frontiers in Language Sciences

TYPE Original Research
PUBLISHED 04 March 2026
DOI 10.3389/flang.2026.1756463

OPEN ACCESS

EDITED BY
Matthew W. Crocker,
Saarland University, Germany

REVIEWED BY
Chia-Wen Lo,
Max Planck Institute for Human Cognitive and Brain Sciences, Germany
Mingjun Wu,
Jiangsu University of Science and Technology, China

*CORRESPONDENCE
Chie Nakamura
cnakamura@waseda.jp

RECEIVED 28 November 2025
REVISED 30 January 2026
ACCEPTED 02 February 2026
PUBLISHED 04 March 2026

CITATION
Nakamura C, Flynn S, Miyamoto Y and Yusa N (2026) Lexical vs. structural cue use in L2 prediction: filler-gap parsing ability shapes learners' information use. *Front. Lang. Sci.* 5:1756463.
doi: 10.3389/flang.2026.1756463

Lexical vs. structural cue use in L2 prediction: filler-gap parsing ability shapes learners' information use

Chie Nakamura[1]*, Suzanne Flynn[2], Yoichi Miyamoto[3] and Noriaki Yusa[4]

[1]School of International Liberal Studies, Waseda University, Tokyo, Japan, [2]Department of Linguistics and Philosophy, Massachusetts Institute of Technology, Cambridge, MA, United States, [3]Graduate School of Humanities, The University of Osaka, Osaka, Japan, [4]Department of English Faculty of Liberal Arts, Miyagi Gakuin Women's University, Miyagi, Japan

This study examines whether second language (L2) sentence processing is governed by the same underlying mechanisms as native language processing or whether it relies on qualitatively distinct mechanisms. Using the visual-world paradigm and permutation analyses, we compared native English speakers and Japanese second language (L2) learners of English in processing globally ambiguous filler-gap dependencies (e.g., Where did Lizzie tell someone that she was going to catch butterflies?). By distinguishing L2 learners based on their comprehension accuracy for unambiguous filler-gap sentences, we identified systematic variation in the mechanisms guiding predictive processing. High-accuracy learners exhibited anticipatory eye-movement patterns comparable to those of native speakers, consistent with the use of structurally guided predictive dependency formation. In contrast, low-accuracy learners also showed predictive behavior, but this prediction was driven primarily by lexical or surface-level regularities rather than structural information. Importantly, neither the structure-based prediction observed in the high-accuracy group nor the lexical cue-based predictive observed in the low-accuracy group can be attributed to direct transfer from Japanese. Together, these results support a gradient view of L2 sentence processing in which qualitatively different predictive mechanisms coexist and may shift as a function of learners' structural computation ability, rather than a simple contrast between non-predictive and native-like processing.

KEYWORDS
eye-tracking, filler-gap dependencies, incremental parsing, L2 sentence processing, permutation analysis

1 Introduction

Sentence comprehension involves incremental parsing (Frazier, 1987; Frazier and Clifton, 1989; Kamide, 2008). Native speakers typically integrate linguistic cues in real time, using prior input to anticipate upcoming linguistic material. Rather than relying solely on bottom-up information (e.g., the linear input of words as they appear), they engage in predictive processing to guide interpretation. This predictive strategy supports more efficient real-time comprehension (Altmann and Kamide, 1999; Omaki et al., 2014). In second language (L2) processing, research has examined the extent to which L2 learners engage in predictive mechanisms similar to those used by native speakers of a particular target language. A central question in this earlier work was whether L2 learners anticipate upcoming input based on contextual and structural cues, or whether they rely more heavily on reactive, integration-based processing in which information is incorporated only after it is encountered.

Debates about the underlying mechanisms of L2 processing were strongly shaped by the Shallow Structure Hypothesis (SSH). SSH proposes that L2 learners tend to rely more heavily on surface-level cues such as word order and lexical associations than on abstract syntactic structures or hierarchical representations, constructing syntactic representations that are shallower or less detailed than those of native speakers (Clahsen and Felser, 2006; see also Clahsen and Felser, 2018; Felser, 2019 for updated reviews). Supporting this view, cross-modal priming and other studies have reported reduced sensitivity to syntactic dependencies, including filler-gap relations, suggesting that real-time syntactic integration is less robust in L2 processing than in L1 processing (Felser and Roberts, 2007; Marinis et al., 2005).

Although SSH does not explicitly address predictive processing in L2 comprehension, it has often been taken to imply that L2 learners may be limited in their ability to engage in structure-based anticipation because they underutilize abstract syntactic information. At the same time, because SSH concerns the depth of syntactic representations rather than the timing of processing, it remains compatible with the possibility that L2 learners engage in predictive processing, but that their predictions are guided more by surface level cues than by hierarchical syntactic structure. This interpretation highlights an important limitation of the predictive vs. non-predictive debate in L2 processing. Qualitative differences between native speakers and L2 learners may arise not from whether prediction occurs, but from the types of cues that guide anticipatory behavior in each group.

In the context of this shift in focus, recent work has moved beyond the earlier predictive vs. non-predictive framing and moved toward more fine-grained questions about what L2 learners predict, when do they predict it, and what cues support their anticipatory behavior (Grüter and Rohde, 2021; Kaan and Grüter, 2021). This emerging gradient refers to the view that predictive processing is not a categorial yes or no property but rather exists along a continuum, varying in strength and timing depending on how efficiently learners can recruit available linguistic cues in real time. Empirical findings illustrate this variability. Some studies report weaker or less consistent predictive behavior in L2 learners when prediction depends on verb semantics, thematic roles, or grammatical gender (Grüter et al., 2012; Hopp, 2013; Lew-Williams and Fernald, 2010; Grüter et al., 2017). Yet other work indicates that L2 learners can engage in robust predictive parsing when reliable cues are available. They have been shown to anticipate direct object positions based on filler-gap constraints (Omaki and Schulz, 2011), to use prosodic cues flexibly for referential prediction (Nakamura et al., 2021), and to exploit semantic and classifier cues during anticipatory looks (Grüter et al., 2020). Together, these findings suggest that L1 and L2 comprehenders both engage in predictive processing but might differ in the types of cues they rely on during anticipation.

A critical next step, therefore, is to examine whether L2 learners can recruit structure-based cues in contexts where native speakers reliably do so. Filler-gap dependencies offer a particularly revealing test case. Prior work on sentences such as (1), in which a wh-phrase can be linked either to the main clause (MC) verb (tell, i.e., Where did Lizzie tell someone?) or to the embedded clause (EC) verb (catch, i.e., Where was she going to catch butterflies?), has shown that native English speakers exhibit a robust preference for the MC interpretation (Frazier and Clifton, 1989; Omaki et al., 2014). This preference has been attributed to the Active-Gap-Filling Hypothesis and the principle of Minimal Dependency Formation (Phillips, 2006), which assume that comprehenders construct dependencies incrementally and prioritize structurally higher attachment sites. Related evidence from other languages shows similar structurally driven preferences in L1 processing (Nakamura et al., 2025a).

(1) Where did Lizzie tell someone__ that she was going to catch butterflies__?

In L2 processing, prior studies have shown that learners are also capable of incremental dependency formation. For example, Omaki and Schulz (2011) demonstrated that L2 learners, like native speakers, process filler-gap dependencies incrementally and are sensitive to island constraints, indicating an ability to construct syntactic dependencies in real time. However, because their study did not examine structurally ambiguous configurations such as (1), this evidence does not address whether learners rely on when multiple interpretations are possible. Thus, while these findings suggest that L2 learners can build filler-gap dependencies incrementally, it remains unclear whether their predictive mechanisms are guided by the same structural cues that drive native speakers' active gap filling.

In addition to investigating the group level differences between native speakers and L2 learners, we also aim to investigate how different types of cues are used within the L2 learner population. Recent research indicates that the ability to compute syntactic dependencies does not always align with global proficiency, and that fine grained grammatical knowledge can be a better predictor of the cues learners rely on during comprehension (Hopp, 2013; Tanner et al., 2013; Cunnings, 2017). These findings suggest that individual differences in predictive behavior may reflect differences in the stability of specific structural representations rather than differences in overall proficiency, highlighting the need to examine variability with the L2 group rather than treating learners as a uniform population.

To investigate this possibility, we tested a relatively homogeneous group of upper intermediate to advanced L2 learners with comparable L2 learning background, allowing us to examine variation in sentence processing mechanisms independently of broad proficiency differences. Within this group, we assessed learners' accuracy on unambiguous filler-gap items such as (2) as an index of structural computation ability, that is, how reliably they construct syntactic dependencies in real time during comprehension. Unambiguous filler-gap sentences such as (2) require learners to identify the grammatically licensed gap position, for which there is only a single structurally valid analysis. Crucially, the correct interpretation in these cases cannot be derived from lexical associations or surface-level heuristics alone. As such, this measure provides a targeted index of learners' consistency in forming syntactic dependencies during online processing, capturing variability in underlying processing mechanisms that is not directly observable through standard proficiency tests.

(2) Where did Lizzie tell someone how she was going to catch butterflies__?

By combining between-group (L1 vs. L2) and within-group comparison between high vs. low accuracy L2 learners on items such as (2), and by testing how this variation in structural computation ability shapes predictive eye movement patterns during the comprehension of globally ambiguous wh dependencies such as (1), our design provides a critical test of whether L2 sentence processing reflects parsing mechanisms that differ qualitatively from those used by native speakers, or whether L2 processing develops along a gradient path toward native like, structure-based prediction. To answer these questions, we employed a visual-world eye-tracking paradigm, which tracks participants' eye movements as they process spoken wh-questions and provides fine-grained temporal data about their moment-by-moment interpretive commitments. To determine when these commitments emerged, we conducted cluster-based permutation analyses, a data-driven statistical method that identifies time windows in which visual attention significantly diverges between competing interpretations.

2 Methods

2.1 Participants

We recruited 51 native Japanese speakers learning English as their L2 and 25 native English speakers as a control group. All participants reported normal or corrected-to-normal vision[1]. The L2 participants were undergraduate students at Waseda University in Tokyo, residing in Japan at the time of testing, and had studied English for at least 6 years. None reported current immersion in an English-speaking environment. Their proficiency ranged from B2 to C1 on the Common European Framework of References for Languages (CEFR), indicating upper-intermediate to advanced competence. Importantly, this relatively restricted proficiency range allowed us to examine variation in sentence processing mechanisms within a broadly comparable proficiency group, rather than differences attributable to overall L2 proficiency level.

Within the L2 group, we assessed individual differences in structural sensitivity using accuracy on unambiguous filler-gap questions (4b). One L2 participant with exceptionally low accuracy on unambiguous filler-gap questions (0.08), more than three standard deviations below the group mean (0.71) was excluded. The final dataset therefore included 50 L2 participants (later divided into high-and low-accuracy subgroups of 27 and 23, respectively) and 25 native speakers.

1 Participants were prospectively recruited between July 2024 to October 2024. All participants provided written informed consent before taking part in the experiment. The study was approved by the Ethics Review Committee on Research with Human Subjects of Waseda University, under protocol number 2024-130. Only adult participants who were 18 years old or older were included in the study.

2.2 Materials and design

The experiment investigated how L2 learners and native speakers process structurally ambiguous wh-questions involving filler-gap dependencies. Participants' eye movements were recorded while they listened to target wh-questions and viewed a corresponding visual display (Figure 1). Each trial consisted of a context sentence, followed by a target question, allowing us to examine the real-time resolution of structural ambiguity during sentence comprehension.

2.3 Stimuli and experimental conditions

The context sentences introduced two events corresponding to the MC and the EC in the subsequent target wh-question (e.g., telling and catching). We manipulated the order in which these events were introduced across two context conditions: an embedded clause-main clause (EC-MC) condition (3a), in which the EC event was introduced first and the MC event second, and a main clause-embedded clause (MC-EC) condition (3b), in which the MC event preceded the EC event. This manipulation was included to control for recency-based biases and to ensure that interpretation preferences reflected syntactic processing rather than memory-related effects.

(3)

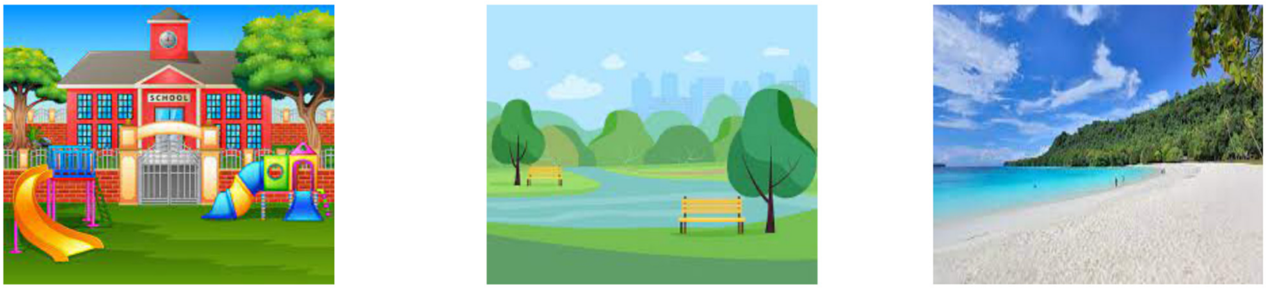
a. EC-MC event condition:
Lizzie caught butterflies in the park using her net. In the afternoon, she saw her friend in the schoolyard and told him about it.

b. MC-EC event condition:
Lizzie saw her friend in the schoolyard. She told him, "I caught butterflies in the park using my net".

The target wh-question was presented in two conditions: an Ambiguous condition (4a), in which the fronted where-phrase could plausibly be associated with either the matrix verb (*tell*) or the embedded verb (*catch*), and an Unambiguous condition (4b), in which the structure of the question constrained interpretation to a single syntactic position. In the Unambiguous question condition, *how* replaced the complementizer *that* in the EC, grammatically restricting the interpretation to the MC event (i.e., *tell*, referring specifically to the schoolyard).

(4)

a. Ambiguous question condition:
Where did Lizzie tell someone that she was going to catch butterflies?



a. Main clause (MC) interpretation b. Embedded clause (EC) interpretation c. Distracter

FIGURE 1
Three pictures presented with (4). **(a)** Main clause (MC) interpretation; **(b)** Embedded clause (EC) interpretation; **(c)** Distracter.

Answer: Ambiguous (could refer to the schoolyard or the park)

b. Unambiguous question condition:
Where did Lizzie tell someone how she was going to catch butterflies?
Answer: Unambiguous (refers only to the schoolyard, the telling location)

2.4. Procedure and task

Each trial followed a fixed sequence. First, a context sentence introducing two events was presented auditorily twice, and the same sentence was displayed simultaneously as text on the computer screen, to ensure comprehension of the discourse context. This was followed by a 3,000 ms preview of a visual display containing three pictures corresponding to the MC interpretation, the EC interpretation, and a distracter (Figure 1). Participants then heard the target wh-question sentence while viewing the visual display. Eye movements were recorded from the onset of the target question until sentence offset. At the end of each trial, participants indicated their interpretation by pressing a key corresponding to one of the three pictures. The procedure is illustrated in Figure 2.

2.5. Counterbalancing and item distribution

The experiment employed a 2 x 2 within-participants design with Context (EC-MC vs. MC-EC) and Question Type (Ambiguous vs. Unambiguous) as factors, yielding four experimental conditions. Each participant completed 24 target items and 36 filler items. Questions in the filler items did not involve structural ambiguity (e.g., *Where did the police officer arrest someone?*). Each sentence set appeared in only one condition for a given participant, and ambiguous and unambiguous versions of the same item were counterbalanced across experimental lists. The position of the three pictures (MC interpretation, EC interpretation, and distractor) was also counterbalanced across items, such that no picture type appeared consistently in a given screen location.

2.6. Eye-tracking apparatus and experimental setting

Eye movements were recorded using a tower-mounted EyeLink 1,000 plus eye-tracking system (SR Research), sampling at 1,000 Hz. Participants were seated approximately 60 cm from the monitor in a quiet laboratory room. Participants' head position was stabilized using a chin rest. Eye-tracker calibration was performed using a 9-point calibration at the beginning of the experiment and repeated as necessary to ensure recording accuracy. Eye movements were recorded continuously throughout each trial, from the onset of the target question until the participant provided a response.

3 Results

We first report overall behavioral results based on key-press responses, followed by the eye-tracking results. Analyses are presented first for the overall comparison between native speakers and all L2 learners, and then for high- and low-accuracy L2 subgroups, defined by comprehension accuracy on unambiguous filler-gap questions.

3.1 Key-press responses

3.1.1 Native vs. all L2 learners

The key-press responses were analyzed to assess the participants' final interpretation of the wh-question's intended referent in each condition by comparing response patterns between native speakers and L2 learners. Accuracy on filler questions was high for both groups, with native speakers achieving 91.9% (SD = 4.8) and L2 learners achieving 83.4% (SD = 7.5), indicating that all the participants remained engaged and attentive throughout the task. For unambiguous target questions (4b), native speakers demonstrated 89.5% accuracy (SD = 7.9), while L2 learners performed at 69.6% accuracy (SD = 19.9). These results suggest that although both groups were generally able to process the structure, L2 learners experienced greater difficulty resolving

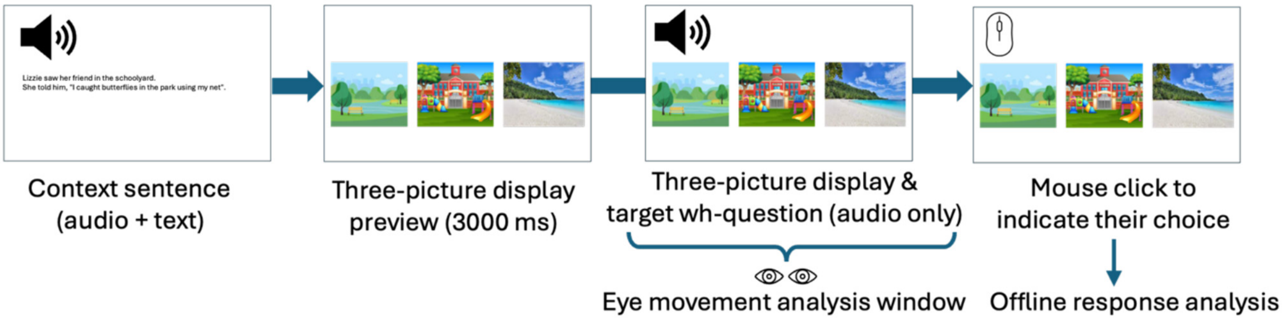


FIGURE 2
Trial procedure.

the filler-gap dependency, even when disambiguating cues were present. Accuracy on these unambiguous items was subsequently used as an index of individual differences in structural sensitivity within the L2 group.

Crucially, in the ambiguous target questions (4a), where there was no objectively correct answer regarding the two possible event locations, we examined the proportion of MC vs. EC picture selections. Figure 3a illustrates native speakers' choices, revealing a clear preference for the MC interpretation. Native speakers selected the MC picture 84% of the time and the EC picture only 15.3% of the time. A logistic regression confirmed that the likelihood of choosing the MC option was significantly higher than chance ($\beta = 1.99$, SE $= 0.30$, $z = 6.78$, $p < 0.001$). Figure 3b shows responses from all L2 learners, who selected the MC picture chosen 70.3% of trials and the EC picture on 26.5%. A logistic regression likewise indicated that L2 learners were significantly more likely to select the MC picture ($\beta = 2.48$, SE $= 0.47$, $z = 5.22$, $p < 0.001$). These results suggest that L2 learners, like native speakers, displayed a robust overall preference for the MC interpretation over the EC interpretation when sentence was structurally ambiguous.

3.1.2 Control analysis: effects of context event order

To assess whether participants' interpretation choices were influenced by the order in which events were introduced in the context sentence, we conducted a control analysis on key-press responses reported in Section. 3.1.1 using binomial generalized linear mixed-effects models (Baayen et al., 2008). Context order (MC–EC vs. EC–MC) was entered as a fixed effect, with random intercepts for participants and items.

For native speakers, context order did not significantly affect interpretation choices ($\beta = -0.48$, SE $= 0.40$, $z = -1.17$, $p = 0.24$), indicating that the strong preference for the MC interpretation cannot be attributed to discourse presentation order or recency effects. The same analysis conducted for L2 learners likewise revealed no reliable effect of context order ($\beta = -0.30$, SE $= 0.20$, $z = -1.46$, $p = 0.14$), suggesting that their MC preference was not driven by the linear order of events in the discourse context. Together, these results confirm that the key-press preferences reported above reflect structural interpretation biases rather than superficial discourse-order effects.

3.1.3 High- vs. low-accuracy L2 learners

To examine whether the individual differences in the stability of learners' structural representations influenced their interpretation of ambiguous filler-gap dependencies, we compared key-press responses between the high- and low-accuracy L2 subgroups. These groups were defined based on participants' accuracy on unambiguous filler-gap questions, divided at the median score of 0.75. Importantly, both subgroups demonstrated solid performance on filler items, indicating comparable task engagement and general comprehension (high-accuracy group mean 87.0%, low-accuracy group mean 79.2%).

Despite their comparable performance on filler items, the two subgroups showed markedly different response patterns on the critical ambiguous items. As shown in Figure 4, the high-accuracy group selected the MC interpretations on 83% of the trials, compared to 55.4% in the low-accuracy group. Conversely, the low-accuracy group chose the EC interpretation more frequently (40.6%) compared to the higher-accuracy group (14.5%). A logistic mixed effects model run separately for each subgroup confirmed this pattern. Only the high accuracy group reliably preferred the MC interpretation over the EC interpretation ($\beta = 2.09$, SE $= 0.38$, $z = 5.56$, $p < 0.001$). In contrast, the low accuracy group showed only a numerical MC preference that did not reach significance ($\beta = 0.25$, SE $= 0.21$, $z = 1.20$, $p = 0.23$).

3.2 Eye-tracking results

3.2.1 Permutation analysis: native vs. all L2 learners

To compare fixation patterns between native speakers and all the L2 learners, we conducted a cluster-based permutation analysis to identify time windows during which looks to the MC and EC pictures diverged significantly. This approach was adopted given that the timing of predictive and integrative processes is known to vary substantially across populations and languages

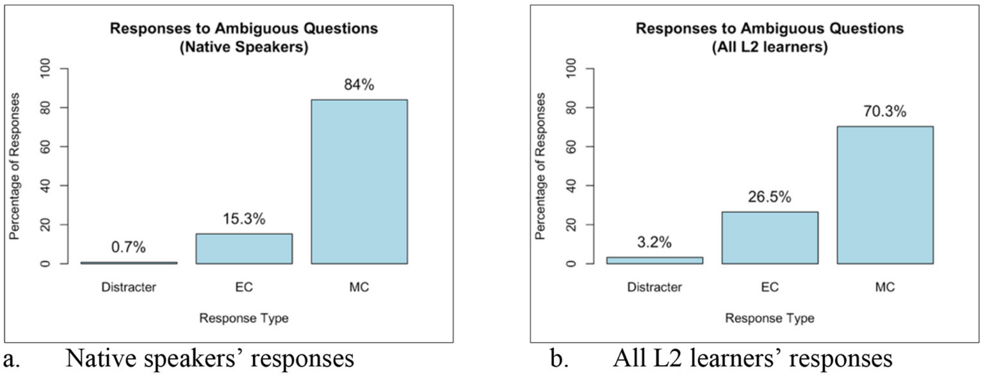


FIGURE 3
Key-press responses for ambiguous questions: **(a)** native speakers; **(b)** all L2 learners.

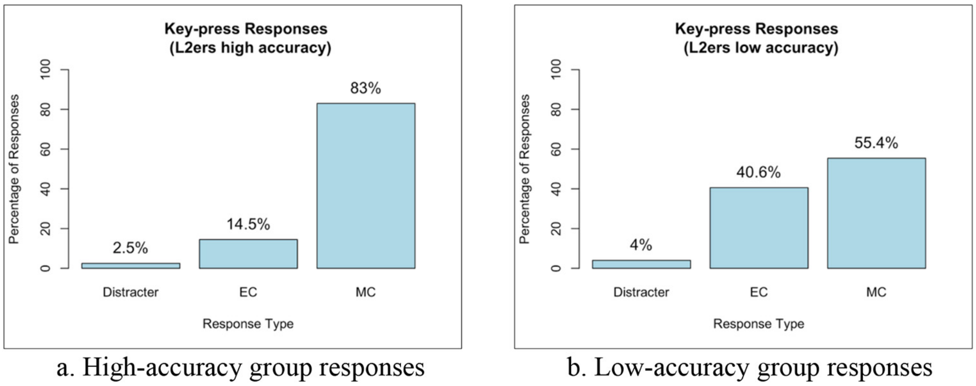


FIGURE 4
Key-press responses for ambiguous questions: **(a)** high-accuracy L2 learners; **(b)** low-accuracy L2 learners.

(Chan et al., 2018). It is possible that fixed, pre-specified time windows may not be appropriate when comparing groups that differ in processing speed or linguistic experience, because the time window in which native speakers show a reliable effect may not be correspond to the timing of the same effect in L2 learners. Cluster-based permutation analyses provide a principled way to identify temporally contiguous effects without imposing arbitrary time windows, while appropriately controlling for multiple comparisons across time.

Following Chan et al. (2018), fixation data were segmented into consecutive 20 ms time bins across a 3,000 ms analysis window, time-locked to the onset of the question. For each time bin, we fit a binomial logistic regression model predicting fixation behavior as a function of picture type (MC vs. EC). The resulting z-statistic associated with picture type was extracted for each time bin. Adjacent bins exceeding the cluster-forming threshold of $p < 0.05$ were grouped into temporal clusters. Cluster-level significance was assessed using a permutation-based Monte Carlo procedure (1,000 permutations). In each permutation, picture-type labels of MC and EC were randomly reassigned within time bins, and the same bin-wise regression and clustering procedure was applied. For each permuted dataset, only the maximum absolute summed cluster-level statistic was retained, yielding a null distribution of maximal cluster statistics. An observed clusters were considered statistically significant if their summed z-statistic exceeded 95% of the values in this null distribution, corresponding to a two-tailed cluster-level significance threshold of $p < 0.05$, intrinsically correcting for multiple comparisons across time. All permutation analyses were conducted using custom scripts implemented in R (R Core Team). Bin-wise logistic regression models were fit using base R functions, and visualization of fixation time-course data was carried out using ggplot 2. Details of the permutation analysis procedure are provided in Appendix A.

Figure 5 shows the proportion of looks to MC and EC pictures for native speakers (top) and L2 learners (bottom), time-locked to the onset of the target ambiguous questions. The blue line indicates looks to the MC picture, and the red line indicates looks to the EC picture. Gray shaded regions mark time windows identified as significant by the permutation analysis. Orange downward bars indicate individual 20 ms time bins that did not reach significance prior to cluster formation.

The permutation analysis revealed that both native speakers and L2 learners showed a greater proportion of looks to the MC picture than the EC picture, but the timing and pattern of this divergence differed across groups. For native speakers, a single significant cluster was observed, during which fixations on the

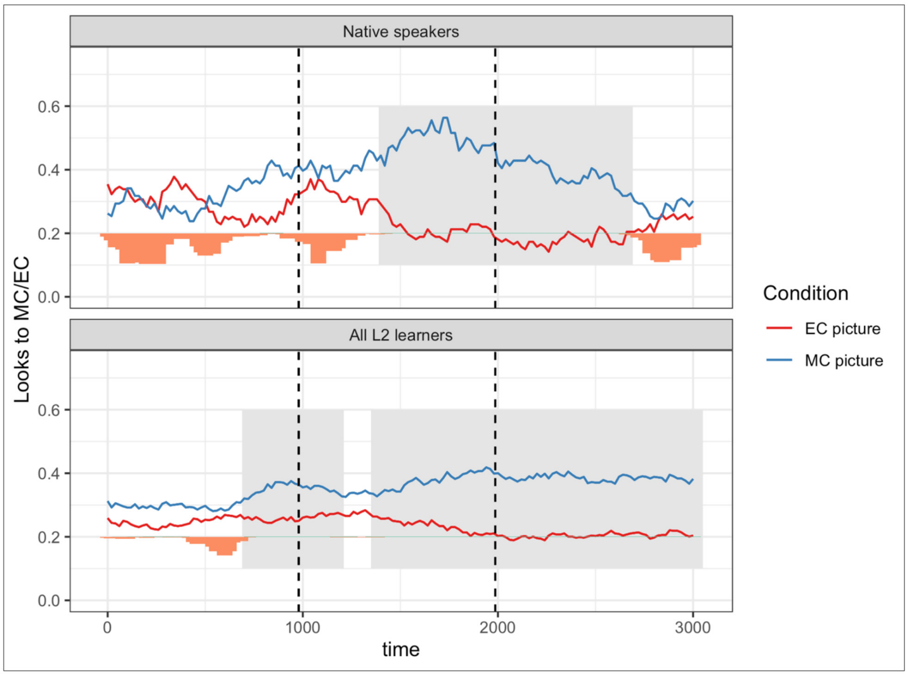


FIGURE 5
Proportion of looks to the MC and EC pictures for native speakers **(top)** and all L2 learners **(bottom)**. The first and second vertical lines indicate the earliest possible onset of the matrix verb (e.g., *tell*, at 979 ms) and the embedded clause verb (e.g., *catch*, at 1986), respectively. The Gray-shaded areas mark time windows in which the proportion of looks to the two pictures differed significantly, based on permutation analysis.

TABLE 1 Significant clusters per group identified by cluster-based permutation *tests*.

Group	Cluster number	Time window (ms)	Observed sum *z*	*p*-value
Native speakers	72	1,440–2,640	255	<0.001
All L2 leaners	126	740–1,160	80.6	<0.001
All L2 learners	138	1,400–3,000	512	<0.001

Time windows indicate the duration over which the significant cluster was observed.

MC picture exceeded those on the EC picture between 1,440 ms and 2,640 ms (total duration: 1,200 ms). L2 learners showed a similar effect but in two distinct time clusters: the first occurred between 740 ms and 1,160 ms (total duration: 420 ms), and the second between 1,400 ms and 3,000 ms (total duration: 1,600 ms). A summary of significant clusters for each group is provided in Table 1.

The early cluster observed in L2 learners reflects increased fixations on the MC location prior to the onset of the matrix verb (i.e., *tell*). Because no structurally informative cue is available at this point in the sentence to distinguish MC from EC interpretations, this effect cannot be attributed to structural parsing. Instead, it is best explained in terms of learners' sensitivity to distributional regularities in the input rather than to structural cues. The early shift toward the MC location appears to reflect reliance on lexical contingencies between the fronted *where*-phrase and the recurrent matrix verb *tell* in the martials. All experimental sentences used *tell* as the matrix verb, and eight filler items also included *tell* in the same position. Although the experimental and filler items were intermixed, *tell* appeared as the matrix verb in approximately half of the total trials (32 out of 60, see Appendix C). L2 learners therefore appear to have detected this distributional regularity and used it to anticipate the upcoming verb. By associating the fronted *where*-phrase with an expected occurrence of *tell*, learners showed an early bias toward the corresponding location (i.e., the MC picture) before the verb was acoustically realized, indicating anticipatory behavior driven by lexical expectations rather than structural constraints.

In contrast, native speakers did not exhibit anticipatory fixations toward the MC location prior to the onset of the matrix verb. Instead, their fixations shifted toward the MC picture only after hearing *tell*, indicating that they waited for a structurally informative cue before committing to a particular interpretation. Although this shift occurred later than the initial anticipatory effect observed in L2 learners, it nonetheless preceded the onset of the EC, suggesting that native speakers engaged in predictive processing once a grammatically licit attachment site became available, even though the sentence remained globally ambiguous.

This timing of native speakers' MC-directed fixations is consistent with accounts such as the Active Gap-Filling Hypothesis and the principle of Minimal Dependency Formation, which propose that comprehenders construct filler-gap dependencies incrementally at the earliest grammatically permitted position. In the present materials, the matrix verb *tell* provides the lexical and syntactic information required to license integration at the matrix clause position, even though a lower, grammatically possible gap site in the EC remains available later in the sentence.

Notably, the onset of native speakers' MC preference closely aligns with the second significant cluster observed in L2 learners. Native speakers' effect began at 1,440 ms, while L2 learners showed a comparable MC-directed shift at 1,400 ms. This suggests that, despite their earlier reliance on lexical or distributional contingencies, L2 learners also used the matrix verb as a structurally informative cue, committing to the MC at a point where a grammatically licit gap became available, as native speakers do. Consistent with this interpretation, L2 learners did not shift their fixations to the EC picture during the auditory presentation of the EC phrase (e.g., *catch butterflies*, beginning around 1,986 ms). Their sustained preference for the MC location during this interval indicates continued commitment to the MC interpretation rather than simple tracking of the most recently mentioned verb.

Taken together, these eye-tracking results suggest that L2 learners initially show anticipatory behavior driven by distributional regularities in the input, but subsequently converge on a structurally guided interpretation once the matrix verb provides a grammatically informative cue. The parallel timing of the later MC-directed shift in both groups suggests that structurally driven dependency formation mechanisms similar to those observed in native speakers can support L2 sentence processing when appropriate structural cues are accessible.

3.2.2 Linear mixed-effect analysis (robustness check)

To assess the robustness of the cluster-based permutation results, we conducted additional generalized linear mixed-effects analyses over pre-specified time windows defined relative to key linguistic events in the sentence. Three time windows were examined: (i) from the onset of the target question to the mean onset of the matrix verb *tell* (TW1), (ii) from the onset of *tell* to the mean onset of the embedded verb *catch* (TW2), and (iii) from the onset of *catch* to the mean sentence offset (TW3). For each time window, looks to the MC and EC pictures were analyzed using binomial generalized linear mixed-effects model with Picture (MC vs. EC) as a fixed effect and random intercepts for participants and items, as well as by-item random slopes for Picture.

For native speakers, no significant difference between looks to the MC and EC pictures was observed in TW1 ($\beta = 0.19$, $p = 0.65$), indicating no anticipatory bias prior to the matrix verb. In contrast, a strong preference for the MC picture emerged in TW2($\beta = 1.53$, $p < 0.001$) and was maintained in TW 3 ($\beta = 1.19$, $p = 0.001$), consistent with structurally guided prediction once a grammatically licit gap position became available.

L2 learners, by contrast, showed a significant preference for the MC picture in all three time windows. An MC preference was already present in TW 1 ($\beta = 0.50$, $p = 0.040$), prior to the onset of the matrix verb, and strengthened in TW2 ($\beta = 1.02$, $p = 0.009$) and TW3 ($\beta = 1.26$, $p < 0.001$). This pattern indicates that L2 learners exhibit anticipatory behavior earlier in the sentence, followed by sustained commitment to the MC interpretation.

Crucially, this pattern mirrors the results of the cluster-based permutation analysis. For native speakers, MC-directed fixations emerge only after the matrix verb provides a structurally informative cue, whereas for L2 learners, an MC preference was already evident prior to the matrix verb onset and persists throughout the sentence. Together, these findings confirm that the permutation results are robust. The linear mixed-effects analyses over pre-specified time windows replicate the key patterns identified by the cluster-based permutation analysis, while the permutation approach provides a more temporally fine-grained characterization of when predictive and integrative processes emerge. Taken together, the results support the conclusion that L2 learners initially rely on distributional regularities but subsequently converge on structurally guided dependency formation (Full model specifications and statistical outputs of the LME analyses are reported in Appendix B).

3.2.3 Permutation analysis: high-accuracy vs. low-accuracy L2 learners

While the aggregate L2 learner pattern revealed both an early distributional bias and a later structurally guided shift, this analysis does not establish whether these effects were uniformly present across individuals. It is possible that early distributional effect is driven primarily by learners who rely strongly on surface-level contingencies, whereas others show little or no such anticipation and instead exhibit clearer evidence of structure-based prediction. Such heterogeneity is expected if L2 learners differ in their ability to compute syntactic dependencies reliably in unambiguous structures, where only a single grammatical analysis is available. This possibility is motivated by prior work demonstrating individual differences in L2 syntactic processing (Hopp, 2013; Cunnings, 2017) and by the present finding that accuracy on unambiguous filler-gap items varies substantially within the L2 group, suggesting differences in structural sensitivity that are not reducible to general English proficiency.

Distributional contingencies, such as the frequent pairing of *where* and *tell* in the materials, are surface-based and do not require computing a structural dependency. If early distributional anticipation reflects weaker structural computation ability, that is, difficulty consistently deriving the correct syntactic analysis even when the structure is unambiguous, then this early effect should be most pronounced among L2 learners who show lower accuracy on unambiguous filler-gap items. In contrast, L2 learners with more reliable structural computation ability, as indexed by high accuracy on unambiguous items where the filler gap relation permits only one grammatical parse, are expected to rely less on surface-level associations and to show prediction patterns that more closely resemble those of native speakers. To evaluate this possibility, we conducted cluster-based permutation analyses separately for the high-accuracy and low-accuracy L2 subgroups, using the same procedure as in the preceding analyses.

Figure 6 shows the proportion of looks to MC and EC pictures for the high-accuracy group (top panel) and the low-accuracy group

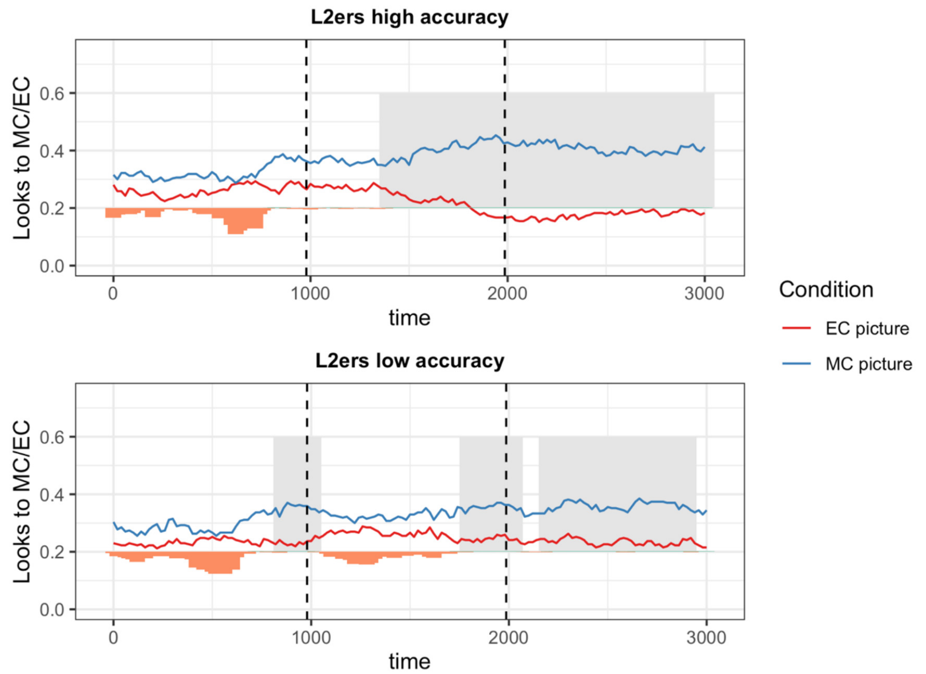


FIGURE 6
Proportion of looks to the MC and EC pictures High-accuracy group **(top)** and Low-accuracy group **(bottom)**. The first and second vertical lines indicate the earliest possible onset of the matrix verb (e.g., *tell*, at 979 ms) and the embedded clause verb (e.g., *catch*, at 1986), respectively. The gray-shaded areas mark time windows in which the proportion of looks to the two pictures differed significantly, based on permutation analysis. Orange downward bars indicate individual 20 ms time bins that did not reach significance ($p < 0.05$) prior to cluster formation.

TABLE 2 Significant clusters per group identified by cluster-based permutation tests.

Group	Cluster number	Time window (ms)	Observed sum *t*	*p*-value
High-accuracy	65	1,400–3,000	475.51	<0.001
Low-accuracy	43	860–1,000	26.51	<0.001
Low-accuracy	82	1,800–2,020	34.54	<0.001
Low-accuracy	91	2,200–2,560	59.31	<0.001
Low-accuracy	96	2,940–3,000	12.20	0.008

Time windows indicate the duration over which the significant cluster was observed.

(bottom panel), time-locked to the onset of the target ambiguous questions. As in Figure 5, the blue line represents looks to the MC picture and the red line represents looks to the EC picture, with gray shading indicating time windows identified as significant by the permutation analysis.

The permutation analysis performed separately for the two L2 subgroups revealed clear differences in the timing patterns of their fixations in processing ambiguous wh questions. The high accuracy group showed a single significant cluster in which fixations on the MC picture exceeded those on the EC picture, closely mirroring the native speaker pattern (1,400 ms to 3,000 ms; total duration, 1,600 ms).

In contrast, the low-accuracy group showed a markedly different pattern, with four distinct significant clusters distributed across the time course. The earliest cluster emerged between 860 ms and 1,000 ms (total duration: 140 ms), reflecting an anticipatory bias toward the MC location prior to the onset of the matrix verb. As in the comparison between native speakers and the full L2 group, this early effect cannot be attributed to syntactic prediction because no structural information was available at this point to license a gap. Instead, it is best explained by sensitivity to lexical co-occurrence patterns, such as the repeated pairing of the fronted *where* with the matrix verb *tell* across trials, suggesting that the low-accuracy learners' earliest prediction was guided by surface-level contingencies rather than by abstract syntactic structure.

The remaining clusters in the low-accuracy group occurred later in the sentence, from 1,800 ms to 2,020 ms (total duration: 220 ms), from 2,200 ms to 2,560 ms (total duration: 360 ms), and from 2,940 ms to 3,000 ms (total duration: 60 ms). Although these clusters broadly overlap with the time window in which native speakers and high-accuracy learners showed a preference for the MC interpretation, the effect in the low-accuracy group was fragmented across short intervals rather than expressed as a single sustained shift. A summary of significant clusters for each subgroup is provided in Table 2.

4 Discussion

The present study examined whether L2 learners engage in predictive parsing when processing structurally ambiguous filler-gap dependencies, and the extent to which their behavior resembles that of native speakers. When L2 learners were analyzed as a single group (hereafter, *the combined L2 group*), both key-press and eye-tracking measures for the ambiguous items showed that they, like native speakers, exhibited a robust preference for the MC interpretations. Crucially, this MC preference cannot be attributed to the selection of the "correct" interpretation, since both MC and EC interpretations were always grammatically available in our ambiguous materials. Instead, the consistent preference for the MC interpretation suggests that L2 learners, like native speakers, attempted to posit a gap at the earliest grammatically licit position, a hall mark of active gap filling in incremental sentence comprehension.

At the same time, the combined L2 analysis also revealed a clear difference from native speakers in the timing and cues that guided prediction. Whereas native speakers showed anticipatory looks toward the MC location only after hearing the matrix verb, the combined L2 group shifted their gaze toward the MC location even before any structural cue licensed an MC gap. Because no grammatical information supported an MC analysis at this point in the sentence, this early anticipatory shift cannot be attributed to a structure-based prediction. A plausible interpretation is that the aggregated L2 pattern reflects sensitivity to distributional regularities in the materials, particularly the frequent pairing of the fronted *where* with the matrix verb *tell*. Viewed at the group level, this early bias suggests that L2 learners initially relied on surface-level lexical contingencies rather than structural constraints. It is important to note that this interpretation reflects the averaged pattern obtained from the combined analysis and does not imply that all L2 learners contributed to this early bias in the same way; as the subgroup analyses indicate, this early effect appears to have been primarily driven by the low-accuracy learners, a point we that we elaborate on below.

Crucially, the combined L2 analysis also showed that the later portion of the fixation time course ultimately converged with the native speaker pattern. Both native speakers and the combined L2 group shifted their gaze toward the MC location once the matrix verb was encountered and maintained this preference throughout the EC region. Because this sustained preference cannot be explained by lexical associations alone, it suggests that once a structurally informative cue became available, L2 learners were able to use it to construct a grammatically licensed dependency. When viewed at the level of the combine L2 group, therefore, the pattern appears hybrid: early distributional, lexical cue-based anticipation followed by structurally guided active gap filling once the matrix verb became available.

However, the subgroup analyses revealed that the early and late effects observed in the combined L2 group were not expressed uniformly across all learners. Instead, they reflected two distinct processing profiles that differed in structural computation ability. High-accuracy L2 learners, whose performance on the unambiguous filler-gap items indicates more reliable structural computation ability, showed a fixation time course that closely resembled that of native speakers: they showed no early distributional effect and exhibited a single sustained MC preference once the matrix verb provided a structurally informative cue. Low-accuracy L2 learners, in contrast, showed an early distributional bias toward the MC location before any structural cue was available, and their later MC preference was delayed relative to native speakers and high-accuracy learners, and was fragmented across shorter clusters. These differences indicate that weaker structural computation ability is associated with greater reliance on surface-level lexical contingencies and delayed use of structural cues. Importantly, this suggests why the combined L2 analysis showed a hybrid pattern: the early lexical effect reflected the behavior of the low accuracy subgroup, while the later structural effect reflected the behavior of both groups, although more robustly for high accuracy learners.

Additional support for the interpretation that low-accuracy group rely more heavily on surface-level cues comes from the key press responses. While the high-accuracy group, like native speakers, showed a clear preference for the MC interpretation at the final decision stage, the low-accuracy group did not. Although eye-tracking data revealed MC commitment during online processing in the low-accuracy group, this commitment was not reflected in their key press choices. Instead, the low-accuracy group showed no reliable preference for the MC interpretation over the EC alternative. One possible explanation is that their structural representations are less robust, which makes their final decision highly sensitive to recency based lexical cues, that is, cues stemming from the most recently encountered portion of the sentence. Because the EC relevant material appears at the end of the question, this late occurring lexical information may have exerted a strong influence on the final choice. For example, after hearing "catch butterflies" in (4a), some low-accuracy learners selected the EC answer option, suggesting that the lexical recency cue may have overridden their earlier MC expectation. This seems to be the most plausible explanation of the observed pattern and is compatible with the idea that when structural representations are not firmly established, surface-level cues can pull comprehenders toward a competing interpretation, leading to a less stable overall commitment that differs between the real time fixation pattern and the final explicit choice.

To account for this L2 processing profile, we propose that L2 learners rely on a cue weighting mechanism in which the use of predictive cues is modulated by the learner's structural computation ability. Learners draw on a broad set of cues in real time, including cues that are not grammatically informative but non-etheless carry predictive value, such as distributional regularities and lexical co-occurrence patterns. L2 learners with weak or unstable structural representations place greater weight on these readily accessible surface-level cues, resulting in earlier non-structural predictions and a less robust structural commitment later in the sentence. In contrast, L2 learners with stronger structural computation ability weight structurally informative cues more heavily, and their resulting prediction pattern resembles that of native speakers. The idea that L2 learners with weaker structural computation ability would rely on information that native speakers do not use for prediction, and would in fact commit to a predictive interpretation even earlier than native speakers, might appear counterintuitive, particularly given proposals that L2 comprehension is constrained by working memory limitations (Cunnings and Felser, 2013; Cunnings, 2017; Kim and Christianson, 2017). However, a cue weighting perspective provides a rational explanation, since the early prediction observed in the low accuracy group is not structurally guided but instead arises from surface-level lexical associations. When structural computation is insufficiently stable, drawing on surface-level cues that are immediately available in the input allows learners to generate predictive commitments under real-time processing constraints (for frequency-based accounts of cue reliance, see Ellis, 2002).

This interpretation shares certain assumptions with the SSH (Clahsen and Felser, 2006), which holds that L2 learners rely less on abstract syntactic representations and more on lexical-semantic information during processing. However, our findings extend this hypothesis by showing that these differences emerge specifically in predictive processing. More importantly, our results reveal a pattern not explicitly predicted by SSH: the nature of predictive parsing in L2 learners depends on the stability or reliability of their structural representations. High-accuracy learners, whose computation of filler-gap dependencies was more reliable, displayed native-like structural prediction. Low-accuracy learners, in contrast, relied more heavily on surface-level contingencies such as lexical co-occurrence patterns, resulting in predictive processing that was less guided by structural information and more susceptible to moment-to-moment lexical cues.

This graded pattern suggests a developmental view in which L2 processing shifts from frequency-based prediction toward structure-based prediction as structural representations become more stable. This interpretation aligns with convergence-based accounts (Hopp, 2013; Cunnings, 2017), which hold that L2 learners can approximate native-like parsing mechanisms when their underlying grammatical representations are sufficiently robust. Rather than a categorical deficit in syntactic prediction, the present findings point to a continuum in which cue use changes as L2 grammatical knowledge becomes more established.

Importantly, our results also indicate that the L2 pattern observed here cannot be explained as a simple transfer of L1 processing strategies. In Japanese, the native language of our L2 participants, wh-phrases remain *in situ* and the interpretation of question depends on clause-final cues such as the question particle (Kuroda, 1965; Saito, 2017). Processing studies show that native Japanese speakers often postpone interpretation until they encounter this clause-final marker (Nakamura et al., 2025b; though see Aoshima et al., 2004; for evidence that Japanese speakers actively posit gap sites). In contrast, Japanese L2 learners of English in our study did not delay interpretation. Instead, they showed early commitment to the MC interpretation. This indicates that they adopted a more predictive parsing strategy characteristic of English wh-dependency resolution, rather than relying on a processing strategy typical of Japanese.

For future research, a central methodological challenge concerns how best to characterize individual differences in structural computation ability. While our high–low grouping used in the present study, based on accuracy for unambiguous filler–gap sentences, provides a theoretically motivated proxy for structural computation ability, future progress will require more fine-grained and independent measures of structural sensitivity that can better isolate the specific structural computational factors underlying individual variation in L2 predictive processing. Future research should also examine the generality of these cue-weighting patterns across syntactic structures and processing contexts. In particular, experimentally manipulating distributional environments will allow direct tests of whether early surface-driven predictions reflect the reliability of L2 learners' syntactic computation during line sentence processing. Finally, longitudinal or training-based studies may help illuminate how individual differences in structural representations develop over time and how these changes contribute to the transition from frequency-based to structure-based prediction in L2 comprehension.

Data availability statement

The raw data supporting the conclusions of this article is publicly available in the OSF repository at: https://osf.io/xywvs/.

Ethics statement

The studies involving humans were approved by the Ethics Review Committee on Research with Human Subjects of Waseda University, under protocol number 20244-130. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

CN: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. SF: Conceptualization, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing. YM: Conceptualization, Funding acquisition, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing. NY: Conceptualization, Funding acquisition, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This research was funded by JSPS KAKENHI Grant Number 21KK0006 (Fostering Joint International Research) and by the JSPS Core-to-Core Program, A. Advanced Research Networks "International Research

Network for the Human Language Faculty" (Grant Number JPJSCCA20210001).

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/flang.2026.1756463/full#supplementary-material

References

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

Aoshima, S., Phillips, C., and Weinberg, A. (2004). Processing filler-gap dependencies in a head-final language. *J. Mem. Lang.* 51, 23–54. doi: 10.1016/j.jml.2004.03.001

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005

Chan, A., Yang, W. C., Chang, F., and Kidd, E. (2018). Four year old Cantonese speaking children's online processing of relative clauses: a permutation analysis. *J. Child Lang.* 45, 174–203. doi: 10.1017/S0305000917000198

Clahsen, H., and Felser, C. (2006). Grammatical processing in language learners. *Appl. Psycholinguist.* 27, 3–42. doi: 10.1017/S0142716406060024

Clahsen, H., and Felser, C. (2018). Some notes on the shallow structure hypothesis. *Stud. Second Lang. Acquis.* 40, 693–706. doi: 10.1017/S0272263117000250

Cunnings, I. (2017). Parsing and working memory in bilingual sentence processing. *Biling. Lang. Cogn.* 20, 659–678. doi: 10.1017/S1366728916000675

Cunnings, I., and Felser, C. (2013). The role of working memory in the processing of reflexives. *Lang. Cogn. Process.* 28, 188–219. doi: 10.1080/01690965.2010.548391

Ellis, N. C. (2002). Frequency effects in language processing: a review with implications for theories of implicit and explicit language acquisition. *Stud. Second Lang. Acquis.* 24, 143–188. doi: 10.1017/S0272263102002024

Felser, C. (2019). Structure-sensitive constraints in non-native sentence processing. *J. Eur. Second Lang. Assoc.* 3, 12–22. doi: 10.22599/jesla.52

Felser, C., and Roberts, L. (2007). Processing wh-dependencies in a second language: a cross-modal priming study. *Second Lang. Res.* 23, 9–36. doi: 10.1177/0267658307071600

Frazier, L. (1987). "Sentence processing: a tutorial review," in *Attention and Performance 12: The Psychology of Reading*, ed. M. Coltheart (Lawrence Erlbaum Associates, Inc.), 559–586.

Frazier, L., and Clifton, C., Jr. (1989). Successive cyclicity in the grammar and the parser. *Lang. Cogn. Process.* 4, 93–126. doi: 10.1080/01690968908406359

Grüter, T., Lau, E., and Ling, W. (2020). How classifiers facilitate predictive processing in L1 and L2 Chinese: the role of semantic and grammatical cues. *Lang. Cogn. Neurosci.* 35, 221–234. doi: 10.1080/23273798.2019.1648840

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

Grüter, T., and Rohde, H. (2021). Limits on expectation-based processing: Use of grammatical aspect for co-reference in L2. *Appl. Psycholinguist.* 42, 51–75. doi: 10.1017/S0142716420000582

Grüter, T., Rohde, H., and Schafer, A. J. (2017). Coreference and discourse coherence in L2: the roles of grammatical aspect and referential form. *Linguist. Approach. Bilingual.* 7, 199–229. doi: 10.1075/lab.15011.gru

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: relations between lexical and syntactic variability. *Second Lang. Res.* 29, 33–56. doi: 10.1177/0267658312461803

Kaan, E., and Grüter, T. (eds.). (2021). "Prediction in second language processing and learning: advances and directions," in *Prediction in Second Language Processing and Learning* (Amsterdam: John Benjamins), 1–18.

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Lang. Linguist. Compass* 2, 647–670. doi: 10.1111/j.1749-818X.2008.00072.x

Kim, J. H., and Christianson, K. (2017). Working memory effects on L1 and L2 processing of ambiguous relative clauses by Korean L2 learners of English. *Second Lang. Res.* 33, 365–388. doi: 10.1177/0267658315623322

Kuroda, S. Y. (1965). *Generative grammatical studies in the Japanese language (Doctoral dissertation)*. Massachusetts Institute of Technology.

Lew-Williams, C., and Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *J. Mem. Lang.* 63, 447–464. doi: 10.1016/j.jml.2010.07.003

Marinis, T., Roberts, L., Felser, C., and Clahsen, H. (2005). Gaps in second language sentence processing. *Stud. Second Lang. Acquis.* 27, 53–78. doi: 10.1017/S0272263105050035

Nakamura, C., Flynn, S., Miyamoto, Y., and Yusa, N. (2025a). "Comparing the processing of wh-question in Japanese-English bilinguals: a cross linguistic eye-tracking study. [Paper presentation]," in *Proceeding of the 15th International Symposium on Bilingualism* (Donostia–San Sebastián).

Nakamura, C., Flynn, S., Miyamoto, Y., and Yusa, N. (2025b). Proximity loses: Real-time resolution of ambiguous wh-questions in Japanese. *Languages* 10:288. doi: 10.3390/languages10120288

Nakamura, C., Harris, J., and Jun, S. A. (2021). "Adaptation to atypical contrastive accent: the L2 advantage," in *Proceedings of the 45th Annual Boston University Conference on Language Development (BUCLD 45)*, 568–583.

Omaki, A., Davidson White, I., Goro, T., Lidz, J., and Phillips, C. (2014). No fear of commitment: children's incremental interpretation in English and Japanese wh-questions. *Lang. Learn. Dev.* 10, 206–233. doi: 10.1080/15475441.2013.844048

Omaki, A., and Schulz, B. (2011). Filler-gap dependencies and island constraints in second-language sentence processing. *Stud. Second Lang. Acquis.* 33, 563–588. doi: 10.1017/S0272263111000313

Phillips, C. (2006). The real-time status of island phenomena. *Language*, 82, 795–823. doi: 10.1353/lan.2006.0217

Saito, M. (2017). Japanese wh-phrases as operators with unspecified quantificational force. *Lang. Linguist.* 18, 1–25. doi: 10.1075/lali.18.1.01sai

Tanner, D., Mclaughlin, J., Herschensohn, J., and Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingual. Lang. Cogn.* 16, 367–382. doi: 10.1017/S1366728912000302