GLOSSA psycholinguistics

Jackson, C. N., Hopp, H., & Grüter, T. (2025). Does syntax guide semantic predictions in L1 and L2 processing? *Glossa Psycholinguistics*, 4(1): 19, pp. 1–36. DOI: https://doi.org/10.5070/G6011.38086

eScholarship University of California

Does syntax guide semantic predictions in L1 and L2 processing?

Carrie N. Jackson, Penn State University, US, cnj1@psu.edu
Holger Hopp, Technische Universität Braunschweig, DE, h.hopp@tu-braunschweig.de
Theres Grüter, University of Hawai'i at Mānoa, US, theres@hawaii.edu

In two visual world experiments with L1 and L2 German speakers, this study investigates how listeners use semantic cues on the verb to predict either a post-verbal object in subject-first SVO sentences or a post-verbal subject in adverb-first verb-second (AdvVS(O)) sentences, with the goal of assessing to what extent the syntax of an unfolding sentence constrains the activation of potential upcoming referents. In all sentence types, both L1 and L2 listeners looked at the referent of the post-verbal argument earlier when the verb was semantically constraining than when it was not. Predictive looks were slower overall in the L2 group, but not attenuated in syntactic configurations absent from their L1, namely, adverb-first verb-second (AdvVS(O)) sentences. Both groups were more likely to fixate plausible subjects than objects on hearing a transitive verb following a sentence-initial adverb, suggesting that prediction of an upcoming argument was constrained not only by semantic association with the verb, but also by the syntactic and thematic structure of the unfolding sentence. These results underscore how sentence-level cues, like word order and the distribution of grammatical and thematic roles, can guide the use of local, namely lexical-level, cues during semantically-cued prediction in both L1 and L2 processing.

1 Introduction

Over the last fifteen years, there has been an explosion of research on predictive processing in both first language (L1) and second language (L2) speakers, and a growing interest in whether predictive processing constitutes an underlying mechanism to support language comprehension and learning (see Bovolenta & Marsden, 2022; Huettig, 2015, 2025; Kuperberg & Jaeger, 2016; Ryskin & Nieuwland, 2023, for recent reviews). There is ample evidence that L1 and L2 speakers can pre-activate upcoming linguistic information, based on the linguistic cues they have already encountered in the input. However, a number of factors can modulate speakers' engagement in prediction, including the reliability and utility of cues for prediction in light of specific task demands and speakers' knowledge base, skills and experience (e.g., Ito & Pickering, 2021; Kaan, 2014; Kuperberg & Jaeger, 2016, among others). Notably, most studies that have investigated the use of semantic information to generate expectations about upcoming words in a sentence – especially among L2 speakers – have focused on listeners' use of verbal semantics to anticipate the post-verbal object in simple SVO sentences with a canonical agent-verb-patient order (e.g., Chambers & Cooke, 2009; Dijkgraaf et al., 2017; Ito, Corley, et al., 2018). A smaller number of studies – all with L1 speakers – have extended the investigation to other constructions in English (e.g., Kukona et al., 2011, Exp. 2) and to languages with different underlying word orders, such as Dutch (Dahan & Tanenhaus, 2004) and Tagalog (Bondoc & Schafer, 2022; Sauppe, 2016). Little is known about how L2 speakers use verbal semantics to predict post-verbal arguments in constructions or languages that do not follow the canonical SVO/agent-verb-patient order. This is of particular interest in cases where the non-canonical order under investigation in the L2 is not present in the learner's L1. To the best of our knowledge, such scenarios have not yet been investigated in L2 predictive processing.

To address this gap in the literature, the present study uses the visual world paradigm to investigate prediction among L1 and L2 German listeners in the context of subject-initial sentences with SVO word order as well as non-subject-initial adverb-first verb-second (AdvVS(O)) sentences in which the verb precedes the agentive subject. By considering both word orders in tandem, we probe how listeners use verbal semantics to anticipate upcoming referents that bear different grammatical and thematic roles across different word orders. In doing so, the present study addresses two key questions in the L1 and L2 prediction literature regarding the use of semantic information for prediction: (i) whether such prediction is driven primarily by semantic association or whether it is also guided by syntactic context (i.e., word order and the distribution of grammatical and thematic roles) (e.g., Bondoc & Schafer, 2022; Chow et al., 2016; Kukona et al., 2011; Pickering & Garrod, 2013; Sauppe, 2016), and (ii) whether syntax similarly guides prediction in an L2, especially in syntactic configurations that are absent in the L1.

1.1 The use of verbal semantics in L1 and L2 prediction

Many studies that investigate the use of semantic information for prediction involve the visual world paradigm, in which participants listen to target sentences while researchers track their eye gazes to different images on a computer screen (for review, see Huettig et al., 2011). Studies using this paradigm have shown that listeners use linguistic information in conjunction with the visual context to build strong expectations for upcoming referents, thereby looking at a target object even before it is mentioned (see Huettig, 2015, for review). In a seminal study, Altmann and Kamide (1999) found that L1 English listeners looked at an image of the target object (e.g., cake) more than unrelated distractors upon hearing a semantically-constraining verb in sentences like (1a), because from this set of objects, only the cake was a plausible object of the verb *eat*. In contrast, they looked at multiple objects upon hearing a semantically-neutral verb in sentences like (1b), because more than one object on the screen was a plausible object of the verb *move*.

(1) a. Constraining verb
The boy will eat the cake.

b. Neutral verb
The boy will move the cake.

Following Altmann and Kamide, many studies have further probed how listeners use verbal semantics to anticipate the upcoming object in sentences with SVO word order, but only a few have considered semantically-cued prediction in other word orders (see Bondoc & Schafer, 2022; Dahan & Tanenhaus, 2004; Sauppe, 2016, for exceptions).

Although the primary focus of their study was not the impact of syntax or word order on prediction, Dahan and Tanenhaus (2004) exploited grammatical properties of Dutch to study semantically-cued prediction in non-subject-initial AdvVS sentences, in which the verb precedes the subject. Dutch is an SOV language that requires verb-second (V2) word order in main clauses. As seen in examples (2a) and (2b), this means that the finite verb always appears in second position in main clauses, even when another constituent occurs sentence initially. In such cases, the subject appears after the verb. Further, when a clause contains only an inflected lexical verb, that verb appears in second position, as in (2a). In constructions with an auxiliary or modal verb, by contrast, the inflected auxiliary appears in second position, while the non-finite lexical verb remains in clause-final position, as in (2b).

(2) a. AdvVS-constraining verb
Nog nooit klom een bok zo hoog.
never before climbed a goat so high
'Never before climbed a goat so high.'

b. AdvVS-neutral verb
Nog nooit is een bok zo hoog geklommen.
never before is a goat so high climbed
'Never before has a goat climbed so high.'

Paralleling research with SVO sentences, Dahan and Tanenhaus found that L1 Dutch listeners looked at the target referent (i.e., *bok* 'goat') earlier in sentences with a semantically-constraining verb (i.e., *klom* 'climbed') versus a neutral auxiliary verb (e.g., *is* 'is'), suggesting that they successfully anticipated a post-verbal subject, based solely on the semantic (thematic) constraints provided by the verb.

Turning to L2 research, a growing number of studies show that L2 speakers also reliably use semantic information to predict upcoming arguments in sentences with SVO word order (e.g., Chambers & Cooke, 2009; Dijkgraaf et al., 2017; Ito, Corley, et al., 2018), even though L2 processing, in general, is more resource-intensive (e.g., Declerck & Kormos, 2012; Francis et al., 2018). Notably, L1 Japanese-L2 English listeners can successfully use verbal semantics to anticipate an upcoming object in SVO sentences, even though their L1 Japanese has SOV word order (Ito, Pickering, et al., 2018). Further, Dijkgraaf et al. (2017) found that L1 Dutch-L2 English listeners used semantic information for prediction to similar degrees in both of their languages, and the strength of their predictions in the L2 were indistinguishable from those made by L1 English listeners. Ito, Corley, et al. (2018) also reported that L1 English and L2 English listeners with a variety of L1s used semantic information for prediction to similar degrees under normal listening conditions, and that both groups experienced a similar delay in prediction when they performed a concurrent memory task, which induced greater cognitive load. These parallels suggest that the mechanisms underlying the use of semantic information for prediction are largely similar in L1 and L2 speakers (see also Chambers & Cooke, 2009; Kaan, 2014).

However, more recent studies point to potential quantitative differences between L2 and L1 speakers in their use of semantic information for prediction when L2 speakers are faced with processing more complex sentences or integrating multiple types of information. Whereas the studies described above all relied on simple sentences with SVO word order, when processing SVO sequences embedded in more syntactically complex sentences, such as relative clauses, L2 listeners exhibited a delay in prediction relative to their processing of syntactically simple sentences (Chun et al., 2021) and a delay relative to L1 listeners (Chun & Kaan, 2019; see also Dijkgraaf et al., 2019; Ito, Pickering, et al., 2018). Paralleling these differences, Corps et al. (2023) found that while L2 English listeners engaged in simple associative semantic prediction (e.g., looking at an image of a tie after hearing the verb *wear*) with a speed similar to L1 English listeners, they were slower than L1 English listeners to use top-down contextual cues of a male versus female voice to focus their gaze on a stereotypically masculine object (e.g., tie) versus a stereotypically feminine object (e.g., dress). Together, such findings have led researchers to hypothesize that

semantically-cued prediction may be qualitatively similar for L1 and L2 speakers, but that it may be more vulnerable to pressures from other resource demands, whether they be non-linguistic (e.g., working memory load) or linguistic (e.g., syntactic context, cross-linguistic influence) (see Ito & Pickering, 2021; Kaan & Grüter, 2021; Schlenter, 2023, for further discussion). The present study contributes to a better understanding of the characteristics of predictive processing in L1 and L2 speakers by investigating how semantically-cued prediction may vary across different syntactic contexts, namely, subject-first SVO versus adverb-first AdvVS(O), especially when one of these word orders is absent in the L1.

1.2 Mechanisms underlying L1 and L2 prediction

There is growing consensus in the psycholinguistic literature that multiple mechanisms may underlie predictive processing among both L1 and L2 speakers. Even proposals that argue for a single underlying mechanism acknowledge that multiple linguistic and non-linguistic factors influence the scope – and even the presence or absence – of predictive processing across different populations and in different contexts (e.g., Knoeferle & Crocker, 2007; Kuperberg, 2007; Pickering & Gambi, 2018; Pickering & Garrod, 2013; Ryskin & Niewland, 2023, among others. See Huettig, 2015, for a comprehensive review). While different models vary in their assumptions regarding the types and number of mechanisms that drive prediction during language processing, in broad terms, all models acknowledge that prediction can be driven by semantic priming through the spreading activation of related concepts. For instance, upon hearing the verb *arrest*, a listener would activate closely related concepts, such as *robber, gunman, policeman, bank, jail,* etc. In turn, this would prime listeners to look at images of these concepts in a visual world display (e.g., Kukona, et al., 2011; see also Chow, et al., 2016). This type of simple associative mechanism may be less resource intensive and more automatic than predictive mechanisms that take a larger number of linguistic and non-linguistic factors into account (e.g., Chow et al., 2016; Pickering & Gambi, 2018, among others). Yet such associative mechanisms can also be noisy and potentially lead to erroneous predictions (e.g., Kukona et al., 2011). Further, among L2 speakers, even such associative mechanisms may be slower and potentially more resource demanding, relative to L1 speakers, due to the more complex processes of lexical activation in the bilingual lexicon (Dijkgraaf et al., 2019).

Alongside prediction by associative priming, models of predictive processing also incorporate routes through which comprehenders can actively anticipate and pre-activate upcoming linguistic input, based on various types of bottom-up and top-down information that they encounter as a sentence unfolds. Such information can include grammatical or morphosyntactic information, like case marking (e.g., Hopp, 2015; Kamide et al., 2003; Schlenter & Felser, 2021), phonological or orthographic forms (e.g., Ito, Pickering, et al., 2018; Martin et al., 2013), contextual information, like the objects presented in a visual world display (e.g., Knoeferle & Crocker, 2007; Tanenhaus et al., 1995) or speaker voice (e.g., Corps et al., 2023; Kamide, 2012), as well as the event structure

and higher-level meaning that people build up through the combined meaning of multiple words over time (e.g., Altmann & Kamide, 2007; Chambers & San Juan, 2008; Kim & Osterhout, 2005). However, making predictions via this route likely requires greater cognitive resources. Therefore, such higher-level information may contribute to predictive processing more selectively, namely, only when such information is maximally reliable and beneficial for comprehension (e.g., Brothers et al., 2017; Huettig, 2015; Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018). The use of multiple and higher-level cues may be even more constrained among L2 speakers, precisely because L2 processing, in general, is more resource demanding, leaving fewer resources to devote to prediction (see, e.g., Ito & Pickering, 2021; Kaan, 2014, for further discussion).

There is evidence that both associative and more complex predictions occur in tandem. For instance, using the visual world paradigm, Kukona et al. (2011) found that when L1 English listeners heard a sentence fragment like *Toby arrests the…* they looked at an image of a plausible patient (e.g., crook), but also an image of a plausible agent (e.g., policeman) even though the subject role in the sentence had already been filled by *Toby*. Such competition effects were attenuated, but not eliminated entirely, when participants listened to passive sentences, like *Toby was arrested by the…* , where participants looked at the plausible agent (e.g., policeman), after *by*, but they still looked at the plausible patient (e.g., crook) more often than at unrelated distractors. Kukona and colleagues suggested that when there are strong associative connections between a verb and its arguments, this can lead to early activation of these arguments, even when such local thematic priming conflicts with sentence-level constraints, like word order and the distribution of grammatical and thematic roles. Further, the relative weight of such local – namely, lexical-level – versus sentence-level constraints can vary, based on the availability and strength of these different constraints in a sentence.

The relative weight of local- versus sentence-level cues during predictive processing can also vary based on listeners' linguistic experience and skills (Borovsky et al., 2012, 2013). For instance, in a visual world experiment with a heterogeneous group of adult bilinguals fluent in English, Peters et al. (2018) tracked participants' eye movements to four images while listening to sentences like *The pirate chases the ship*: the target (ship), a distractor semantically associated with the agent (treasure chest), an action-related distractor, i.e., a plausible object for the verb without considering the subject (cat), and an unrelated distractor (bone). Similar to the monolingual English participants in Borovsky et al. (2012), these bilingual participants' looks to the action-related distractor (cat) temporarily increased after hearing the verb (*chases*), even though the referent was an unlikely continuation in the context of the preceding sentence fragment as a whole (*The pirate chases…*). Of note, these local semantic association effects were greater among participants who self-identified as non-native speakers of English and participants with lower English vocabulary skills. Peters and colleagues concluded that less proficient bilingual speakers are more likely to activate not only the most probable referent, given cumulative information from the preceding sentence fragment, but also referents that are only locally coherent, that is,

semantically associated with the immediately preceding word, suggesting that simple associative prediction may play a greater role among less proficient language users.

1.3 Word order and argument structure in German

Investigating how L1 and L2 German listeners use verbal semantics to predict post-verbal arguments across different word orders provides an ideal environment in which to probe the extent to which prediction proceeds via simple semantic association versus the use of higher-level information. The present study takes advantage of the verb-second (V2) property of German. Like Dutch, German exhibits V2 word order in main clauses, meaning that the finite verb always appears in second position in main clauses, regardless of whether the subject or another constituent occurs sentence initially. In instances where a non-subject constituent occurs sentence initially, the grammatical subject appears after the finite verb, as in (3).

(3) Im Sommer springt der Frosch ins Wasser.
in.the summer jumps.V the frog.SUB into.the water
'In the summer the frog jumps into the water.'

This contrasts with languages with more rigid SVO word order, like English, where the subject always precedes the verb, even in instances where another constituent appears sentence initially (*In summer, the frog jumps*…).

At least at lower proficiency levels, L2 German speakers with L1 English, an SVO language, will continue to place the subject before the finite verb in adverb-first sentences (e.g., **Im Frühling der Frosch springt ins Wasser*; Jackson & Ruf, 2017; Pienemann, 1998). Doing so maintains the canonical agent-verb-patient order of SVO sentences from their L1 English, even though it results in an ungrammatical verb-third sentence in German. Even at more advanced proficiency levels, L2 German speakers with L1 English exhibit continued difficulties with V2 word order and are less likely to produce non-subject initial sentences, compared to L1 German speakers (Carroll et al., 2000; O'Brien & Féry, 2015). In the context of L2 processing and prediction, less experience with the V2 property of German, and its consequence that the verb precedes the subject in non-subject first utterances, also means that L2 German speakers with English or another SVO language as their L1 have less experience anticipating post-verbal subjects. Thus, one primary question driving the present study is whether L2 speakers will predict a post-verbal subject when listening to precisely such AdvVS(O) sentences in German.

1.4 Present study

In two visual world experiments, the present study examines whether differences in word order modulate the use of semantic information for prediction among L1 and L2 German listeners. In Experiment 1, we presented participants with SVO and AdvVS sentences in order to investigate their use of semantically informative cues on the verb to predict the post-verbal argument. In

Experiment 2, which focuses specifically on non-subject-initial AdvVS(O) sentences, we vary the argument structure of the verb. Participants were presented with transitive (AdvVSO) and intransitive (AdvVS) sentences in order to tease apart whether semantic cues on the verb lead to the anticipation of *any* compatible argument via simple semantic association or whether the preceding syntactic context constrains anticipation to the next *grammatically possible* referent, in this case, the subject. In so doing, the present study addresses the following research questions:

- Does word order (SVO vs. AdvVS) affect (L1 and L2) listeners' use of verbal semantics to predict a post-verbal argument? (Experiment 1)
    - Is L2 listeners' semantically-cued prediction attenuated in word order configurations not present in the L1?
- In non-subject-initial sentences (AdvVS and AdvVSO), do (L1 and L2) listeners use verbal semantics to predict the first syntactically required argument, namely, the *subject*? (Experiment 2)
    - Within word order configurations not present in their L1, do L2 listeners rely more than L1 listeners on bottom-up semantic association versus top-down cues, like syntactic context, to predict post-verbal arguments?

2 Experiment 1

In Experiment 1, participants listened to SVO and intransitive AdvVS sentences in both constraining-verb and neutral-verb conditions, while viewing visual scenes containing four images, all of which could serve as a potential post-verbal subject or object after an auxiliary verb, but only one of which could serve as a plausible post-verbal subject or object after the finite lexical verb. The goal of Experiment 1 was to replicate the findings with AdvVS sentences from Dahan and Tanenhaus (2004) with L1 German listeners and extend this work to L2 German listeners whose L1 does not allow lexical verbs to precede the subject in declarative main clauses. For L1 German listeners, we expect to see a similar use of semantic cues for predicting post-verbal arguments in both SVO and AdvVS sentences that contain semantically-constraining finite verbs versus semantically-neutral auxiliary verbs. If L2 German listeners similarly use semantic cues to predict post-verbal arguments in both SVO and AdvVS sentences with semantically-constraining verbs, this would suggest that L2 speakers can engage in semantically-cued prediction even with word orders that are not present in their L1 (e.g., Ito, Pickering, et al., 2018).

2.1 Methods

2.1.1 Participants

Thirty-four L2 German speakers living in the United States and 34 L1 German speakers living in Germany participated in the study in exchange for payment. One L2 German speaker was excluded because they were an L1 speaker of Pennsylvania Dutch, another V2 language, and another L2

German speaker was excluded due to excessive track-loss during the eye-tracking task. Two L1 German speakers were excluded because they reported significant exposure to an additional language besides German in early childhood. Thus, all analyses were based on 32 L2 German participants (17 female; 15 male) and 32 L1 German participants (22 female; 7 male; 3 gender not provided). Most L2 participants spoke English as their first language (27) but other L1s included Chinese (2), Russian (1) and Spanish (2). The L2 participants started learning German in school at age 4 or older. As an overall measure of L2 German proficiency, the L2 participants completed a 30-point grammar and vocabulary test (Goethe Institut, 2004). The results revealed a wide range in proficiency from A2-C2 level. See **Table 1** for complete biographical information.

Table 1: Experiment 1: Participant information.

	L1 Speakers (n = 32)		**L2 Speakers (n = 32)**	
	M(SD)	***Range***	***M(SD)***	***Range***
Age (years)	22.4 (2.0)	20–28	21.7 (2.8)	18–31
Age of first exposure to German (years)	NA	NA	13.4 (3.5)	4–20[1]
Total time in a German-speaking environment (months)	NA	NA	6.8 (14.6)	0–78
Goethe test score (out of 30)	NA	NA	16.6 (5.0)	9–28

2.1.2 Materials

This subsection first describes the stimuli that were used in the primary eye-tracking task, including norming procedures for these stimuli. Subsequently, we describe the stimuli and procedures for all ancillary tasks (AdvVS production task, grammaticality judgment task, L2-to-L1 translation task) that were used to assess participants' knowledge of the targeted structures and vocabulary in the primary eye-tracking task.

2.1.2.1 Eye-tracking task

The experimental stimuli consisted of 16 SVO and 16 AdvVS sentence pairs. As seen in examples (4) and (5), in constraining-verb sentences, the lexical verb appeared in V2 position, whereas in neutral-verb sentences, an auxiliary verb appeared in V2 position and the lexical verb appeared in sentence-final position. (The complete list of experimental materials for both experiments is available on OSF, https://doi.org/10.17605/OSF.IO/TJYPQ.) In both conditions, a sentence-medial adverb (e.g., *täglich* 'daily') was included after the verb to extend the temporal window for capturing predictive effects.

[1] Three participants reported limited exposure to German through school-based instruction of 1–2 hours a week in early elementary school. Given that this exposure was limited and, in all other ways, these participants had background and proficiency profiles similar to the other participants, they were included in the final pool of participants.

(4) a. SVO-constraining verb
Simone füttert täglich den Hund im Garten.
Simone.SUB feeds.V daily the dog.OBJ in.the garden

b. SVO-neutral verb
Simone soll täglich den Hund im Garten füttern.
Simone.SUB should.V-MOD daily the dog.OBJ in.the garden feed.V
'Simone feeds/should feed the dog in the garden daily.'

(5) a. AdvVS-constraining verb
Im Sommer springt täglich der Frosch ins Wasser.
in.the summer jumps.V daily the frog.SUB into.the water

b. AdvVS-neutral verb
Im Sommer wird täglich der Frosch ins Wasser springen.
in.the summer will.V-MOD daily the frog.SUB into.the water jump.V
'In summer the frog jumps/will jump into the water daily.'

The initial adverbial phrase in AdvVS sentences did not provide additional semantic cues to the target referent. Similarly, the subject in SVO sentences was always a proper noun. Each sentence was paired with a visual display containing four images (see **Figure 1**). In constraining-verb sentences, only one of the four referents in the image display was a plausible continuation of the sentence – the grammatical subject in AdvVS sentences and the direct object in SVO sentences. In contrast, any referent from the image display was a plausible continuation in neutral-verb sentences. The log frequency of the lexical verbs and target nouns for the SVO sentences (verb: $M = 2.9$, $SD = 0.4$; target noun: $M = 3.2$, $SD = 0.6$) and AdvVS sentences (verb: $M = 3.0$, $SD = 0.5$; target noun: $M = 3.1$, $SD = 0.8$) did not differ from each other, based on values from SUBTLEX-DE (Brysbaert et al., 2011; all $ps > .5$).

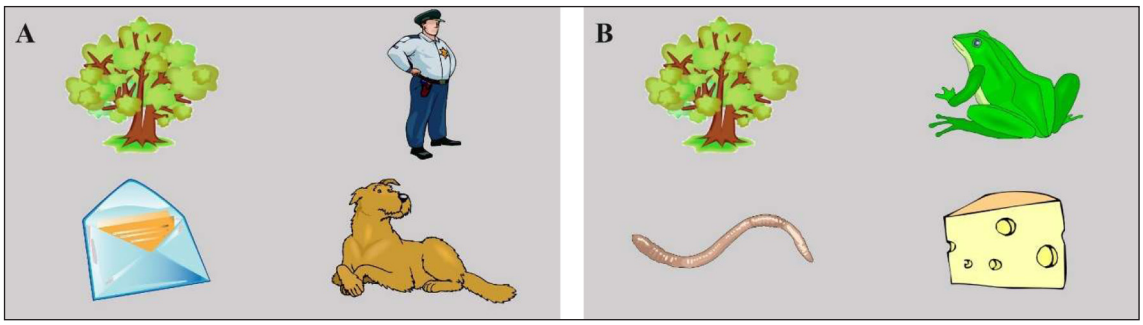

Figure 1: (A) Image display for sentences in (4). **(B)** Image display for sentences in (5).

Each visual display contained clipart images (from clipart.com) of the target and three distractor referents, with the distractors containing a mix of animate and inanimate referents. All four nouns had the same grammatical gender, so that participants could not use the

gender-marked determiner (e.g., *der*) as an additional cue to identify the target referent. Separate repeated-measures ANOVAs for SVO and AdvVS sentences revealed that the log frequency of the target noun (SVO: $M = 3.2$, $SD = 0.6$; AdvVS: $M = 3.1$, $SD = 0.8$) did not differ from the log frequency of the distractor nouns (SVO: $M = 2.8$, $SD = 0.7$; AdvVS: $M = 2.6$, $SD = 0.6$), based on values from SUBTLEX-DE (Brysbaert et al., 2011; all $ps > .09$). The position of the target image was counterbalanced across all items.

To test for semantic predictability, the 32 experimental items and their respective image displays were distributed across two lists, such that each list contained eight sentences from each condition. Twenty German native speakers, none of whom participated in the main experiment, read the first half of each experimental sentence (e.g., *Simone füttert täglich…*) and chose which image in the image display came next in the sentence. For SVO sentences, the participants chose the target 97.0% of the time in constraining-verb sentences ($SD = 6.6$) but only 40.0% of the time in neutral-verb sentences ($SD = 22.2$), a difference that was statistically significant, $t(15) = 9.51$, $p < .001$. For AdvVS sentences, the participants chose the target 93.5% of the time in constraining-verb sentences ($SD = 11.7$) but only 33.0% of the time in neutral-verb sentences ($SD = 19.2$; $t(15) = 11.21$, $p < .001$).

The 32 experimental items were divided into two lists, so that participants saw eight constraining-verb and eight neutral-verb sentences for each sentence type (AdvVS and SVO), but only one version of any given sentence pair. They were interspersed with 32 filler items that were similar in length and structure to the experimental items. Filler sentences included a variety of transitive and intransitive verbs, an equal number of subject-first and adverb-first sentences, and an equal number of sentences where the finite verb in V2 position was the lexical verb versus an auxiliary verb. To disguise verbal semantics as a predictive cue to the post-verbal target referent in the experimental items, approximately one third of the filler sentences included other potentially predictive cues to the post-verbal referent (e.g., the subject), and in approximately half of the filler sentences, the target referent did not immediately follow the finite verb. Experimental and filler items were presented in pseudorandomized order, so that participants never saw more than one experimental item from the same condition in a row.

All sentences were recorded by a female native speaker of German at a sampling rate of 44.1 kHz. She spoke at a slow, but still natural, speech rate. The mean length of the critical region from verb onset to target noun onset was longer for constraining-verb sentences (SVO: $M = 1605$ ms, $SD = 161$ ms; AdvVS: $M = 1586$ ms, $SD = 164$ ms) than neutral-verb sentences (SVO: $M = 1346$ ms, $SD = 144$ ms; AdvVS: $M = 1340$ ms, $SD = 102$ ms) but there was no significant difference between SVO and AdvVS sentences within either condition (all $ps > .7$).

2.1.2.2 AdvVS production task

After the eyetracking task, participants completed a production task to measure their knowledge of V2 word order in non-subject initial sentences during language production (see Jackson & Ruf, 2017, for similar procedures). The task consisted of a series of 20 pictures that were presented using E-Prime 3.0 software (Psychology Software Tools, Pittsburgh, PA). Participants were instructed to orally describe each picture, beginning their sentence with the phrase provided. For nine sentences, this phrase was an adverbial phrase. Three sentences each began with a temporal phrase (e.g,. *am Dienstag* 'on Tuesday'), a locative phrase (e.g., *auf dem Tisch* 'on the table') or a manner phrase (e.g., *mit dem Löffel* 'with the spoon'). For these nine sentences, participants received one point if they placed the verb immediately after the adverbial phrase, thereby producing a grammatical V2 sentence. The task also included eight sentences in which the provided phrase was the subject in the picture, and three sentences in which the phrase provided was the direct object in the picture.

2.1.2.3 Grammaticality judgment task

After the AdvVS production task, participants completed a grammaticality judgment task to measure their knowledge of the V2 property of German as it relates to verb placement in SVO and AdvVS sentences like those used in the eye-tracking task. In this task, participants read 32 sentences and rated how acceptable these sentences sounded on a 4-point scale (1 = *definitiv inakzeptabel* 'definitely unacceptable'; 4 = *definitiv akzeptabel* 'definitely acceptable'). This task was presented via Qualtrics (Qualtrics, Provo, UT). Eight sentences from each of the four experimental conditions in the eye-tracking task (SVO-constraining verb; SVO-neutral verb; AdvVS-constraining verb; AdvVS-neutral verb) were modified to create ungrammatical versions of each item that violated the V2 word order rule in German (see examples (6) and (7)).

(6) SVO-constraining verb; Ungram
*Simone täglich füttert den Hund im Garten.
Simone daily feeds the dog in.the garden
'Simone feeds the dog daily in the garden.'

(7) AdvVS-constraining verb; Ungram
*Im Sommer der Frosch springt täglich ins Wasser.
in.the summer the frog jumps daily into.the water
'In the summer the frog jumps daily into the water.'

Items were distributed across two lists, so that participants saw either the grammatical or the ungrammatical version of any given item, but not both. Participants, thus, saw four sentences in each of the eight conditions, for a total of 16 grammatical and 16 ungrammatical sentences. The

sentences were presented in a randomized order.[2] For scoring purposes, constraining-verb and neutral-verb sentences were grouped together according to overall sentence type (grammatical vs. ungrammatical; AdvVS vs. SVO).

2.1.2.4 L2-to-L1 translation task

As a final task, all L2 participants completed an L2-to-L1 translation task in order to measure their knowledge of the lexical verbs included in the eye-tracking task. In this task, the L2 participants saw the 32 target nouns and the 32 lexical verbs from the experimental items in the eye-tracking task in their infinitive form. They were prompted to provide the English equivalent. This task was presented via Qualtrics (Qualtrics, Provo, UT).

2.1.3 Procedure

Prior to starting the study, participants provided written consent for their participation. Then they completed a language background questionnaire. This was followed by the eye-tracking task. After the eye-tracking task, the L1 participants completed the AdvVS production task and the grammaticality judgment task. The L2 participants first completed the German proficiency task, followed by the AdvVS production task. Then they completed the L2-to-L1 vocabulary translation task, followed by the grammaticality judgment task.

The eye-tracking task was conducted on SMI RED eye-tracking systems, which recorded eye movements from the right eye with a sampling rate of 60 Hz. Participants were instructed to listen to the sentences while looking at four images on the computer screen and use the mouse to "click on the object that was mentioned in the sentence as quickly as possible". The experiment began with six practice trials and a 9-point calibration with a subsequent validation. Calibrations were accepted if the mean deviation of visual acuity was less than 1 degree. Visual scenes were presented for 2000 ms prior to the onset of the sentence and remained on the screen until participants pressed the space bar at the conclusion of each sentence. There was an 800 ms interval showing a blank screen separating each trial from the next trial.

2.1.4 Data analysis

Four areas of interest (AOIs) were coded for each visual scene, with each AOI encompassing the quadrant in the display containing the respective object. Analyses included all fixations to one of these four AOIs. Saccades, blinks and looks outside of these AOIs were excluded prior to data analysis. Fixation and mouse click data were exported into R version 4.3.1 (R Core Team, 2023)

[2] In one list, one of the grammatical SVO-neutral verb items was accidentally presented in its ungrammatical version. Thus, participants who received this list saw three grammatical SVO-neutral verb sentences and five ungrammatical SVO-neutral verb sentences.

for preprocessing and data analysis. Prior to analysis, trials in which more than 30% of fixations were missing, due to track loss, were removed, leading to the exclusion of 1.3% of L1 participant trials and 4.9% of L2 participant trials. One additional L2 participant was excluded because fewer than four trials per experimental condition remained. Prior to analysis, the fixation data were time-locked to the onset of the finite verb (e.g., *springt* 'jump'), the point at which participants could begin to predict the target referent in constraining-verb sentences. Given that the two sentence types differed in structure (AdvVS vs. SVO) and contained different verbs and target referents, all analyses were conducted separately for each sentence type.

The primary research question asked whether word order affects the use of semantic cues from the verb for predicting post-verbal arguments. To identify the point at which the proportion of looks to the target diverged in constraining-verb versus neutral-verb conditions for each sentence type (AdvVS, SVO) and for each group (L1, L2), we used divergence point analyses, applying the non-parametric bootstrap procedure in Stone et al. (2021). An initial divergence point was established by identifying the first measurement point in a run of at least 12 consecutive measurement points (i.e., 200 ms) in which the proportion of looks to the target significantly differed between constraining-verb and neutral-verb sentences, as determined by unpaired t-tests. Next, a bootstrap distribution of divergence points was estimated through resampling the original dataset 2000 times and repeating the divergence point calculation after each resample. This procedure was conducted separately for each sentence type within each participant group. (The complete set of experimental scripts for both experiments is available on OSF.)

Given the research showing that L2 proficiency may impact L2 predictive processing (e.g., Chambers & Cooke, 2009), we explored in a secondary analysis whether L2 proficiency modulated the size of the prediction effect – operationalized as the proportion of looks to the target in constraining-verb versus neutral-verb sentences – among the L2 speakers. For these analyses, we defined a temporal window that encompassed fixations from 200 ms after the onset of the verb to 200 ms after the onset of the target noun. This time window allowed us to explore the size of prediction effects across the entire prediction window. Results were analyzed in R 4.3.1 (R Core Team, 2023) using linear mixed-effects models with the lme4 package version 1.1-34 (Bates et al., 2015) and lmerTest package version 3.1-3 (Kuznetsova et al., 2017). Separate analyses were conducted for AdvVS and SVO sentences, with Constraint (Constraining-verb vs. Neutral-verb), contrast coded –0.5 and 0.5, and L2 Proficiency, centered at the sample mean, as fixed effects. The random effects structure for each analysis was determined by starting with the maximum structure justified by the experimental design (Barr et al., 2013), with slopes being removed due to non-convergence (based on which variables contributed the least amount of unique variance to the random effects structure).

2.2 Results

As seen in **Table 2**, the results from the AdvVS production task and the grammaticality judgment task revealed a range in productive and receptive knowledge of German AdvVS word order

among the L2 participants, but most participants scored at or close to ceiling on these tasks.[3] Results from the L2-to-L1 translation task demonstrated that the L2 participants knew, on average, 86.7% of the lexical verbs from the eye-tracking stimuli.[4] Turning to the eye-tracking task, participants were highly accurate in clicking on the target at 98.6% across both groups (see **Table 2**), so data from all trials were included in the analyses reported below.

Table 2: Experiment 1: Ancillary task results.

	L1 Speakers		**L2 Speakers**	
	M(SD)	***Range***	***M(SD)***	***Range***
AdvVS Production (out of 9)	8.9 (0.4)	7–9	8.1 (1.8)	2–9
Mean GJT rating (1–4)				
AdvVS-Gram	3.8 (0.23)	3.1–4.0	3.5 (0.4)	2.6–4.0
AdvSV-Ungram	1.5 (0.59)	1.0–3.1	1.8 (0.9)	1.0–3.8
SVAdvO-Gram	3.9 (0.15)	3.6–4.0	3.5 (0.4)	2.6–4.0
SAdvVO-Ungram	1.4 (0.58)	1.0–3.1	1.8 (0.9)	1.0–3.9
Translation accuracy: Nouns (% correct)	NA	NA	94.2 (3.2)	84.3–96.9
Translation accuracy: Verbs (% correct)	NA	NA	86.7 (13.1)	53.1–100
Mouse-click accuracy (% correct)	98.3 (2.5)	90.6–100	98.8 (1.9)	93.8–100

Note. For the grammaticality judgment task, 1 = definitely unacceptable; 4 = definitely acceptable.

2.2.1 Divergence point analyses

Figure 2a displays the mean proportion of looks to the target, along with the bootstrapped means and CIs, for each group and sentence type, aligned and time-locked to the onset of the finite verb. **Figure 2b** displays the distribution of bootstrapped divergence points for each group and sentence type. As seen in **Figure 2a**, for both L1 and L2 listeners, the mean divergence points and their 95% confidence intervals (CI) were before the onset of the target in constraining-verb sentences for both sentence types (L1 SVO: M = 807 ms, 95% CI [765, 833]; L2 SVO: M = 1174 ms, 95% CI [1071, 1326]; L1 AdvVS: M = 896 ms, 95% CI [816, 1003]; L2 AdvVS: M = 1318 ms, 95% CI [1207, 1445]).

[3] One L1 German speaker did not complete the grammaticality judgment task. So the L1 speaker results are based on results from 31 participants.

[4] Excluding items for which the L2 listeners did not supply an appropriate translation for the target verb would have removed 13.4% of the data. In a second set of analyses, in which such items were removed, we found the same pattern of results for the divergence point analyses (see OSF for these supplemental analyses). Since translation tasks tend to underestimate receptive vocabulary knowledge, as participants also need to come up with an appropriate L1 translation, we decided to maximize statistical power and used the full data set in all analyses reported here.

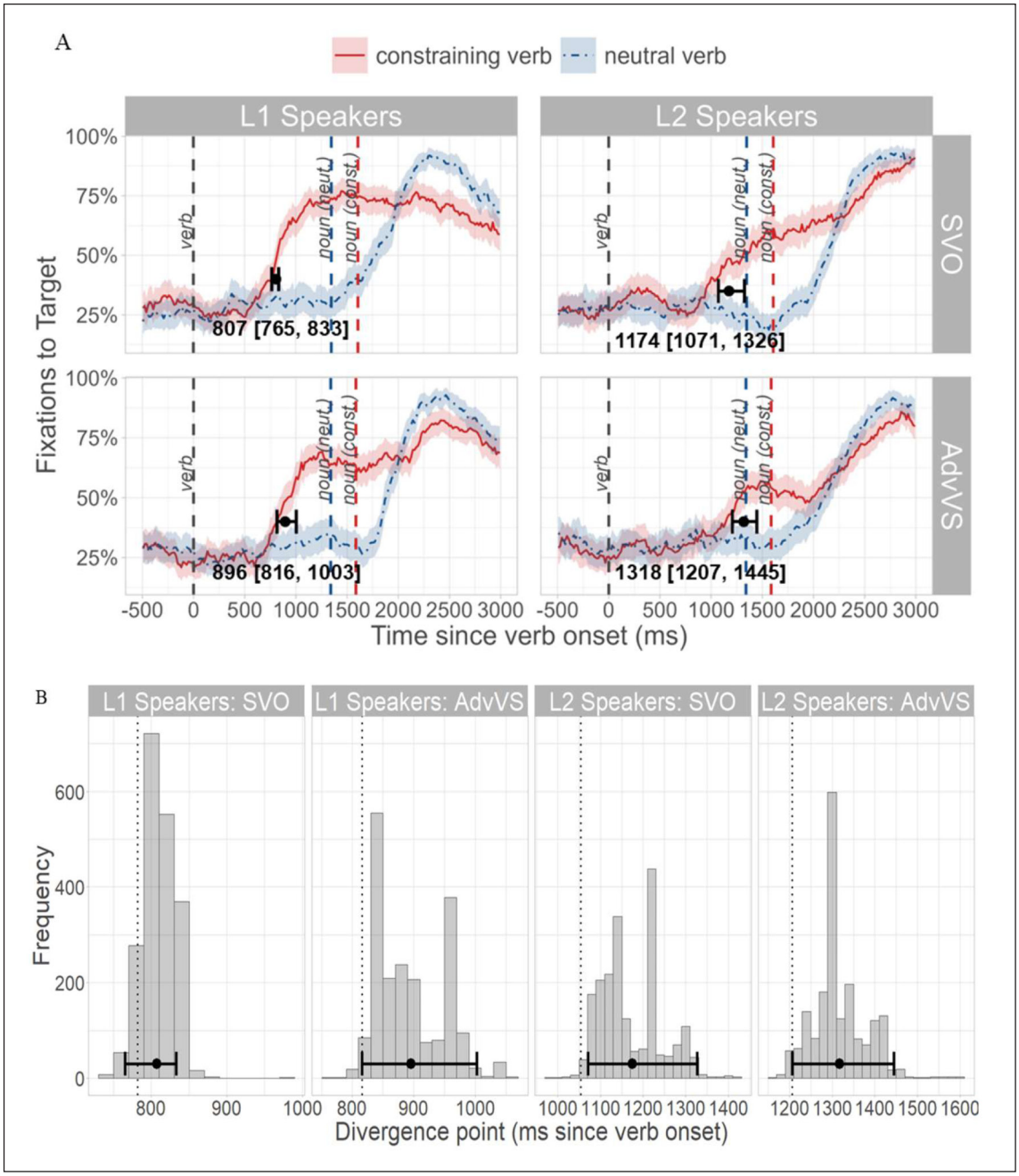

Figure 2: Experiment 1: **(A)** Proportion of looks to target referent for constraining-verb and neutral-verb sentences by group and sentence type. 0 ms is verb onset. Vertical lines indicate mean verb and target noun onsets. Error bands on the fixation curves show 95% confidence intervals. Points with error bars overlaid on the fixation curves represent the bootstrapped mean divergence point and its 95% confidence interval. **(B)** Bootstrap distribution of divergence points for each group and sentence type. The x-axis illustrates the distribution of divergence points based on 2000 bootstraps. The y-axis illustrates the number of resamples where a given divergence point was observed. Points with error bars represent the bootstrapped mean divergence point and its 95% confidence interval. Dotted vertical lines represent the original divergence points prior to resampling.

To compare the divergence points between the L1 and L2 listeners for each sentence type, we bootstrapped the difference between their divergence points (see the supplemental materials on OSF for figures illustrating the distribution of differences). The mean difference in divergence points between the L1 and L2 listeners for AdvVS sentences was 421 ms, 95% CI = [267, 578]. The mean difference in divergence points between the L1 and L2 listeners for SVO sentences was 367 ms, 95% CI = [255, 527]. Neither CI crossed zero, indicating that the L2 listeners were reliably slower than the L1 listeners for both sentence types. However, the CIs for both sentence types largely overlapped, indicating that the L2 slowdown was similar for both sentence types.

2.2.2 Time window analyses for proficiency effects within the L2 group

For the full results, see Tables S1 and S2 in the supplemental materials on OSF. In the time window analyses (200ms post-verb onset to 200 ms post-target noun onset) for both sentence types, there was a robust effect of Constraint, with the proportion of looks to the target greater in constraining-verb sentences than in neutral-verb sentences (SVO: $t = 4.52, p < .001$; AdvVS: $t = 3.43, p < .001$). There was a main effect of Proficiency, with more looks to the target overall as proficiency increased (SVO: $t = 2.04, p = .051$; AdvVS: $t = 2.54, p = .011$). There were no significant interactions between Constraint and L2 Proficiency (SVO: $t = 1.90, p = .068$; AdvVS: $t = 1.81, p = .072$), indicating that the proportion of looks to the target in constraining-verb sentences did not reliably differ from looks to the target in neutral-verb sentences as a function of L2 proficiency. However, we acknowledge that, given the relatively small sample size of L2 participants, the absence of any significant interactions may be due to the lack of statistical power sufficient to detect such interactions.

2.3 Discussion

Both L1 and L2 German listeners were faster to look at the post-verbal object in SVO sentences with constraining verbs, as opposed to neutral verbs, and did so prior to target noun onset, paralleling results from previous research (e.g., Chambers & Cooke, 2009; Dijkgraaf et al., 2017; Ito, Corley, et al., 2018). The same pattern emerged in AdvVS sentences, in that both L1 and L2 German listeners were faster to look at the post-verbal subject in AdvVS sentences with constraining verbs, as opposed to neutral verbs, and did so prior to target noun onset, paralleling previous work with L1 listeners on semantic prediction in Dutch AdvVS sentences (Dahan & Tanenhaus, 2004). The L2 listeners were slower than the L1 listeners to anticipate the target referent for both SVO and AdvVS sentences, similar to other studies that directly compared the use of semantic cues for prediction by L1 vs. L2 speakers (e.g., Chun & Kaan, 2019; Dijkgraaf et al., 2019; Ito, Corley, et al., 2018). Critically, the mean difference in divergence points between the L1 and L2 listeners overlapped for each sentence type, suggesting that the L2 listeners exhibited no additional prediction delay with AdvVS sentences as opposed to SVO sentences, even though this required them to anticipate a post-verbal subject in a syntactic configuration not present

in their L1. Further, in neither SVO nor AdvVS sentences did the proportion of looks to the target in constraining-verb versus neutral-verb sentences vary as a function of L2 proficiency. Although one must interpret such null effects with caution, this suggests that L2 proficiency did not modulate the magnitude of L2 semantically-cued prediction. In all, the findings suggest that L2 speakers – like L1 speakers – can use semantic cues for prediction across different word orders.

However, in AdvVS sentences, the target referent was also the only semantically-plausible referent associated with the lexical verb. Thus, the increased looks to the target referent after a semantically-constraining verb could stem from simple semantic association, rather than reflecting evidence that the participants specifically predicted a post-verbal *subject* (e.g., Chow et al., 2016; Kukona et al., 2011). To tease apart these two possibilities, we designed Experiment 2 to include transitive AdvVSO sentences, alongside visual scenes that depicted two possible post-verbal referents, namely, a plausible (agent) subject and a plausible (patient) object of the lexical verb. Given that Experiment 1 was also the first L2 study to our knowledge to examine the semantic prediction of post-verbal subjects, rather than post-verbal objects, we also included AdvVS sentences in Experiment 2, to test whether we could replicate these results from Experiment 1 with a new group of L2 speakers.

3 Experiment 2

In Experiment 2, participants listened to intransitive AdvVS and transitive AdvVSO sentences with the same experimental design as in Experiment 1. The goal of Experiment 2 was to replicate and build on the results for AdvVS sentences from Experiment 1. More specifically, through the inclusion of transitive AdvVSO sentences and image displays that depicted both a plausible subject (agent) and a plausible object (patient), Experiment 2 could tease apart whether listeners use semantic cues from the verb to predict *any* verbal argument via simple semantic association or whether syntax guides expectations and leads listeners to specifically predict the *subject* (agent) referent. If just any plausible verbal argument is activated upon hearing the verb (e.g., Kukona et al., 2011), predictive looks to the subject referent should be attenuated in AdvVSO sentences with constraining verbs, as opposed to neutral verbs, relative to what is found for AdvVS sentences. More importantly, within AdvVSO sentences, we would expect looks to the potential subject (agent) and object (patient) following the verb to be similarly distributed even in the constraining-verb condition, as both are semantically associated with the verb. However, if prediction of an upcoming argument is also constrained by the syntax of the unfolding AdvVSO sentence, we expect more looks to the plausible subject (agent) than to the plausible object (patient) for constraining-verb sentences, compared to neutral-verb sentences, in which any animate referent could plausibly follow the verb. Further, if we see a similar pattern of results with AdvVS and AdvVSO sentences among both L1 and L2 listeners, this would suggest that the use of semantic information for prediction is qualitatively similar for both L1 and L2 speakers. However, if the

pattern of results among L1 and L2 listeners differs, this would be consistent with the proposal that the use of higher sentence-level cues for prediction may be more difficult for L2 speakers, leading to a greater reliance on simple associative mechanisms (e.g., Ito & Pickering, 2021).

3.1 Methods

3.1.1 Participants

Thirty-one L2 German speakers living in the United States and 38 L1 German speakers living in Germany participated in the study in exchange for payment. One L2 German speaker was excluded because they were an L1 speaker of Dutch, another V2 language, one L2 German speaker was excluded due to excessive track-loss during the eye-tracking task, and two L2 German speakers were excluded because they knew fewer than 50% of the lexical verbs from the experimental stimuli, as measured by the L2-to-L1 translation task. Three L1 German speakers were excluded because they did not list German as an L1 on the language background questionnaire, one was excluded because they reported significant exposure to English in early childhood, and two were excluded because they never clicked on the correct target referent during the eye-tracking task. Thus, all analyses are based on 27 L2 German participants (11 female; 15 male; 1 non-binary) and 32 L1 German participants (25 female; 7 male). Most L2 participants spoke English as (one) of their first languages (24), but other L1s included Arabic (1), Konkani (1), Portuguese (1), Spanish (1) and Yoruba (1). The results from the German proficiency task (Goethe Institut, 2004) revealed a range in proficiency from A2 level to C1 level. This range was similar to the proficiency level in Experiment 1, although a t-test revealed that the L2 participants in Experiment 2 were, overall, slightly more proficient than the L2 participants in Experiment 1 ($t(57) = 2.70$, $p = .034$). See **Table 3** for complete biographical information.

Table 3: Experiment 2: Participant information.

	L1 Speakers (n = 32)		**L2 Speakers (n = 27)**	
	M(SD)	***Range***	***M(SD)***	***Range***
Age (years)	22.1(1.7)	20–27	24.7 (7.8)	18–48
Age of first exposure to German (years)	NA	NA	13.3 (3.7)	4–19[5]
Total time abroad (months)	NA	NA	11.5 (6.9)	0–31
Goethe test score (out of 30)	NA	NA	19.3 (4.5)	10–27

[5] Similar to Experiment 1, two participants reported limited exposure to German through school-based instruction of 1–2 hours a week in early elementary school. Given that this exposure was limited, and, in all other ways, these participants had similar background and proficiency profiles as the other participants, they were included in the final pool of participants.

3.1.2 Materials

3.1.2.1 Eye-tracking task

The experimental stimuli consisted of 16 AdvVS sentence pairs and 16 AdvVSO sentence pairs. The AdvVS sentences, their image displays and the audio recordings were the same as those used in Experiment 1. As seen in example (8), in AdvVSO sentences, both the subject and object were animate.

(8) a. AdvVSO-constraining verb

In	der Nacht	erschießt	plötzlich	der Jäger	einen Tiger	im	Dschungel.
in	the night	shoots.V	suddenly	the hunter.SUB	a Tiger.OBJ	in.the	jungle

b. AdvVSO-neutral verb

In	der Nacht	muss	plötzlich	der Jäger	einen Tiger	im
in	the night	must.V-MOD	suddenly	the hunter.SUB	a Tiger.OBJ	in.the

Dschungel	erschießen.
jungle	shoot.V

'At night, the hunter shoots/must shoot the tiger suddenly in the jungle.'

The log frequencies of the lexical verbs and target nouns for AdvVSO sentences (verb: $M = 2.7$, $SD = 0.6$; target noun: $M = 2.7$, $SD = 0.8$) did not differ from the corresponding log frequencies in the AdvVS sentences, based on values from SUBTLEX-DE (Brysbaert et al., 2011; all $ps > .1$).

As seen in **Figure 3**, the image display for AdvVSO sentences included an image of both the agent subject (henceforth referred to as the *agent*) and the patient object (henceforth referred to as the *patient*). The remaining two referents were inanimate. As with the AdvVS sentences, all four nouns had the same grammatical gender, so that participants could not use the gender-marked determiner (e.g., *der*) as an additional cue to identify the target referent. A repeated-measures ANOVA revealed that the log frequency of the agent noun ($M = 2.7$, $SD = 0.8$) did not differ from the log frequency of the patient noun ($M = 2.6$, $SD = 0.7$) or the two inanimate distractor nouns ($M = 2.6$, $SD = 0.6$), based on values from SUBTLEX-DE (Brysbaert et al., 2011; $p > .5$). The positions of the agent and patient were counterbalanced across all items.

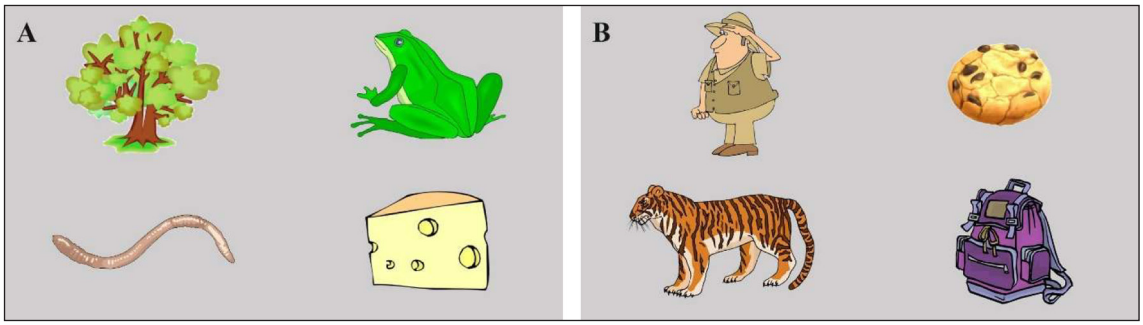

Figure 3: (A) Image display for sentences in (5) (Experiment 1). **(B)** Image display for sentences in (8).

To test for semantic predictability, the 16 AdvVSO items and their respective image displays were divided into two lists, such that each list contained eight constraining-verb and eight neutral-verb items. Twenty-one German native speakers, none of whom participated in the main experiment, read the first half of each experimental sentence (e.g., *In der Nacht erschießt plötzlich…*) and they were prompted to choose which image in the display came next in the sentence. For AdvVSO sentences, the participants chose the agent 78.8% of the time in constraining-verb sentences (SD = 17.3), but only 26.0% of the time in neutral-verb sentences (SD =27.1), a difference that was statistically significant ($t(15) = 8.09$, $p < .001$).

The 32 experimental items were divided into two lists, so that participants saw eight constraining-verb and eight neutral-verb sentences for each sentence type (AdvVS and AdvVSO), but no more than one version of any given sentence pair. They were interspersed with 32 filler items that were similar in length and structure to the experimental items. Twelve filler items from Experiment 1 were modified so that the image display contained two referents from the sentence, similar to the set-up of the image display for AdvVSO sentences. All other filler items and their image displays were identical to Experiment 1. As in Experiment 1, the experimental and filler items were presented in pseudorandomized order, so that participants never saw more than one experimental item from the same condition in a row.

The AdvVSO sentences were recorded by the same female native speaker of German from Experiment 1, with a sampling rate of 44.1 kHz. In AdvVSO sentences, the mean length of the critical region from verb onset to noun onset was longer for constraining-verb sentences (M = 1642 ms, SD = 165) than neutral-verb sentences (M = 1344 ms, SD = 219), but there was no significant difference between the AdvVS and AdvVSO sentences for either condition (all $ps > .3$).

3.1.2.2 Additional tasks

The format of this task was identical to the L2-to-L1 translation task in Experiment 1. It contained the 32 target nouns and 32 lexical verbs from the AdvVS and AdvVSO sentences and the 16 patient nouns from the AdvVSO sentences. The AdvVS production task and the grammaticality judgment task were identical to those in Experiment 1.

3.1.3 Procedure

The following changes were made to the procedure from Experiment 1. In the eye-tracking task, the participants were instructed to click on *the first object* that was mentioned in the sentence, since in Experiment 2, both the agent and the patient in AdvVSO sentences were depicted in the visual display. As the L1 German participants in Experiment 1 were at ceiling on the AdvVS production task and the grammaticality judgment task, only the L2 participants in Experiment 2 completed these two tasks. All other procedures were identical to Experiment 1.

3.1.4 Data analysis

Data extraction and preprocessing were identical to Experiment 1. Track loss led to the exclusion of 0.1% of L1 participant trials and 2.6% L2 participant trials. One L2 participant was excluded because fewer than four trials per experimental condition remained. As in Experiment 1, all analyses were conducted separately for each sentence type (AdvVS vs. AdvVSO).

Following the same procedures as in Experiment 1, we initially used divergence point analyses to identify the onset of predictive looks to the target referent for constraining-verb sentences versus neutral-verb sentences for each sentence type (AdvVS and AdvVSO) and each participant group (L1 and L2 listeners). We then used linear mixed-effects models to explore whether L2 proficiency modulated the size of any prediction effects, following the same procedures as in Experiment 1. To specifically address whether listeners use verbal semantics to predict the syntactically required post-verbal argument (RQ2), however, it is critical to also investigate potential competition effects between the agent (e.g., *Jäger* 'hunter') and the patient (e.g,. *Tiger* 'tiger') in AdvVSO sentences. Therefore, we conducted a second divergence point analysis in which we compared the onset of predictive looks to these two images in AdvVSO-constraining verb sentences and AdvVSO-neutral verb sentences, respectively.

3.2 Results

As seen in **Table 4**, results from the AdvVS production task and the grammaticality judgment task revealed a range in productive and receptive knowledge of German AdvVS word order among the L2 participants, but most participants scored at, or close to, ceiling on these tasks. The results from the L2-to-L1 translation task demonstrated that all L2 participants knew, on average, 81.1% of the lexical verbs from the eye-tracking stimuli. T-tests revealed that the L2 participants in Experiment 2 scored significantly higher than the L2 participants in Experiment 1 on the AdvVS production task, $t(57) = 2.07$, $p = .043$. No other differences were statistically significant (all $ps > .15$). Turning to the eye-tracking task, participants were highly accurate in clicking on the target noun at 95.2% across both groups (see **Table 4**), so data from all trials were included in the analyses reported below.[6]

[6] Excluding items for which the L2 listeners did not know the target verbs would have removed 20.6% of the data. As in Experiment 1, we decided to maximize statistical power and used the full data set in all analyses reported here. In a second set of analyses, in which such items were removed, we found the same pattern of results for the divergence point analyses, except for the comparison of constraining-verb sentences versus neutral-verb sentences for AdvVSO sentences among the L2 listeners (see OSF for these supplemental analyses). For AdvVSO sentences among the L2 listeners, none of the 2000 bootstraps converged, revealing no reliable divergence point in the relevant time period. This underscores the point that any divergence point for AdvVSO sentences among the L2 speakers from the analyses using the full data set should be interpreted with caution.

Table 4: Experiment 2: Ancillary task results.

	L1 Speakers		L2 Speakers	
	M(SD)	***Range***	***M(SD)***	***Range***
AdvVS Production (out of 9)	NA	NA	8.9 (0.5)	7–9
Mean GJT rating				
AdvVS-Gram	NA	NA	3.7 (0.4)	2.6–4.0
AdvSV-Ungram	NA	NA	1.8 (0.9)	1.0–3.8
SVAdvO-Gram	NA	NA	3.4 (0.4)	2.6–4.0
SAdvVO-Ungram	NA	NA	1.6 (0.8)	1.0–4.0
Translation accuracy: Nouns (% correct)	NA	NA	94.2 (6.1)	76.0–100
Translation accuracy: Verbs (% correct)	NA	NA	81.1 (17.9)	50.0–100
Mouse-clicking accuracy (% correct)	95.0 (5.1)	75–100	95.4 (9.0)	53.1–100

Note. For the grammaticality judgment task, 1 = definitely unacceptable; 4 = definitely acceptable.

3.2.1 Divergence point analyses (constraining-verb vs. neutral-verb sentences)

As seen in **Figure 4a**, for L1 listeners, the mean divergence points and their 95% confidence intervals (CI) were before the onset of the target in constraining-verb sentences for both sentence types (AdvVS: M = 906 ms, 95% CI [765, 1105]; AdvVSO: M = 1055 ms, 95% CI [986, 1224]). For L2 listeners, the mean divergence point in AdvVS constraining-verb sentences was before the target onset, and the bootstrapped divergence points occurred within the range of 200 ms of the target onset, as demonstrated by the righthand tail of the 95% CI (AdvVS: M = 1512 ms, 95% CI [1207, 1768]; see also **Figure 4b**), and, thus, prior to when listeners would launch saccades to the target referent as a result of processing auditory information from the target noun (Huettig et al., 2011).[7] In AdvVSO constraining-verb sentences, the mean divergence points and their 95% confidence intervals (CI) for L2 listeners were before the target onset (AdvVSO: M = 1416 ms, 95% CI [1258, 1630]). However, as seen in **Figure 4b**, only 163 of the 2000 bootstraps converged. Therefore, this result should be interpreted with caution.

[7] An anonymous reviewer observed that there is a bimodal distribution of bootstrapped divergence points for the L2 speakers for AdvVS sentences, in contrast to both sentence types in Experiment 1. While this may be an indication of individual differences in divergence points across participants, discussing the potential for individual differences in L1 and L2 semantically-cued prediction is beyond the scope of the present study. However, it would be an interesting avenue for future research.

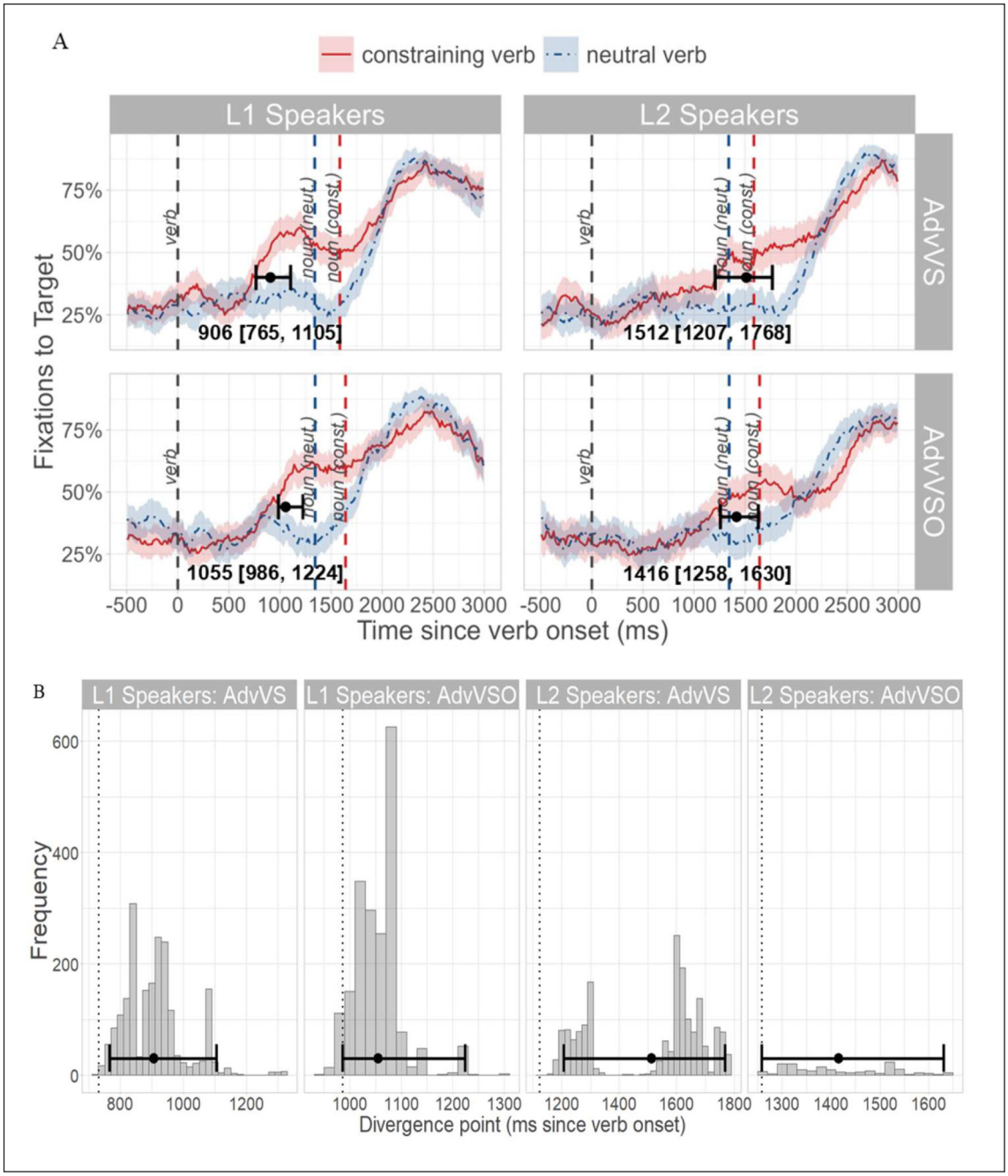

Figure 4: Experiment 2: **(A)** Proportion of looks to target referent for constraining-verb and neutral-verb sentences by group and sentence type. 0 ms is verb onset. Vertical lines indicate mean verb and target noun onsets. Error bands on the fixation curves show 95% confidence intervals. Points with error bars overlaid on the fixation curves represent the bootstrapped mean divergence point and its 95% confidence interval. **(B)** Bootstrap distribution of divergence points for each group and sentence type. The x-axis illustrates the distribution of divergence points based on 2000 bootstraps. The y-axis illustrates the number of resamples where a given divergence point was observed. Points with error bars represent the bootstrapped mean divergence point and its 95% confidence interval. Dotted vertical lines represent the original divergence points prior to resampling.

The mean difference in divergence points between the L1 and L2 listeners for AdvVS sentences was 606 ms, 95% CI = [187, 932] (see the supplemental materials on OSF for figures illustrating the distribution of differences). The mean difference in divergence points between the L1 and L2 listeners for AdvVSO sentences was 361 ms, 95% CI = [204, 578]. Neither CI crossed zero, indicating that the L2 listeners were reliably slower than the L1 listeners for both sentence types. However, the CIs for both sentence types largely overlapped, indicating that the L2 slowdown was similar for both sentence types.

3.2.2 Time window analyses for proficiency effects within the L2 group

For the full results, see Tables S5 and S6 in the supplemental materials on OSF. In the time window analysis (200 ms post-verb onset to 200 ms post-target noun onset) for both sentence types, there was a robust effect of Constraint in both models, with the proportion of looks to the target referent greater in constraining-verb sentences than in neutral-verb sentences (AdvVS: $t = 3.70$, $p < .001$; AdvVSO: $t = 2.47$, $p = .014$). No robust effects of L2 Proficiency emerged (AdvVS: $t = –0.44$, $p = .662$; AdvVSO: $t = 1.90$, $p = .058$). There were no significant interactions between Constraint and L2 Proficiency (AdvVS: $t = 1.40$, $p = .162$; AdvVSO: $t = –0.18$, $p = .861$), indicating that the proportion of looks to the target in constraining-verb sentences did not reliably differ from looks to the target in neutral-verb sentences as a function of L2 proficiency. However, we acknowledge that, given the relatively small sample size of L2 participants, the absence of any significant interactions may be due to the lack of statistical power sufficient to detect such interactions.

3.2.3 Divergence point analyses (AdvVSO sentences: agent vs. patient)

To illustrate potential competition effects between the agent (e.g., *Jäger* 'hunter') and the patient (e.g,. *Tiger* 'tiger') in AdvVSO sentences, **Figure 5** shows fixations to the agent versus the patient. As seen in **Figure 5a**, for both L1 and L2 listeners, the mean divergence points and their 95% confidence intervals (CI) were before the onset of the agent noun in constraining-verb sentences (L1: $M = 988$ ms, 95% CI [884, 1054]; L2: $M = 1455$ ms, 95% CI [1224, 1581]). In contrast, for the L2 listeners the mean divergence point and 95% CIs occur after the onset of the agent noun in neutral-verb sentences ($M = 2060$ ms, 95% CI [1921, 2176]). As seen in **Figure 5b**, for the L1 listeners, the mean divergence point and the majority of bootstrapped divergence points also occur after the onset of the agent noun in neutral-verb sentences ($M = 1583$ ms, 95% CI [782, 1751]). However, a small number of bootstrapped divergence points occur prior to the onset of the agent noun, such that the lower 95% CI extends into the prediction window.

The mean difference in divergence points between the L1 and L2 listeners for constraining-verb sentences was 467 ms, 95% CI = [221, 629] (see the supplemental materials on OSF for figures illustrating the distribution of differences). The mean difference in divergence points between the L1 and L2 listeners for neutral-verb sentences was 475 ms, 95% CI = [255, 1309]. Neither CI crossed zero, indicating that the L2 listeners were reliably slower than the L1 listeners for both sentence types. Again, the CIs for both constraining-verb and neutral-verb sentences largely overlapped, indicating that the L2 slowdown was similar for both sentence types.

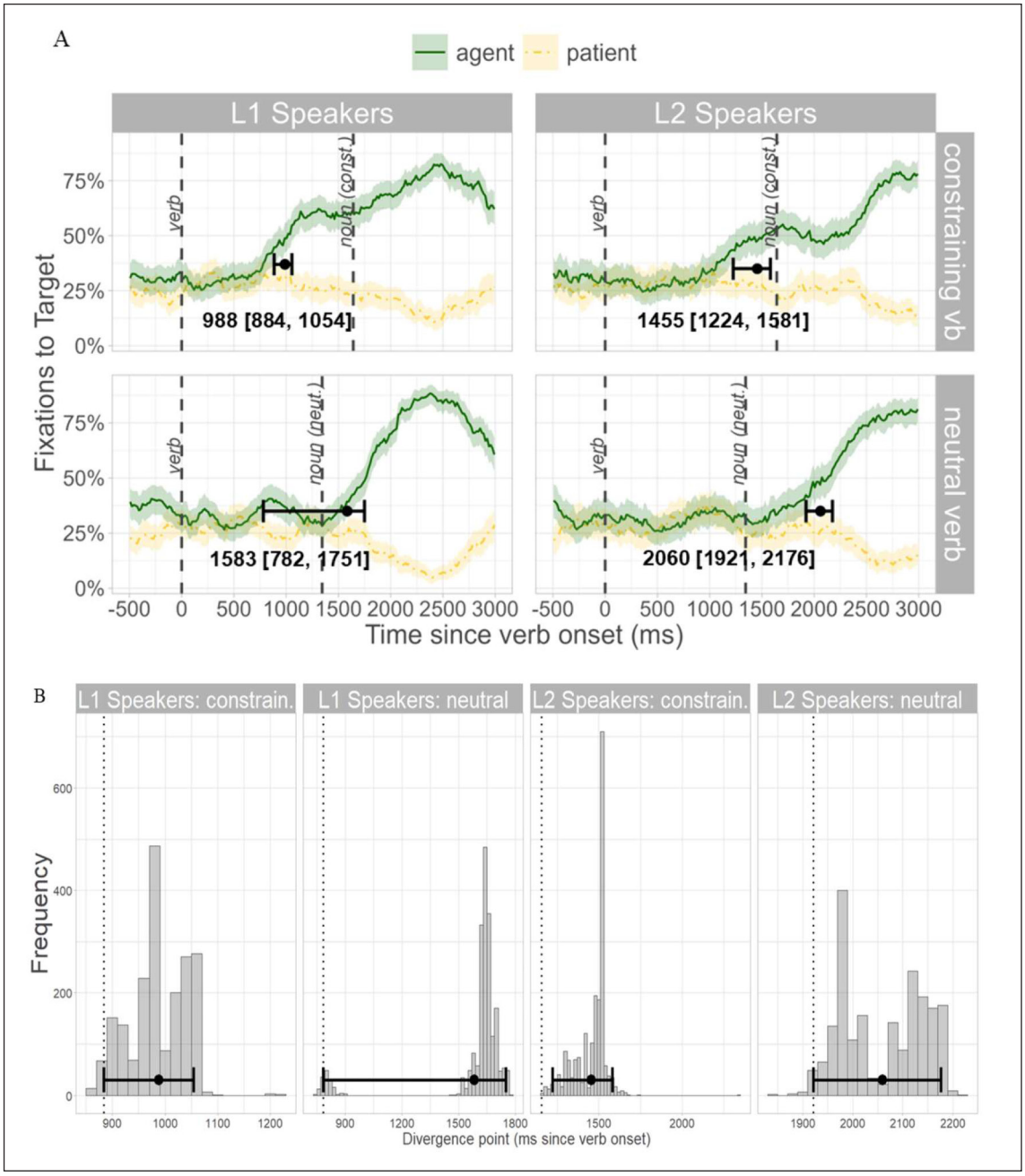

Figure 5: Experiment 2: **(A)** Proportion of looks to agent versus patient in AdvVSO sentences, by group and constraint. 0 ms is verb onset. Vertical lines indicate mean verb and agent noun onsets. Error bands on the fixation curves show 95% confidence intervals. Points with error bars overlaid on the fixation curves represent the bootstrapped mean divergence point and its 95% confidence interval. **(B)** Bootstrap distribution of divergence points for each group and sentence type. The x-axis illustrates the distribution of divergence points based on 2000 bootstraps. The y-axis illustrates the number of resamples where a given divergence point was observed. Points with error bars represent the bootstrapped mean divergence point and its 95% confidence interval. Dotted vertical lines represent the original divergence points prior to resampling.

3.3 Discussion

Similar to Experiment 1, both L1 and L2 German listeners were faster to look at the post-verbal subject in AdvVS sentences with constraining verbs, as opposed to neutral verbs. Both groups were also faster to look at the post-verbal subject (agent) in AdvVSO sentences with constraining verbs, as opposed to neutral verbs, and did so prior to target noun onset, although this effect was less reliable among the L2 listeners. More importantly, a second set of divergence point analyses with AdvVSO sentences revealed that both the L1 and L2 listeners were more likely to look at the agent than the patient, and did so prior to noun onset only in AdvVSO sentences with a semantically-constraining verb. Looks to the agent did not reliably diverge from looks to the patient until after target noun onset in neutral-verb sentences in either listener group. This suggests that both the L1 and L2 listeners' predictive eye movements were not guided by semantic association with the verb alone. Rather, they preferentially looked at the referent that was most likely to be mentioned next in the given syntactic configuration. Notably, this preference was found in both L1 and L2 listeners.

As in Experiment 1, the mean difference in divergence points between the L1 and L2 listeners was similar across all analyses, suggesting that the L1 and L2 listeners used qualitatively similar mechanisms for prediction and that any differences between the two groups were only quantitative in scope. As in Experiment 1, in neither AdvVS nor AdvVSO sentences did the proportion of looks to the target referent in constraining-verb versus neutral-verb sentences vary as a function of L2 proficiency. Although one must interpret such null effects with caution, this suggests that L2 proficiency did not modulate the magnitude of L2 semantic prediction – and, more specifically, the magnitude of predicting a post-verbal subject – in this experiment.

4 General discussion

The primary goal of the present study was to investigate whether L1 and L2 German listeners use semantic cues from the verb to predict post-verbal arguments across different word orders, especially in a word order not present in L2 listeners' L1 (Experiment 1), and whether such prediction is guided by syntactic context (i.e., word order and the distribution of grammatical and thematic roles) among both L1 and L2 listeners (Experiment 2). In Experiment 1, both L1 and L2 listeners were faster to look at the post-verbal object in SVO sentences and the post-verbal subject in AdvVS sentences when the sentences contained a semantically-constraining finite verb, as opposed to a semantically-neutral auxiliary verb. Experiment 2 replicated the L1 and L2 results with AdvVS sentences in Experiment 1 and extended this pattern to AdvVSO sentences. More importantly, a second set of analyses with just AdvVSO sentences revealed that only in sentences with a semantically-constraining finite verb – but not in sentences with a neutral auxiliary verb – both the L1 and L2 listeners looked to the agent more than the patient prior to hearing the first postverbal noun. In all analyses across both experiments, although the L2 listeners were reliably slower to make predictions than the L1 listeners, the mean difference in divergence points between the two groups was similar, suggesting that any differences in

prediction between the L1 and L2 listeners were quantitative, not qualitive, in nature. Further, across both experiments, the proportion of looks to the target referent in constraining-verb versus neutral-verb sentences among L2 listeners did not vary as a function of L2 proficiency.

4.1 Word order and argument structure constrain predictive processing

Experiment 1 showed that listeners use the semantics of a verb in V2 position to anticipate the post-verbal argument in German, regardless of whether that argument is a subject or an object. Further, the results with AdvVSO sentences in Experiment 2 suggest that the use of semantic information to predict argument roles cannot stem only from semantic association, in contrast to other research (e.g., Chow et al., 2016; Kukona et al., 2011). Critically, in constraining-verb AdvVSO sentences, both L1 and L2 listeners looked to the target (agent) subject more than the (patient) object prior to the onset of the post-verbal noun. In other words, both groups preferentially anticipated a plausible subject. This is more compatible with the hypothesis that, even in contexts in which strong associative connections exist between a verb and its arguments, speakers use sentence-level cues, like word order and argument structure, for prediction, at least when these are sufficiently strong and reliable.

This appears to contrast with the results for English active sentences from Kukona et al. (2011, Experiment 1), in which L1 English listeners looked at both a plausible agent (e.g., policeman) and a plausible patient (e.g., crook) when listening to sentence fragments like *Toby arrests the…* up until the onset of the target noun phrase. The present results are more in line with Kukona and colleagues' results for English passive sentences, like *Toby was arrested by…* (Experiment 2), in which competition effects between plausible agent and plausible patient referents were attenuated by the time participants encountered the word *by*. Kukona and colleagues attributed this difference in active versus passive sentences to the fact that their passive sentences provided participants with additional syntactic cues (i.e., *was* and *by*) to constrain the target referent, as well as more time for these additional cues to influence prediction. As in Kukona and colleagues' Experiment 2, sentences in the present study contained an additional word (a semantically neutral adjective, e.g., *täglich*, 'daily') between the verb and the post-verbal argument, thus providing more time for the integration of multiple cues and for predictive effects to emerge, compared to Kukona and colleagues' Experiment 1. We also note that in both Kukona and colleagues' passive sentences in their Experiment 2 and in our AdvVS(O) sentences, the post-verbal argument and target of prediction was the *agent* of the event. We hypothesize that agents may be privileged as targets of anticipation, as they are the most critical thematic role in establishing event representations (e.g., Bornkessel & Schleswesky, 2006).

Further support for this hypothesis comes from recent work with L1 Tagalog speakers (Bondoc & Schafer, 2022; Sauppe, 2016). Tagalog is a verb-initial language in which the order of the subsequent arguments is based on thematic roles (agent versus patient) and morphological markers on the verb that confer a more prominent status on one of the two arguments, which is referred to as the *pivot* (see Sauppe, 2016, and references cited therein for further detail).

The pivot argument canonically and most frequently appears clause-finally, while the non-pivot argument appears immediately after the verb, leading to both VSO and VOS sentences. Sauppe (2016) showed that regardless of morphological pivot markers on the verb, and, thus, regardless of the most likely upcoming word order, participants were more likely to look at the most plausible agent after hearing the verb. Sauppe concluded that Tagalog speakers initially use verbal semantics to anticipate potential referents based on their thematic roles, rather than using morphological cues to identify the syntactic function (i.e., pivot role) and word order of the target referents (see also Bondoc & Schafer, 2022, for similar results). Together, these studies and our study highlight how agents may be more prominent than patients, leading listeners to identify the most plausible agent early on in sentence processing (e.g., Bornkessel & Schlesewsky, 2006). This line of research underscores the point that the mechanisms underlying semantically-cued prediction among both L1 and L2 listeners go beyond the use of simple semantic association to include, at the very least, the consideration of grammatical and thematic roles and their distribution across different word orders and argument structures.

To further disentangle the contribution of grammatical versus thematic roles in prediction, future studies should consider including sentences with verbs that require thematic roles other than agents and patients, such as implicit causality verbs associated with stimulus and experiencer roles, as in (9).

(9) a. An Halloween verängstigt die Hexe die Frau im Dorf.
'At Halloween the witch frightens the woman in the village.'

b. An Halloween fürchtet die Frau die Hexe im Dorf.
'At Halloween the woman fears the witch in the village.'

In both (9a) and (9b), the event-causing stimulus (i.e., *Hexe* 'witch') is considered more prominent than the experiencer (i.e., *Frau* 'woman'), even though *Hexe* is the grammatical subject in (9a), but the grammatical object in (9b) (e.g., Hartshorne, 2014). If prediction is driven primarily by the prominence of thematic roles, then one would expect listeners to look at an event-causing stimulus, like *Hexe*, after hearing the verb in both (9a) and (9b). However, if prediction is driven primarily by the anticipation of the next grammatically possible referent, namely, the subject, then one would expect decreased looks to an event-causing stimulus, like *Hexe*, after hearing the verb in (9b) versus (9a). Such research could uncover additional evidence regarding the exact mechanisms that drive prediction, and which types of linguistic information are most crucial for prediction, during L1 and L2 processing.

4.2 L1 versus L2 predictive processing

Despite the delay in divergence points for the L2 listeners versus the L1 listeners, no qualitative differences in predictive processing emerged, paralleling other L2 studies that investigated the use of semantic cues for prediction (e.g., Chun & Kaan, 2019; Chun et al., 2021; Dijkgraaf et al., 2017, 2019; Ito, Corley, et al., 2018; Ito, Pickering, et al., 2018). Importantly, in contrast to

previous L2 studies, which have all used sentences with SVO word order, the L2 listeners here engaged in semantically-cued prediction even with AdvVS(O) sentences – a word order that is not part of their L1 and that diverges from a more canonical agent-verb-patient word order. Further, like the L1 listeners, the L2 listeners always fixated on the most plausible post-verbal argument, even in AdvVS(O) sentences, when the syntactic configuration required listeners to use sentence-level cues to constrain their predictions to the subject (agent) of the target sentence, rather than using verbal semantics alone to anticipate any plausible post-verbal referent. Together, these results underscore the point that L2 speakers do not always privilege simple semantic association, instead using both syntactic context and argument structure to guide prediction.

Although the absence of any interaction between L2 proficiency and the use of verbal semantics in Experiments 1 and 2 must be interpreted with caution, given the relatively small sample size in both experiments, this result adds to the growing number of studies showing that L2 proficiency may not always modulate L2 prediction. In line with other recent L2 research, we tentatively hypothesize that L2 speakers will generate predictions based on those cues that are most reliable in the input, given their L2 knowledge and experience more broadly (see Kaan & Grüter, 2021, for further discussion). For the L2 listeners in this study, verbal semantics were a reliable cue for anticipating the most plausible and syntactically expected post-verbal argument across different word orders, regardless of L2 proficiency. This included the anticipation of post-verbal subjects in AdvVS(O) sentences – even in the absence of any experience using verbal semantics to anticipate post-verbal subjects from their L1.

5 Conclusion

Using the visual world paradigm, the present study expands the scope of L1 and L2 prediction research by investigating how listeners use verbal semantics to predict different post-verbal arguments in both subject-initial SVO sentences and non-subject initial AdvVS(O) sentences. The results highlight the fact that, for both L1 and L2 speakers, the use of semantic cues during prediction is not based solely on semantic associations, but also involves the anticipation of the grammatically most likely upcoming argument, based on the syntax of the sentence as it unfolds. This suggests that syntax and argument structure, in addition to lexical semantics, reliably guide prediction across a variety of word orders, even when these word orders require listeners to anticipate upcoming referents with different grammatical and thematic roles. Further, L2 speakers engage in similar predictive processes, regardless of whether they are listening to sentences with SVO word order or other word orders that differ from those in their L1. In terms of the underlying mechanisms of prediction, the present study demonstrates that differences that may exist between L1 and L2 semantically-cued prediction cannot be reduced to L2 speakers over-relying on simple associative mechanisms. In all, the study underscores the conclusion that both L1 and L2 speakers integrate bottom-up and top-down cues when generating expectations about upcoming linguistic input in real-time comprehension.

Abbreviations

AdvVS = Adverb-Verb-Subject word order

AdvVSO = Adverb-Verb-Subject-Object word order

Gram = grammatical

L1 = first language

L2 = second language

OBJ = Direct object

SUB = Subject

SVO = Subject-Verb-Object word order

Ungram = ungrammatical

V = Verb

V-MOD = Modal verb

V2 = Verb-second

Data accessibility statement

All experimental materials and data that the authors have used and have the right to share are available at https://doi.org/10.17605/OSF.IO/TJYPQ. Supplementary files with additional analyses from Experiments 1 and 2 are available at https://doi.org/10.17605/OSF.IO/TJYPQ.

Ethics and consent

This study was approved by the Institutional Review Board at Penn State University (protocol #13949), and all participants provided informed consent.

Acknowledgements

We would like to thank Valérie Keppenne, Shane Cummings, Felicity Sarnoff, Carly Levy, Maura Jaeger, Rebecca Adler, Naomi Kailasam, Lena Milz, Emilia Peinecke, and Leah Wildner for their help with data preparation and data collection.

Competing interests

The authors have no competing interests to declare.

Authors' contributions

Carrie N. Jackson: Conceptualization, Methodology, Data collection, Formal analysis, Writing – Original draft, Writing – Review & Editing, Project administration.

Holger Hopp: Conceptualization, Methodology, Data collection, Writing – Review & Editing.

Theres Grüter: Conceptualization, Methodology, Data collection, Writing – Review & Editing.

ORCiD IDs

Carrie N. Jackson: https://orcid.org/0000-0003-0284-850X

Holger Hopp: https://orcid.org/0000-0003-3026-3349

Theres Grüter: https://orcid.org/0000-0001-6354-9787

References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Altmann, G. T., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518. https://doi.org/10.1016/j.jml.2006.12.004

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. https://doi.org/10.48550/arXiv.1406.5823. https://doi.org/10.18637/jss.v067.i01

Bondoc, I. P., & Schafer, A. J. (2022). Differential effects of agency, animacy, and syntactic prominence on production and comprehension: Evidence from a verb-initial language. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *76*(4), 302–326. https://psycnet.apa.org/doi/10.1037/cep0000280. https://doi.org/10.1037/cep0000280

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review*, *113*(4), 787–821. https://psycnet.apa.org/doi/10.1037/0033-295X.113.4.787. https://doi.org/10.1037/0033-295X.113.4.787

Borovsky, A., Burns, E., Elman, J. L., & Evans, J. L. (2013). Lexical activation during sentence comprehension in adolescents with history of specific language impairment. *Journal of Communication Disorders*, *46*(5–6), 413–427. https://doi.org/10.1016/j.jcomdis.2013.09.001

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, *112*(4), 417–436. https://doi.org/10.1016/j.jecp.2012.01.005

Bovolenta, G., & Marsden, E. (2022). Prediction and error-based learning in L2 processing and acquisition: A conceptual review. *Studies in Second Language Acquisition*, *44*(5), 1384–1409. https://doi.org/10.1017/S0272263121000723

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2017). Goals and strategies influence lexical prediction during sentence comprehension. *Journal of Memory and Language*, *93*, 203–216. https://doi.org/10.1016/j.jml.2016.10.002

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology, 58*(5), 412–424. https://doi.org/10.1027/1618-3169/a000123

Carroll, M., Murcia-Serra, J., Watorek, M., & Bendiscioli, A. (2000). The relevance of information organization to second language acquisition studies: The descriptive discourse of advanced adult learners of German. *Studies in Second Language Acquisition*, *22*(3), 441–466. https://doi.org/10.1017/S0272263100003065

Chambers, C. G., & Cooke, H. (2009). Lexical competition during second-language listening: Sentence context, but not proficiency, constrains interference from the native lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 1029–1040. https://psycnet.apa.org/doi/10.1037/a0015901. https://doi.org/10.1037/a0015901

Chambers, C. G., & San Juan, V. (2008). Perception and presupposition in real-time language comprehension: Insights from anticipatory processing. *Cognition*, *108*(1), 26–50. https://doi.org/10.1016/j.cognition.2007.12.009

Chow, W. Y., Smith, C., Lau, E., & Phillips, C. (2016). A "bag-of-arguments" mechanism for initial verb predictions. *Language, Cognition and Neuroscience*, *31*(5), 577–596. https://doi.org/10.1080/23273798.2015.1066832

Chun, E., Chen, S., Liu, S., Chan, A., Kaan, E., & Grüter, T. (2021). Influence of syntactic complexity on second language prediction. In E. Kaan & T. Grüter (Eds.), *Prediction in second language processing and learning* (pp. 69–90). John Benjamins. https://doi.org/10.1075/bpa.12.04chu

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

Corps, R. E., Liao, M., & Pickering, M. J. (2023). Evidence for two stages of prediction in non-native speakers: A visual-world eye-tracking study. *Bilingualism: Language and Cognition*, *26*(1), 231–243. https://doi.org/10.1017/S1366728922000499

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 498. https://doi.org/10.1037/0278-7393.30.2.498

Declerck, M., & Kormos, J. (2012). The effect of dual task demands and proficiency on second language speech production. *Bilingualism: Language & Cognition, 15*(4), 782–796. https://doi.org/10.1017/S1366728911000629

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

Francis, A. L., Tigchelaar, L. J., Zhang, R., & Zekveld, A. A. (2018). Effects of second language proficiency and linguistic uncertainty on recognition of speech in native and nonnative competing speech. *Journal of Speech, Language, and Hearing Research, 61*(7), 1815–1830. https://doi.org/10.1044/2018_JSLHR-H-17-0254

Goethe Institut. (2004). *Einstufungstest [Placement test]*. Available at: https://www.goethe.de/en/spr/kup/tsd.html (accessed September 2020).

Hartshorne, J. K. (2014). What is implicit causality? *Language, Cognition and Neuroscience*, *29*(7), 804–824. https://doi.org/10.1080/01690965.2013.796396

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

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. https://doi.org/10.1017/9781009245470

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

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

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

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

Jackson, C. N., & Ruf, H. T. (2017). The priming of word order in second language German. *Applied Psycholinguistics*, *38*(2), 315–345. https://doi.org/10.1017/S0142716416000205

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

Kamide, Y. (2012). Learning individual talkers' structural preferences. *Cognition*, *124*(1), 66–71. https://doi.org/10.1016/j.cognition.2012.03.001

Kamide, Y., Scheepers, C., & Altmann, G. T. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, *32*, 37–55. https://doi.org/10.1023/A:1021933015362

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*(2), 205–225. https://doi.org/10.1016/j.jml.2004.10.002

Knoeferle, P., & Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language*, *57*(4), 519–543. https://doi.org/10.1016/j.jml.2007.01.003

Kukona, A., Fang, S. Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition*, *119*(1), 23–42. https://doi.org/10.1016/j.cognition.2010.12.002

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

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

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Martin, C. D., Thierry, G., Kuipers, J. R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, *69*(4), 574–588. https://doi.org/10.1016/j.jml.2013.08.001

O'Brien, M. G., & Féry, C. (2015). Dynamic localization in second language English and German. *Bilingualism: Language and Cognition*, *18*(3), 400–418. https://doi.org/10.1017/S1366728914000182

Peters, R. E., Grüter, T., & Borovsky, A. (2018). Vocabulary size and native speaker self-identification influence flexibility in linguistic prediction among adult bilinguals. *Applied Psycholinguistics*, *39*(6), 1439–1469. https://doi.org/10.1017/S0142716418000383

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

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347. https://doi.org/10.1017/S0140525X12001495

Pienemann, M. (1998). *Language processing and second language development: Processability theory*. John Benjamins. https://doi.org/10.1075/sibil.15

Psychology Software Tools, Inc. (2016). *E-Prime 3.0* [Computer software].

Qualtrics. (2020). *Qualtrics* [Computer software].

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation. https://www.r-project.org/

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

Sauppe, S. (2016). Verbal semantics drives early anticipatory eye movements during the comprehension of verb-initial sentences. *Frontiers in Psychology*, *7*, 95. https://doi.org/10.3389/fpsyg.2016.00095

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 linguistic domains: Evidence from German. In E. Kaan & T. Grüter (Eds.), *Prediction in second language processing and learning* (pp. 47–68). John Benjamins. https://doi.org/10.1075/bpa.12.03sch

Stone, K., Lago, S., & Schad, D. J. (2021). Divergence point analyses of visual world data: Applications to bilingual research. *Bilingualism: Language and Cognition*, *24*(5), 833–841. https://doi.org/10.1017/S1366728920000607

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634. https://doi.org/10.1126/science.7777863