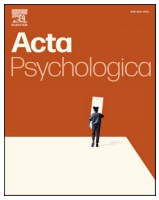
Acta Psychologica 266 (2026) 106948

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy



The use of lexico-semantics and gender marking as cues to prediction in L1 and L2 German

Pernelle Lorette

Department of Psycholinguistics, Universität Mannheim, Anglistik I, Schloss, 68161, Mannheim, Germany

ARTICLE INFO

Keywords:
Prediction
First and second language processing
Visual world paradigm
Verb semantics
Gender marking

ABSTRACT

In three visual-world eye-tracking studies, we investigated the individual and combinatory use of lexico-semantic information of the verb and gender marking of the determiner to predict the referent of an upcoming object among L1 and L2 German speakers. Importantly, in our design random, the linguistic input randomly varied in terms of the type of information (lexico-semantic, morphosyntactic, or a combination thereof) that could be used as predictive cue, most closely reflecting real-world language input compared to most previous studies. Cluster-based permutation analyses revealed prediction effects based on lexico-semantic cues in both L1 and L2, although for placement verbs, effects were only found in L1. Gender cues and the combination of lexico-semantic and gender cues also led to prediction effects solely in L1. The results not only confirm more limited predictive processing in L2 compared to L1 reported in previous studies, but also suggest a moderating role of type of cue in L1 predictive processing. These findings are discussed in the light of the good-enough approach (Ferreira et al., 2002; Ferreira & Patson, 2007) and the shallow structure hypothesis (Clahsen & Felser, 2006, 2018).

1. Introduction

Prediction – i.e., the pre-activation or anticipation of upcoming input based on prior information – is broadly recognised as a central mechanism of language processing (Huettig, 2025; MacDonald, 2013; Pickering & Gambi, 2018). Language comprehenders can draw on a range of linguistic and non-linguistic cues to generate predictions, including lexico-semantic information (e.g. Altmann & Kamide, 1999), morphosyntactic information such as case (e.g., Hopp, 2015) or gender marking (e.g., Dussias et al., 2013), prosodic (Weber et al., 2006) and phonological (Kukona, 2020), and visual information (e.g. Knoeferle et al., 2005). Both first (L1) and second language (L2) speakers engage in predictive processing, but prediction in L2 tends to be less robust and stable (Kaan & Grüter, 2021; for reviews see Kaan, 2014; Schlenter, 2023). However, prediction effects are not always observed, even in L1 (Ito et al., 2017a; Kochari & Flecken, 2019; Nieuwland et al., 2018; Riordan et al., 2015), and the ubiquity of prediction in language processing is still much debated (e.g., Huettig & Mani, 2016; Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018). According to an influential theoretical approach, two routes can lead to prediction: prediction-by-association and prediction-by-production (Pickering & Gambi, 2018). The prediction-by-association route relies on automatic semantic or associative priming mechanisms. The prediction-by-production route, on the other hand, is an optional route that engages covert activation of the production system to simulate the production of the input, with lower-level (e.g., form-related) representations activated after higher-level ones (e.g., semantic representations). Prediction-by-association is thought to be less effective but also less cognitively demanding and more rapidly engaged than prediction-by-production. The associative route may therefore be particularly relied upon in demanding contexts, such as L2 processing (e.g., Schlenter, 2023). In this contribution, we aim to explore L1 and L2 speakers' ability to predict, particularly focussing on the use of lexico-semantic and morphosyntactic cues in L2 prediction. We partially replicate Van Bergen and Flecken's (2017) study on prediction in Dutch based on the semantic contrast between *zetten* ('put.$_{\text{STAND}}$') vs. *legen* ('put.$_{\text{LIE}}$') – which we however conduct in German – and extend their study to include other (non-placement) verbs as well as gender marking as a predictive cue. Furthermore, we also extend the investigation of predictive language processing to contexts where the input randomly alternates between providing semantic cues, gender cues, combined cues, or no cues, which more closely reflects the irregularity of the input in the real world compared to the bulk of previous studies.

E-mail address: p.lorette@uni-mannheim.de.

https://doi.org/10.1016/j.actpsy.2026.106948
Received 18 November 2025; Received in revised form 23 April 2026; Accepted 24 April 2026
Available online 29 April 2026

2. Using lexico-semantic and morphosyntactic cues predictively

When processing an auditory sentence such as *The boy will eat the cake* in a visual context where only one edible object can be seen among other objects, Altmann and Kamide (1999) showed in their seminal study that L1 English speakers make anticipatory eye-movements to the edible object upon processing the verb semantics, i.e., before the noun *cake* is encountered in the auditory input. Their study employed the visual world paradigm (VWP), where participants listen to spoken sentences while viewing a visual scene containing a limited set of objects, including the target (referred to in the sentence) and competitors or distractors. Not only L1 speakers but also L2 speakers have been found to generate predictions based on lexico-semantic information (e.g., Dijkgraaf et al., 2017, 2019), although L2 prediction based on semantic cues tends to be weaker and/or delayed compared to L1 (e.g., Chun & Kaan, 2019; Kim & Grüter, 2021; Schlenter, 2023; but see Dijkgraaf et al., 2017).

Besides semantics, comprehenders can also exploit morphosyntactic cues to predict upcoming referents. Gender marking on prenominal articles or adjectives can enable prediction of an upcoming noun in L1 (Dahan et al., 2000; Huettig & Janse, 2016; Otten & Van Berkum, 2009; Wicha et al., 2004) and L2 (Dussias et al., 2013; Foucart et al., 2014; Hopp & Lemmerth, 2018). However, other studies report no reliable use of gender marking to predict an upcoming noun in L1 (Kochari & Flecken, 2019; Nieuwland et al., 2018) or in L2 (Grüter et al., 2012; Lew-Williams & Fernald, 2010). In particular, L2 speakers who lack strong explicit knowledge of gender assignment (Hopp, 2013, 2016), whose L1 does not encode grammatical gender, or who have lower L2 proficiency levels (Dussias et al., 2013; Foucart & Frenck-Mestre, 2011, 2012) are less likely to rely on gender marking to anticipate upcoming nouns. Crucially, most VWP studies reporting predictive use of gender marking by L2 speakers employed visual displays with only one (or occasionally two) gender competitor(s) besides the target object (Dussias et al., 2013; Hopp & Lemmerth, 2018). This design likely reduces visual processing load and lexical competition compared to the classical four-object displays, thereby leaving more cognitive resources available to engage in prediction. It is thus unclear whether L2 speakers would engage in prediction in cognitively more demanding contexts that better approximate real-world settings. Moreover, many of these VWP studies employed tasks requiring an explicit motor response (e.g., "click on the target"), which may enhance predictive behaviour by increasing task salience and encouraging strategic anticipation, compared to passive "look-and-listen" paradigms (see Huettig & Tanenhaus, 2025 for a discussion of task effects in the VWP).

Beyond gender marking, other morphosyntactic information can be used to generate prediction, such as case marking, verb morphology, and classifiers. While L1 speakers use case marking as predictive cue, it proves to be difficult to use for L2 speakers (Frenck-Mestre et al., 2019; but see Schlenter & Felser, 2021), even when they have good explicit knowledge of the L2 case-marking system (Mitsugi & Macwhinney, 2016). When aided by additional cues such as prosody, however, L2 speakers – especially highly proficient ones – can exploit case marking predictively, although the generated predictions are not consistently correct (Henry et al., 2020). Contrary to case marking and classifiers, verb morphology (number) appears easier for (Dutch) L2 speakers to exploit, although L2 prediction effects remain slower and weaker than in L1 (Koch et al., 2021, 2023).

Different types of cues can also be incrementally integrated to generate prediction. Kamide, Altmann, and Haywood (2003) showed that Japanese L1 speakers can simultaneously integrate case-marking and lexico-semantic information within a single noun phrase (NP) to predict the referent of another NP (study 3). They also found that English L1 speakers can combine semantic information from two different NPs to predict a subsequent NP (study 2). However, their design did not allow to demonstrate the incremental integration of different types of cues *across* multiple NPs to predict the referent of a later NP — a process that would, in fact, be most efficient for language processing in many real-world situations. Imagine, for instance, interacting with a French-speaking child who says "Je veux manger le …" ("I want to eat the.-$_{MASC}$") while sitting at a table covered with grammatically feminine and grammatically masculine food items and (non-edible) toys. Kamide, Scheepers, and Altmann (2003) sought to address this limitation. Using German sentences with both SVO (subject–verb–object) and OVS (object–verb–subject) word orders, they found that L1 German speakers could combine morphosyntactic information – i.e., case marking – from one NP with the verb semantics to predict the referent of another NP. Yet, it is unclear whether their findings are generalizable, as the visual information used in their display may have biased participants toward interpreting one NP as the agent of the sentence. Thus, participants may have predicted that the other NP is the patient solely by using the case-marking and semantic information of a single NP rather than by truly integrating the verb lexico-semantics and the case-marking of the first NP. To control for this potential bias, Hopp (2015) adapted Kamide, Scheepers, and Altmann's (2003) materials. His (limited) sample of 12 L1 speakers showed predictive looks to the target by integrating the case marking of one NP with the lexico-semantics of the verb to predict an upcoming NP.

3. Factors affecting the use of lexico-semantic and morphosyntactic cues in L2

3.1. Type of cue

Besides the L1 German group, Hopp (2015) also tested English-L1/German-L2 late bilinguals and found that neither low- nor high-proficiency L2 speakers were able to combine case-marking with lexico-semantics to predict an upcoming NP in L2 German. Instead, L2 speakers relied solely on lexico-semantic cues, interpreting the first NP as agent and predicting that the second NP would be the theme, irrespective of whether the first NP was marked for nominative (agent) or accusative (theme). A similar asymmetry between the use of lexico-semantic and morphosyntactic cues in L1 and L2 processing was observed in Grüter et al.'s (2020) study on the use of classifier as grammatical form class and as semantic cues to predict in Mandarin. While L1 speakers used classifiers primarily as grammatical cues, L2 speakers primarily relied on classifier semantics to generate predictions.

This stronger reliance on lexico-semantics over morphosyntax in L2 aligns with the *Shallow Structure Hypothesis* (Clahsen & Felser, 2006, 2018). This hypothesis builds on the more general *good-enough approach to language comprehension* (Ferreira et al., 2002; Ferreira & Patson, 2007), according to which comprehenders often rely on semantic heuristics rather than detailed grammar-driven analyses, particularly when processing is constrained by limited time or cognitive resources (Christianson, 2016). Within this framework, the *Shallow Structure Hypothesis* posits that L2 comprehenders rely even more heavily on non-grammatical than on grammatical information compared to L1 comprehenders. Accordingly, semantic and pragmatic cues tend to be prioritised over morphosyntactic information during L2 comprehension. To the extent that predictive processing relies on the rapid extraction and integration of information from the unfolding input to generate predictions about the upcoming input, a reduced sensitivity to grammatical information may limit L2 comprehenders' ability to use morphosyntactic cues to generate expectations about upcoming linguistic material. As a result, predictions in L2 comprehension may rely relatively more on lexico-semantic information than on morphosyntactic structure.

3.2. Cross-linguistic influence

However, Hopp's (2015) conclusion that his findings support a tendency to rely on lexico-semantics rather than on morphosyntax in L2 may also be interpreted through the lens of cross-linguistic influence (CLI). Several findings suggest that using morphosyntactic information

as predictive cue is particularly challenging for L2 speakers whose L1 does not encode the same morphosyntactic features. While Hopp's (2015) L2 speakers could transfer the lexico-semantic cues from their L1, the morphosyntactic cues (i.e., case-marking) were absent from the L2 speakers' L1 and could thus not be transferred, which may have constrained their use of this type of cue for prediction. Evidence consistent with this interpretation comes from Schlenter and Felser (2021), who tested (advanced to near-native) L1-Russian/L2-German bilinguals (i.e., two languages using case marking). They showed predictive processing based on verb semantics in L2 German (study 1) and were also able to use case marking in L2 German (study 2) as predictive cue, albeit with what the authors call "reduced certainty" (p. 63) and a delayed time course compared to the L1 speakers. However, Frenck-Mestre et al. (2019) did not replicate this effect: their L2 speakers of Korean failed to use case marking predictively in L2, regardless of whether their L1 encoded this feature — note that the L2 speakers' overall proficiency level is unclear from the provided information.

CLI also seems to affect prediction based on lexico-semantics, although, to our knowledge, only one study has investigated this so far. In a visual world eye-tracking study, Van Bergen and Flecken (2017) examined how L1 speakers and L2 speakers of Dutch use the semantic contrast between putting an object in a vertical versus a horizontal position — a semantic distinction lexicalised in both Dutch (*zetten*, 'put.$_{\text{STAND}}$' vs. *leggen*, 'put.$_{\text{LIE}}$') and German (*stellen*, 'put.$_{\text{STAND}}$' vs. *legen*, 'put.$_{\text{LIE}}$'), but not in English or French (*put* or *mettre* 'put' is used regardless of the final position of the object). Dutch L1 speakers predictively used placement-verb semantics, directing their gaze toward the target object before its auditory onset. Among L2 speakers, only German-Dutch bilinguals, whose L1 also encodes this contrast, showed what the authors termed "hesitant predictions", with faster integration of semantic information upon hearing the target object in predictable than non-predictable trials — suggesting some pre-activation of the target. In contrast, French-Dutch and English-Dutch bilinguals did not use this cue predictively and integrated verb semantics only after hearing the object noun, with no difference between predictable and non-predictable trials. These findings suggest that L2 speakers' use of lexico-semantic cues in prediction may also be limited to cases where such cues are available in the L1.

The notion that L1–L2 transferability shapes the type of cues L2 speakers rely on for processing (or indeed, predicting) input can be accounted for by the (*Unified*) *Competition Model* (Bates & MacWhinney, 1981; MacWhinney, 1997; MacWhinney et al., 2017). This model does not assume a stable bias toward lexico-semantic over morphosyntactic cues in L2, but rather that cue weighting, and hence cue strength, depends on the availability and reliability of cues in the input, with cue strengths transferring from the L1 to the L2. As a result, cues that are absent or unreliable in the L1 are expected to be weaker and less readily exploited during L2 processing. Again, as predictive processing relies on the ability to rapidly exploit available cues to anticipate upcoming input, L2 comprehenders may be less likely to generate predictions based on cues that are weak or unreliable in their L1.

3.3. Proficiency

One individual dimension that may affect the cognitive effort associated with language processing, and hence engagement in predictive processing, is proficiency — although findings on its role remain inconsistent. Some studies report proficiency effects on lexico-semantic prediction (Peters et al., 2015) whereas others do not (Dijkgraaf et al., 2019; Kim & Grüter, 2021) — and some even find no predictive use of semantics among advanced L2 speakers (Ito et al., 2017b). Proficiency effects have also been reported on prediction based on morphosyntactic cues. Regarding case marking, Hopp (2015) reported that L2 proficiency affected the speed of information integration in speech, but not prediction per se. Beyond overall L2 proficiency, Hopp (2013) further showed that L2 speakers with stronger knowledge of the L2 gender system predicted more reliably based on gender marking than those with weaker knowledge. Similarly, Foucart and Frenck-Mestre's (2011) EEG study revealed that German-L1/French-L2 bilinguals were sensitive to gender violations in French — indicating pre-activation of gender-appropriate lexical candidates. However, when gender was incongruent across L1 and L2, this sensitivity emerged only among highly proficient speakers, even though both proficiency groups showed explicit knowledge of L2 noun gender in offline tasks. Proficiency may also interact with L1–L2 similarity in explaining predictive processing. For instance, Dussias et al. (2013) found that for L2 speakers whose L1 lacks grammatical gender, proficiency did modulate their ability to use gender predictively. In contrast, L2 speakers whose L1 marks gender on articles predicted based on gender marking in L2 regardless of proficiency level — although this effect was restricted to feminine forms, leaving the absence of effects for masculine forms unexplained. The interaction between proficiency and L1–L2 similarity is also reflected in Hopp and Lemmerth's (2018) findings: highly proficient L2 speakers predicted based on gender marking regardless of lexical congruency (whether translation equivalents share the same gender across L1 and L2) and syntactic congruency (whether gender is morphosyntactically realised in the same way in L1 and L2), whereas less proficient L2 speakers used gender marking predictively only for lexically congruent nouns in syntactically incongruent contexts. Lexical congruency effects have, however, also been observed at relatively high proficiency levels (Morales et al., 2016).

Taken together, these findings suggest that proficient L2 speakers can engage in predictive processing using both lexico-semantic and morphosyntactic cues, provided they are sufficiently familiar with the relevant feature — either due to its presence in the L1 or due to robust knowledge of the target-language system. Inconsistencies in proficiency effects in the literature may stem from the fact that general proficiency measures might fail to capture knowledge of a specific (lexico-semantic or morphosyntactic) feature, or from the possibility that familiarity with a transferable feature from the L1 may already be at ceiling, which would render overall proficiency irrelevant.

3.4. Available cognitive resources

Prediction also seems to depend on available cognitive resources. High working memory capacity (WMC) has been shown to facilitate prediction based on gender cues (Huettig & Janse, 2016) or semantic cues (Ding et al., 2023) in L1 processing. Given that L2 processing typically requires more cognitive effort (Hopp, 2022), limited cognitive resources might affect predictive processing even more in L2 contexts. Indeed, WMC has been found to facilitate prediction based on grammatical number (Koch et al., 2021, 2023), whereas no such effect emerged for semantic prediction in cognitively demanding contexts (Chun & Kaan, 2019). Prediction effects also appear delayed under higher cognitive load for both L1 and L2 speakers (Allison et al., 2025; Ito et al., 2018; but see Chun & Kaan, 2019). These findings suggest that predictive processing draws on working memory resources and that the extent to which comprehenders engage in prediction depends on their available cognitive resources at the moment.

All the factors mentioned above likely interact. As Foucart (2021, p. 107) explains, "once proficiency increases and cognitive resources are less taxed during L2 processing, the impact of cross-linguistic differences on prediction mechanisms may be reduced, resulting in native-like prediction in L2". One possibility is that predictive processing emerges only when sufficient cognitive resources – affected among others by proficiency, cross-linguistic differences and working memory capacity – are available to support the generation of predictions. According to the *Utility Hypothesis* (Kuperberg & Jaeger, 2016), predictions are generated only when doing so is useful, that is, when they optimize the efficiency of linguistic processing. If prediction generation requires excessive metabolic resources relative to its potential benefit, the process would not take place — including in L1. This framework may help explain why

predictive behaviour is often less robustly and less consistently observed in L2 as it involves greater processing demands than in L1, especially at lower proficiency levels.

4. The present study

In summary, previous research on L2 speakers' ability to use different types of cues to predict the referent of an upcoming noun remains inconsistent. Some studies suggest that L2 speakers rely on morphosyntactic cues only when these are transferable from their L1 (Dussias et al., 2013; Hopp & Lemmerth, 2018), whereas others report a more general difficulty in using morphosyntactic information in L2 (Frenck-Mestre et al., 2019; Grüter et al., 2020; Hopp, 2015), although some evidence points to successful exploitation of L2 morphosyntactic cues (Koch et al., 2021, 2023; Schlenter & Felser, 2021). Lexico-semantic cues appear more readily usable for prediction in L2, although typically in a weaker or delayed manner relative to L1. Yet whether L2 speakers can use lexico-semantic cues that are absent from the L1 remains unclear due to limited evidence — so far investigated in only one study focussing on the semantics of placement verbs (Van Bergen & Flecken, 2017). Addressing this question is essential to disentangle the respective contributions of CLI and cue type to predictive processing in L2. To this end, we investigate the use of semantic cues and grammatical gender marking in German by L1 German speakers and French learners of German. Crucially, we contrast semantic cues that are transferable from the L1 (verbs with direct translation equivalents in both languages) with semantic cues that are absent from the L1, namely German placement verbs, which encode object position (*stellen* 'put.$_{\text{STAND}}$' vs. *legen* 'put.$_{\text{LIE}}$'), unlike their French counterpart (*mettre* 'put'), which do not specify the final position of the object. Both French and German mark grammatical gender, with masculine and feminine categories in both languages (and an additional neuter category in German), realised on determiners and adjectives.

Finally, both L1 and L2 speakers' predictive behaviour has mainly been investigated in studies focussing on one type of cue at a time. However, this reflects real-world language processing only to a limited extent, as the sources of information that can serve as predictive cues are far from constant and may alternate unpredictably across cue types in naturalistic contexts. Moreover, evidence that L1 speakers integrate different types of cues from different constituents – i.e., combine information from different levels of representation – to generate predictions stems almost exclusively from studies on case-marking combined with the semantics of highly constraining verbs (Hopp, 2015; Kamide, Altmann, & Haywood, 2003; Kamide, Scheepers, & Altmann, 2003). Thus, even in L1, the relative contribution of lexico-semantic and morphosyntactic cues to prediction remains to be fully elucidated.

Consequently, our research questions and hypotheses are:

1) *Do German L1 speakers generate predictions during online sentence comprehension in German based on morphosyntactic and lexico-semantic cues when these cue types vary unpredictably across sentences?*

Hypothesis 1a. Theoretically, the *Good-Enough* approach to language comprehension (Ferreira et al., 2002; Ferreira & Patson, 2007) posits that L1 speakers rely more on semantic heuristics than on grammar-driven processing, particularly if processing is constrained by limited time or processing resources (Christianson, 2016). While this approach was originally formulated to explain input integration rather than predictive processing, it can inform how comprehenders more readily integrate specific types of cues over others, which, if integrated quickly enough, have the potential to be used to generate predictions about upcoming input. In a context where predictive cues, if available, alternately stem from morphosyntax or lexico-semantics, we might expect L1 speakers to reliably engage in prediction when lexico-semantic cues are available, but not to show consistent predictive behaviour based on morphosyntactic cues — contrary to previous evidence of prediction based on morphosyntax obtained in contexts where only morphosyntactic cues were informative (e.g., Dahan et al., 2000; Huettig & Janse, 2016; Otten & Van Berkum, 2009; Wicha et al., 2004).

Hypothesis 1b. Based on the *prediction-by-association* vs. *prediction-by-production account* (Pickering & Gambi, 2018), prediction effects are expected to be more readily observed when associative mechanisms (semantic or associative priming) can support prediction, since such mechanisms are assumed to be automatic and cognitively not very demanding. On the other hand, predictions that rely on the covert activation of the comprehender's production system require additional processing resources and are optional. Accordingly, lexico-semantic cues that are associatively or semantically related to the target word (non-placement-verb semantics) are expected to yield stronger predictive effects compared to either lexico-semantic cues that are not associatively related with the target (placement-verb semantics) or morphosyntactic cues (gender marking), as these require the engagement of the production system.

2) *Do German L2 speakers generate predictions during online sentence comprehension in German based on morphosyntactic and lexico-semantic cues when the former are present in the L1 while the latter are absent in the L1?*

Hypothesis 2a. According to the *Shallow Structure Hypothesis* (Clahsen & Felser, 2006, 2018), L2 speakers rely more on semantic than grammatical information. Based on this hypothesis, we would therefore expect that L2 speakers will use verb semantics to generate predictions – even for placement verbs, which are not transferable from the L1 – whereas gender marking will not be used predictively, even when noun gender is congruent across L1 and L2.

Hypothesis 2b. In contrast, the (*Unified) Competition Model* (Bates & MacWhinney, 1981; MacWhinney, 1997; MacWhinney et al., 2017) predicts the opposite pattern. Based on this model, we would predict that gender marking (for nouns with congruent gender across L1 and L2) will serve as a predictive cue for L2 speakers since this cue is available and reliable both in the L1 and in the L2, whereas placement verb semantics would not reliably be used for L2 prediction given its absence in L1, leading to the transfer of an extremely weak cue strength in L2 processing.

Hypothesis 2c. Based on the *prediction-by-association* vs. *prediction-by-production account* (Pickering & Gambi, 2018), we might expect L2 speakers (who have not reached very advanced proficiency levels) to rely primarily on associative mechanisms to generate prediction (non-placement-verb semantics), since the covert engagement of the production system is likely to be too cognitively demanding and may therefore not efficiently support comprehension.

3) *Can L1 speakers and L2 speakers of German combine verb semantics and gender marking across different constituents to form predictions during online sentence comprehension in German?*

Hypothesis 3. As previous research has shown that L1 speakers can integrate different types of information from one NP to anticipate the referent of an upcoming NP (Hopp, 2015; Kamide, Altmann, & Haywood, 2003; Kamide, Scheepers, & Altmann, 2003), we expect them to be able to combine semantic and morphosyntactic cues *across* different constituents to predict upcoming information. In contrast, for L2 speakers, given their general difficulty in exploiting morphosyntactic cues, the combination of semantic and gender cues is predicted to be too

demanding, and thus we do not expect evidence of successful cue combination.

5. Experiment 1 — L1 speakers (materials from Van Bergen and Flecken, 2017)

5.1. Methods

5.1.1. Participants

Following a convenience sampling approach, the study was advertised among L1-German students at the author's university. Participants qualified as L1-German speaker if they had started acquiring German before age three. The 28 L1 German participants received course credit. The data of six participants were excluded because of technical or comprehension issues during data collection. The final sample includes 22 L1 German speakers (all female, mean age = 20.4, *SD* = 1.8), which is similar to the sample size of similar visual-world-paradigm eye-tracking studies reported above (e.g. study 3 in Kamide, Altmann, & Haywood, 2003; study 2 in Kamide, Scheepers, & Altmann, 2003). However, our design included three times more items per condition than these previous studies, leading to higher power. All participants had normal or corrected-to-normal vision and did not report any auditory impairment.

5.1.2. Materials

We adapted Van Bergen and Flecken's (2017) items to the purpose of our study, which were provided by Monique Flecken. We developed 96 German SVO sentences in the past tense, such as *Die Frau legte vor einer halben Stunde den durchschnittlich großen Kerzenständer auf den Tisch* (*The woman put.*$_{LIE.PAST}$ *half an hour ago the.*$_{MASCULINE}$ *average-sized.*$_{MASCULINE}$ *chandelier on the table*, 'The woman put the average-sized chandelier on the table half an hour ago'). They all followed the structure: subject noun + verb (marked for past tense) + seven-syllable temporal phrase + definite determiner (marked for accusative case[1] and feminine or masculine gender) + five-syllable adjective phrase (marked for accusative case and feminine or masculine gender) + target object noun (which, depending on the condition, could or could not be predicted) + locational phrase. Each sentence was paired with a four-image display — see Materials in Supplementary Materials, accessible via https://osf.io/dg5cq/. Forty-eight images of common household objects (placed on a table or a chair) created by Van Bergen and Flecken (2017) were reproduced but were recombined into original displays for this study. Four experimental conditions and a baseline condition were constructed to contrast the information available in the input that enables prediction of the sentence-final object.

For the 32 items of the "Semantic + Gender Condition", each display contained one object drawn once lying and once standing (SL object), another object in a lying position (L object), and another object in a standing position (S object) — see Fig. 1. The L objects are prototypically conceptualised as lying for German speakers as they cannot stand on their own and are thus used with *legen* ('put.$_{LIE}$') due to their intrinsic properties. Such intrinsic property does not exist for S objects in German, but all S objects in our displays have a natural base they rest on and should thus be used with *stellen* ('put.$_{STAND}$'). In this condition, the target was one of the SL objects and was grammatically feminine or masculine. Its duplicate was represented in the other position (standing or lying) and regarded as the gender competitor. The position competitor shared its position but not its gender with the target. The distractor was an object sharing neither its position nor its gender with the target (always neuter gender). In this condition, the target could only be anticipated by maintaining the semantic information from the verb and combining it with the gender information from the determiner of the following NP — i.e., the target cannot be uniquely predicted from either cue alone.

The "Semantic Condition" included 24 items. In these sentences, the lexico-semantic information of the verb was constraining (i.e., enabled to predict the object of the sentence). Two sub-conditions were distinguished. In the "Semantic-PlacementV Condition", the twelve sentences involved a placement verb (either *stellen* 'put.$_{STAND}$' or *legen* 'put.$_{LIE}$') and the position of the target object was different from the position of the three other objects in the display. In the twelve sentences of the "Semantic-OtherV Condition", another constraining verb (e.g., *aß*, 'ate') was used that enabled prediction of the target object[2] (e.g., only one object in the display was edible.), following standard VWP designs in which verb semantics restrict the set of possible referents (e.g., Kamide, Altmann, & Haywood, 2003). The specific verbs were selected by the author on the basis of this prior literature to meet the following criteria: they belong to the "core vocabulary for German learners"[3] (Tschirner & Möhring, 2020); they were semantically compatible with exactly one object in the visual display; and they were semantically incompatible with the remaining three objects, based on the selectional restrictions of the verbs – i.e., the semantic constrains on their permissible arguments. The "Gender Condition" also included 24 items. Here, the verb semantics was non-constraining (e.g., *sah*, 'saw') but the gender marking of the determiner (feminine or masculine) enabled to predict the object (e.g., the target is grammatically feminine while the three other objects in the display are masculine). In the 16 items in the "No-prediction condition", neither the verb semantics (e.g., *vergaß*, 'forgot') nor the gender marking of the determiner enable to predict the object of the sentence (i.e., all objects are either feminine or masculine).[4] All sentences were recorded by a female L1 German speaker and processed in Audacity to align the verb onset of all sentences 700 ms after sentence onset, the determiner onset after 4000 ms and the target object onset after 5500 ms.

5.1.3. Procedure

The study was conducted using an Eyelink Portable Duo eye-tracker in the eye-tracking lab at the author's department. Participants were seated approximately 70 cm from a 1920 × 1080 resolution screen. Eye movements from the dominant eye (identified via the Miles test) were recorded at a 1000 Hz sampling rate with a head stabiliser. After providing written consent to participate in this study (approved by the ethics committee of the author's university on March 2022, approval number 17/2022), participants completed the German version of the LexTALE (Lemhöfer & Broersma, 2012), which will not be further discussed in this contribution. Participants were then familiarised with the drawings of the objects involved in the experimental task. Each object was presented on the screen for 2500 ms while participants heard its label (with gender-marked definite article in the nominative case) via two external speakers. The labels had been recorded one by one by the speaker who also recorded the experimental items. The experimental task began with a 9-point calibration. We aimed for accuracy below 0.5° (and repeated calibration throughout the experiment if needed).

[1] Note that determiners are also marked for case in German. In all our sentences, accusative case was used, so feminine-marked determiners consistently had the form die and masculine-marked determiners consistently had the form den. Accusative-marked determiners are syncretic with nominative in the feminine, but not in the masculine. However, follow-up analyses revealed no difference in the predictive use of gender marking for feminine versus masculine target objects, so we do not further pursue this distinction here.

[2] Note that most targets had feminine or masculine gender, but seven items included a neuter target. This could not be avoided because the choice of objects was limited by the objects included in Van Bergen and Flecken's (2017) materials. However, this should not bias the results since the gender of the target is irrelevant in this condition.

[3] Listed in the 2000 most commonly used words of German.

[4] Throughout the task, when one object was represented in two positions in the display, it was not always depicting the target object, thus not enabling to predict that, once an object was represented twice, the target was necessarily one of these two pictures.

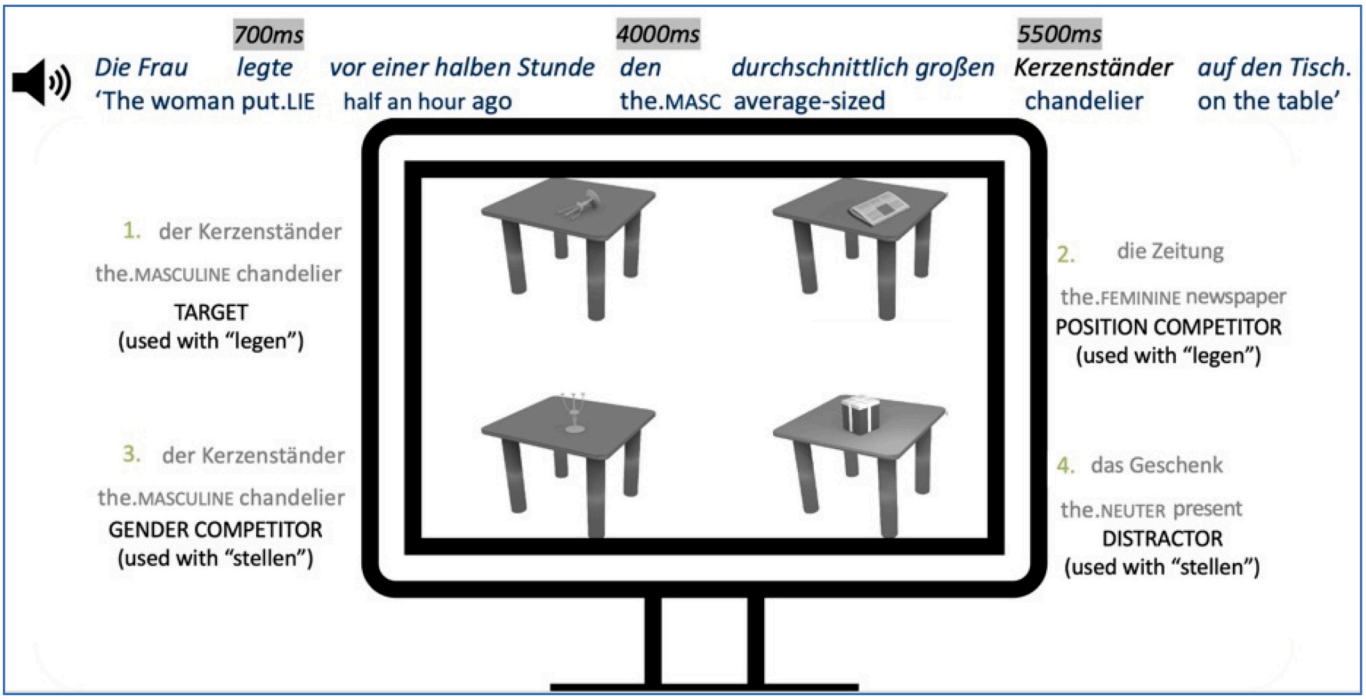


Fig. 1. Example of a display paired with an auditory sentence in the "Semantic + Gender Condition", Experiment 1. Participants would only see the display on the screen (without annotations) and hear the sentence over speakers.

Participants were instructed to listen carefully to the sentences while looking at the screen. They were informed that they would occasionally be asked questions about what they had heard or seen. Before beginning the main task, they were exposed to two practice items and given the opportunity to ask any remaining questions. The experimenter then remained quietly seated behind a two-meter-high partition within the same room. The experimental phase consisted of 96 items, each trial comprising a 500 ms fixation cross, followed by a 3000-ms preview of the visual display, followed by the auditory sentence, which was presented via two external speakers while the display remained on screen. Twenty-five percent of the items were followed by an attention-check question to be answered via a button box. Items were presented in random order and in four blocks, allowing for three self-timed breaks. During the second break, participants completed a socio-demographic questionnaire.

After the experimental task, participants completed a "recall task" designed to verify both their identification of the pictograms and their knowledge of the grammatical gender of the depicted objects. They were presented again with the drawing of each object used in the study and asked to type the definite article and the label of that object on the keyboard. In a final debriefing, participants were asked what they thought the study investigated and whether they had any additional comment. The entire procedure lasted approximately 55 min.

5.1.4. Data pre-processing and analysis

Attention questions were answered with 96% accuracy, indicating that participants paid attention during the experiment. No participant was excluded due to low accuracy in the attention checks. Regarding responses from the recall task, 92% of the objects were correctly identified (with correct article). However, the outcomes revealed slight confusion patterns, including *die Pfanne* ('the.FEM pan'), which 12 participants recalled as *der Topf* ('the.MASC'), *die Milchtüte* ('the.FEM milk carton'), which 8 participants recalled as *der Milchkarton* or *der Milchbeutel* ('the.MASC milk carton'), and *der iPod* ('the.MASC iPod'), which 8 participants recalled as *das iPod* ('the.NEUTER iPod'). While the grammatical gender of *Milchtüte* was not decisive for any item, the gender of *Pfanne* and *iPod* was decisive for a few items in the Semantic + Gender in the Gender Condition. Moreover, subsequent visual analyses of the raw data revealed that objects that were relatively big in size compared to other objects and depicted in an unusual position (i.e. lying chandelier, lying lamp and standing megaphone) particularly attracted participants' gaze. However, further analyses with a restricted dataset excluding items with wrongly remembered objects or with "gaze-attracting" objects did not show meaningful shifts in the revealed results, except for the significant cluster in the Gender condition turning from significant to marginally significant, arguably due to differences in statistical power, and one cluster in the Semantic + Gender conditions turning from marginally significant to significant. Therefore, the results reported below include the full dataset. The results for the restricted dataset are reported in Fig. 1 in the Supplementary Data.

The eye-tracking data were down-sampled to 500 Hz and exported using the sample report in the Data Viewer software (SR Research Ltd., version 2.2.38). Equally sized areas of interest (AOIs) were designed around each of the four surfaces (table or chair) on which each object was placed. Each sample was coded as looks toward one of the four objects (target object, position competitor, gender competitor, distractor) or looks outside of these regions. Prediction was operationalised as a higher proportion of looks to the target in each of the experimental conditions compared to the No-prediction condition in the prediction window. In the Semantic + Gender Condition and the Semantic Condition, the prediction window ranged from 200 milliseconds (ms) after verb onset to 200 ms after object onset — as it is assumed that speech takes 200 ms to influence eye-movement control (Salverda et al., 2014). In the Gender Condition, the prediction window ranged from 200 ms after determiner onset to 200 ms after object onset. Note that, following Van Bergen and Flecken (2017), we also analysed eye movements in the integration window, i.e. from object onset to 700 ms after object onset. In this way, even in the absence of a "pure" prediction effect before the object appears in the input, we could investigate any potential advantage in terms of integration of the incoming linguistic input with the previously encountered information when the referent of the object could be pre-activated. This would thus still indicate pre-activation of the object of the sentence.

Data were analysed in R, version 4.2.1 (R Core Team, 2022) and

RStudio, version 2022.12.0 + 353 (RStudio Team, 2020) — see analysis scripts in Supplementary Materials. The VRWPre package (Porretta et al., 2017) was used for data pre-processing and visualisation. Blinks (7.2%) and off-screen samples (1.58%) were marked as trackloss and events containing more than 25% of trackloss were excluded. Due to important patterns in residuals which could not be reduced with different model specifications, the data could not be analysed with the originally planned GAMMS, which would have enabled to statistically compare the time-course of effects in different conditions. Instead, cluster-based permutation analyses (CPA) involving *t* statistics (Maris & Oostenveld, 2007) were performed with the EyetrackingR package (Forbes et al., 2023) based on an adapted version of Koch et al.'s (2021) script. This non-parametric procedure compares the proportion of looks to the target in the No-prediction condition with the proportion of looks to the target in each of the experimental conditions (Semantic + Gender, Semantic-PlacementV, Semantic-OtherV, and Gender) in order to identify time clusters showing significant anticipation due to the presence of prediction cues. Condition was contrast-coded, with No-prediction as baseline level. First, the participant-averaged gaze data are segmented in 50 ms bins, and a paired *t*-test is conducted between the proportions of looks to the target in the experimental condition and in the No-prediction condition in each bin. Contiguous time bins showing a significant difference are then grouped together to form preliminary clusters. Then, to assess whether these clusters could have emerged by chance, 2500 permutations are run. In each permutation, condition labels are randomly reassigned, the same *t*-test procedure is repeated, and new clusters are identified. By comparing the preliminary clusters to this distribution of random clusters from the permuted datasets, a probability estimate is obtained for each preliminary cluster, indicating whether the cluster resulted due to chanced. Clusters found to be statistically significant in the prediction or integration time window were interpreted as evidence of anticipatory processing.[5] Four CPAs (comparing each of the experimental condition with the No-prediction condition) were run separately.

5.2. Results

The different CPAs run in the prediction-and-integration window (i.e., from 200 ms after verb onset [900 ms] or determiner onset [4200 ms] to 700 ms after object onset [6200 ms]) revealed significant clusters detailed in Table S1 in the Supplementary Data. Fig. 2 visualises the data from the prediction-and-integration window including the (marginally) significant clusters identified in the CPAs. Compared to the No-prediction Condition, the Semantic Condition with placement verbs revealed two significant cluster (3000–4100 ms, $p = .013$ and 4600–6100 ms, $p < .001$), the Semantic Condition with other constraining verbs revealed one significant cluster (1600–6100 ms, $p < .001$), the Gender Condition revealed one significant cluster (5400–6100 ms, $p = .025$) and the Semantic + Gender Condition revealed two marginally significant clusters (4600–5200 ms, $p = .069$ and 5400–6000 ms, $p = .056$). Thus, the semantics of placement verbs, the semantic of non-placement verbs, and the gender marking of the determiner allowed some pre-activation of the target object, showing that L1 speakers used these cues predictively. Strong conclusions cannot be drawn about the combination of semantic + gender cues, although the data may indicate some trend toward the predictive use of these combined cues. Although CPAs do not enable to statistically test for the onset and offset of effects, the mean difference of proportion of looks to target between each experimental condition and the No-prediction condition in the significant clusters, together with the computed Cohen's *dz* – reported in Table S1 – suggests that the semantics of placement verbs (Fig. 2, panel A) yields small prediction effects (mean proportion difference in each cluster = 0.1 and 0.12, Cohen's $dz = 0.91$ and 0.74, respectively), whereas the semantics of other constraining verbs (panel B) yields stronger effects (mean proportion difference = 0.25, Cohen's $dz = 1.53$). Similarly, the gender marking of determiners (panel C) appears to yield rather limited pre-activation of the object of the sentence (mean proportion difference = 0.09, Cohen's $dz = 0.69$). Finally, a marginally significant trend suggests that predicting based on the combination of the semantics of placement verb with the gender marking of determiners might happen to a small extent (panel D), but this marginally significant trend requires further investigation.

5.3. Discussion

Experiment 1 investigated whether L1 German speakers predict the object of an SVO sentence based on (a) the semantics of placement verbs, (b) the semantics of other verbs, (c) the gender marking of the determiner, or (d) the combination of semantics and gender marking as predictive cue in a context where these sources of information alternate in serving as the informative predictive cue. Participants launched predictive eye movements to the target object based on the semantics of "other" (non-placement) verbs and, although to a small extent, based on the semantics of placement verbs. This finding is consistent with well-established evidence on the use of verb semantics as predictive cues (e.g., Altmann & Kamide, 1999). It also replicates Van Bergen and Flecken's (2017) results, though under conditions where placement verb semantics was less available and reliable as a prediction cue. Specifically, in our design, less than half of the items contained a (constraining) placement verb, whereas van Bergen and Flecken used (constraining) placement verbs (*leggen*, 'put.$_{\text{LIE}}$' or *zetten*, 'put.$_{\text{STAND}}$') in two-thirds of their stimuli, with the remaining items containing another (non-constraining) placement verb (*plaatsen*, 'put'). This difference likely increased the overall salience and reliability of placement verb semantics as a prediction cue in their study, or may even have heightened participants' awareness of the study focus, potentially altering their natural processing behaviour. Although the strength of Van Bergen and Flecken's (2017) effects can only be estimated descriptively from their visualisations (see Fig. 2, p. 31), the magnitude of the effect in our data appears comparable to theirs. Our findings thus extend the available evidence suggesting that the content of predictions can also concern perceptual features of objects (Rommers et al., 2013).

Gender marking was also used to pre-activate the target object, as the proportion of looks to the target increased when gender marking could be used as a prediction cue. However, this effect was small. Although absence of reliable gender-based prediction in L1 has been reported previously (Kochari & Flecken, 2019; Nieuwland et al., 2018), our findings contrast with studies showing clear, robust effects of gender-based prediction (Dahan et al., 2000; Huettig & Janse, 2016; Otten & Van Berkum, 2009; Wicha et al., 2004). Several methodological factors may explain this discrepancy. First, our displays included more competitors than most previous VWP studies (e.g., Dussias et al., 2013), thereby increasing competition. Second, the overall cue availability and reliability of gender marking were lower in our materials: gender marking could serve as a prediction cue in only 20% of the items, compared to at least 50% in earlier work (Hopp & Lemmerth, 2018; Huettig & Janse, 2016). Third, our sentences, and thus our prediction windows, were substantially longer than comparable studies (e.g., Hopp, 2015; Van Bergen & Flecken, 2017). While longer sentences may allow more time for prediction to occur, they also impose greater demands on attention and working memory, potentially reducing resources available for predictive processing (e.g., Ding et al., 2023; Ito et al., 2018) and potentially preventing prediction due to excessive

[5] The cluster-based permutation analyses reported here are group-level, participant-based analyses: time courses were first aggregated by participant, and statistical inference is therefore based on between-participant variability. As a consequence, item-level variability is not modelled in this framework, which is a known limitation of standard CPA approaches. The present results should therefore be interpreted as population-level effects across participants rather than item-level effects.

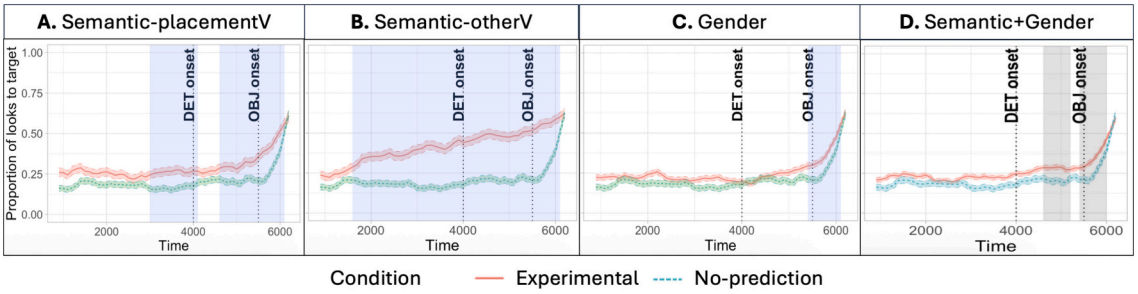


Fig. 2. Proportion of looks to the target picture over time (prediction-and-integration window) in the experimental conditions and the No-prediction condition. Vertical dotted lines represent the determiner onset (4000 ms) and the object onset (5500 ms). Error bands show 95% confidence intervals. Shaded areas indicate time clusters during which there was a significant (blue) or marginally significant (grey) difference between experimental and No-prediction conditions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

metabolic costs (Kuperberg & Jaeger, 2016).

Beyond the use of lexico-semantic cues and gender cues individually, the data also offer preliminary indications that L1 speakers may be able to combine verb semantics and gender marking across constituents to predict the referent of the object of a sentence. However, these effects were only marginally significant. As previous studies have demonstrated the ability to integrate lexico-semantic and morphosyntactic information across constituents for prediction (Hopp, 2015; Kamide, Altmann, & Haywood, 2003; Kamide, Scheepers, & Altmann, 2003), it is plausible that the relatively long, slow-paced, and information-rich sentences in our experiment strained participants' cognitive resources, limiting the extent to which predictive processing could occur.

Experiment 1 is not without limitations. First, the task design was constrained by our adaptation of materials from Van Bergen and Flecken (2017), which led to a few difficulties: (1) while the size of the AOIs was constant within displays, the size of the relevant objects on the chair or table varied within displays, potentially biasing eye movements; and (2) some pictograms were not unambiguously recognised, as indicated by the recall task results. Moreover, we did not measure participants' cognitive resources, despite evidence that working memory capacity influences predictive use of gender cues (Huettig & Janse, 2016; Otten & Van Berkum, 2008). Including such measures could help explain variability in the individual and combinatory use of lexico-semantic and morphosyntactic information to form predictions.

In sum, Experiment 1 provides initial evidence that L1 speakers use both semantic and gender information to generate predictions. The semantics of non-placement verbs appears to trigger a (descriptively) rapid and substantial increase in proportion of looks to the target upon verb onset, whereas the patterns for the other types of cues show less consistent increases. The findings thus raise new questions about why placement verb semantics led to what appears as only modest predictive behaviour and why gender marking did not elicit a stronger increase in predictive looks to the target. In order to further explore these questions and to overcome the limitations of Experiment 1, Experiment 2 was designed with new materials for the prediction task, the inclusion of a working memory measure, and the addition of an L2 sample.

6. Experiment 2 — L1 and L2 speakers (own materials)

6.1. Methods

6.1.1. Participants

Following a convenience sampling approach, the study was advertised among L1-German students at the author's university as well as through social media posts and e-mails targeting adult German speakers and French-German sequential bilinguals. Forty-five adults who did not participate in Experiment 1 took part in exchange for monetary compensation or course credits. Data from six participants were excluded due to comprehension or technical issues during data collection. The final sample, described in Table 1, includes 20 German L1 speakers and 19 L1-French/L2-German sequential bilinguals (L2 speakers). Participants qualified as L1-German speaker if they had started acquiring German before age three and as L2-German speakers if they had not started learning it before age six. Note that they were only invited if they declared "being at least able to hold a basic daily conversation in German (at least at B2 level)". All participants were residing in Germany, and L2 speakers had been residing in Germany for at least four months (mean length of stay = 2.8 years, SD = 3.3). The self-reported proficiency levels were advanced among L1 speakers and ranged from upper-intermediate to advanced among L2 speakers. The mean scores on the lexical test LexTALE[6] (Lemhöfer & Broersma, 2012) confirmed a substantial proficiency difference between L1 speakers (M = 91, SD = 8) – corresponding to a (very) advanced average proficiency level – and L2 speakers (M = 64, SD = 6) – corresponding to an upper-intermediate (CEFR-B2) average proficiency level, t (32) = 10.4, p < .001. A Welsh t-test did not reveal any significant difference between the working memory scores obtained by the L1 and L2 speakers, t (31.7) = 0.99, p = .329. All participants had normal or corrected-to-normal vision and did not report any auditory impairment.

6.1.2. Materials

As in Experiment 1, we developed 96 SVO sentences, such as *Mia sah eben die kleine Lampe in dem Abstellraum* (*Mia see.*$_{PAST}$ *just the.*$_{FEMININE}$ *small.*$_{FEMININE}$ *lamp in the storeroom*, 'Mia just saw the small lamp in the storeroom'), which all followed the structure: subject noun + verb (marked for past tense) + two-syllable temporal phrase + definite determiner (marked for accusative case and feminine or masculine gender) + two-or-tree-syllable adjective phrase (marked for accusative case and feminine or masculine gender) + target object noun (which, depending on the condition, could or could not be predicted) + locational phrase. The length of the sentences was reduced compared to Experiment 1 to avoid artificially long sentences with structurally complex constituents, potentially straining auditory working memory. Each sentence was paired with a four-image display. New objects were drawn for this study to overcome the limitations of Van Bergen and Flecken's (2017) materials highlighted above and to be able to include only objects that are feminine or masculine in both French and German (i.e., congruent gender). Forty-two black-and-white drawings of common household objects were included — see pictures, displays and sentences in Supplementary Materials. In order to avoid size and perspective differences between the objects, they were not placed on a support and were all of similar size, contrary to Experiment 1 and Van Bergen and Flecken's (2017) study. The same conditions were used as in Experiment 1. Twenty-four stimuli were developed for each of the three

[6] Note that due to technical issues during data collection, three L2 speakers and two L1 speakers have missing data for the LexTALE.

Table 1
Sample descriptives.

Sample	N	Age		Female	Working memory [0–14]		Self-rated proficiency [0–5]		LexTALE proficiency [0–100]	
		M	*SD*	%	*M*	*SD*	*M*	*SD*	*M*	*SD*
L1 speakers	20	27.9	8.4	35	8.3	2.3	4.9	0.2	91	8
L2 speakers	19	24.1	3.3	63	7.7	1.4	3.9	0.7	64	6

Note. Working memory measured by the backward digit span test from Wechsler Adult Intelligence Scale-Revised in the participants' L1 (WAIS-R, Wechsler, 1981). Minimum and maximum scores possible on each variable reported between squared brackets.

experimental conditions and the No-prediction condition — see Supplementary Materials. The displays and sentences in each condition were further created following the same criteria as in Experiment 1, except that half of the items in the "gender" condition involved a non-constraining placement verb (e.g., *legen* '*put.*$_{LIE}$' with all four objects on the display being in horizontal position) and the other half involved another non-constraining verb (e.g., *sah*, 'saw'). All sentences were recorded by a female L1 German speaker and processed in Audacity to align the verb onset of all sentences 700 ms after sentence onset, the determiner onset after 2200 ms and the target object onset after 2900 ms.

6.1.3. Procedure

The procedure closely followed that of Experiment 1, with the exception of a minor adaptation to the familiarisation task and the inclusion of a brief measure of working memory capacity. Prior to the experimental task, participants were familiarised with the drawings to minimise the risk of misinterpretation or confusion due to unfamiliar vocabulary (especially for the L2 speakers). The experimenter presented each object individually, naming it aloud (in German) with its definite determiner, which was marked for grammatical gender. Subsequently, the same objects were shown again one by one in a random order, and participants were asked to name each object along with its corresponding determiner. Objects that were initially mislabelled were presented again until participants named them correctly. This procedure was repeated across three sets of objects, covering a total of 42 objects. Only after participants had correctly named all items together with their determiners were they allowed to proceed to the experimental phase. This ensured that all participants correctly identified the intended objects depicted in each picture and knew both the lexical label and grammatical gender of each object in standard German, a prerequisite for predictive processing. After the experimental task, participants completed a backward digit span test in their L1 (*WAIS-R*, Wechsler, 1981), comprising 14 sequences across seven levels (two trials per level). Numbers were played from pre-recorded files at a rate of one digit per second. The test began with two-digit sequences, increasing by one digit per level, and was terminated if both trials at a given level were failed.

6.1.4. Data pre-processing and analysis

All participants answered between 22 and 24 attention questions correctly, except for three L2 speakers who gave only 20 correct answers. However, their data were not disregarded since a qualitative analysis of their eye movements showed that they sustained attention throughout the experiment. The eye-tracking data were preprocessed and analysed using the same procedure as in Experiment 1. Blinks (4.5%) and off-screen samples (0.9%) were marked as track loss and events containing more than 25% of track loss were excluded.

6.2. Results

Among L1 speakers, CPAs revealed significant time clusters in the prediction-and-integration time window (i.e., from 200 ms after verb onset [900 ms] or determiner onset [2400 ms] to 700 ms after object onset [3600 ms]) in all four experimental conditions — see Table S2 in Supplementary Data for a summary of the CPA outcomes and Fig. 3 for visualisations. These significant clusters indicate that L1 speakers pre-activated the objects of the sentence in the presence of all four types of prediction cue. Fig. 3 shows that the semantics of placement verbs (panel A, top row) yielded a weak and late increase in predictive looks to the target prediction effects (significant cluster: 2800–3700 ms, $p < .001$, mean difference in proportion of looks to target between the experimental and No-prediction condition = 0.13, Cohen's $dz = 1.32$), whereas the semantics of other constraining verbs (panel B, top row) yielded a stronger increase upon verb onset (significant cluster: 1300-3700 ms, $p < .001$, mean difference in proportion of looks to target = 0.28, Cohen's $dz = 1.65$). Similarly, the gender marking of determiners (panel C, top row) yields a weak increase in proportion of looks to the target (significant cluster: 2500–3350, $p = .001$, mean difference in proportion of looks to target = 0.11, Cohen's $dz = 0.94$). Finally, L1 speakers are able to combine the semantics of placement verb with the gender marking of determiners (panel D, top row) to predict the target object (significant cluster: 1950-3350 ms, $p < .001$, mean difference in proportion of looks to target = 0.16, Cohen's $dz = 1.19$).

Among L2 speakers, the only type of cue used to predict the upcoming object was the semantics of "other" (non-placement) verbs (significant cluster: 2500–3700 ms, $p = .002$, mean difference in proportion of looks to target = 0.13, Cohen's $dz = 0.85$). L2 speakers did not use the semantics of placement verbs nor the gender marking of determiners to predict the object of the sentence. Moreover, the significant cluster toward the end of the integration window in the Semantic + Gender Condition (2450–3700 ms, $p = .036$) indicates that L2 speakers looked more to the target upon hearing it in the No-prediction condition than in the Semantic + Gender Condition. This stems from the fact that L2 speakers continued looking to both versions of SL object (i.e., the object depicted in both standing and lying position) until the end of the sentence, suggesting difficulty for our L2 speakers (i.e., predominantly at upper-intermediate level) to integrate the semantics of the object when both a lying and a standing version of the object were depicted in the display. Note that descriptive visual analyses of the raw data did not reveal any meaningful pattern based on WMC nor based on L2 proficiency[7] — see Fig. S2 (WMC), Fig. S3 (self-rated proficiency), and Fig. S4 (LexTALE score) in Supplementary Data.

6.3. Discussion

Experiment 2 investigated L1 German speakers' and L1-French/L2-German sequential bilinguals' individual and combinatory use of lexico-semantic and morphosyntactic cues (i.e., gender marking) to

[7] We report only visual descriptive explorations since these variables were not part of our research questions and the subsample sizes (high group vs low group) are very small. For instance, exploratory CPAs in the high-LexTALE score group ($n = 8$) revealed the same pattern as in the entire L2 sample, i.e., no significant cluster was revealed except in the Semantic_OtherV condition. These exploratory analyses on a tiny small sample size are thus not further discussed. Note that exploratory linear mixed models in the last 500-ms bins similarly did not provide any evidence of the effect of Lextale scores on proportion of looks (see Table S3 in Supplementary Data) and exploratory CPAs including only those L2 speakers with a Lextale score in the B2 range (60 to 80% score) revealed the same patterns as the CPAs run on the entire L2 sample (see Table S4 in Supplementary Data).

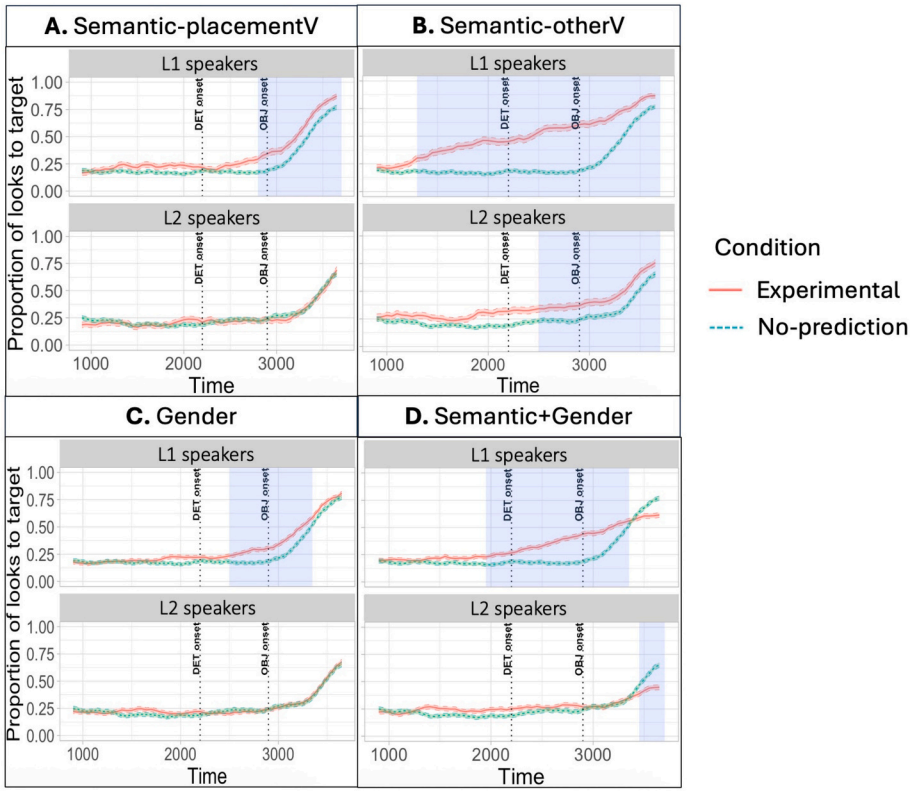


Fig. 3. Proportion of looks to the target picture over time (prediction-and-integration window) in each of the experimental conditions compared to the No-prediction condition (panels A to D) among L1 speakers (top) and L2 speakers (bottom). Vertical dotted lines represent the determiner onset (2200 ms) and the object onset (2900 ms). Error bands show 95% confidence intervals. Shaded areas indicate time clusters during which the CPAs revealed a significant (blue) or marginally significant (grey) difference between experimental and No-prediction conditions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

predict the referent of an upcoming noun in German in a context where these sources of information alternate in serving as informative predictive cue. As in Experiment 1, the L1 speakers used verb semantics to predict the object of an SVO sentence, although, "other" (non-placement) verbs (e.g., *essen*, 'eat'; *entsperren*, 'unlock') yielded stronger increase in predictive looks to the target than placement verbs (*legen*, 'put$_{LIE}$', *stellen*, 'put$_{STAND}$'). These trends will need to be confirmed in future research, but we speculate that this observed discrepancy may reflect different underlying mechanisms, namely *prediction-by-association* – as "other" (non-placement) verbs are semantically or associatively related to the target and may thus prime it – vs. *prediction-by-production* – as the semantics of placement verbs does not prime the target. Future studies will have to clarify whether L1 speakers' prediction based on semantic cues is indeed more limited when the cue does not semantically prime the target (e.g., *put.$_{STAND}$–plant*) than when it does (e.g., *water–plant*). However, a limitation of our design is that in the Semantic condition with non-placement verbs, the target was the only object that could plausibly serve as the patient of the action (e.g., only one object could be watered). In contrast, in the Semantic condition with placement verbs, nearly all four displayed objects could potentially serve as the patient of a put.$_{STAND}$ or put.$_{LIE}$ action — except for objects that can only lie (e.g., a sausage). This potentially required participants to engage in more detailed processing of the visual information to assess the spatial properties of each object. Since it has been shown that increased visuospatial working memory load weakens semantic prediction effects (Allison et al., 2025), the additional visual processing demand in the Semantic-placementV condition may have taxed the cognitive resources available for prediction and contribute to the smaller effects observed for placement verbs.

In addition, the results provide evidence for L1 speakers' use of gender marking to predict, although this effect is relatively modest. The observation that gender marking leads to lower proportions of predictive looks to the target than the semantics of (non-placement) verbs will need to be statistically confirmed in future research. If confirmed, it would align with the *Good-Enough* approach to language comprehension, which posits that the parser tends to rely more on semantic heuristics than on syntactic computation when the context does not allow for a detailed analysis of the input, e.g., with long and complex sentences. Two factors may have contributed to the observed weakness of gender-based prediction compared to previous findings. First, the relatively low reliability of gender cues in our design compared to earlier work may have weakened the strength of this cue, leading participants to rely more on other cues. Second, the prediction window was shorter in the Gender condition than in the Semantic condition, possibly providing insufficient time for robust predictive eye-movements before the onset of the object. Experiment 3 was thus designed to further investigate these speculations.

Beyond the separate use of lexico-semantic cues and gender cues, Experiment 2 also shows that L1 speakers can combine the verb semantics with the gender information of the determiner to predict the referent of the object of a sentence. This extends previous evidence (Hopp, 2015; Kamide, Altmann, & Haywood, 2003; Kamide, Scheepers, & Altmann, 2003) that L1 speakers can combine lexico-semantic information and morphosyntactic information *across* different constituents to form predictions — even when the constituents of the sentences are relatively complex. Although speculative, the tendency for combined cues to elicit a higher proportion of predictive looks than either cue alone may suggest that, once one type of information (semantic or morphosyntactic) has been processed, integrating the other does not impose a substantially higher cost for L1 comprehenders.

Experiment 2 also replicated previous findings (e.g., Chun & Kaan, 2019; Kim & Grüter, 2021; Schlenter, 2023; but see Dijkgraaf et al.,

2017) showing that L2 speakers can use the semantics of constraining verbs to predict the referent of an upcoming word. However, our data do not reveal evidence that L2 speakers whose L1 does not encode the semantic distinction between placing an object vertically or horizontally, used placement verb semantics to generate predictions, consistent with Van Bergen and Flecken (2017). One limitation that will need to be accounted for in future research, however, is that we did not control for their explicit (receptive or productive) knowledge of the *stellen/legen* semantic contrast — although no participant mentioned in the post-experimental debriefing that they did not understand the content of the sentences. Finally, our data did not provide any evidence that L2 speakers used gender marking, either alone or in combination with placement verb semantics, to generate predictions. This was the case despite gender congruency across L1 and L2 for all nouns included in the task, as well as participants' explicit knowledge of noun gender — recall that they were allowed to proceed to the prediction task only after successfully naming all objects with their correct gender in the familiarisation task. While aligning with previous findings (Grüter et al., 2012; Lew-Williams & Fernald, 2010), our results contrasts with studies reporting L2 speakers' predictive use of gender marking (Dussias et al., 2013; Hopp & Lemmerth, 2018) – although L2 speakers' selective and unexplained use of feminine but not masculine gender found in Dussias et al. (2013) raises questions about the generalisability of the results – and other morphosyntactic cues encoded in the L1 (e.g., Schlenter & Felser, 2021). It is worth noting, however, that while all nouns included in our task had congruent gender across German and French, grammatical gender is not always congruent across these two languages, which might have prevented L2 speakers to rely on this cue altogether (but see Hopp & Lemmerth, 2018). Overall, our findings suggest that prediction in L2 is far from systematic.

To investigate the unexpectedly weak prediction effects based on gender cues among L1 speakers, Experiment 3 was designed to test whether increasing the reliability of gender cues in the task and extending the prediction window would lead to stronger predictive effects among L1 speakers.

7. Experiment 3 — L1 speakers (gender marking as only prediction cue)

7.1. Methods

7.1.1. Participants

Following a convenience sampling approach, the study was advertised among L1-German students at the author's university. Thirty adults who did not participate in Experiment 1 or 2 took part in return for monetary reward or course credits. However, the data of five participants were excluded because of comprehension or technical issues during data collection. The final sample includes 25 German L1 speakers (L1 speakers), including 21 females, 11 males and 2 participants identifying as gender diverse, with a mean age of 27.6 (SD = 11.4) and a mean working memory of 8 (SD = 2.1) as measured by a backward digit span test (WAIS-R, Wechsler, 1981) where the maximum score is 14. All participants had normal or corrected-to-normal vision and did not report any auditory impairment.

7.1.2. Materials, procedure, data pre-processing and analysis

The materials (sentences + displays) from the Gender condition and the No-prediction condition from Experiment 2 were reused (24 items per condition), with one adaptation: in order to make the prediction window longer, the two-syllable adjective phrase between the determiner and the target object was lengthened to five syllables (i.e., as in Experiment 1). The verb onset of all sentences was aligned at 700 ms after sentence onset, the determiner onset at 2200 ms and the target object onset at 3700 ms. An example item is *Noah sah gestern die ziemlich neue Zahnbürste in dem Badezimmer* (*Noah see.*$_{PAST}$ *yesterday the.*$_{FEMININE}$ *rather new.*$_{FEMININE}$ *toothbrush in the bathroom*, 'Noah saw yesterday the rather new toothbrush in the bathroom'). This means that half of the items had a gender cue enabling prediction and half of the item did not, raising the reliability of the gender cue to 50% compared to 20% in Experiment 1 and Experiment 2. The procedure, data pre-processing and analysis were the same as in Experiment 1.

7.2. Results

As visualised in Fig. 4 (Panel A) and reported in Table S5 in the Supplementary Data, the CPA comparing the proportion of looks to the target in the prediction-and-integration window (i.e., from 200 ms after determiner onset [2400 ms] to 700 ms after object onset [4300 ms]) between the Gender condition and the No-prediction condition only revealed a significant cluster ranging from 4200 ms to 4500 ms, p = .047 (mean difference in proportion of looks to target between the experimental and No-prediction condition = 0.1, Cohen's dz = 0.7). This suggests some (weak) pre-activation of the target object. Although remaining at the descriptive level, the pattern observed in the raw data could be interpreted as an integration advantage of gender cues rather than "pure" prediction of the target given the very late differentiation between the proportion of looks to the target between the Gender and the No-prediction conditions. As the size of this effect, as reflected in Cohen's dz, is – contrary to expectations – weaker than in Experiment 2, further exploratory analyses were conducted. CPAs did not reveal any effect of masculine versus feminine gender (Dussias et al., 2013). We also exploratorily investigated a potential effect of working memory. Based on median split ($median$ = 8), a high-working-memory group (n = 10, mean working memory score = 9.9, SD = 1.2) and a low-working-memory group (n = 15, mean working memory score = 6.7, SD = 1.4) were created based on the scores obtained in the backward digit span test.[8] These analyses revealed a marginally significant cluster ranging from 3300 ms to 3600 ms, p = .089 in the high-working-memory group and a marginally significant cluster ranging from 4300 to 4500, p = .088 in the low-working-memory group. Although only marginally significant, we report these results given the exploratory nature of the analyses and the small sample sizes precluding high statistical power, although they will need further exploration in future research.

7.3. Discussion

Experiment 3 examined L1 German speakers' use of grammatical gender as a cue to predict the referent of an upcoming noun. To strengthen the reliability of gender marking as a prediction cue, it was made predictive in 50% of the items (compared to 20% in Experiment 2), with this would lead this cue to be more strongly relied on. Additionally, the prediction window was lengthened (matching that of Experiment 1), thereby allowing more time for predictive processing. Surprisingly, these manipulations, which were expected to foster predictive behaviour, did not appear to strengthen predictive behaviour relative to Experiment 2. As discussed in Experiment 1, the lengthened adverbial phrases may have imposed additional cognitive demands, such that the cost-benefit trade-off of engaging in predictive processing was no longer worthwhile, even for L1 speakers. Alternatively, the increased structural complexity of the constituents may have encouraged a shift toward a "prediction-by-association" route, which is not efficient for prediction based on morphosyntactic cues. Most previous research demonstrating predictive use of gender marking has employed relatively simple and short sentences, such as *Click on the.*$_{MASC/FEM}$ *[target object]* (Dahan et al., 2000), *Look at the.*$_{MASC/FEM}$ *[target object]* (Huettig & Janse, 2016) or *Where is the.*$_{MASC/FEM/NEUT}$ *yellow [target object]* (Hopp & Lemmerth, 2018). One exception is Dussias et al.

[8] CPAs cannot include continuous independent variables, hence the reduction of the continuous variable "working memory score" into a dichotomous one.

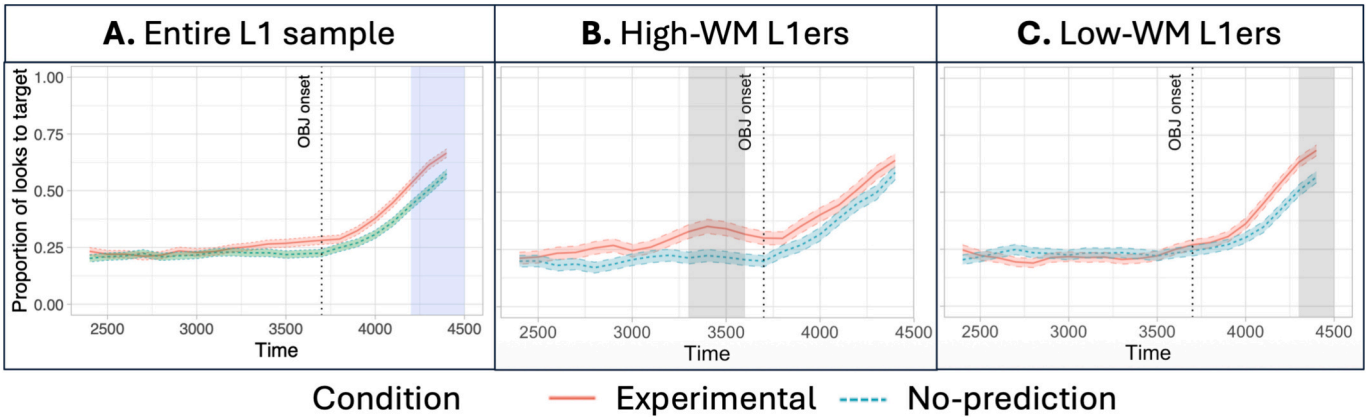


Fig. 4. Proportion of looks to the target picture over time (prediction-and-integration window) in each of the Gender Condition compared to the No-prediction condition among the entire sample of L1 speakers (Panel A), the subsample of L1 speakers with high working memory capacity (Panel B), and the subsample of L1 speakers with low working memory capacity (Panel C). Vertical dotted lines represent the object onset (3700 ms). Error bands show 95% confidence intervals. Shaded areas indicate time clusters during which there was a significant (blue) or marginally significant (grey) difference between experimental and No-prediction conditions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(2013), who employed more complex sentences (e.g., *The student was drawing the.*$_{MASC/FEM}$ *[target object] that he saw yesterday*), but the simultaneous visual scene contained only two object — lowering cognitive demands linked to visual processing and to selection among more potential candidates.

Our findings suggest that once linguistic input becomes more complex – which arguably is a feature of natural language use – comprehenders, even in their L1, may engage less in predictive processing based on gender cues than suggested in previous studies. This aligns with previous arguments that prediction in language comprehension is no as ubiquitous or automatic as sometimes assumed (e.g., Huettig & Mani, 2016; Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018). Although the limited sample size and the marginal nature of the results preclude firm conclusions, Experiment 3 tentatively indicate that L1 speakers with higher working memory capacity may have used gender marking to pre-activate the target object before its auditory onset, whereas those with lower working memory capacity did not. Future research will need to confirm this moderating role of working memory in the use of gender cue (Huettig & Janse, 2016), particularly in contexts involving longer and syntactically complex sentences (see Chun & Kaan, 2019 for evidence of prediction based on semantic cues in complex sentences).

8. General discussion

The present study examined L1 and L2 German speakers' use of lexico-semantic and morphosyntactic (gender) cues, individually and in combination, to predict the referent of an upcoming noun when these cue types vary unpredictably across sentences. Across three eye-tracking experiments, we systematically manipulated the type of available cue (semantic and/or morphosyntactic) to assess their impact on predictive processing in L1 and L2, while taking into account the transferability of these cues from the L1 to the L2 and their reliability in the task.

Experiments 1 and 2 showed that L1 speakers use the verb semantics to predict the referent of an upcoming noun, confirming well-established findings (e.g. Altmann & Kamide, 1999). Constraining "other" (non-placement) verbs (e.g., *gießen* 'water', *entsperren* 'unlock') elicited a strong increase in proportion of predictive looks to the target, whereas placement verbs (*stellen* 'put.$_{STAND}$', *legen* 'put.$_{LIE}$') yielded only a weak increase. This pattern suggests that, while semantic cues can support anticipatory processing, the strength of this support may depend on properties of the visual or linguistic context. In particular, scenes associated with placement verbs may increase the number of plausible continuations or require more detailed spatial encoding, potentially taxing cognitive resources available for prediction (Allison et al., 2025). Furthermore, predictive effects may be modulated by the semantic or associative relationship between verb and object, with stronger associative links facilitating anticipatory activation (e.g., Huettig, 2015; Pickering & Gambi, 2018). This pattern supports Hypothesis 1b and aligns with the *Prediction-by-association* vs. *Prediction-by-production* framework. When semantic cues do not support associative priming — as is the case for the semantics of placement verbs, prediction likely requires engagement of the comprehender's production system, a mechanism that is more resource-demanding and may therefore be engaged to a lesser degree. Beyond lexico-semantic cues, our study shows that L1 speakers also use grammatical gender predictively even in contexts where the source of prediction cues alternates between lexico-semantics and morphosyntax, thus disconfirming Hypothesis 1a. However, gender marking yielded only a weak increase in proportion of looks to the target. This pattern again supports Hypothesis 1b and suggests that L1 speakers can engage in prediction-by-production when associative mechanisms cannot lead to prediction, but that this more demanding process leads to smaller predictive effects reflected in anticipatory eye movements. Together, these findings suggest that prediction in L1 comprehension relies on multiple mechanisms whose engagement depends on cue properties and processing demands. L1 speakers appear able to generate predictions based on both lexico-semantic and morphosyntactic information, with tentative indication of a more limited predictive behaviour based on cues that are not associatively related to the target.

Regarding L2 speakers, they showed (limited) predictive effects based on non-placement verb semantics, but no evidence of predictive use of either placement verbs or gender marking. This suggests that prediction in L2 is far from systematic, emerging primarily when cues permit relatively automatic associative activation. Unlike L1 speakers, whose predictive behaviour based on gender marking was weak but observed, L2 speakers showed no evidence of predictive behaviour based on gender marking at all, despite its availability in their L1 and the lexical gender congruency across L1 and L2 for the words used in the task. This strong imbalance between the use of lexico-semantic and morphosyntactic cues in L2 partially supports Hypothesis 2a and chimes with the *Shallow Structure Hypothesis* (Clahsen & Felser, 2006, 2018). However, the observed patterns align even more closely with the *Prediction-by-association* vs. *Prediction-by-Production* framework (Pickering & Gambi, 2018), since L2 speakers only used lexico-semantic cues that allow predictions to arise through semantic or associative priming, while cues that would require engagement of the production system did not yield predictive effects — supporting Hypothesis 2c. Note that, although a larger scale study specifically designed to investigate the

effect of proficiency would need to confirm this pattern, proficiency was not found to modulate predictive behaviour within the proficiency range present in our L2 sample (mainly reflecting upper-intermediate level).

Furthermore, our results indicate that L1 speakers, but not L2 speakers, can integrate semantic and morphosyntactic cues across constituents, confirming Hypothesis 3. That is, L1 speakers were able to maintain semantic information provided by one constituent and combine it with morphosyntactic information from a subsequent constituent to generate predictions about the upcoming referent. Finally, Experiment 3 exploratively examined whether increasing the reliability of the gender cue and extending the prediction window would enhance L1 speakers' predictive use of morphosyntactic information. Contrary to expectations, predictive effects based on gender marking were not increased relative to Experiment 2. Altogether, our findings do not provide clear evidence for or against the (*Unified*) *Competition Model*. The present design did not allow a decisive test of cue-transfer predictions: the different types of semantic cue already differed in strength in L1, and L2 speakers showed no predictive use of either non-transferable semantic cues or transferable gender cues. Moreover, increasing the reliability of gender marking in Experiment 3 did not enhance reliance on this cue among L1 speakers. Together, these patterns prevent a clear interpretation from a competition-based perspective. However, it is plausible that other processes have obscured potential cue-reliability effects. In particular, the greater complexity and length of the adverbial phrase in Experiment 3 and the associated increase in processing cost may have outweighed any benefit of engaging in predictive processing, supporting the view that prediction is resource-dependent and not mandatory — as posited by the *Utility Hypothesis*[9] (Kuperberg & Jaeger, 2016). Exploratory analyses further suggest that participants with higher working-memory capacity were more likely to use gender marking predictively, consistent with accounts linking individual cognitive resources to prediction.

A further limitation of our design concerns differences in perceptual and informational salience between cue types. In the present study, morphosyntactic cues were realised as gender marking on determiners, which are phonologically reduced and unstressed, whereas semantic cues were conveyed by full lexical verbs, which are longer, stressed, and lexically contentful. We acknowledge that such differences may have contributed to the observed asymmetry between cue types — although this reflects naturally occurring properties of spoken language. Thus, the predictive use of linguistic cues should not be understood as a binary function of cue type (semantic vs. morphosyntactic), but as graded and sensitive to multiple cue-specific properties, such as perceptual salience, semantic transparency, regularity, etc. Even within morphosyntactic markers, different cues may vary in their predictive strength, as illustrated by previous work (e.g., Koch et al., 2021, 2023, showing that regularity of verb morphology affects its predictive use). While our design aimed to reflect naturalistic language input, future work could systematically manipulate such properties to further disentangle their contribution to predictive processing in L1 and L2 comprehension.

Taken together, the present findings provide a nuanced picture of prediction in L1 and L2 and contribute to ongoing debates about the mechanisms and ubiquity of prediction in language comprehension. First, they underscore that even in L1, anticipatory mechanisms are modulated by cue properties and appear sensitive to task demands and cognitive resources (Huettig & Mani, 2016; Kuperberg & Jaeger, 2016) — although further research specifically manipulating these factors is needed to fully clarify their role. Importantly, the present design introduced variability in the availability and type of predictive cues across sentences, a situation that more closely approximates natural language comprehension than highly constrained experimental paradigms. Under such conditions, predictive behaviour appears more limited than expected, suggesting that the prevalence of prediction reported in highly constrained contexts may overestimate its ubiquity in everyday comprehension. Second, the results indicate that, even in sentences with long and complex constituents and with varying types of predictive cues, L2 prediction based on semantic cues occurs, although it appears less pronounced than in L1. At the same time, our data provide no evidence that gender marking, even when transferable from the L1, is used predictively in L2 under such conditions. This pattern may reflect reduced automaticity in syntactic processing, weaker lexical representations, increased cognitive demands, or stronger difficulty to engage the production system in L2. Although null results warrant cautious interpretation, they nonetheless serve an important role in preventing the field from assuming that prediction is an automatic or ubiquitous mechanism of language processing and instead highlight its dependence on cue properties, processing demands, and listener resources.

[9] Note that there is no indication that participants engaged in conscious, strategic cost-benefit reasoning during the task. During debriefing, participants were asked what they believed the study investigated, and none reported having noticed that the upcoming target object could be anticipated in some sentences, suggesting that predictive behaviour, when observed, was not driven by explicit strategies but rather reflects implicit processing mechanisms.

CRediT authorship contribution statement

Pernelle Lorette: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author, who is an English L2 speaker, used ChatGPT (OpenAI, 2023) and Le Chat (Mistral AI, 2025) solely to improve the form, not to generate any content. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

Declaration of competing interest

I have no competing interest to declare.

Acknowledgements

I thank the three anonymous reviewers for their valuable comments on the previous version of this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.actpsy.2026.106948.

Data availability

The data, analysis scripts and materials supporting the findings of this study are openly available in OSF at https://osf.io/dg5cq/

References

Allison, C., Huettig, F., Fernandez, L., & Lachmann, T. (2025). Visuospatial working memory load reduces semantic prediction in the visual world. *Language, Cognition and Neuroscience*, 1–10. https://doi.org/10.1080/23273798.2025.2522272

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

Bates, E., & MacWhinney, B. (1981). Second-language acquisition from a functionalist perspective: Pragmatic, semantic and perceptual strategies. *Annals of the New York Academy of Sciences, 379*(1), 190–214. https://doi.org/10.1111/j.1749-6632.1981.tb42009.x

Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good-enough, underspecified, or shallow language processing. *Quarterly Journal of Experimental Psychology, 69*(5), 817–828. https://doi.org/10.1080/17470218.2015.1134603

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

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied PsychoLinguistics, 27*(1), 3–42. https://doi.org/10.1017/S0142716406060024

Clahsen, H., & Felser, C. (2018). Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition, 40*(3), 693–706. https://doi.org/10.1017/S0272263117000250

Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory and Language, 42*(4), 465–480. https://doi.org/10.1006/jmla.1999.2688

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

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

Ding, J., Zhang, Y., Liang, P., & Li, X. (2023). Modulation of working memory capacity on predictive processing during language comprehension. *Language, Cognition and Neuroscience, 38*(8), 1133–1152. https://doi.org/10.1080/23273798.2023.2212819

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

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*(1), 11–15. https://doi.org/10.1111/1467-8721.00158

Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension: The 'good enough' approach. *Language and Linguistics Compass, 1*(1–2), 71–83. https://doi.org/10.1111/j.1749-818X.2007.00007.x

Forbes, S., Dink, J., & Ferguson, B. (2023). *EyetrackingR (version 0.2.1)* [computer software].

Foucart, A. (2021). Language prediction in second language: Does language similarity matter? In E. Kaan, & T. Grüter (Eds.), *12. Bilingual Processing and Acquisition* (pp. 92–114). John Benjamins Publishing Company. https://doi.org/10.1075/bpa.12.05fou.

Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Bilingualism: Language and Cognition, 14*(3), 379–399. https://doi.org/10.1017/S136672891000012X

Foucart, A., & Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language, 66*(1), 226–248. https://doi.org/10.1016/j.jml.2011.07.007

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

Frenck-Mestre, C., Kim, S. K., Choo, H., Ghio, A., Herschensohn, J., & Koh, S. (2019). Look and listen! The online processing of Korean case by native and non-native speakers. *Language, Cognition and Neuroscience, 34*(3), 385–404. https://doi.org/10.1080/23273798.2018.1549332

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

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research, 28*(2), 191–215. https://doi.org/10.1177/0267658312437990

Henry, N., Jackson, C. N., & Hopp, H. (2020). Cue coalitions and additivity in predictive processing: The interaction between case and prosody in L2 German. *Second Language Research*. https://doi.org/10.1177/0267658320963151, 026765832096315.

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

Hopp, H. (2015). Semantics and morphosyntax in predictive L2 sentence processing. *International Review of Applied Linguistics in Language Teaching, 53*(3). 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. (2022). Second language sentence processing. *Annual Review of Linguistics, 8*(2022), 235–256. https://doi.org/10.1146/annurev-linguistics-030821-054113

Hopp, H., & Lemmerth, N. (2018). Lexical and syntactic congruency in l2 predictive gender processing. *Studies in Second Language Acquisition, 40*(1), 171–199. https://doi.org/10.1017/S0272263116000437

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

Huettig, F. (2025). *Looking ahead: The new science of the predictive mind*. Cambridge University Press.

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience, 31*(1), 80–93. https://doi.org/10.1080/23273798.2015.1047459

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

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

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

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

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., & Grüter, T. (2021). Prediction in second language processing and learning. In E. Kaan, & T. Grüter (Eds.), *Prediction in second language processing and learning: Advances and directions* (pp. 1–24). 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

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

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

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition, 95*(1), 95–127. https://doi.org/10.1016/j.cognition.2004.03.002

Koch, E. M., Bulté, B., Housen, A., & Godfroid, A. (2021). Using verb morphology to predict subject number in L1 and L2 sentence processing: A visual-world eye-tracking experiment. *Journal of the European Second Language Association, 5*(1), 115–132. https://doi.org/10.22599/jesla.79

Koch, E. M., Bulté, B., Housen, A., & Godfroid, A. (2023). The predictive processing of number information in subregular verb morphology in a first and second language. *Applied PsychoLinguistics, 44*(5), 750–783. https://doi.org/10.1017/S014271642300022X

Kochari, A. R., & Flecken, M. (2019). Lexical prediction in language comprehension: A replication study of grammatical gender effects in Dutch. *Language, Cognition and Neuroscience, 34*(2), 239–253. https://doi.org/10.1080/23273798.2018.1524500

Kukona, A. (2020). Lexical constraints on the prediction of form: Insights from the visual world paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(11), 2153–2162. https://doi.org/10.1037/xlm0000935

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

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

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language, 63*(4), 447–464. https://doi.org/10.1016/j.jml.2010.07.003

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology, 4*, 226. https://doi.org/10.3389/fpsyg.2013.00226

MacWhinney, B. (1997). Second language acquisition and the competition model. In A. M. De Groot, & J. F. Kroll (Eds.), *Tutorials in bilingualism. Psycholinguistic perspectives* (pp. 113–142). Lawrence Erlbaum Associates Publishers.

MacWhinney, B., Hickmann, M., Veneziano, E., & Jisa, H. (2017). A unified model of first and second language learning. In *Sources of variation in first language acquisition: Languages, contexts, and learners* (pp. 287–310). John Benjamins.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Mitsugi, S., & Macwhinney, B. (2016). The use of case marking for predictive processing in second language Japanese. *Bilingualism: Language and Cognition, 19*(1), 19–35. https://doi.org/10.1017/S1366728914000881

Morales, L., Paolieri, D., Dussias, P. E., Kroff, J. R. V., Gerfen, C., & Bajo, M. T. (2016). The gender congruency effect during bilingual spoken-word recognition. *Bilingualism: Language and Cognition, 19*(2), 294–310. https://doi.org/10.1017/S1366728915000176

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., … Ito, A. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife, 7*, Article e33468. https://doi.org/10.7554/eLife.33468

Otten, M., & Van Berkum, J. J. (2008). Discourse-based word anticipation during language processing: Prediction or priming? *Discourse Processes, 45*(6), 464–496. https://doi.org/10.1080/01638530802356463

Otten, M., & Van Berkum, J. J. (2009). Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain Research, 1291*, 92–101. https://doi.org/10.1016/j.brainres.2009.07.042

Peters, R. E., Gruter, T., & Borovsky, A. (2015). Anticipatory and locally coherent lexical activation varies as a function of language proficiency. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 37. https://escholarship.org/uc/item/1pf4g78m.

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

Porretta, V., Kyröläinen, A., Van Rij, J., & Järvikivi, J. (2017). *VWPre: Tools for preprocessing visual world data [R package version, 1(0)]*.

R Core Team. (2022). *R: A language and environment for statistical computing* [computer software]. R Foundation for Statistical Computing.

Riordan, B., Dye, M., & Jones, M. N. (2015). Grammatical number processing and anticipatory eye movements are not tightly coordinated in English spoken language comprehension. *Frontiers in Psychology, 6*, 590. https://doi.org/10.3389/fpsyg.2015.00590

Rommers, J., Meyer, A. S., Praamstra, P., & Huettig, F. (2013). The contents of predictions in sentence comprehension: Activation of the shape of objects before they are referred to. *Neuropsychologia, 51*(3), 437–447. https://doi.org/10.1016/j.neuropsychologia.2012.12.002

RStudio Team. (2020). *RStudio: Integrated development for R* [Computer software]. RStudio, PBC http://www.rstudio.com/.

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language, 71*(1), 145–163. https://doi.org/10.1016/j.jml.2013.11.002

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

Schlenter, J., & Felser, C. (2021). Second language prediction ability across different linguistic domains: Evidence from German. In E. Kaan, & T. Grüter (Eds.), *Vol. 12. Prediction in second language processing and learning* (pp. 47–68). John Benjamins Publishing Company. https://doi.org/10.1075/bpa.12.03sch.

Tschirner, E., & Möhring, J. (2020). *A frequency dictionary of German: Core vocabulary for learners* (2nd ed.). Routledge.

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

Weber, A., Grice, M., & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition, 99*(2), B63–B72. https://doi.org/10.1016/j.cognition.2005.07.001

Wechsler, D. (1981). *Manual for the Wechsler adult intelligence scale-revised (WAIS-R)*. Psychological Corporation.

Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience, 16*(7), 1272–1288. https://doi.org/10.1162/0898929041920487

Huettig, F., & Tanenhaus, M. K. (2025). Rethinking task importance in the visual world paradigm. *Brain Research, 1867*, Article 149965. https://doi.org/10.1016/j.brainres.2025.149965