempsey et al., 2020) or show that is very difficult to detect (Prasad & Linzen, 2019). In L2 speakers, the evidence is again mixed (Kaan et al., 2019; Hopp, 2020). Kaan et al. examined adaptation to garden-path sentences (filler-gap wh-subordinates and ambiguous coordination) in L1 and L2 English speakers, in a study using

self-paced reading. Adaptation was only found in the L1 English group, and only for the 'easier' structure (coordination). Therefore, learning from prediction error during processing was difficult, not just for L2 speakers but for L1 speakers too, who only showed adaptation to the less cognitively demanding of the two structures examined. Such a finding is arguably compatible with evidence that L2 speakers have difficulty recovering from garden-path sentences in a way that is reflective of the difficulty faced by children in their L1. In Pozzan & Trueswell's (2016) study, participants listened to instructions and carried them out while interacting with a visual scene. Error rates on garden-path sentences (showing an inability to revise the initial parse) were similar for 5-year-old L1 speakers and adult L2 learners when there were no referential cues supporting the target interpretation. On the other hand, Hopp (2020, Experiment 1) found that L2 German speakers *could* adapt to garden-path sentences (specifically, to the intransitive use of optionally transitive verbs such as *play*) if the sentences provided an unambiguous cue flagging the correct interpretation, in the form of case marking (e.g., in "The boy played and he pleased the parents with the music", the verb *played* would be followed by *and* and the pronoun *he* in the nominative case, signalling the start of a new clause)⁵. In sum, as is the case for prediction mechanisms, we see that there is variation both among L1 and L2 speakers in the extent to which they can adapt to specific syntactic structures.

Another potential source of variability in adaptation is that prediction depends on context: Evidence suggests that the extent to which predictions are made during language comprehension depends on the overall reliability of context as a source of prediction (Delaney-Busch et al., 2019), and that predictions stop being formulated when cues become unreliable (Brothers et al., 2017). In a self-paced reading study by Brothers et al. (2017, Experiment 2),

⁵ However, participants continued to show garden-path effects in response to optionally transitive verbs when these were embedded in other sentences, without unambiguous case marking. This suggests that they did not adapt to the intransitive use of optionally transitive verbs in general, but rather to the nominative pronoun signalling the start of a new clause (Hopp, 2020).

the global validity of predictive cues affected the extent of prediction found. Participants read critical sentences that all had highly predictable completions (i.e., in highly constrained sentential contexts), and a set of highly constraining filler sentences that were manipulated by either having expected or unexpected completions. Participants who saw the filler sentences with the expected completions showed an effect of predictability on the critical items (i.e., reduced reaction times for predictable completions), while the group who saw the filler items with unexpected completions did not show a statistically significant prediction effect on the predictable completion critical items. This suggests that the overall likelihood of disconfirmed predictions had led the group who had experienced the unpredictable completions to abandon the use of sentential constraint as a cue.

The sensitivity of prediction to cue reliability means that prediction error may sometimes result in abandonment or temporary suppression of predictive mechanisms, instead of leading to adaptation (Brothers et al., 2017; Hopp, 2016; Husband & Bovolenta, 2020; Lau et al., 2013; Van Heugten et al., 2012; Wlotko & Federmeier, 2011). Hopp (2016) manipulated cues that had high predictive value (gender marking in German) so as to make them unreliable. L1 speakers, who had been using them to anticipate upcoming referents in a visual world eye-tracking paradigm, stopped anticipatory looking when gender marking became unreliable. Hopp argues that this explains why L2 German speakers in previous studies (Hopp, 2013, 2016) only used gender predictively if they consistently produced accurate gender marking; for those who did not (meaning that they had non-target representations), prediction lead to error, which perhaps led to it being abandoned. If a cue is not a reliably predictive cue, it is arguably an important part of the learning process to stop using it (as it would cause inefficient processing). On the other hand, language users can rapidly adapt to input that is seemingly inconsistent, if they can identify new reliable cues in it which have predictive value. KroczeK & Gunter (2017) exposed listeners to speech by speakers who differed in the relative

probabilities of specific syntactic structures they used (OSV / SVO word order in German). While structure usage was not consistent across speakers, listeners developed distinct expectations for syntactic continuations depending on the speaker, as each speaker manifested reliable structure usage. These findings suggest that language users are constantly evaluating the reliability of cues and potential cues, abandoning them if they are no longer reliable and tuning in to new ones which reliably predict upcoming input.

Conclusion

We have seen that the picture, when it comes to prediction in comprehension and learning, is extremely nuanced. Prediction itself consists of different processes, with different levels of complexity, from basic priming mechanisms to preactivation of specific features. Prediction abilities can vary depending on a large number of factors, in L1 and L2 speakers alike. Variability between L1 and L2 speakers – and even among L1 and L2 speakers – increases as prediction becomes more complex. Sensitivity to word predictability is the most robust type of prediction, showing the least difference between L1 and L2 speakers. However, even at the most complex level of prediction (i.e., preactivation of specific features), L2 speakers of sufficient proficiency have the potential to predict in a native speaker-like fashion, though higher proficiency does not *necessarily* lead to native speaker-like prediction. In both groups of language users, then, prediction is modulated by cognitive abilities, processing speed, and various aspects of proficiency. More research is needed to investigate the complex relationships between these factors. The role of cognitive and linguistic individual differences, is clearly a burgeoning area of interest in the language learning sciences (e.g., Bolibaugh & Foster, 2021; Buffington et al., 2021; Pili-Moss, 2021; Riches & Jackson, 2018; Walker et al., 2020; including special issues dedicated to the topic such as those edited by Andringa & Dabrowska, 2019; Roberts & Meyer, 2012), and investigating the role of individual cognitive

abilities in prediction and, specifically, error-based learning, in L1 and L2 (or Lx) speakers would constitute a timely extension of this agenda. Such research could shed more light on some of the factors we have reviewed, such as the varying effect of L2 proficiency (for instance, helping to clarify the relative contributions of knowledge representation and 'procedural' proficiency in enabling prediction). Individual differences focused research could also help to address broader questions relating to the explanatory power of processing-based accounts of language acquisition (e.g., Havron et al., 2021).

There is ample evidence – from empirical studies on priming and adaptation – that prediction error can be one source of learning, and such evidence is compatible with the predictions made by computational models employing error-based learning. Adaptation to syntactic structure is observed in both L1 and L2 speakers, but again, there is a great deal of variation. Cue reliability affects the extent to which comprehenders make predictions, and factors such as the complexity of the specific syntactic structures encountered can influence the degree of adaptation that can take place, both in L1 and L2 speakers. The extent to which prediction error could also support the acquisition of complex contingencies, as suggested by the *Hypothesis Testing* proposal, remains an open question to be investigated empirically. More generally, further research will be needed to investigate the question of which kind of linguistics properties can be learned by prediction error, and through which specific mechanisms. This review has focused primarily on error-based learning during the online processing of syntactic structure, mostly evidenced by syntactic adaptation. Other strands of research, however, have used different paradigms to investigate the effect of prediction error on L2 acquisition, such as research on declarative memory formation and vocabulary learning briefly mentioned at the start of this review (e.g., DeLoof et al., 2018). In reality, these mechanisms - error-based learning during online syntactic processing, and enhanced declarative memory formation driven by prediction error - are unlikely to operate in isolation,

yet the relationship between them is still unclear. A promising avenue for further research will be to examine the connections between them, for instance, by investigating the extent to which error-based learning during incremental sentence processing involves adjusting *existing* representations, and to what extent it may rely on the formation of *new* declarative memories (e.g., see Bernolet et al., 2016, for evidence that syntactic priming is enhanced by explicit memory of the prime sentence). In turn, this understanding may help to address the question of what can and cannot be actually learned (rather than merely consolidated) through prediction error.

In light of the evidence we have reviewed, there is clearly no simple answer to the question of whether impaired prediction in L2 (or Lx) speakers could be a (qualitative or quantitative) hindrance for acquisition. However, the role of prediction in L2 acquisition is a fruitful avenue for future research, best approached with an awareness of its many nuances and complexities.

References

- Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, 73(3), 247–264. [https://doi.org/10.1016/s0010-0277\(99\)00059-1](https://doi.org/10.1016/s0010-0277(99)00059-1)
- Ambridge, B., & Brandt, S. (2013). Lisa filled water into the cup: The roles of entrenchment, pre-emption and verb semantics in German speakers' L2 acquisition of English locatives. *Zeitschrift Für Anglistik Und Amerikanistik*, 61(3), 245–263. <http://pcwww.liv.ac.uk/~ambridge/Papers/German.pdf>
- Andringa, S., & Curcic, M. (2015). HOW EXPLICIT KNOWLEDGE AFFECTS ONLINE L2 PROCESSING: Evidence from Differential Object Marking Acquisition. *Studies in Second Language Acquisition*, 37(02), 237–268. <https://doi.org/10.1017/S0272263115000017>
- Andringa, S., & Dąbrowska, E. (2019). Individual differences in first and second language ultimate attainment and their causes. *Language Learning*, 69, 5–12. <https://doi.org/10.1111/lang.12328>

- Arai, M., van Gompel, R. P. G., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cognitive Psychology*, *54*(3), 218–250. <https://doi.org/10.1016/j.cogpsych.2006.07.001>
- Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, *17*(3), 364–390. [https://doi.org/10.1016/0010-0285\(85\)90013-1](https://doi.org/10.1016/0010-0285(85)90013-1)
- Bock, K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*(3), 355–387.
- Bock, K., Dell, G. S., Chang, F., & Onishi, K. H. (2007). Persistent structural priming from language comprehension to language production. *Cognition*, *104*(3), 437–458. <https://doi.org/10.1016/j.cognition.2006.07.003>
- Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: transient activation or implicit learning? *Journal of Experimental Psychology. General*, *129*(2), 177–192. <https://doi.org/10.1037//0096-3445.129.2.177>
- Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one’s age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, *112*(4), 417–436. <https://doi.org/10.1016/j.jecp.2012.01.005>
- Boyd, J. K., & Goldberg, A. E. (2011). LEARNING WHAT NOT TO SAY: THE ROLE OF STATISTICAL PREEMPTION AND CATEGORIZATION IN A-ADJECTIVE PRODUCTION. *Language*, *87*(1), 55–83. <http://www.jstor.org/stable/23011540>
- Branigan, H. P., & Messenger, K. (2016). Consistent and cumulative effects of syntactic experience in children’s sentence production: Evidence for error-based implicit learning. *Cognition*, *157*, 250–256. <https://doi.org/10.1016/j.cognition.2016.09.004>
- Brod, G., Hasselhorn, M., & Bunge, S. A. (2018). When generating a prediction boosts learning: The element of surprise. *Learning and Instruction*, *55*, 22–31. <https://doi.org/10.1016/j.learninstruc.2018.01.013>
- Brothers, T., Swaab, T. Y., & Traxler, M. J. (2017). Goals and strategies influence lexical prediction during sentence comprehension. *Journal of Memory and Language*, *93*, 203–216. <https://doi.org/10.1016/j.jml.2016.10.002>
- Buffington, J., Demos, A. P., & Morgan-Short, K. (n.d.). THE RELIABILITY AND VALIDITY OF PROCEDURAL MEMORY ASSESSMENTS USED IN SECOND LANGUAGE ACQUISITION RESEARCH. *Studies in Second Language Acquisition*, 1–28. <https://doi.org/10.1017/S0272263121000127>
- Chang, F. (2002). Symbolically speaking: a connectionist model of sentence production. *Cognitive Science*, *26*(5), 609–651. https://doi.org/10.1207/s15516709cog2605_3

- Chang, F. (2009). Learning to order words: A connectionist model of heavy NP shift and accessibility effects in Japanese and English. *Journal of Memory and Language*, 61(3), 374–397. <https://doi.org/10.1016/j.jml.2009.07.006>
- 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>
- Chang, F., Janciauskas, M., & Fitz, H. (2012). Language adaptation and learning: Getting explicit about implicit learning. *Language and Linguistics Compass*, 6(5), 259–278. <https://doi.org/10.1002/lnc3.337>
- Chang, F., Kidd, E., & Rowland, C. F. (2013). Prediction in processing is a by-product of language learning [Review of *Prediction in processing is a by-product of language learning*]. *The Behavioral and Brain Sciences*, 36(4), 350–351. <https://doi.org/10.1017/S0140525X12002518>
- Chow, W.-Y., Lau, E., Wang, S., & Phillips, C. (2018). Wait a second! delayed impact of argument roles on on-line verb prediction. *Language, Cognition and Neuroscience*, 33(7), 803–828. <https://doi.org/10.1080/23273798.2018.1427878>
- Chow, W.-Y., Momma, S., Smith, C., Lau, E., & Phillips, C. (2016). Prediction as memory retrieval: timing and mechanisms. *Language, Cognition and Neuroscience*, 31(5), 617–627. <https://doi.org/10.1080/23273798.2016.1160135>
- Curcic, M., Andringa, S., & Kuiken, F. (2019). The Role of Awareness and Cognitive Aptitudes in L2 Predictive Language Processing: Role of Awareness and Aptitudes in L2 Prediction. *Language Learning*, 69, 42–71. <https://doi.org/10.1111/lang.12321>
- 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. (2009). *Electrophysiological explorations of linguistic pre-activation and its consequences during online sentence processing*. Unpublished doctoral dissertation, UC San Diego. <https://escholarship.org/uc/item/4q7520sb>
- DeLong, K. A., Groppe, D. M., Urbach, T. P., & Kutas, M. (2012). Thinking ahead or not? Natural aging and anticipation during reading. *Brain and Language*, 121(3), 226–239. <https://doi.org/10.1016/j.bandl.2012.02.006>
- DeLong, K. A., Troyer, M., & Kutas, M. (2014). Pre-processing in sentence comprehension: Sensitivity to likely upcoming meaning and structure: Pre-processing in sentence comprehension. *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.

- De Loof, E., Ergo, K., Naert, L., Janssens, C., Talsma, D., Van Opstal, F., & Verguts, T. (2018). Signed reward prediction errors drive declarative learning. *PloS One*, *13*(1), e0189212. <https://doi.org/10.1371/journal.pone.0189212>
- Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, *109*(2), 193–210. <https://doi.org/10.1016/j.cognition.2008.07.008>
- Dempsey, J., Liu, Q., & Christianson, K. (2020). Convergent probabilistic cues do not trigger syntactic adaptation: Evidence from self-paced reading. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *46*(10), 1906–1921. <https://doi.org/10.1037/xlm0000881>
- 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>
- Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Memory and Language*, *20*(6), 641.
- Elman, J. L. (1990). Finding Structure in Time. *Cognitive Science*, *14*(2), 179–211. https://doi.org/10.1207/s15516709cog1402_1
- Elman, J. L. (1993). Learning and development in neural networks: the importance of starting small. *Cognition*, *48*(1), 71–99. [https://doi.org/10.1016/0010-0277\(93\)90058-4](https://doi.org/10.1016/0010-0277(93)90058-4)
- Farmer, T., Fine, A., Yan, S., Cheimariou, S., & Jaeger, F. (2014). Error-Driven Adaptation of Higher-Level Expectations During Reading. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *36*(36). <https://escholarship.org/uc/item/00t2m3wr>
- Fazekas, J., Jessop, A., Pine, J., & Rowland, C. (2020). Do children learn from their prediction mistakes? A registered report evaluating error-based theories of language acquisition. *Royal Society Open Science*, *7*(11), 180877. <https://doi.org/10.1098/rsos.180877>
- Ferreira, V. S., & Bock, K. (2006). The functions of structural priming. *Language and Cognitive Processes*, *21*(7-8), 1011–1029. <https://doi.org/10.1080/016909600824609>
- Fine, A. B., & Jaeger, T. F. (2016). The role of verb repetition in cumulative structural priming in comprehension. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *42*(9), 1362–1376. <https://doi.org/10.1037/xlm0000236>
- Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid Expectation Adaptation during Syntactic Comprehension. *PloS One*, *8*(10), e77661. <https://doi.org/10.1371/journal.pone.0077661>

- Foraker, S., Regier, T., Khetarpal, N., Perfors, A., & Tenenbaum, J. (2009). Indirect evidence and the poverty of the stimulus: the case of anaphoric one. *Cognitive Science*, 33(2), 287–300. <https://doi.org/10.1111/j.1551-6709.2009.01014.x>
- Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *The Quarterly Journal of Experimental Psychology Section A*, 33(4), 465–495. <https://doi.org/10.1080/14640748108400804>
- 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.
- Frank, S. L. (2013). Uncertainty Reduction as a Measure of Cognitive Load in Sentence Comprehension. *Topics in Cognitive Science*, 5(3), 475–494. <https://doi.org/10.1111/tops.12025>
- Frank, S. L., & Hoeks, J. (2019). The interaction between structure and meaning in sentence comprehension: Recurrent neural networks and reading times. *Proceedings of the 41st Annual Conference of the Cognitive Science Society*, 337–343. <https://doi.org/10.31234/osf.io/mks5y>
- Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2013). Word surprisal predicts N400 amplitude during reading. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 878–883. <https://repository.ubn.ru.nl/bitstream/handle/2066/119221/119221.pdf>
- Frenck-Mestre, C. (2002). An on-line look at sentence processing in the second language. In R. Herrida & J. Altarriba (Eds.), *Bilingual Sentence Processing* (pp. 217–236). North Holland.
- Frenck-Mestre, C., German, E. S., & Foucart, A. (2014). Qualitative Differences In Native And Nonnative Semantic Processing as Revealed by ERPs. In R. R. Heredia & J. Altarriba (Eds.), *Foundations of Bilingual Memory* (pp. 237–255). Springer New York. https://doi.org/10.1007/978-1-4614-9218-4_12
- Goodkind, A., & Bicknell, K. (2018). Predictive power of word surprisal for reading times is a linear function of language model quality. *Proceedings of the 8th Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2018)*, 10–18. <https://www.aclweb.org/anthology/W18-0102.pdf>
- Greve, A., Cooper, E., Kaula, A., Anderson, M. C., & Henson, R. (2017). Does prediction error drive one-shot declarative learning? *Journal of Memory and Language*, 94, 149–165. <https://doi.org/10.1016/j.jml.2016.11.001>
- Greve, A., Cooper, E., Tibon, R., & Henson, R. N. (2019). Knowledge is power: Prior knowledge aids memory for both congruent and incongruent events, but in different ways. *Journal of Experimental Psychology. General*, 148(2), 325–341. <https://doi.org/10.1037/xge0000498>

- 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., & Schafer, A. (2014). The role of discourse-level expectations in non-native speakers' referential choices. *Proceedings of the Annual Boston University Conference on Language Development*. <https://par.nsf.gov/biblio/10028988>
- 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>
- Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research*, 30(3), 251–266. <https://doi.org/10.1023/a:1010490917575>
- Hartsuiker, R. J., Kolk, H. H. J., & Huiskamp, P. (1999). Priming Word Order in Sentence Production. *The Quarterly Journal of Experimental Psychology Section A*, 52(1), 129–147. <https://doi.org/10.1080/713755798>
- Hartsuiker, R. J., & Westenberg, C. (2000). Word order priming in written and spoken sentence production. *Cognition*, 75(2), B27–B39. [https://doi.org/10.1016/s0010-0277\(99\)00080-3](https://doi.org/10.1016/s0010-0277(99)00080-3)
- Havron, N., Babineau, M., Fiévet, A.-C., Carvalho, A., & Christophe, A. (2021). Syntactic prediction adaptation accounts for language processing and language learning. *Language Learning*, 71(4). <https://doi.org/10.1111/lang.12466>
- Havron, N., de Carvalho, A., Fiévet, A.-C., & Christophe, A. (2019). Three- to Four-Year-Old Children Rapidly Adapt Their Predictions and Use Them to Learn Novel Word Meanings. *Child Development*, 90(1), 82–90. <https://doi.org/10.1111/cdev.13113>
- Hintz, F., Meyer, A. S., & Huettig, F. (2014). The influence of verb-specific featural restrictions, word associations, and production-based mechanisms on language-mediated anticipatory eye movements. *The 27th Annual CUNY Conference on Human Sentence Processing*. https://pure.mpg.de/rest/items/item_1949864/component/file_1949863/content
- Hirsh-Pasek, K., & Golinkoff, R. M. (1996). The intermodal preferential looking paradigm: A window onto emerging language comprehension. In D. McDaniel, C. McKee, & H. S. Cairns (Eds.), *Language, speech, and communication. Methods for assessing children's syntax* (pp. 105–124). The MIT Press. <https://psycnet.apa.org/record/1997-97174-005>
- 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. *IRAL, International Review of Applied Linguistics in Language Teaching: Revue Internationale de Linguistique Appliquée Enseignement Des Langues. Internationale Zeitschrift Fur*

- Angewandte Linguistik in Der Spracheroziehung*, 53(3). <https://doi.org/10.1515/iral-2015-0014>
- Hopp, H. (2016). Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research*, 32(2), 277–307. <https://doi.org/10.1177/0267658315624960>
- 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>
- Hopp, H. (2020). Morphosyntactic adaptation in adult L2 processing: Exposure and the processing of case and tense violations. *Applied Psycholinguistics*, 41(3), 627–656. <https://doi.org/10.1017/S0142716420000119>
- 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., & Guerra, E. (2015). Testing the limits of prediction in language processing: Prediction occurs but far from always. *The 21st Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP 2015)*.
- 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>
- Husband, E. M., & Bovolenta, G. (2020). Prediction failure blocks the use of local semantic context. *Language, Cognition and Neuroscience*, 35(3), 273–291. <https://doi.org/10.1080/23273798.2019.1651881>
- 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., 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>
- Jackson, C. N. (2018). Second language structural priming: A critical review and directions for future research. *Second Language Research*, 34(4), 539–552. <https://doi.org/10.1177/0267658317746207>
- Jackson, C. N., & Hopp, H. (2020). Prediction error and implicit learning in L1 and L2 syntactic priming. *International Journal of Bilingualism*, 24(5–6), 895–911. <https://doi.org/10.1177/1367006920902851>

- Jackson, C. N., & Ruf, H. T. (2017). The priming of word order in second language German. *Applied Psycholinguistics*, 38(2), 315–345. <https://doi.org/10.1017/S0142716416000205>
- Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, 127(1), 57–83. <https://doi.org/10.1016/j.cognition.2012.10.013>
- Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism*, 4(2), 257–282. <https://doi.org/10.1075/lab.4.2.05kaa>
- Kaan, E., & Chun, E. (2018a). Syntactic Adaptation. In K. D. Federmeier & D. G. Watson (Eds.), *Psychology of Learning and Motivation* (Vol. 68, pp. 85–116). Academic Press. <https://doi.org/10.1016/bs.plm.2018.08.003>
- Kaan, E., & Chun, E. (2018b). Priming and adaptation in native speakers and second-language learners. *Bilingualism: Language and Cognition*, 21(2), 228–242. <https://doi.org/10.1017/S1366728916001231>
- Kaan, E., Futch, C., Fuertes, R. F., Mujcinovic, S., & de la Fuente, E. Á. (2019). Adaptation to syntactic structures in native and nonnative sentence comprehension. *Applied Psycholinguistics*, 40(1), 3–27.
- Kaan, E. & Grüter, T. (forthcoming). Prediction in second language processing and learning: Advances and directions. In: Kaan, E. and Theres Grüter (eds.) *Prediction in second language processing and learning*, John Benjamins.
- 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>
- Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan. <http://www.math.chalmers.se/~ulfp/Review/fastslow.pdf>
- Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49(1), 133–156. [https://doi.org/10.1016/S0749-596X\(03\)00023-8](https://doi.org/10.1016/S0749-596X(03)00023-8)
- Kaschak, M. P. (2007). Long-term structural priming affects subsequent patterns of language production. *Memory & Cognition*, 35(5), 925–937. <https://doi.org/10.3758/bf03193466>
- Kaschak, M. P., & Borreggine, K. L. (2008). Is long-term structural priming affected by patterns of experience with individual verbs? *Journal of Memory and Language*, 58(3), 862–878. <https://doi.org/10.1016/j.jml.2006.12.002>
- Kaschak, M. P., Kutta, T. J., & Jones, J. L. (2011). Structural priming as implicit learning: cumulative priming effects and individual differences. *Psychonomic Bulletin & Review*, 18(6), 1133–1139. <https://doi.org/10.3758/s13423-011-0157-y>

- Kaschak, M. P., Loney, R. A., & Borreggine, K. L. (2006). Recent experience affects the strength of structural priming. *Cognition*, 99(3), B73–B82. <https://doi.org/10.1016/j.cognition.2005.07.002>
- Kim, H., & Grüter, T. (2021). PREDICTIVE PROCESSING OF IMPLICIT CAUSALITY IN A SECOND LANGUAGE: A VISUAL-WORLD EYE-TRACKING STUDY. *Studies in Second Language Acquisition*, 43(1), 133–154. <https://doi.org/10.1017/S0272263120000443>
- Kroczek, L. O. H., & Gunter, T. C. (2017). Communicative predictions can overrule linguistic priors. *Scientific Reports*, 7(1), 17581. <https://doi.org/10.1038/s41598-017-17907-9>
- Kuperberg, G. R., Brothers, T., & Wlotko, E. W. (2020). A Tale of Two Positivities and the N400: Distinct Neural Signatures Are Evoked by Confirmed and Violated Predictions at Different Levels of Representation. *Journal of Cognitive Neuroscience*, 32(1), 12–35. https://doi.org/10.1162/jocn_a_01465
- 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., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, 62, 621–647. <https://doi.org/10.1146/annurev.psych.093008.131123>
- 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>
- 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
- Leal, T., Slabakova, R., & Farmer, T. A. (2017). THE FINE-TUNING OF LINGUISTIC EXPECTATIONS OVER THE COURSE OF L2 LEARNING. *Studies in Second Language Acquisition*, 39(3), 493–525. <https://doi.org/10.1017/S0272263116000164>
- Ledoux, K., Traxler, M. J., & Swaab, T. Y. (2007). Syntactic priming in comprehension: evidence from event-related potentials. *Psychological Science*, 18(2), 135–143. <https://doi.org/10.1111/j.1467-9280.2007.01863.x>
- 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>
- Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, 18(3), 193–198.

- Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology*, 88, 22–60. <https://doi.org/10.1016/j.cogpsych.2016.06.002>
- 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, S. A., & Shillcock, R. C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science*, 14(6), 648–652. https://doi.org/10.1046/j.0956-7976.2003.psci_1480.x
- McDonough, K., & Trofimovich, P. (2015). Structural priming and the acquisition of novel form-meaning mappings. In T. Cadierno & S. Wind Eskildsen (Eds.), *Usage-based perspectives on second language learning* (pp. 105–123). De Gruyter Mouton Berlin.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. In *Proceedings of the ICLR Workshop*. <http://arxiv.org/abs/1301.3781>
- 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://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_1406626
- 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.
- Monsalve, I. F., Frank, S. L., & Vigliocco, G. (2012). Lexical surprisal as a general predictor of reading time. *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 398–408. <https://dl.acm.org/doi/10.5555/2380816.2380866>
- Montero-Melis, G., & Florian Jaeger, T. (2020). Changing expectations mediate adaptation in L2 production. *Bilingualism: Language and Cognition*, 23(3), 602–617. <https://doi.org/10.1017/S1366728919000506>
- Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, 17(2), 357–374. <https://doi.org/10.1017/s0305000900013817>
- Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., Ferguson, H. J., Fu, X., Heyselaar, E., Huettig, F., Matthew Husband, E., Ito, A., Kazanina, N., Kogan, V., Kohút, Z., Kulakova, E., Mézière, D., Politzer-Ahles, S., Rousselet, G., ... Von Grebmer Zu Wolfsturn, S. (2020). Dissociable effects of prediction and integration during language comprehension: evidence from a large-scale study using brain potentials.

Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 375(1791), 20180522. <https://doi.org/10.1098/rstb.2018.0522>

- 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. <https://doi.org/10.7554/eLife.33468>
- 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., Ferguson, H. J., Busch-Moreno, S., Fu, X., Tuomainen, J., Kulakova, E., Matthew Husband, E., ... Huettig, F. (2017). Limits on prediction in language comprehension: A multi-lab failure to replicate evidence for probabilistic pre-activation of phonology. In *bioRxiv* (p. 111807). <https://doi.org/10.1101/111807>
- O'Grady, W. (2005). *Syntactic carpentry: an emergentist approach to syntax*. Lawrence Erlbaum.
- Otten, M., & Van Berkum, J. J. A. (2008). Discourse-Based Word Anticipation During Language Processing: Prediction or Priming? *Discourse Processes*, 45(6), 464–496. <https://doi.org/10.1080/01638530802356463>
- Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Learning in reverse: eight-month-old infants track backward transitional probabilities. *Cognition*, 113(2), 244–247. <https://doi.org/10.1016/j.cognition.2009.07.011>
- Peters, R., Grüter, T., & Borovsky, A. (2018). Vocabulary size and Native Speaker self-identification influence flexibility in linguistic prediction among adult bilinguals. *Applied Psycholinguistics*, 39(6), 1439–1469. <https://doi.org/10.1017/S0142716418000383>
- Phillips, C., & Ehrenhofer, L. (2015). The role of language processing in language acquisition. *Linguistic Approaches to Bilingualism*, 5(4), 409–453.
- 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>
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *The Behavioral and Brain Sciences*, 36(4), 329–347. <https://doi.org/10.1017/S0140525X12001495>
- Pili-Moss, D. (2021). Cognitive Predictors of Child Second Language Comprehension and Syntactic Learning. *Language Learning*. <https://doi.org/10.1111/lang.12454>
- Pinker, S. (2009). *Language Learnability and Language Development*. Harvard University Press.

- Potts, R., Davies, G., & Shanks, D. R. (2019). The benefit of generating errors during learning: What is the locus of the effect? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 45(6), 1023–1041. <https://doi.org/10.1037/xlm0000637>
- Potts, R., & Shanks, D. R. (2014). The benefit of generating errors during learning. *Journal of Experimental Psychology. General*, 143(2), 644–667. <https://doi.org/10.1037/a0033194>
- Pozzan, L., & Trueswell, J. C. (2016). Second language processing and revision of garden-path sentences: a visual word study*. *Bilingualism: Language and Cognition*, 19(3), 636–643. <https://doi.org/10.1017/S1366728915000838>
- Prasad, G., & Linzen, T. (2019). *Rapid syntactic adaptation in self-paced reading: detectable, but requires many participants*. <https://doi.org/10.31234/osf.io/9ptg4>
- Rabagliati, H., Gambi, C., & Pickering, M. J. (2016). Learning to predict or predicting to learn? *Language, Cognition and Neuroscience*, 31(1), 94–105. <https://doi.org/10.1080/23273798.2015.1077979>
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical Conditioning II: Current Research and Theory*, 2, 64–99.
- Riches, N. and Jackson, L. (2018), Individual Differences in Syntactic Ability and Construction Learning: An Exploration of the Relationship. *Language Learning*, 68: 973-1000. <https://doi.org/10.1111/lang.12307>
- Robenalt, C., & Goldberg, A. E. (2016). Nonnative Speakers Do Not Take Competing Alternative Expressions Into Account the Way Native Speakers Do: L2 Learners and Competition Dynamics. *Language Learning*, 66(1), 60–93. <https://doi.org/10.1111/lang.12149>
- Roberts, L. and Meyer, A. (2012), Individual Differences in Second Language Learning: Introduction. *Language Learning*, 62: 1-4. <https://doi.org/10.1111/j.1467-9922.2012.00703.x>
- Rowland, C. F., Chang, F., Ambridge, B., Pine, J. M., & Lieven, E. V. M. (2012). The development of abstract syntax: evidence from structural priming and the lexical boost. *Cognition*, 125(1), 49–63. <https://doi.org/10.1016/j.cognition.2012.06.008>
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928. <https://www.ncbi.nlm.nih.gov/pubmed/8943209>
- Schwanenflugel, P. J., & LaCount, K. L. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 14(2), 344–354. <https://doi.org/10.1037/0278-7393.14.2.344>
- Schwanenflugel, P. J., & White, C. R. (1991). The Influence of Paragraph Information on the Processing of Upcoming Words. *Reading Research Quarterly*, 26(2), 160–177. <https://doi.org/10.2307/747980>

- Sekerina, I. A. (2015). Predictions, fast and slow. *Linguistic Approaches to Bilingualism*, 5(4), 532–536. <https://doi.org/10.1075/lab.5.4.16sek>
- Shin, J.-A., & Christianson, K. (2012). Structural Priming and Second Language Learning. *Language Learning*, 62(3), 931–964. <https://doi.org/10.1111/j.1467-9922.2011.00657.x>
- Stahl, A. E., & Feigenson, L. (2017). Expectancy violations promote learning in young children. *Cognition*, 163, 1–14. <https://doi.org/10.1016/j.cognition.2017.02.008>
- Stanovich, K. E., & West, R. F. (1981). The effect of sentence context on ongoing word recognition: Tests of a two-process theory. *Journal of Experimental Psychology. Human Perception and Performance*, 7(3), 658. <https://psycnet.apa.org/record/1982-07095-001>
- Stanovich, K. E., & West, R. F. (1983). On priming by a sentence context. *Journal of Experimental Psychology. General*, 112(1), 1–36. <https://doi.org/10.1037//0096-3445.112.1.1>
- Szewczyk, J. M., & Schriefers, H. (2013). Prediction in language comprehension beyond specific words: An ERP study on sentence comprehension in Polish. *Journal of Memory and Language*, 68(4), 297–314. <https://doi.org/10.1016/j.jml.2012.12.002>
- Traxler, M. J., & Foss, D. J. (2000). Effects of sentence constraint on priming in natural language comprehension. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 26(5), 1266–1282. <https://doi.org/10.1037//0278-7393.26.5.1266>
- Trenkic, D., Mirkovic, J., & Altmann, G. T. M. (2014). Real-time grammar processing by native and non-native speakers: Constructions unique to the second language*. *Bilingualism: Language and Cognition*, 17(2), 237–257. <https://doi.org/10.1017/S1366728913000321>
- Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 31(3), 443–467. <https://doi.org/10.1037/0278-7393.31.3.443>
- Van Heugten, M., Dahan, D., Johnson, E. K., & Christophe, A. (2012). Accommodating syntactic violations during online speech perception. *Poster Presented at the 25th Annual CUNY Conference on Human Sentence Processing, May*, 14–16. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.670.6140&rep=rep1&type=pdf>
- Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, 83(2), 176–190. <https://doi.org/10.1016/j.ijpsycho.2011.09.015>
- Van Schijndel, M., & Linzen, T. (2018). Modeling garden path effects without explicit hierarchical syntax. In T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 2603–2608). Austin,

- Walker, N., Monaghan, P., Schoetensack, C. and Rebuschat, P. (2020), Distinctions in the Acquisition of Vocabulary and Grammar: An Individual Differences Approach. *Language Learning*, 70: 221-254. <https://doi.org/10.1111/lang.12395>
- Weber, K., Christiansen, M. H., Indefrey, P., & Hagoort, P. (2019). Primed From the Start: Syntactic Priming During the First Days of Language Learning. *Language Learning*, 69(1), 198–221. <https://doi.org/10.1111/lang.12327>
- Wehbe, L., Vaswani, A., Knight, K., & Mitchell, T. (2014). Aligning context-based statistical models of language with brain activity during reading. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 233–243. <https://www.aclweb.org/anthology/D14-1030.pdf>
- Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: an event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, 16(7), 1272–1288. <https://doi.org/10.1162/0898929041920487>
- Wlotko, E. W., & Federmeier, K. D. (2011). Flexible implementation of anticipatory language comprehension mechanisms. *Journal of Cognitive Neuroscience*, 23(Supplement 1), 233.