

ORIGINAL ARTICLE

Reduced false memory effects for predictable words in L2 speakers of German: evidence from self-paced reading and recognition memory

Katja I. Haeuser^{1,2}  and Theres Grüter³ 

¹Department of Psychology, Saarland University, Saarbrücken, Germany, ²Collaborative Research Center “Information Density & Linguistic Encoding” (CRC 1102), Saarland University, Saarbrücken, Germany and

³Department of Second Language Studies, University of Hawai‘i at Mānoa, Honolulu, HI, USA

Corresponding author: Katja I. Haeuser; Email: khaeuser@coli.uni-saarland.de

(Received 25 February 2025; revised 27 August 2025; accepted 3 September 2025)

Abstract

Previous research has demonstrated that predictable words that are not presented linger in memory and lead to false recognition in subsequent memory tests. However, little is known about these effects among second language learners, a population that is known for engaging less in prediction. Here, we used a self-paced reading and word recognition memory test to examine encoding differences and subsequent memory effects in groups of L1 and L2 speakers of German. For initial reading, results showed no group differences in the size of the predictability effect, possibly because group differences in attention allocation during reading masked predictability effects. For recognition memory, L2 learners showed reduced rates of false remembering for predictable words (after correcting for response bias), and they were also less likely to false-alarm to predictable words with high subjective memory confidence, similar to L1 speakers. In addition, L2 learners showed reduced recognition memory for previously presented words. Taken together, these results are consistent with models arguing that lexical-semantic entries are less firmly represented in the L2 lexicon, which in turn lowers pre-activation of predictable referents during L2 sentence processing and leads to the formation of less distinct memory representations for previously encoded information.

Keywords: Bilingualism; false memory; prediction; reading; sentence comprehension

According to a recent study, nearly half of the world’s population is bilingual (Grosjean, 2021). For example, at least 59% of all Europeans speak at least one foreign language—a number that has been steadily increasing over the past decades, according to recent polls (www.europa.eu). The large proportion of bilingual speakers across the world stands in stark contrast to the relatively small proportion of studies in cognitive psychology and psycholinguistics that have focused on how

The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

the human mind represents and processes multiple languages versus just a single one. For example, the role of prediction, that is, our ability to anticipate upcoming linguistic input before it is perceived, has been studied extensively for several decades among monolingual native (L1) adults (e.g., Kuperberg & Jäger, 2016; Ryskin & Nieuwland, 2023) but did not start to be investigated systematically among bilinguals and second language (L2) users until the past decade (e.g., Kaan, 2014; Schlenter, 2023). Similarly, there is a much larger body of research on language and memory among monolinguals than there is among bilinguals and L2 speakers. In this paper, we draw on these previous strands of research to investigate how engagement in prediction in German-speaking L1 and L2 speakers affects their recognition memory. Specifically, we investigate whether potentially reduced engagement in prediction during L2 processing may lead to reduced false memory effects for predictable words among L2 users.

Prediction in first language (L1) processing

It is well known that native speakers of a language leverage the linguistic context to anticipate upcoming linguistic input (for reviews, see Huettig, 2015; Kuperberg & Jaeger, 2016; Ferreira & Chantavarin, 2018; Ryskin & Nieuwland, 2023). In fact, there is now evidence that predictive processing occurs at all levels of linguistic representation, that is, syntactic (e.g., Staub & Clifton, 2006), semantic (e.g., Altmann & Kamide, 1999; Borovsky et al., 2012; Federmeier & Kutas, 1999; Kamide et al., 2003), and lexical (DeLong et al., 2014; Ito et al., 2016; DeLong et al., 2005). Even in the absence of strong semantically predictive information, comprehenders make use of word form information conveyed by prenominal modifiers or articles to (successfully) predict the identity of an upcoming referent (e.g., Cholewa et al., 2019; also see Boudewyn et al., 2015; DeLong et al., 2005; Fleur et al., 2020; Haeuser et al., 2022; Wicha et al., 2003; Van Berkum et al., 2005). The consequence of pre-activating or predicting words during comprehension is that predicted referents are easier to process than unpredicted ones (for reviews, see Van Petten & Luka, 2012; Staub, 2015), which could be useful to language processing by freeing up mental resources that would otherwise be bound to incremental, piece-by-piece processing of language. For example, predictable referents show shorter fixation times than unpredictable ones (e.g., Ehrlich & Rayner, 1981) and are more likely to be skipped during initial reading (e.g., Staub, 2015). In addition, predictable words show reduced N400 event-related potentials (ERP) components (e.g., Kutas & Federmeier, 2011), which suggests that less mental effort is needed to integrate predictable words into an unfolding sentence context. In sum, prediction is a critical component not only in general theories of cognition (e.g., Clark, 2013) but also in L1 language processing (e.g., Shain, 2024; Staub, 2024).

Prediction in second language (L2) processing

What about prediction in a language that is not native? An intuitive hypothesis is that L2 speakers of a language should be less likely to engage in prediction, possibly because they have less entrenched lexical representations in that language than L1

users do (e.g., Bialystok et al., 2008; Kroll & Stewart, 1994), or because resource limitations that are inherent to L2 language processing put a cap on engagement in top-down processing (e.g., Foucart et al., 2016; Grüter et al., 2017; Kaan et al., 2010). Notably, this is consistent with what much of the previous literature on prediction in L2 has found. While we know that L2 speakers can and do use linguistic and non-linguistic context to anticipate upcoming referents (for recent reviews, see Bovolenta & Marsden, 2022; Hopp, 2022; Kaan & Grüter, 2021; Schlenter, 2023), there is also evidence that L2 speakers may use linguistic cues less readily, more slowly, and/or to a different extent. For example, early work that focused on phonological and gender cues provided by definite articles found no or reduced effects of prediction (Lew-Williams & Fernald, 2010; Martin et al., 2013), even among learners with near-native proficiency (Grüter et al., 2012; but see Dussias et al., 2013; Foucart et al., 2014). While prediction at the word form level is debated on for L1 processing (e.g., DeLong et al., 2017; Nieuwland et al., 2018; Urbach et al., 2020; Yan et al., 2017), there is now evidence for differences in L2 prediction even when it comes to prediction based on semantics. For example, Dijkgraaf and colleagues (2019) used a visual-world eye-tracking paradigm to demonstrate that L1-Dutch/L2-English bilinguals were less likely to fixate competitor images that were semantically related to a highly predictable referent in their L2 than in their L1. Similarly, Ito and colleagues (2018) found that both L1 and L2 (L1 Japanese) speakers of English looked at a target referent that was highly predictable from the discourse context (e.g., cloud, following “The tourists expected rain when the sun went behind the . . .”) before that target appeared in the acoustic input; however, L2 speakers did so more slowly compared to L1 speakers. Thus, findings from a variety of methods and different levels of linguistic representation indicate that the scope of prediction is reduced or slowed down in L2 processing.

False remembering in L1

While initial studies on prediction focused on the immediate processing consequences of prediction (e.g., during sentence comprehension), recent research has begun to investigate whether prediction affects the memory representations that language users store and maintain after engaging in prediction. Notably, studies in this vein of research show that predictable words that are disconfirmed tend to “linger” or persist in memory and can lead to false remembering in downstream recognition memory tests (e.g., Chung & Federmeier, 2023; Haeuser & Kray, 2024; Haeuser & Kray, 2022; Hubbard et al., 2019; Hubbard & Federmeier, 2024; Rich & Harris, 2021; Rommers & Federmeier, 2018; also see Roediger & McDermott, 1995; Roediger et al., 2001; Chang & Brainerd, 2021). For example, a seminal study by Hubbard and colleagues (2019) showed that predictable words that were never actually encountered elicited high rates of false remembering in a subsequent recognition memory test, where comprehenders incorrectly endorsed previously predictable words (henceforth, lures) as “old” (i.e., after reading a sentence such as, “Be careful, the top of the stove is very dirty.”, comprehenders attested to having read “hot”). Notably, false-alarm rates to predictable lures in that and many other studies were much higher than false-alarm rates to “new” words, that is, words that

were not previously seen and not predictable. This illustrates that false remembering of predictable words cannot be merely attributed to response bias, that is, a general bias to endorse words as previously seen, which is a common source of distortion in recognition memory tasks (e.g., Brainerd et al., 2008; Fraundorf et al., 2019). Previous studies have also shown that false memory judgments to predictable lures are often made with high subjective confidence (i.e., “sure old” response; see Hubbard & Federmeier, 2024; Haeuser & Kray, 2024) or in conditions when participants indicate that they vividly remember having seen the predictable word earlier on (e.g., Norman & Schacter, 1997; Schacter et al., 1997). Healthy aging older adults, for example, frequently endorse lure words as “old” while indicating they are very confident of their false memory judgment (e.g., Balota et al., 1999; Dodson et al., 2007; Greene et al., 2024; Haeuser & Kray, 2024; Kensinger & Schacter, 1999; Tun et al., 1998). These findings attest to the fact that comprehenders often vividly “remember” seeing the lure earlier on, and that different participant groups may show systematic differences in the strength of the false memory illusion (e.g., Greene et al., 2024).

Interestingly, the findings from the prediction literature stand against the backdrop of a much larger research field on false remembering which has evolved around the so-called DRM paradigm (coined after seminal work by James Deese in the 1950s and Roediger & McDermott (1995), who made the paradigm popular in memory research). In the DRM paradigm, participants encode lists of content words (e.g., “bed, rest, dream, awake, pillow . . .”) that are all semantically related to a particular lure word (e.g., “sleep”). In subsequent recognition tests, participants frequently indicate that they have seen the critical lure word earlier during encoding, or they falsely recall the lure in subsequent free recall tests (for recent reviews on DRM-based false remembering, see Gallo, 2010; Chang & Brainerd, 2021; for cognitive theories on false memory effects, see Brainerd et al., 2008; Howe, 2008; Howe et al., 2009; Holliday & Weekes, 2006; Reyna & Brainerd, 2016; Roediger et al., 2001; Sommers & Lewis, 1999). Mirroring the findings from the prediction literature, false memory judgments to DRM lures are frequently made with high subjective confidence (e.g., Norman & Schacter, 1997; Roediger & McDermott, 1995, Experiment 2; Tun et al., 1998; see also Greene et al., 2024), and their frequency of occurrence goes above and beyond effects that could be explained by mere response bias alone (Roediger & McDermott, 1995; Brainerd et al., 2008; Chang & Brainerd, 2021; Gallo, 2010). The striking difference between DRM-based and prediction-based false remembering is that, for DRM lists, lure words are entrenched by means of repeated presentation of semantic associates, whereas for predictable words in sentences, participants really only get one “shot” at the critical lure. Despite this difference, very recent research indicates that prediction-based false memory can be explained by means of similar, if not the same, cognitive processes that come to bear during false remembering of DRM words, namely activation of critical lures during encoding and monitoring failures during memory retrieval (e.g., Haeuser & Kray, 2024; for activation-monitoring theories of DRM-based false remembering, see Hicks & Hancock, 2002; Howe, 2008; Howe et al., 2009; Roediger et al., 2001).

False memory in L2

In light of the consistently observed attenuation in L2 speakers' engagement in prediction, an obvious hypothesis for the outcome of a false memory experiment on predictable words is that L2 speakers should show reduced rates of false remembering compared to L1 speakers. We know of only one previous study that has investigated prediction-based false remembering from an L2 perspective before (Foucart et al., 2016). However, and in contrast to the hypothesis, that study found relatively equivalent rates of false remembering in L2 speakers. In that study, late French-Spanish bilinguals ($n = 22$) listened to constraining sentences in Spanish (e.g., “El pirata tenía el mapa secreto, pero nunca encontró . . .”, “The pirate had the secret map, but he never found . . .”) which continued with either the predictable (e.g., “el tesoro”, “the treasure”) or an unpredictable noun (e.g., “la gruta”, “the cave”) and later completed a recognition memory test. Importantly, during the listening phase, the critical noun was always completely muted so that only the prenominal gender-marked article indicated whether the upcoming noun was the predictable one or not (Lemhöfer & Broersma, 2012). ERPs recorded during the listening phase showed a more negative deflection for articles that were inconsistent with the gender of the predictable noun, suggesting that listeners showed early brain responses when predictions were not met. Results from the “old”/“new” recognition memory test showed that predictable nouns were false-alarmed to more often when participants had previously heard the sentence with the prediction-consistent gender-marked article (e.g., “encontró el” vs “encontró la”), with mean false-alarm rates of 49% versus 43%, respectively (note that differences in response bias were not taken into account in that study). As these findings for L2 speakers were very similar to those from an earlier study with L1 speakers of Spanish using the same materials (Foucart et al., 2015), the authors concluded that late bilinguals, much like native speakers, use anticipation processes during L2 comprehension that lead to the creation of memory traces for expected but not-encountered words. In other words, prediction elicits false remembering in L2 as much as it does in L1, contrary to the hypothesis stated above.

Interestingly, the results from the Foucart et al. (2016) study stand in stark contrast to previous DRM research on bilingualism, which has consistently found that L2 speakers show reduced, and not equivalent, rates of false remembering, compared to L1 speakers (see e.g., Anastasi et al., 2005; Arndt & Beato, 2017; Bialystok et al., 2020; Sahlin et al., 2005; Suarez & Beato, 2023; for mini-review, see Suarez & Beato, 2021). For example, Arndt & Beato (2017) used a within-subject design on L1-English/L2-Spanish speakers to study false recognition memory of DRM lures when using English and Spanish word lists. All L2 groups had relatively low proficiency in their second language. Across three experiments, the results showed that false recognition memory effects were lower when participants were tested in their L2 compared to their L1. The authors concluded that processing word lists in an L2 seems to create less robust false memory representations than processing similar word lists in an L1.¹

Past studies have offered two critical explanations for the reduced false memory effect among L2 speakers. One explanation holds that L2 participants show overall reduced-activation levels when processing their L2, owing to the fact that their L2

lexical network is less interconnected and/or less entrenched (e.g., Arndt & Beato, 2017; Anastasi et al., 2005; Futrell et al., 2020; Pawley & Syder, 1983; for reviews, see Suarez & Beato, 2021; Beato et al., 2023; for compatible theories on L1-based false remembering, see Howe, 2008; Roediger et al., 2001). Qualitatively, this account is very similar to accounts in the L2 literature that have attributed reduced engagement in prediction to reduced levels of lexical-semantic activation and/or spreading within the L2 lexicon (e.g., Ito & Pickering, 2021). Therefore, we will subsume both accounts into one and refer to them as *reduced-activation* accounts. The *reduced-activation* account was originally brought forward by empirical studies reporting on L2-based reductions in false memory (Suarez & Beato, 2021; Beato et al., 2023). However, this account also directly falls out of theories arguing that L2 speakers predict less when processing their L2 (e.g., Bovolenta & Marsden, 2022; Hopp, 2022; Kaan & Grüter, 2021; Schlenter, 2023). Both lines of research suggest that L2 speakers activate lexical representations less distinctly during initial processing, which makes it less likely for them to predict or pre-activate critical lures during initial encoding. As such, L2 speakers should also be less likely to experience interference from lures in the memory test (e.g., what the DRM literature has referred to as “monitoring failures” where residue activation for critical lures is mistaken for genuine activation; see Roediger et al., 2001; Howe, 2008). Consequently, *reduced-activation* accounts predict that L2 speakers should show reduced rates of false remembering overall.

The second explanation for reduced DRM effects in L2 holds that, due to L2 speakers being more experienced at juggling multiple activated representations at a time because of speaking two languages, they should be at an advantage when it comes to correctly suppressing activation from critical lures during encoding and/or memory retrieval (e.g., *The word “sleep” looks familiar but I know I saw “bed”, not “sleep”*). We will refer to this account as the *bilingual-advantage* account (e.g., Bialystok et al., 2020; Bialystok et al., 2008; Morales et al., 2013). The *bilingual-advantage* account was brought forward in an empirical study by Bialystok and colleagues (2020) who demonstrated that L2 speakers show reduced false memory effects for lure words in the DRM paradigm, when compared to L1 speakers (see Experiments 2 and 3 of that paper). According to the authors, L2 speakers are better at avoiding false alarms to lures due to their improved abilities in selective attention and executive control, which should arise from L2 speakers’ continuous experience at selecting from activated representations in the two languages (Bialystok et al., 2020). Of note, the *bilingual-advantage* account predicts that L1–L2 differences could arise both at encoding and retrieval. During both processing stages, L2 speakers are better at suppressing activation from co-activated lures.

The *reduced-activation* and *bilingual-advantage* accounts have in common that they both predict reduced false memory effects in L2 speakers, although they differ in the hypothesized cognitive mechanisms behind these (reduced) effects. The two accounts make opposite predictions, however, for recognition memory of previously seen, old, words: If reduced rates of false remembering are a consequence of L2 speakers activating lexical representations less distinctly (i.e., *reduced-activation* account), then L2 speakers should also show reductions in correct recognition memory for old words (because old words, too, should have been activated less distinctly). If, on the other hand, reduced false memory rates in L2 result from L2

speakers being experienced at suppressing unwanted activation from lures, then they should show L1-equivalent rates of correct recognition memory for old words (see Bialystok et al., 2020).

The present study

The goal of the present study is (i) to test the central prediction of both the *reduced-activation* and the *bilingual-advantage* accounts that false memory effects will be reduced among L2 compared to L1 speakers, and (ii) to distinguish between the two accounts by probing for memory of both critical lures and previously seen old words. We do so using an experimental paradigm from previous research with L1 speakers (e.g., Haeuser & Kray, 2024; Haeuser & Kray, 2022), including a self-paced reading (SPR) task followed by a recognition memory test, with L1 and advanced L2 speakers of German. During SPR, participants read German sentences constructed to constrain their expectations toward a predictable noun, with half of the critical sentences continuing with the predictable noun, and the other half continuing with an unpredictable but semantically plausible noun. After a brief distractor task, participants completed a surprise (i.e., previously unannounced) recognition memory test, in which they were presented with previously seen, “old” words (i.e., previously seen predictable and unpredictable words), “lures” (i.e., predictable but not actually presented words), as well as “new” words, and had to indicate whether the presented words were “old” or “new”. In order to investigate vivid recollection versus mere familiarity with an item, participants were also asked to indicate how confident they were in their response (i.e., sure vs maybe old and new; for rationale, see Haeuser & Kray, 2024).

Hypotheses and predictions

For SPR, both the *reduced-activation* and the *bilingual-advantage* accounts would expect reduced predictability effects in L2, albeit for different reasons. According to the *reduced-activation* account, L2 speakers are less likely to engage in prediction when processing their second language, so they should show less facilitation for predictable words and less processing difficulty for unpredictable words. The *bilingual-advantage* account would also expect a reduced predictability effect in L2, because L2 speakers should be better at suppressing interfering activation from predictable words when reading unpredictable sentences.

For the recognition memory test, both accounts expect reduced rates of false remembering for predictable lures in L2 compared to L1 speakers, over and above possible group-wise differences in response bias (i.e., false alarms to “new” words), and possibly, differences between L1 and L2 regarding their subjective memory confidence (e.g., Dodson et al., 2007; Greene et al., 2024; Haeuser & Kray, 2024; Hubbard & Federmeier, 2024; Norman & Schacter, 1997; also see Pawley & Syder, 1983). However, diverging predictions emerge from the two accounts when it comes to correct recognition memory for previously seen, “old” words (henceforth, *true* recognition memory; see Roediger et al., 2001; Roediger & McDermott, 1995; Chang & Brainerd, 2021; Gallo, 2010). The *reduced-activation* account predicts lower rates of true recognition memory for L2 compared to L1 speakers, as L2 speakers would activate lexical representations for “old” words less distinctly during initial

Table 1. Background information on the L1 and L2 speaker group

	L1 (<i>n</i> = 44)		L2 (<i>n</i> = 47)	
	Mean	Range	Mean	Range
AoA German (in years)	0.6	0–4	17	5–34
LexTale German Score	88	47–98	68	49–98
Immersion duration (in years)	–	–	4	0–21
Current usage of German (1–100%)	–	–	73.67	15–95
Self-rated overall proficiency (1–10)	–	–	7.60	4–10
Self-rated reading proficiency (1–10)	–	–	8.40	6–10
Self-rated writing proficiency (1–10)	–	–	7.30	5–10
Self-rated speaking proficiency (1–10)	–	–	7.13	2–10

Note. We define age of acquisition (AoA) as the age at which participants reported to first learning German.

encoding. In contrast, the *bilingual-advantage* account does not predict any differences between L1 and L2 speakers for true recognition memory.

Method

Participants

A total of 91 L1 (*n* = 44, 25m, 19f, 0nb, *Mean age* = 27 years) and L2 (*n* = 47, 55m, 22f, 0nb, *Mean age* = 25 years) speakers of German was recruited through social media and flyer advertisement. The study was run online (see below for details) in the time period from July until October 2023. L1 speakers were recruited from the first author’s institution and reported learning German from birth and using German as their primary language in everyday life.² The L2 group consisted of relatively advanced L2 learners of German, who were foreign-language students with a diverse set of L1s at universities in Germany (L1s: Arabic (5), English (4), Chinese (2), French (5), Hindi (7), Indonesian (2), Italian (2), Japanese (1), Kinyarwanda (1), Malayalam (1), Persian (3), Polish (1), Portuguese (1), Russian (1), Serbian (1), Spanish (1), Turkish (1), Telugu (1), Ukranian (2), Hungarian (1), and Vietnamese (1)). Table 1 summarizes relevant background information on the L1 and L2 groups, including self-reported age of acquisition (AoA), immersion duration, usage of German, self-ratings of overall and subskills in German on a 10-point scale, as well as scores from the LexTALE German test as an independent measure of German proficiency (Lemhöfer & Broersma, 2012). None of the participants reported a history of reading or writing disabilities and/or taking neuropsychiatric medication at the time of testing.

Materials

Materials for self-paced reading

All materials are included on this paper’s OSF project page under the link <https://osf.io/e56k4/>. Materials for the SPR task consisted of 32 constraining German

Table 2. Descriptive means (and SD) of properties of critical nouns in the self-paced reading task

Condition	Length (<i>N</i> characters)	Frequency (SUBTLEX-DE)	Cloze probability L1 sample	Cloze probability L2 sample
Predictable	5.94 (2.09)	2.84 (0.64)	0.94 (.01)	0.85 (.12)
Unpredictable	6.56 (2.00)	2.52 (0.64)	< .001	< .001

Note. Frequency estimates are log-per million values from the movie subtitle corpus (Brysbaert et al., 2011).

sentence frames (e.g., *Im Krankenhaus vernäht die Stationsschwester mit schneller Hand die ...*, “At the hospital, the nurse stitches up quickly the ...”), which continued in a relatively predictable (e.g., *Wunde*, “injury”) and relatively less predictable (but somewhat plausible) noun (e.g., *Hose*, “pants”), yielding a total of $n = 64$ experimental sentences. Predictable and unpredictable nouns were matched with respect to length, $t(62) = -1.23$, $p = .22$, and log-per-million word frequency, $t(60) = 1.94$, $p = .06$ (frequency values were taken from the movie subtitle database SUBTLEX-DE; see Brysbaert et al., 2011).

A cloze probability pretest conducted with 33 German L1 speakers (*Mean age* = 28 years) and 17 L2 learners of German (*Mean age* = 29 years, *AoA range* = 3–23), who did not participate in the main experiment, had shown relatively high cloze probabilities for predictable nouns and near-zero cloze probabilities for unpredictable nouns (see Table 2). On average, cloze probabilities for predictable nouns were lower among L2 than L1 speakers ($M = 0.85$, *range* = 0.56–1.00, and $M = 0.94$, *range* = 0.73–1.00, respectively), a significant difference, $t(62) = -3.24$, $p < .01$. Table 2 shows descriptive statistics and by-group cloze probabilities for predictable and unpredictable nouns used in the experiment.

All predictable and unpredictable sentences contained two- or three-word sentence continuations after the noun that continued the sentence in a plausible fashion (e.g., “in the room next door”), in order to allow for spill-over effects, which frequently occur in self-paced sentence reading. Note that the spill-over region was identical for predictable and unpredictable versions of an item. The experimental sentences were distributed across two lists using a Latin square design, such that one participant was presented either with the predictable or unpredictable version of an item during the experiment, but not both (i.e., a total of 16 predictable and 16 unpredictable items per subject). Each list additionally contained 25 moderately constraining filler sentences (taken from the Potsdam sentence corpus; see Kliegl et al., 2006), which were included to ensure that participants continued to generate predictions in the course of the experiment (despite frequently having predictions disconfirmed; see e.g., Delaney-Busch et al., 2019; Fine et al., 2013). All experimental sentences were followed by yes/no comprehension questions, inserted to ensure that participants were reading the experimental sentences for comprehension. Comprehension questions were identical for predictable and unpredictable versions of an item and mostly probed information provided in the sentence context.

Materials for word recognition memory

Materials for the recognition memory task consisted of 64 “old” nouns (i.e., 32 previously seen predictable and 32 unpredictable nouns in SPR), 40 “new” nouns (i.e., nouns not previously seen in SPR), and 32 “lure” nouns (i.e., nouns previously predicted but not seen; for recognition materials, see <https://osf.io/e56k4/>). A single participant saw a total of 88 nouns in the recognition task (i.e., 32 previously seen, old nouns (16 predictable, 16 unpredictable), 40 new nouns, and 16 lures). “New” nouns were obtained from a subset of German nouns listed in the movie subtitle database (SUBTLEX-DE; Brysbaert et al., 2011) that, in a previous step, had been cleared of all nouns used in the experiment. From that list, the experimenters chose “new” nouns that roughly fell in the same length and frequency range as the “old” and “lure” nouns. As a result, old-predictable, old-unpredictable, new and lure nouns in the experiment were matched both with respect to word length ($F(3, 132) = 1.69, p = .17$) and frequency ($F(3, 129) = 2.46, p = .07$). Table 3 shows descriptive statistics for words used in the recognition task.

All recognition nouns were presented on two experimental lists, with list composition dependent on list administration in the SPR task: Participants who saw the unpredictable version of an experimental sentence during SPR were presented with the previously seen, “old” unpredictable noun, as well as the potentially predictable “lure” in the recognition task. Likewise, participants who saw the predictable version of a sentence in the SPR task were presented with the previously seen “old” predictable word during the recognition task (but not the lure).

Procedure

The study was conducted online using the experimental platform LabVanced (Finger et al., 2017), a JavaScript web application that allows for online implementation of behavioral research. Participation was only possible through a PC (Windows or Linux), Mac, or tablet; cell phone or iPhone participation was disabled. The study consisted of five major sections. The first section was a brief questionnaire (~10 minutes) in which participants reported on their language background. The language background questionnaire was followed by a non-cumulative word-by-word SPR task (~15–20 minutes). Following the SPR task, there was a brief retention interval (~five minutes), in which participants solved simple math problems. The math task served the primary purpose of clearing participants’ short-term memory before they proceeded to the noun recognition memory task. The fourth section of the experiment was the noun recognition task (~15 minutes), which was followed by the German LexTale (Lemhöfer & Broersma, 2012). All participants completed the study in the specified order.

In the SPR task, participants read the experimental sentences on a screen word by word (each word was only displayed once and was not replaced by dashes later on, that is, there was no mask, no moving-window reading). Each word was presented in the center of a white screen using Lucinda 18pt font and stayed on the screen until participants pushed the space bar, which revealed the next word in the sentence. Participants were instructed to read the sentences as quickly and thoroughly as

Table 3. Descriptive means (and SD) of properties of nouns used in the recognition task

Condition	Length (N characters)	Log-frequency (SUBTLEX-DE)
Old-predictable	5.94 (2.09)	2.84 (0.64)
Old-unpredictable	6.56 (2.00)	2.52 (0.64)
New	5.60 (1.08)	2.64 (0.29)
Lure	5.94 (2.09)	2.84 (0.64)

possible and to answer all comprehension questions as accurately as possible by pressing the “S” (Yes, correct) and “L” (No, incorrect) keys on the keyboard. The experimental task was preceded by ten practice sentences to allow participants to get used to the word-by-word reading task.

In the (surprise) noun recognition memory task, participants saw single nouns appear on the screen in a central position using Lucinda 18pt font, separated by a central fixation cross displayed for 5000 ms. Participants were instructed to hover their middle and index fingers of each hand over the S, D and J, K keys on the keyboard. They were instructed to press the S or D key for all “new” nouns, and the J or K key for all “old” nouns, and to additionally indicate their confidence with their recognition rating (*D* = Maybe New, *S* = Sure New, *J* = Maybe Old, *K* = Sure Old). The confidence judgments were included in order to obtain a subjective estimate of how accurately participants judge the veracity of their memory judgments (e.g., Greene et al., 2024). For example, previous research on L1 processing had shown that, counterintuitively, memory intrusions are often committed with high confidence, suggesting that participants are misguided in evaluating the strengths or weaknesses of their memories (e.g., Greene et al., 2024; Haeuser & Kray, 2024).

Participants were instructed that they did not need to memorize the key combinations, because every trial contained a schematic display of the response options (SD, JK), and what they represented, at the bottom of the screen. Participants were instructed to respond to each word as quickly and accurately as possible. About two-thirds of all participants completed the experiment within 30–70 minutes (average finish time = 62 minutes). We did not exclude participants based on longer experimental durations. On average, the L2 speakers took longer than the L1 speakers ($M = 73$ minutes vs $M = 51$ minutes, respectively).

Data analysis

All analysis scripts and data files are available on OSF (<https://osf.io/e56k4/>). We report the results for the word-by-word SPR task and subsequent word recognition task in two separate sections. To statistically analyze the reading and main recognition data, we ran linear mixed effects models and generalized linear mixed models (LMERs and GLMERs; Baayen et al., 2008) specified in the lme4 library (Bates et al., 2015) in R (R Core Team 2022; version 4.1.3), which were run on unaggregated trial-by-trial data. All models included random intercepts for subjects and items and were initially fit with the fullest by-item and by-subject random slope

structure warranted by the design (i.e., “maximal” model including all within-subject and within-item predictors, including their interactions whenever specified; Barr et al., 2013). In the case of convergence warnings, models were simplified progressively using the *least variance* approach (e.g., Barr et al., 2013). *P*-values were calculated using the R package *lmerTest* (Kuznetsova et al., 2017).

Results

Self-paced reading

One participant from the L2 group had an accuracy rate below 70% on the comprehension questions and was excluded from further analysis. On average, the remaining participants responded to the comprehension questions very accurately ($M = 0.90$, $range = 0.72-1$), with L2 speakers ($M = 0.88$, $range = 0.72-1$) responding slightly less accurately than L1 speakers ($M = 0.91$, $range = 0.81-0.97$), a significant difference, $t(88) = 2.68$, $p < .01$. Thus, even though participants in both speaker groups showed good comprehension of the sentences overall, L2 speakers seemed to encode fewer details from the sentences into short-term memory.

The critical region for statistical analysis of the RT data consisted of the predictable and unpredictable noun, as well as the two words immediately following the noun to catch spill-over effects common in word-by-word SPR (i.e., the noun, the first word after the noun, $N + 1$, and the second word after the noun, $N + 2$). We fit separate models for each word in the critical region. The outcome variable was reading times, log-transformed to avoid skewness. Fixed effects were predictability (a categorical variable with two levels: predictable vs unpredictable, sum-coded with -1 and 1 to allow for interpretation as main effect; cf. Brehm & Alday, 2022; Schad et al., 2020)³, and speaker group (two levels: L1 and L2, sum-coded with -1 and 1), as well as the interaction between predictability and speaker group. Each model additionally included control predictors for trial number (to offset effects of customization), word position in the sentence, word frequency (i.e., log-per-million frequency values from the SUBTLEX-DE database; Brysbaert et al., 2011), and word length. We did not include a control predictor for scaled reading times of the previous word, as the inclusion of this predictor led to multicollinearity with *predictability* in models Noun+1 and Noun+2. All control predictors were scaled. All models were initially fit with the maximal random slope structure warranted by the design (i.e., by-subject random slopes for group, and by-item random slopes for predictability and group). Models that did not converge were simplified in their random slope structure using the *least variance* approach.

Based on visual inspection of the reading time data, we excluded data points in the L1 group that were faster than 100 ms and slower than 1500 ms (for the Noun and Noun+1 region), and slower than 2000 ms (for the Noun+2 region). For the L2 group, also based on visual inspection of the data, we excluded data points that were faster than 100 ms and slower than 2000 ms across regions. This procedure affected less than 2.7% of all data points in the L2 group, and less than 2.4% of all data points in the L1 group, justifiable rates according to Ratcliff (1993).⁴ Figure 1 shows a partial effects plot of reading times in the critical region.

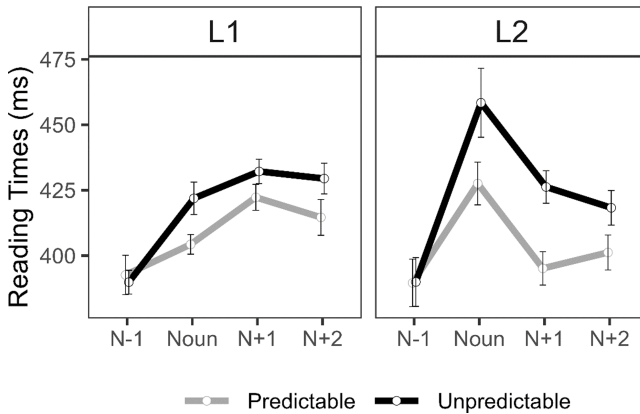


Figure 1. Reading times in the self-paced reading task. Note. Error bars reflect SE.

Noun. We found a main effect of predictability ($b = 0.02$, $SE = 0.01$, $t = 2.37$, $p = .02$), indicating that across groups, unpredictable nouns elicited processing difficulty, as well as a main effect of group, suggesting that L2 speakers read more slowly than L1 speakers overall ($b = 0.08$, $SE = 0.03$, $t = 2.38$, $p = .02$). The interaction between predictability and group did not reach statistical significance ($t = 1.28$, $p = .20$).

Noun+1. There was a main effect of predictability ($b = 0.02$, $SE = 0.01$, $t = 2.97$, $p = .01$), again indicating slowed reading of unpredictable items. The main effect of group was not significant ($t = 1.31$, $p = .20$), nor was the interaction between predictability and group ($t = 1.60$, $p = .11$).

Noun+2. There was a main effect of predictability ($b = 0.02$, $SE = 0.01$, $t = 2.29$, $p = .03$). Much like in the *Noun+1* model, the main effect of group was not significant ($t = 1.53$, $p = .13$), nor was the interaction between predictability and group ($t = -0.004$, $p = .43$).

In sum, two key findings emerged from the analysis of the SPR data: First, L2 speakers read more slowly compared to L1 speakers overall, an effect that was numerically present for all words in the critical region but reached significance only on the noun. Second, unpredictable words led to slower reading times overall, suggesting that participants in both groups experienced integration difficulties when reading sentence continuations that were unpredictable based on context. Critically, we observed no significant interactions between predictability and group. Thus, contrary to the predictions that derive from both the *reduced-activation* and the *bilingual-advantage* accounts, we obtained no evidence from the SPR task that predictability effects are reduced in L2 compared to L1 speakers.⁵

Noun recognition memory

Two participants (one from the L1 and one from the L2 group) were excluded from all following analyses because their hit rate was lower than their false-alarm rate to new items.

Table 4. Average hit and false alarm rates by group in the recognition memory task

	Hits	False alarms	d'
L1 speakers	0.74	0.18	1.70
L2 speakers	0.75	0.28	1.38
Average	0.75	0.23	1.54

We first examined whether L1 and L2 speakers showed systematic differences with respect to response bias. To do this, we computed hits and false alarms, as well as the signal detection measure d' for participants in both groups (d prime). d' is a measure for recognition discrimination that takes into hits and false alarms (e.g., Fraundorf et al., 2019; see Stanislaw & Todorov, 1999). It is calculated as a difference score between z-transformed hit and false-alarm rates (e.g., using the R-formula $qnorm(HR) - qnorm(FAR)$).⁶ Results are presented in Table 4. L2 speakers showed elevated false-alarm rates to new words, $t(85) = -3.35$, $p < .001$, in other words, there was evidence for elevated response bias among L2 speakers. Hit rates to old items did not differ between L1 and L2 speakers, $t(85) = -0.53$, $p = .60$. Altogether, this led to a significantly lower d' in the L2 compared to the L1 group, $t(85) = 2.92$, $p < .01$, suggesting that L2 speakers discriminated less successfully between old and new items.

The between-group difference for false alarms indicates that we cannot directly compare absolute (i.e., uncorrected) rates of true and false memory (to lures) in L1 and L2 speakers, as these rates are systematically distorted (i.e., shifted upward) in L2 speakers as a function of their response bias. As we describe below, our main analysis takes this response bias into account by using effect coding, as it computes recognition memory *effects* that are corrected for response bias.

To statistically analyze recognition memory rates, we used GLMER models using the Bobyqa optimizer, to facilitate convergence. Since there was no difference in hit rates between old predictable and unpredictable words, we collapsed all hits to predictable and unpredictable words into a single “old” category for the statistical analysis. Prior to analysis, recognition data were cleaned minimally by removing, in both groups, RT outliers that were faster than 200 ms and slower than 10000 ms, a procedure that affected less than 1% of all data points.⁷

The outcome variable was trial-by-trial “old” judgments, that is, whether a word was judged as “old” or “new” (coded in 1’s and 0’s). For previously seen (“old”) nouns, old judgments reflect hits. For lures and new items, old judgments reflect false alarms (i.e., participants incorrectly endorsing a word as old when it was, in fact, new). Fixed effects in the model were *word type* (three levels: old, new, and lure; see below for contrast coding), confidence (low vs high confidence, sum-coded with -1 and 1), and group (L1 vs L2, sum coded with -1 and 1 for L1 and L2 speakers, respectively), as well as all two- and three-way interactions between them. Covariates included word length, frequency (log-per-million values from the SUBTLEX-DE database; see Brysbaert et al., 2011), trial number, and recognition response time. All covariates were scaled to facilitate direct comparison between their beta estimates.

For the predictor *word type*, we set up two orthogonal (i.e., statistically independent; Schad et al., 2020) contrasts that take into account response bias. The two contrasts of interest for this study were a) lure vs new (false memory effect), and b) old vs new (true memory effect); we explain these below.

The first contrast *lure vs new* measured the critical false memory effect by comparing false alarms for lures against false alarms to new items. Note that since false alarms to lures are compared to those of new words, this effect is corrected for response bias. If predictable words elicit false recognition memory, false-alarm rates to lures should be consistently higher than those to new words. According to both the *reduced-activation* and *bilingual-advantage* accounts, we expected this effect to interact with speaker group, such that the false memory effect should be reduced in L2 compared to L1 speakers. We also expected the memory effect to interact with confidence, such that L2 speakers should be less likely to express high confidence when false alarming to predictable lures.

The second contrast *old vs new* measured hit rates to old items by taking into account the difference between old and new nouns. We will refer to this contrast as the “true” memory effect—that is, hit responses corrected for response bias. We implemented this contrast to help distinguish between the two accounts of reduced false memory effects in L2 speakers. If reduced false remembering for lures among L2 participants result from overall lower activation when processing words in a second language (i.e., *reduced-activation* account), then L2 speakers should also show reduced rates of true remembering. If, on the other hand, L2 speakers show a reduced false memory effect specifically because they are more experienced at suppressing activation from irrelevant representations (i.e., *bilingual-advantage* account), then there should be no difference between L1 and L2 speakers with respect to true recognition memory.

Results. The final model converged with by-subject and by-noun random intercepts, as well as by-noun random slopes for speaker group, and by-subject random slopes for condition and confidence. Figure 2 shows proportions of items judged as old in L1 and L2 speakers, split up by condition and confidence.

For the contrast *lure vs new* (i.e., false memory effect), the model showed a significant main effect ($b = 1.19$, $SE = 0.18$, $z = 6.71$, $p < .001$), indicating an overall false memory effect (i.e., elevated false-alarm rates to lure compared to new items). There was also a significant two-way interaction between *lure vs new* and group ($b = -0.36$, $SE = 0.11$, $z = -3.27$, $p = .001$), indicating a reduced false memory effect among L2 compared to L1 speakers, consistent with the predictions of both accounts under consideration and in line with previous false memory studies (reviewed in Suarez & Beato, 2021). Irrespective of speaker group, the false memory effect was larger in high- compared to low-confidence judgments, indicated by a significant interaction between *lure vs new* and confidence ($b = 0.51$, $SE = 0.08$, $z = 6.05$, $p < .001$), suggesting that both speaker groups misjudged their memory contents when false-alarming to lures. Notably, the three-way interaction between *lure vs new*, group and confidence reached statistical significance ($b = -0.17$, $SE = 0.08$, $z = -2.10$, $p = .04$). Model splits by speaker groups indicated that the interaction between *lure vs new* and confidence was numerically smaller among L2 ($b = 0.31$, $SE = 0.11$, $z = 2.89$, $p < .004$) compared to L1 speakers ($b = 0.66$, $SE = 0.13$, $z = 5.24$, $p < .001$). This means that L2 speakers were less likely to

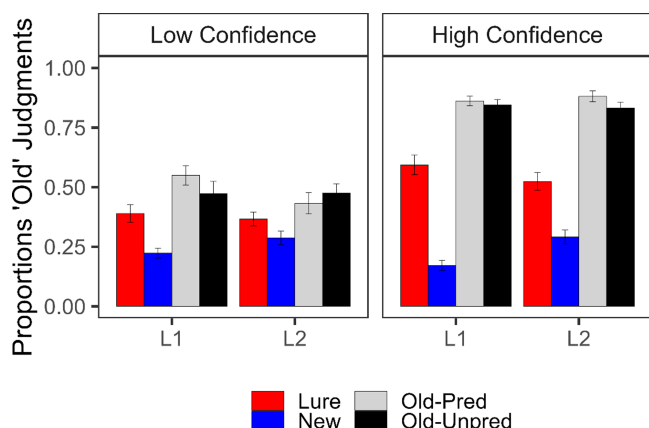


Figure 2. Recognition memory test results. *Note.* Error bars reflect SE. Lure = previously predictable but not presented noun. New = genuinely new noun. Old = previously presented predictable and unpredictable nouns.

assign high subjective confidence to their memory judgments when false alarming to predictable lures.

In sum, there were two key findings for false memory: First, and in line with the *reduced-activation* and *bilingual-advantage* accounts, L2 speakers showed a reduced false memory effect compared to L1 speakers. Second, L2 speakers were less likely to false-alarm to critical lures with high subjective memory confidence, which suggests that predictable lures elicited less vivid remembering in L2 vs L1 speakers.

For the contrast *old vs new* (i.e., true memory effect), there was a significant main effect for *old vs new* ($b = 2.53$, $SE = 0.17$, $z = 14.57$, $p < .001$; see Figure 2), indicating that both groups discriminated consistently between old and new items. However, in line with the *reduced-activation* account, we also found a significant interaction between *old vs new* and group ($b = -0.28$, $SE = 0.11$, $z = -2.55$, $p = .01$), suggesting that L2 speakers showed reduced true memory rates compared to L1 speakers (see Figure 2). This finding supports the hypothesis that reduced false memory rates in L2 result from overall reduced or less distinct activation of lexical representations when processing a second language, and not from greater experience in suppressing activation from unwanted lexical representations as a consequence of bilingualism. There was also a significant, positive-going, interaction between *old vs new* and confidence ($b = 1.13$, $SE = 0.08$, $z = 14.06$, $p < .001$), indicating that, regardless of speaker group, participants showed a larger true memory effect (i.e., better old-new discrimination) in high- compared to low-confidence judgments. The three-way interaction between *old vs new*, group, and confidence was not significant ($b = -1.0$, $SE = 0.08$, $z = -1.23$, $p = .22$).

In sum, there were two key findings for true memory: First, much like false recognition memory, true recognition memory was reduced among L2 speakers, aligning with the prediction of the *reduced-activation* account, which holds that L2 speakers activate lexical representations less distinctly when processing their L2. This overall reduced lexical activation results in reduced rates of false remembering

of predictable lures, but also in lower recognition rates for previously seen, “old” words. Second, irrespective of speaker group, recognition memory for previously seen, “old”, words was elevated in high confidence judgments, suggesting that when remembering previously presented information, participants in both speaker groups were mostly accurate in judging the veracity of their memory contents.⁸

Discussion

We tested L1 and L2 speakers of German in a SPR and recognition memory paradigm to investigate false remembering and its cognitive underpinnings in L2 processing. According to the *reduced-activation* account (e.g., Arndt & Beato, 2017; Anastasi et al., 2005; Howe et al., 2009; Suarez & Beato, 2021), L2 speakers should show reduced false memory effects for predictable words, as they (pre-)activate or predict lexical representations during encoding less distinctly than L1 speakers. According to the *bilingual-advantage* account, reduced false memory effects in L2 arise because L2 speakers are more experienced at suppressing unwanted activation from interfering lexical representations (i.e., predictable words) when comprehending unpredictable sentences. While these two accounts make similar predictions for *false* recognition memory (i.e., reduced false memory effects for L2), and for initial reading of predictable and unpredictable sentences (i.e., a reduced predictability effect for L2), they diverge in their predictions regarding *true* memory for previously seen, “old”, words. While the *reduced-activation* account predicts *reduced* rates of true remembering in L2 compared to L1 speakers (because L2 speakers activate lexical representations less distinctly and this should also pertain to previously seen, “old” words), the *bilingual-advantage* account predicts no particular L1/L2 differences when it comes to true memory.

We found that during initial reading, both L1 and L2 speakers slowed down when reading unpredictable sentence continuations, but against the predictions of both accounts, we did not obtain evidence for a reduced predictability effect in L2 speakers. In the recognition memory test, L2 speakers showed a reduced false memory effect compared to L1 speakers (i.e., lower rates of false remembering of predictable compared to new words, corrected for response bias), consistent with both the *reduced-activation* and *bilingual-advantage* accounts. L2 speakers also showed reduced rates of *true* recognition memory (i.e., lower hit rates to old words when corrected for response bias to new words). This latter findings is only consistent with the *reduced-activation* account, not the *bilingual-advantage* account. In addition, we found that L2 speakers were less likely to false alarm to predictable lures with high subjective memory confidence, suggesting that L2 speakers remembered predictable lures less vividly than L1 speakers. We discuss our findings for SPR and recognition memory below.

Prediction during reading in L1 and L2

In line with a large body of previous research, we found that L2 speakers read sentences more slowly than L1 speakers during SPR (e.g., Conklin et al., 2020; Dirix et al., 2020; Siyanova-Chanturia et al., 2011). In addition, we found a predictability

effect during initial reading of sentences, that is, slower reading times for unpredictable compared to predictable sentence continuations, irrespective of group. As is common in SPR, the predictability effect spilled over onto the following words of the sentence. However, and in contrast to the predictions that derive from both the *reduced-activation* and the *bilingual-advantage* accounts, we did not find evidence for a reduced predictability effect in L2 speakers. This null effect is somewhat surprising, given the relatively large body of research attesting to reduced effects of prediction in L2 speakers (e.g., Schlenter, 2023; Kaan & Grüter, 2021). An obvious question that arises is whether we had enough power to obtain such an effect in the first place. We followed up on this question by conducting a power analysis that started with a set of comparable L1 data (Haeuser & Kray, 2024) and simulated new data for a probable set of L2 speakers, assuming a higher intercept (i.e., slower reading times) and a reduced predictability effect in L2. This power analysis showed that with a total sample size of 91 subjects and 32 items, the present study was sufficiently powered to detect a reduced predictability effect in L2 that is as small as two-thirds of the effect in L1 speakers (we note here that the previous literature has often shown L1–L2 differences in reading speed that are much larger than that; e.g., Conklin et al., 2020; Siyanova-Chanturia et al., 2011).

We believe a more likely reason for the lack of a predictability by group interaction in our data—though speculative in nature—may lie in the potentially distributed nature of between-group differences in L1 and L2 speakers. This distributed nature could not be captured by the analysis of reading times on individual segments. Consider the pattern of reading times illustrated in Figure 1: In the L2 group, there is a clear peak at the noun, indicating a focal integration effort once the predictable/unpredictable noun itself is encountered. By contrast, the reading-time curve in the L1 group appears flatter and more “smeared out,” with no distinct peak on the noun itself. Hence, whereas L2 speakers showed a relatively focal integration effort, L1 speakers seemed to distribute their reading effort across multiple words in the critical region. The reasons for these group differences are unknown, but one speculative explanation could be that L2 readers are less able to hold verbal information in short-term memory, which may put more pressure on them to integrate new information immediately as it comes along, resulting in a costly use of resources at a single point. L1 speakers, by contrast, may be able to distribute attention and processing efforts more efficiently by holding more information in short-term memory, showing no such single-point integration effort. Future studies will be needed to investigate if this finding substantiates for other L1/L2 groups, and to what extent such differences in reading take-up are related to group-wise and individual differences in working memory capacity.

False remembering in L1 and L2

Our initial analysis of L1 and L2 speakers’ performance on the recognition memory task (i.e., *d* prime analysis) demonstrated lower recognition memory among L2 speakers overall, a result that was primarily driven by greater response bias among L2 speakers. In other words, L2 speakers were more likely than L1 speakers to endorse any kind of presented word as “old,” a result that replicates previous

findings from recognition memory research using the DRM paradigm (e.g., Beato & Arndt, 2017). Our primary analysis of the recognition memory data took these differences in response bias into account and showed a reduced false memory effect for predictable words in L2 compared to L1 speakers. In other words, over and above any differences in response bias and in line with both the *reduced-activation* and *bilingual-advantage* accounts, L2 speakers were less likely to false alarm to predictable lures than L1 speakers. This aspect of our findings corroborates and extends previous research using the DRM paradigm showing that false remembering of semantically related information is reduced in L2 compared to L1 processing (Anastasi et al., 2005, Experiments 3 and 4; Sahlin et al., 2005; Howe et al., 2009; Arndt and Beato, 2017; Beato and Arndt, 2021). In contrast, this result is not in line with a previous investigation from the prediction literature (Foucart et al., 2016), which had suggested L1-equivalent rates of false remembering in French-Spanish bilinguals (albeit without direct statistical comparison between L1 and L2 speakers). Importantly, Foucart et al. (2016) did not consider potential differences in response bias in L2 vs L1 speakers, a factor that our study showed to play an important role.

Notably, our results also indicated that L2 speakers showed a qualitatively different pattern than L1 speakers when judging the contents of their memory. Even though both groups were more likely to false alarms to predictable lures with high compared to low subjective memory confidence, this increase was attenuated in L2 speakers, suggesting that L2 speakers experienced less detailed remembering when presented with predictable lures. Hence, L2 speakers showed evidence of a reduced false memory effect for predictable lures. While this aspect of our findings does not help adjudicate between the two L2-processing accounts, that is, *reduced-activation* and *bilingual-advantage*, we note that the totality of our findings is more in line with the *reduced-activation* account. Specifically, we found that true memory rates during word recognition were also reduced among L2 speakers. The most parsimonious explanation for these findings attributes differences between L1 and L2 processing to how strongly these groups activate lexical-semantic information during sentence reading. L2 speakers not only show reduced rates of false memory, indicating less consistent use of sentence context to predict or pre-activate lexical representations⁹, but they may also show more diffuse and less distinct processing of words more generally. This interpretation is consistent with a large body of research on predictive processing in L2 which has consistently demonstrated that L2 speakers show reduced or slowed-down prediction effects. The *reduced-activation* interpretation of false remembering is also consistent with a common interpretation of reductions in false remembering in an L2 within the DRM paradigm. Previous studies in this line of research have hypothesized that L2 speakers' lexical representations may be less robust and less strongly connected to semantic representations in memory (e.g., Beato & Arndt, 2021; Suarez & Beato, 2021), which in turn lowers the likelihood that L2 speakers will automatically activate semantically associated information.

In sum, the totality of results found in the present study are more in line with the *reduced-activation* interpretation of the false memory effect, and not with the *bilingual-advantage* account (also see Digeremenci et al., 2022; Gunnerud et al., 2020; Lehtonen et al., 2018). Specifically, we found that L2 speakers not only show a

reduced false memory effect but also a reduced *true* memory effect—that is, reduced effects of correct recognition memory, when corrected for response bias. Taken together, these findings could indicate that lexical entries in the L2 lexicon may be less strongly connected to semantic representations in memory (e.g., Kroll & Stewart, 1994). This, in turn, may lead to less distinct (i.e., semantically enriched) encoding of presented information and reduce the likelihood of L2 speakers to automatically activate predictable or semantically associated information during language processing.

Conclusion

Previous studies with native speakers showed that predictable but not presented words linger in memory and elicit false remembering. The present study extended this research to L2 speakers of German and found evidence for a reduced false memory effect in L2 compared to L1 speakers. L2 speakers were also less likely to falsely remember predictable lures with high subjective memory confidence, suggesting that predictable words elicited less vivid and detailed remembering in L2 speakers. Notably, L2 speakers also showed reduced levels of accurate remembering for previously presented, “old” words. Our results thus support theories arguing that lexical representations in L2 are less entrenched and less interconnected with semantic information in long-term memory (e.g., Kroll & Stewart, 1994), leading to less distinct (i.e., semantically enriched) lexical activation during L2 processing, and consequently to reduced (true) memory for encountered words, as well as reduced (false) memory for words that were predictable but not encountered.

Replication package. All data files, materials, and analysis scripts can be found on this paper’s project page using the link <https://osf.io/e56k4/>.

Financial support. This work was funded by the *Deutsche Forschungsgesellschaft* DFG, German Research Foundation—Project-ID 232722074—SFB 1102, project A5.

Competing interests. Parts of the results reported in this paper were presented at the ISBPAC 2024 conference in Swansea, UK. The authors declare no competing interests.

Notes

1 Of note, that paper also demonstrated higher (not lower) rates of correct recognition memory for L2-based processing. The authors relate these results to the relatively low proficiency level of their participants, arguing that non-proficient L2 speakers may have translated L2 lists into their L1 during encoding, which may have effectively “doubled” their activation levels for previously seen “old” words.

2 We did not formally assess language status in our L1 group. This group likely does have some command of a foreign language, but as they are living in their L1 environment, they are less likely to regularly use both languages in their everyday lives than the L2 group, which is living immersed in an L2 environment.

3 Note that we replicated all results reported below when rerunning models with cloze probability as a scaled continuous variable.

4 We replicated all findings reported below when using a more stringent outlier treatment procedure in which we mean-replaced (rather than excluded) outliers that were more than 2.5 standard deviations from a subjects’ condition mean.

5 We ran exploratory models on the subset of L2 speakers that investigated whether predictability effects in the SPR task were additionally modulated by proficiency or age of acquisition (see this paper’s OSF, for R-Markdown). For proficiency, we ran three sets of models in which proficiency was operationalized as the

scaled LexTale score, self-rated overall proficiency in German and self-rated reading proficiency in German. Across models, no interactions with predictability or memory surfaced.

6 Here, we computed false-alarm rates by only taking into account new items, not lures. Rates equal to 0 or 1 were replaced by 0.01 and 0.99, respectively.

7 Again, we replicated all findings when using an outlier treatment procedure in which we mean-replaced RT outliers by subject-condition.

8 Again, we ran exploratory models on the data of the L2 speakers that investigated whether false or true memory was additionally modulated by L2-proficiency or age of acquisition (for results, see OSF of this article). No effects emerged.

9 We acknowledge that this account is not directly borne out in our reading time data which showed no reduced predictability effect for L2 speakers. However, as outlined above, we think this null result is more likely to stem from possible L1/L2 differences in integration effort during reading, which our analysis did not capture (i.e., focal integration effort on the noun in L2 speakers, compared to a more distributed integration effort in L1 speakers).

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)
- Anastasi, J., Rhodes, M., Marquez, S., & Velino, V. (2005). The incidence of false memories in native and non-native speakers. *Memory*, 13(8), 815–828. <https://doi.org/10.1080/09658210444000421>
- Arndt, J., & Beato, M. S. (2017). The role of language proficiency in producing false memories. *Journal of Memory and Language*, 95, 146–158. <https://doi.org/10.1016/j.jml.2017.03.004>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Balota, D. A., Cortese, M. J., Duchek, J. M., Adams, D., Roediger III, H. L., McDermott, K. B., & Yerys, B. E. (1999). Veridical and false memories in healthy older adults and in dementia of the Alzheimer's type. *Cognitive neuropsychology*, 16(3–5), 361–384. <https://doi.org/10.1080/026432999380834>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language*, 68, 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Beato, M. S., & Arndt, J. (2021). The effect of language proficiency and associative strength on false memory. *Psychological Research*, 85(8), 3134–3151.
- Beato, M. S., Suarez, M., Cadavid, S., & Albuquerque, P. B. (2023). False memory in between-language conditions: a brief review on the effect of encoding and retrieving in different languages. *Frontiers in Psychology*, 14, 1237471. <https://doi.org/10.3389/fpsyg.2023.1237471>
- Bialystok, E., Craik, F., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(4), 859–873. <https://doi.org/10.1037/a0015638>
- Bialystok, E., Dey, A., Sullivan, M. D., & Sommers, M. S. (2020). Using the DRM paradigm to assess language processing in monolinguals and bilinguals. *Memory & Cognition*, 48, 870–883. <https://doi.org/10.3758/s13421-020-01016-6>
- 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>
- Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2015). Graded expectations: predictive processing and the adjustment of expectations during spoken language comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, 15, 607–624. <https://doi.org/10.3758/s13415-015-0340-0>
- Bovolenta, G., & Marsden, E. (2022). Prediction and error-based learning in L2 processing and acquisition: a conceptual review. *Studies in Second Language Acquisition*, 44(5), 1384–1409. <https://doi.org/10.1017/S0272263121000723>

- Brainerd, C. J., Yang, Y., Reyna, V. F., Howe, M. L., & Mills, B. A. (2008). Semantic processing in “associative” false memory. *Psychonomic Bulletin & Review*, 15, 1035–1053. <https://doi.org/10.3758/PBR.15.6.1035>
- Brehm, L., & Alday, P. M. (2022). Contrast coding choices in a decade of mixed models. *Journal of Memory and Language*, 125, 104334. <https://doi.org/10.1016/j.jml.2022.104334>
- Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Böhlte, J., & Böhl, A. (2011). The word frequency effect. *Experimental Psychology*, 58(5), 412–424. <https://doi.org/10.1027/1618-3169/a000123>
- Chang, M., & Brainerd, C. J. (2021). Semantic and phonological false memory: a review of theory and data. *Journal of Memory and Language*, 119, 104210. <https://doi.org/10.1016/j.jml.2020.104210>
- Cholewa, J., Neitzel, I., Bürgens, A., & Günther, T. (2019). Online-processing of grammatical gender in noun-phrase decoding: an eye-tracking study with monolingual German 3rd and 4th graders. *Frontiers in Psychology*, 10, 2586. <https://doi.org/10.3389/fpsyg.2019.02586>
- Chung, Y. M. W., & Federmeier, K. D. (2023). Read carefully, because this is important! how value-driven strategies impact sentence memory. *Memory & Cognition*, 51(7), 1511–1526.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181–204. <https://doi.org/10.1017/S0140525X12000477>
- Conklin, K., Alotaibi, S., Pellicer-Sánchez, A., & Vilkaite-Lozdienė, L. (2020). What eye-tracking tells us about reading-only and reading-while-listening in a first and second language. *Second Language Research*, 36(3), 257–276.
- Degirmenci, M. G., Grossmann, J. A., Meyer, P., & Teichmann, B. (2022). The role of bilingualism in executive functions in healthy older adults: a systematic review. *International Journal of Bilingualism*, 26(4), 426–449.
- 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., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia*, 61, 150–162. <https://doi.org/10.1016/j.neuropsychologia.2014.06.016>
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117–1121. <https://doi.org/10.1038/nn1504>
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2017). Is there a replication crisis? Perhaps. Is this an example? No: a commentary on Ito, Martin, and Nieuwland (2016). *Language, Cognition and Neuroscience*, 32(8), 966–973. <https://doi.org/10.1080/23273798.2017.1279339>
- 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>
- Dirix, N., Vander Beken, H., De Bruyne, E., Brysbaert, M., & Duyck, W. (2020). Reading text when studying in a second language: an eye-tracking study. *Reading Research Quarterly*, 55(3), 371–397. <https://doi.org/10.1002/rrq.277>
- Dodson, C. S., Bawa, S., & Krueger, L. E. (2007). Aging, metamemory, and high-confidence errors: a misrecollection account. *Psychology and Aging*, 22(1), 122–133. <https://doi.org/10.1037/0882-7974.22.1.122>
- Dussias, P. E., Kroff, J. R. V., Tamargo, R. E. G., & 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 verbal learning and verbal behavior*, 20(6), 641–655. [https://doi.org/10.1016/S0022-5371\(81\)90220-6](https://doi.org/10.1016/S0022-5371(81)90220-6)
- Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: long-term memory structure and sentence processing. *Journal of Memory and Language*, 41(4), 469–495. <https://doi.org/10.1006/jmla.1999.2660>
- Ferreira, F., & Chantavarin, S. (2018). Integration and prediction in language processing: a synthesis of old and new. *Current Directions in Psychological Science*, 27(6), 443–448. <https://doi.org/10.1177/0963721418794491>

- 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>
- Finger, H., Goetze, C., Diekamp, D., Standvoß, K., & König, P. (2017). Labvanced: a unified JavaScript framework for online studies. In *International Conference on Computational Social Science (Cologne)* (pp. 1–3). University of Osnabrück.
- Fleur, D. S., Flecken, M., Rommers, J., & Nieuwland, M. S. (2020). Definitely saw it coming? The dual nature of the pre-nominal prediction effect. *Cognition*, 204, 104335. <https://doi.org/10.1016/j.cognition.2020.104335>
- Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(5), 1461–1469. <https://doi.org/10.1037/a0036756>
- Foucart, A., Ruiz-Tada, E., & Costa, A. (2015). How do you know I was about to say “book”? Anticipation processes affect speech processing and lexical recognition. *Language, Cognition and Neuroscience*, 30(6), 768–780. <https://doi.org/10.1080/23273798.2015.1016047>
- Foucart, A., Ruiz-Tada, E., & Costa, A. (2016). Anticipation processes in L2 speech comprehension: evidence from ERPs and lexical recognition task. *Bilingualism: Language and Cognition*, 19(1), 213–219. <https://doi.org/10.1017/S1366728915000486>
- Fraundorf, S. H., Hourihan, K. L., Peters, R. A., & Benjamin, A. S. (2019). Aging and recognition memory: a meta-analysis. *Psychological Bulletin*, 145(4), 339–371. <http://doi.org/10.1037/bul0000185>
- Futrell, R., Gibson, E., & Levy, R. P. (2020). Lossy-context surprisal: an information-theoretic model of memory effects in sentence processing. *Cognitive Science*, 44(3), e12814. <https://doi.org/10.1111/cogs.12814>
- Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, 38, 833–848. <https://doi.org/10.3758/MC.38.7.833>
- Greene, N. R., Forsberg, A., Guitard, D., Naveh-Benjamin, M., & Cowan, N. (2024). A lifespan study of the confidence-accuracy relation in working memory and episodic long-term memory. *Journal of Experimental Psychology: General*, 153(5), 1336–1360. <https://doi.org/10.1037/xge0001551>
- Grosjean, F. (2021). *Life as a Bilingual: Knowing and Using Two or More Languages*. Cambridge University Press.
- 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. 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>
- Gunnerud, H. L., ten Braak, D., Reikerås, E. K. L., Donolato, E., & Melby-Lervåg, M. (2020). Is bilingualism related to a cognitive advantage in children? A systematic review and meta-analysis. *Psychological Bulletin*, 146(12), 1059–1083.
- Haeuser, K. I., & Kray, J. (2022). Uninvited and unwanted: false memories for words predicted but not seen. In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th annual meeting of the cognitive science society* (pp. 2401–2408). Cognitive Science Society. <https://escholarship.org/uc/item/7w22b8gm>
- Haeuser, K. I., & Kray, J. (2024). Age differences in context use during reading and downstream effects on recognition memory. *Psychology and Aging*, 39(7), 715–730. <https://doi.org/10.1037/pag0000845>
- Haeuser, K. I., Kray, J., & Borovsky, A. (2022). Hedging bets in linguistic prediction: Younger and older adults vary in the breadth of predictive Processing. *Collabra: Psychology*, 8(1), 36945.
- Hicks, J. L., & Hancock, T. W. (2002). Backward associative strength determines source attributions given to false memories. *Psychonomic Bulletin & Review*, 9(4), 807–815. <https://doi.org/10.3758/BF03196339>
- Holliday, R. E., & Weekes, B. S. (2006). Dissociated developmental trajectories for semantic and phonological false memories. *Memory*, 14(5), 624–636. <https://doi.org/10.1080/09658210600736525>
- Hopp, H. (2022). Second language sentence processing. *Annual Review of Linguistics*, 8(1), 235–256. <https://doi.org/10.1146/annurev-linguistics-030821-054113>
- Howe, M. L. (2008). What is false memory development the development of? Comment on Brainerd, Reyna, and Ceci (2008). *Psychological Bulletin*, 134(5), 768–772. <https://doi.org/10.1037/0033-2909.134.5.768>

- Howe, M. L., Wimmer, M. C., Gagnon, N., & Plumpton, S. (2009). An associative-activation theory of children's and adults' memory illusions. *Journal of Memory and Language*, **60**(2), 229–251. <https://doi.org/10.1016/j.jml.2008.10.002>
- Hubbard, R. J., & Federmeier, K. D. (2024). The impact of linguistic prediction violations on downstream recognition memory and sentence recall. *Journal of Cognitive Neuroscience*, **36**(1), 1–23. https://doi.org/10.1162/jocn_a_02078
- Hubbard, R. J., Rommers, J., Jacobs, C. L., & Federmeier, K. D. (2019). Downstream behavioral and electrophysiological consequences of word prediction on recognition memory. *Frontiers in Human Neuroscience*, **13**, 291. <https://doi.org/10.3389/fnhum.2019.00291>
- Huetttig, 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>
- Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: evidence from brain potentials. *Journal of Memory and Language*, **86**, 157–171. <https://doi.org/10.1016/j.jml.2015.10.007>
- Ito, A., & Pickering, M. J. (2021). Automaticity and prediction in non-native language comprehension. In: E. Kaan & T. Grüter (Eds.), *Prediction in Second Language Processing and Learning* (pp. 25–46). John Benjamins.
- 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.
- 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., Dallas, A. C., & Wijnen, F. (2010). Syntactic predictions in second-language sentence processing. In J.-W. Zwart & M. de Vries (Eds.), *Structure preserved. Festschrift in the honor of Jan Koster* (pp. 207–213). John Benjamins. <https://doi.org/10.1075/la.164.23kaa>
- Kaan E., & Grüter, T. (2021) Prediction in second language processing and learning: advances and directions. In: E. Kaan & T. Grüter (Eds.), *Prediction in Second Language Processing and Learning* (pp. 2–24). John Benjamins.
- Kamide, Y., Altmann, G. T., & 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)
- Kensinger, E. A., & Schacter, D. L. (1999). When true memories suppress false memories: effects of aging. *Cognitive Neuropsychology*, **16**(3–5), 399–415. <https://doi.org/10.1080/026432999380852>
- Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: the influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, **135**, 12–35. <https://doi.org/10.1037/0096-3445.135.1.12>
- Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, **33**(2), 149–174. <https://doi.org/10.1006/jmla.1994.1008>
- Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension?. *Language, Cognition and Neuroscience*, **31**, 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>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, **82**(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., De Bruin, A., & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological Bulletin*, **144**(4), 394–425.
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, **44**(2), 325–343.
- 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>

- 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>
- Morales, J., Calvo, A., & Bialystok, E. (2013). Working memory development in monolingual and bilingual children. *Journal of Experimental Child Psychology*, 114(2), 187–202. <https://doi.org/10.1016/j.jecp.2012.09.002>
- Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., et al., (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife*, 7, e33468. <https://doi.org/10.7554/eLife.33468>
- Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition*, 25(6), 838–848.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and Communication* (pp. 191–226). Longman. <https://doi.org/10.4324/9781315836027-12>
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114(3), 510–532. <https://doi.org/10.1037/0033-2909.114.3.510>
- R Core Team (2022). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. <https://www.R-project.org/>
- Reyna, V. F., Corbin, J. C., Weldon, R. B., & Brainerd, C. J. (2016). How fuzzy-trace theory predicts true and false memories for words, sentences, and narratives. *Journal of Applied Research in Memory and Cognition*, 5(1), 1–9. <https://doi.org/10.1016/j.jarmac.2015.12.003>
- Rich, S., & Harris, J. (2021). Unexpected guests: When disconfirmed predictions linger. In *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society* (pp. 2246–2252). Cognitive Science Society.
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4), 803–814. <https://doi.org/10.1037/0278-7393.21.4.803>
- Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: a multiple regression analysis. *Psychonomic bulletin & review*, 8, 385–407. <https://doi.org/10.3758/BF03196177>
- Rommers, J., & Federmeier, K. D. (2018). Lingering expectations: a pseudo-repetition effect for words previously expected but not presented. *NeuroImage*, 183, 263–272. <https://doi.org/10.1016/j.neuroimage.2018.08.023>
- Ryskin, R., & Nieuwland, M. S. (2023). Prediction during language comprehension: what is next?. *Trends in Cognitive Sciences*, 27(11), 1032–1052. <https://doi.org/10.1016/j.tics.2023.08.003>
- Sahlin, B. H., Harding, M. G., & Seamon, J. G. (2005). When do false memories cross language boundaries in English–Spanish bilinguals?. *Memory & Cognition*, 33, 1414–1421. <https://doi.org/10.3758/BF03193374>
- Schacter, D. L., Koutstaal, W., & Norman, K. A. (1997). False memories and aging. *Trends in Cognitive Sciences*, 1(6), 229–236. [https://doi.org/10.1016/S1364-6613\(97\)01068-1](https://doi.org/10.1016/S1364-6613(97)01068-1)
- Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: a tutorial. *Journal of Memory and Language*, 110, 104038. <https://doi.org/10.1016/j.jml.2019.104038>
- Schlenter, J. (2023). Prediction in bilingual sentence processing: How prediction differs in a later learned language from a first language. *Bilingualism: Language and Cognition*, 26(2), 253–267. <https://doi.org/10.1017/S1366728922000736>
- Shain, C. (2024). Word frequency and predictability dissociate in naturalistic reading. *Open Mind*, 8, 177–201. https://doi.org/10.1162/opmi_a_00119
- Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: an eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, 27(2), 251–272. <https://doi.org/10.1177/0267658310382068>
- Sommers, M. S., & Lewis, B. P. (1999). Who really lives next door: creating false memories with phonological neighbors. *Journal of Memory and Language*, 40(1), 83–108. <https://doi.org/10.1006/jmla.1998.2614>
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior research methods, instruments, & computers*, 31(1), 137–149. <https://doi.org/10.3758/BF03207704>

- Staub, A.** (2015). The effect of lexical predictability on eye movements in reading: critical review and theoretical interpretation. *Language and Linguistics Compass*, 9(8), 311–327. <https://doi.org/10.1111/lnc3.12151>
- Staub, A.** (2024). Predictability in language comprehension: prospects and problems for surprisal. *Annual Review of Linguistics*, 11. <https://doi.org/10.1146/annurev-linguistics-011724-121517>
- Staub, A., & Clifton Jr, C.** (2006). Syntactic prediction in language comprehension: evidence from either... or. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(2), 425–436. <https://doi.org/10.1037/0278-7393.32.2.425>
- Suarez, M., & Beato, M. S.** (2021). The role of language proficiency in false memory: a mini review. *Frontiers in Psychology*, 12, 659434. <https://doi.org/10.3389/fpsyg.2021.659434>
- Suarez, M., & Beato, M. S.** (2023). False memory in a second language: the importance of controlling the knowledge of word meaning. *Plos one*, 18(5), e0285747. <https://doi.org/10.1371/journal.pone.0285747>
- Tun, P. A., Wingfield, A., Rosen, M. J., & Blanchard, L.** (1998). Response latencies for false memories: gist-based processes in normal aging. *Psychology and Aging*, 13(2), 230–241. <https://doi.org/10.1037/0882-7974.13.2.230>
- Urbach, T. P., DeLong, K. A., Chan, W. H., & Kutas, M.** (2020). An exploratory data analysis of word form prediction during word-by-word reading. *Proceedings of the National Academy of Sciences*, 117(34), 20483–20494. <https://doi.org/10.1073/pnas.1922028117>
- Van Berkum, J. J., Brown, C. M., Zwitterlood, 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 Petten, C., & Luka, B. J.** (2012). Prediction during language comprehension: benefits, costs, and ERP components. *International Journal of Psychophysiology*, 83(2), 176–190. <https://doi.org/10.1016/j.jpsycho.2011.09.015>
- Wicha, N. Y., Moreno, E. M., & Kutas, M.** (2003). Expecting gender: an event related brain potential study on the role of grammatical gender in comprehending a line drawing within a written sentence in Spanish. *Cortex*, 39(3), 483–508. [https://doi.org/10.1016/S0010-9452\(08\)70260-0](https://doi.org/10.1016/S0010-9452(08)70260-0)
- Yan, S., Kuperberg, G. R., & Jaeger, T. F.** (2017). Prediction (or not) during language processing. A commentary on Nieuwland et al. (2017) and DeLong et al.(2005). *BioRxiv*, 143750.

Cite this article: Haeuser, K. I. & Grüter, T. (2025). Reduced false memory effects for predictable words in L2 speakers of German: evidence from self-paced reading and recognition memory. *Applied Psycholinguistics*. <https://doi.org/10.1017/S0142716425100155>