Applied Psycholinguistics (2026), **47**, e8, 1–26
doi:10.1017/S0142716426100526

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL ARTICLE

First and second language speaker understanding of implausible English passives in discourse contexts

Stephen Skalicky and Victoria Chen

Victoria University of Wellington, New Zealand
Corresponding author: Stephen Skalicky; Email: stephen.skalicky@vuw.ac.nz

(Received 2 February 2025; revised 21 January 2026; accepted 28 January 2026)

Abstract

We examined how first (L1) and second language (L2) English speakers interpret thematic roles when semantic cues conflict with syntactic cues in discourse contexts. We tested three groups: monolingual English speakers, Taiwanese Mandarin-speaking and Japanese-speaking L2 learners of English. Participants read written vignettes containing target sentences with *by*-phrases (e.g., Jerry was robbed *by* the thief). In these sentences, the embedded *by*-phrase noun was either semantically congruent or incongruent with the previous and subsequent discourse context. An offline comprehension task measured interpretations of the agent. Following predictions from the Competition Model and the Shallow Structure Hypothesis, we tested whether L2 speakers would rely more heavily on semantic cues as a general rule, or whether differences in how L1–L2 structures differ within and between participants would influence comprehension of the agents. Results from Bayesian multilevel regression models indicated that L2 speakers were more likely to choose agents based on semantic cues when compared to L1 speakers. No strong effects of L1–L2 cue similarity were obtained. Our results lend new support to the view that L2 speakers rely more heavily on discourse-level semantic cues than L1 speakers, highlighting the need for further research on cue competition in discourse contexts.

Keywords: competition model; discourse processing; L2 sentence processing; shallow structure hypothesis

Introduction

Human sentence processing involves the weighted integration of different linguistic cues (e.g., word order, noun phrase animacy, and semantic plausibility) to arrive at a coherent interpretation of meaning. The Competition Model (Bates & MacWhinney, 1987; Li & MacWhinney, 2012; MacWhinney, 1987) provides a framework for understanding and predicting how cues will influence language

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (https://creativecommons.org/licenses/by-nc-sa/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is used to distribute the re-used or adapted article and the original article is properly cited. The written permission of Cambridge University Press or the rights holder(s) must be obtained prior to any commercial use.

https://doi.org/10.1017/S0142716426100526 Published online by Cambridge University Press

processing and comprehension. Crucially, languages rely on different cues to mark sentence information, such as semantic roles. Moreover, cues differ in their strength and reliability, in that some cues are more reliable than others for determining the meaning of sentences. Competition between cues occurs when cues point to diverging interpretations, such as when English passive sentences subvert canonical SVO word order.

According to the Competition Model, learning an additional language is partially a process of learning how to attend to different processing cues in the target language. Learners will initially map their first language (L1) cues to second language (L2) input, which can result in processing errors when two languages use different cues to mark the same information (e.g., determining the agent in a sentence). Gaining proficiency in an L2 is thus a function of learning the different cues important to the L2, especially if they differ from L1 cues (Bates & MacWhinney, 1987; McManus, 2021).

Evidence from both L1 and L2 processing research has provided empirical evidence attesting to the predictions of the Competition Model across a variety of languages (Bates & MacWhinney, 1987; Harrington, 1987; Kilborn, 1989; Liu et al., 1992; Mcdonald, 1987; Su, 2001; Walter et al., 2024; Zhao et al., 2025). Following this extensive research on cue competition in single-sentence processing, there has been some interest in how these mechanisms operate within broader discourse contexts (Pan & Felser, 2011; Su, 2004). Such research is important because psycholinguistic models of discourse processing suggest that readers and listeners construct mental situation models that are continuously updated with new information (Just & Carpenter, 1980; O'Brien & Cook, 2016; van den Broek & Helder, 2017), and meaning that prior information is considered while reading new sentences.

However, little research to date has investigated the potential for discourse information to act as a cue and/or mitigate potential syntactic cue competition (c.f., Su, 2004). In the current study, we consider both questions, asking whether semantic information in discourse can override a syntactic cue that disagrees with the semantic information. We further examine whether L1–L2 cue similarity interacts with competition between discourse and syntactic cues during L2 sentence processing. To do so, we investigate how English monolinguals and two groups of English L2 speakers (Taiwanese Mandarin L1 and Japanese L1) attend to conflicts between a syntactic cue (agentive *by*-phrases) and semantic plausibility cues when processing English passive sentences. We address the following two key questions: (1) how L1 and L2 English speakers resolve competition between discourse-level semantic cues and sentence-internal syntactic cues, and (2) whether cross-linguistic differences in cue reliability lead to different sentence representations in L2 speakers from different L1 backgrounds.

L2 Sentence processing and cue competition

According to the Competition Model, language users attend to different language-specific cues to extract semantic information, such as thematic roles, from sentences (Bates & MacWhinney, 1987; Li & MacWhinney, 2012; MacWhinney, 1987; Zhao

https://doi.org/10.1017/S0142716426100526 Published online by Cambridge University Press

et al., 2025). These cues include syntactic features (word order, structural position) and morphological markers (case marking, agreement), which vary systematically across languages in their strength and reliability. For example, in English, word order is a strong cue for determining agents in canonical SVO sentences (e.g., *the woman wrote the book*), but is less reliable in noncanonical sentences, such as passives. In these cases, the *by*-phrase in passive sentences (e.g., *The book was written by the woman*) serves as a strong cue for agent identification.

There are many theoretical models and explanations of L1–L2 influence, such as the feature reassembly model, the Bilingual Interactive Activation Plus (BIA+), the Revised Hierarchical Model, and so on (Lardiere, 2009; Tokowicz, 2014). Here, we choose to focus on the Competition Model, because a core prediction of the Competition Model is the influence of cue strength and L1 transfer effects on L2 sentence processing strategies. In this model, L2 learners initially map their L1 cue strategies onto L2 input, which can result in processing difficulties when languages use different cues to mark semantic information (Li & MacWhinney, 2012; Liu et al., 1992; Tokowicz et al., 2024). A large number of empirical studies have applied the Competition Model to the study of L2 sentence processing, exploring how L2 learners adjust their use of linguistic cues based on exposure and L1–L2 similarities. In this research, L2 learners initially apply L1 cues to L2 sentence processing, but this effect attenuates as L2 proficiency increases (Zhao et al., 2025). These effects have been compared across many different language pairs, such as Dutch and English (Mcdonald, 1987), and English and German (Kilborn, 1989). These studies showed that when processing English sentences, English speakers relied more heavily on word order, whereas Dutch and German speakers relied more strongly on morphology, suggesting the influence of L1 cues on L2 (English) sentence processing. Similar results have been attained when testing Japanese learners of English (Harrington, 1987), Mandarin Chinese speakers of English (Liu et al., 1992; Su, 2001; Zhao & Fan, 2021), English, Dutch, Italian, Polish, and Swedish learners of Chinese (Liang et al., 2022), and learners of artificial languages (Walter et al., 2024).

Discourse-level cue competition and L2 sentence processing

While previous research on this theme has focused primarily on sentence-level cue competition, real-world language processing occurs within rich discourse contexts that provide additional pragmatic cues to sentence meaning and inferences (O'Brien et al., 2015). Discourse context can either support or conflict with sentence-level cues, creating different types of competition scenarios (van den Broek & Helder, 2017). When discourse context supports syntactic analysis (plausible scenarios), cue competition may be minimal. However, when discourse context conflicts with syntactic structure (implausible scenarios), speakers must resolve competition between syntactic and semantic cues. Following predictions of the Competition Model, speakers will weigh these cues based on their learned reliability, with L1 and L2 speakers showing different weighting patterns based on their linguistic background and level of target language proficiency.

However, few studies have tested cue competition in discourse settings (cf. Su, 2004). In their study, Su (2004) tested competition among word order, animacy, and discourse context among L1 Chinese and L1 English speakers (i.e., no L2 learners

https://doi.org/10.1017/S0142716426100526 Published online by Cambridge University Press

were involved). Short contexts were provided before target sentences, which varied in their word order (e.g., noun verb noun; verb noun noun) and animacy of noun phrases (e.g., animate-animate; animate-inanimate). The contexts biased either the first or second noun to be the *doer* of the sentence, and participants were asked to listen to the sentence and then indicate the agent. The results showed that discourse information decreased reliance on word order cues, but more so for the Chinese speakers when compared to the English speakers. In other words, discourse-level information mitigated cue competition for L1 speakers, but this effect was stronger in one language (Chinese).

An open question is whether L1–L2 cue transfer effects will emerge when studying L2 processing and comprehension of larger discourse contexts, when compared to the traditional single-sentence methods many Competition Model studies use. The rationale behind this question is that L2 speakers are thought to rely more strongly on discourse-level information when compared to L1 speakers (Cunnings, 2017; Hopp, 2022b). Models such as the Shallow Structure Hypothesis (SSH) posit that L2 speakers must rely more strongly on semantic and pragmatic information because their L2 syntactic parses are more shallow when compared to L1 speakers (Clahsen & Felser, 2006, 2018). Alternative accounts argue that L2 speakers differ not in the quality of the syntactic representation, but instead, in the memory-based retrieval strategies, where L2 speakers give more weight to discourse-level cues when considering the meaning of sentences after initial parsing (Cunnings, 2017).

Regardless of these different explanations, empirical studies attest to a stronger influence of discourse-level cues among L2 speakers when compared to L1 speakers (Crible et al., 2021; Felser & Drummer, 2022; Hopp, 2022a; Pan & Felser, 2011). For example, when processing prepositional phrases that could modify either a VP or an NP positioned within the same sentence, Chinese learners of English were better able to determine the modifier when the discourse context supported the VP or NP referent. In contrast, native English speakers showed no such sensitivity to the discourse context (Pan & Felser, 2011). In a different study, French and Dutch learners of English showed comparability to a native English speaker baseline when processing correct and incorrect uses of English discourse connectives *if* and *when* (Zufferey et al., 2015). However, in the same study, negative transfer effects emerged when learners judged a subset of the same target sentences in isolation. Both groups of learners had lower accuracy when judging English connectives involving cue competition with their L1 equivalent structures (*if* for French and *when* for Dutch). The implication was that supporting information from prior discourse was able to mitigate L1–L2 cue competition for these target structures.

As such, the scope of cue competition shifts when considering whether sentences are processed in isolation or within larger discourse contexts. Following the logic from models such as the SSH, there is a general weighting of semantic over syntactic information for L2 speakers when compared to L1. It follows then that for L2 speakers, discourse information (or, semantic cues) may also partially override localized sentence L1–L2 cue competition in L2 sentence processing. This possibility, however, also recognizes the influence of L2 proficiency. As proficiency in the target language increases, learners begin to rely less on mapping of L1 cues and instead learn how L2 cues operate (Bates & MacWhinney, 1987). Yet, L2 processing

https://doi.org/10.1017/S0142716426100526 Published online by Cambridge University Press

studies show that even highly proficient L2 learners will rely more heavily on semantic versus syntactic information during L2 sentence processing (e.g., Deniz, 2022). In other words, discourse information, such as semantic plausibility, may serve as a stronger cue for L2 speakers when compared to L1 speakers, even at relatively high levels of L2 proficiency.

Passive formation in English, Japanese, and Taiwanese Mandarin

The current study examines how discourse information interacts with L1–L2 cue similarity. To do so, we compare responses to agentive *wh*-questions for English passive *by*-phrases embedded in short story contexts. We collected data from English monolinguals, L1 Mandarin learners of English in Taiwan, and L1 Japanese learners of English. Our study has two primary goals: (1) assess and compare cue weighting strategies among the monolingual and L2 English groups when syntactic and semantic cues conflict in our implausible vignettes, and (2) examine these strategies in light of cross-linguistic differences in cue reliability and L2 proficiency between the Japanese and Mandarin Chinese speakers.

The English passive and cue competition

English passives are considered marked constructions for several reasons. First, unlike their active counterparts, passives display an unusual mapping of grammatical relation and thematic role against the Thematic Hierarchy (Fillmore et al., 1968; Larson, 1988), allowing the theme to override the agent and constitute the subject. Second, through the demotion of the agent and the promotion of the theme to subject status, passive constructions usually display a different word order than canonical active transitives. Third, the passive construction in English and many other Indo-European languages has been shown to have low frequency in discourse and is acquired later than the active. All these factors make the passive an ideal testing ground for cue competition during English sentence processing.

In English, passivization (1) features a fronted theme that occupies the subject position and a demoted agent realized as an optional oblique introduced by the preposition *by*. This preposition constitutes a strong syntactic cue signaling agentive interpretation in passive constructions (Wanner, 2009; Xiao et al., 2006).

(1) Amos was$_{\text{PASS}}$ slapped (*by* Naomi.)$_{\text{agentive}}$ (*by as an agent-introducing preposition*)

However, this preposition is not linked uniquely to the agent semantic role, and can also introduce locative phrases, as in (2):

(2) Roberta stood (*by* the sea.)$_{\text{locative}}$ (*by as a locative-introducing preposition*)

This bifunctionality creates syntactic ambiguity in English passives, because a *by*-phrase may introduce either an agent or a locative/bystander complement. Although the ambiguity may be reduced by animacy, where both phrases co-occur in the same sentence as two adjuncts, in which case the agent *by*-phrase must

https://doi.org/10.1017/S0142716426100526 Published online by Cambridge University Press

precede the adjunct *by*-phrase, as in (3), the ambiguity is salient when the locative phrase is an animate NP with no occurring agent *by*-phrase, as in (4).

(3) Amos was slapped (*by* Naomi)$_{\text{agentive}}$ (*by* the sea.)$_{\text{locative}}$
(4) Camina was attacked (*by* a police officer)$_{\text{locative?/agentive,}}$ who then made the arrest.

Cross-linguistic cue reliability: Japanese

Like English, Japanese employs several valency-decreasing operations, including the *ase* and *niyotte* passives (Iwasaki, 2018; Washio, 1987). As with English passives, these constructions feature a fronted theme in the subject/nominative position and an optional agent-denoting postposition phrase (PP). In Japanese, this PP is introduced by the marker *ni*, as in (5a). Similar to the English preposition *by*, the marker *ni* may also introduce locative PPs, as in (5b). It is also possible for a passive sentence to contain two *ni*-marked adjuncts, one expressing the agent and the other the locative, as in (5c).

(5) Japanese
a. Taro-ga [Nanako-**ni**]$_{\text{agent}}$ ik-ase-rare-ru.
Taro-NOM [Nanako-**DAT**] go-CAU-PASS-IN
"Taro was made by Nanako to go."
b. Taro-ga [Osaka-**ni**]$_{\text{locative}}$ ik-ase-rare-ru.
Taro-NOM [Osaka-**DAT**] go-CAU-PASS-IN
"Taro was made to go to Osaka."
c. Taro-ga (Nanako-**ni**)$_{\text{agent}}$ (Osaka-**ni**)$_{\text{locative}}$ ik-ase-rare-ru.
Taro-NOM (Nanako-**DAT**) (Osaka-**DAT**) go-CAU-PASS-IN
"Taro was made by Nanako to go to Osaka."

The parallel between English *by* and Japanese *ni* means that both languages present speakers with the same cue reliability challenge: the agent-marking morpheme can also signal locative information in certain conditions. If Japanese learners of English transfer their L1 cue weighting strategies to English, the resulting English *by* should be a similarly less reliable marker of agentive semantic roles, potentially leading to similar competition resolution patterns as native English speakers when syntactic and semantic cues conflict.[1]

Cross-linguistic cue reliability: Mandarin Chinese

Mandarin Chinese presents a different cue reliability pattern that should lead to direct transfer effects in L2 English processing. Like Japanese and English, Mandarin possesses a detransitivizing construction commonly analyzed as a passive (Chu, 1973; Hashimoto, 1988; Pan & Hu, 2021). Similar to the English passive, this construction is characterized by fronting of the theme and demotion of the agent to an optional preposition phrase introduced by the marker *bei* (6).

https://doi.org/10.1017/S0142716426100526 Published online by Cambridge University Press

(6) Mandarin
Zhangsan$_{\text{theme}}$ **bei** **(Lisi)$_{\text{agent}}$** **da-le.**
Zhangsan BEI LISI **hit-PRF**
"Zhangsan was hit (by Lisi)."

Similar to the English preposition *by*, the Mandarin passive marker *bei* serves as a reliable cue for indicating the agent role (Li et al., 1993). However, this marker differs crucially from English *by* in its cue reliability: *bei* is not compatible with locative NPs. This is because, as in (7a), locative obliques in Mandarin are introduced by a distinct preposition: *zai,* as seen in (7b):

(7) Mandarin
a. * Zhangsan$_{\text{theme}}$ **bei** hebian$_{\text{locative}}$ **bei** Lisi$_{\text{agent}}$ da-le.
Zhangsan **bei** riverside **bei** Lisi hit-PRF
(Intended: "Zhangsan was hit by Lisi by the river.")
b. Zhangsan$_{\text{theme}}$ **zai** hebian$_{\text{locative}}$ **bei** Lisi$_{\text{agent}}$ da-le.
Zhangsan LOC hebian **bei** Lisi hit-PRF
"Zhangsan was hit by Lisi by the river."

This cross-linguistic difference in cue reliability has important implications for our current study. Mandarin speakers know the passive marker *bei* is a highly reliable cue for agentive interpretation with no locative ambiguity. When processing English passives, this L1 experience should lead Mandarin speakers to weigh the syntactic cue (*by*-phrase = agent) more heavily than Japanese speakers, who have learned that the equivalent marker (*ni*) is less reliable due to its locative ambiguity. The result should be a stronger agentive bias for English *by* for Mandarin Chinese when compared to Japanese speakers.

Hypotheses

Based on our review of the theoretical predictions made by the SSH and Competition Model, as well as the typological differences among English, Japanese, and Mandarin Chinese, at least two hypotheses are possible. We formalize these as Hypotheses 1 and 2.

Hypothesis 1. *Semantic cues will be stronger for L2 speakers, and syntactic cues will be stronger for L1 speakers.*

Hypothesis 2. *English L1 speakers and Japanese L2 speakers of English have a weaker agentive bias for* by *when compared to Mandarin L2 speakers of English.*

In addition to investigating these two competing hypotheses, we consider whether reliance on semantic versus syntactic cues becomes more target-like with higher L2 proficiency.

https://doi.org/10.1017/S0142716426100526 Published online by Cambridge University Press

Materials

Our study design used short text vignettes containing three sentences each. Using vignettes allowed us to embed different target sentences within the same discourse contexts. In one condition, semantic information in the discourse and syntactic cues in the target sentence both implicated the same event participant as the agent of an action. In another condition, the semantic and syntactic cues were in contrast, pointing towards two different NPs.

Vignette design

In order to create competition between discourse-level information and the agentive *by*-phrase, we created 20 three-sentence stories using transitive verbs biased towards a completed action. The stories were divided so that 12 included animate human objects and eight included inanimate objects as themes (this was done primarily to introduce variation in the nouns and verbs used across the stories). Each story had four versions, two target versions (by-phrase with plausible or implausible agents) and two filler versions (a temporal or locative prepositional phrase). Animate stories featured three participants: a named theme (*Jerry*), an unnamed bystander (*his wife*), and an unnamed agent (*a thief*). Inanimate stories included a named bystander (*Mary*) interacting with an object theme (*a flower*) and a named agent (*John*). All stories followed the same structure: the first sentence established a past timeframe using past progressive, the second sentence reported someone learning about the event, and the final sentence reinforced the agent's role. We controlled for word frequency, balanced gender in names, standardized sentence lengths, and avoided repetition of names, verbs, and locations across stories. Examples are shown in Table 1.

The two conditions critical to our study were stories in which an agentive interpretation of the *by*-phrase agreed or conflicted with the semantic information of the vignette. In each vignette, a specific event participant was described in a way that primed expectations for that event participant to be the agent of the main action. For example, a thief is described as moving behind some unsuspecting newlyweds, or John is described as walking past a garden whilst distracted by his mobile phone. In each case, these actions position those event participants to be the agents of the subsequent action (the thief commits the robbery; John crushes a flower). If the *by*-phrase included those event participants (e.g., *the flower was crushed by John*), an agentive reading of the *by*-phrase was congruent with these expectations (plausible text condition). If the *by*-phrase included a different event participant (e.g., *Jerry was robbed by his bride*), an agentive reading would conflict with these expectations (implausible text condition). Crucially, the target sentence in the implausible scenario remains true and accurate under a locative reading (e.g., near his bride, near Mary), but is false if interpreted as agentive. For the animate vignettes, these expectations were further enhanced via semantic links between the nouns and verbs used in the *by*-phrases (e.g., *thief* has stronger semantic links to *robbery* when compared to *bride*). The filler conditions were included to prevent participants from identifying the focus of the study (i.e., the *by*-phrase) and encourage them to pay attention to all aspects of each vignette.

https://doi.org/10.1017/S0142716426100526 Published online by Cambridge University Press

Table 1. Examples of story stimuli and their different conditions, with **key verbs** bolded

Animate noun example		
Sentence 1. Last night, when Jerry and his bride were gambling at a new casino, a thief moved to stand behind Jerry.	Phrase after **robbed** . . .	Condition
	. . . by the thief.	plausible (agentive)
Sentence 2. Jerry's family later found out that Jerry was **robbed** . . .	. . . by his bride.	implausible (bystander)
	. . . before dinner	Filler (temporal)
	. . . at the casino.	Filler (locative)
Sentence 3. The thief was eventually arrested.		
Inanimate noun example		
Sentence 1. Last Monday, when Mary was planting a flower in her garden, John walked towards Mary while looking at his phone.	Phrase after **crushed** . . .	Condition
	. . . by John.	plausible (agentive)
Sentence 2. Mary's brother later heard that the flower was **crushed** . . .	. . . by Mary.	implausible (bystander)
	. . . before breakfast.	Filler (temporal)
	. . . in the garden.	Filler (locative)
Sentence 3. Mary told John to be more careful.		

Norming study

We conducted a norming study to verify whether the implausible vignettes triggered a noticeable conflict between semantic and syntactic information. This was achieved by gathering ratings of plausibility for vignettes in the four different conditions. We chose plausibility ratings because we wanted to test whether the events described in each vignette seemed believable, with the expectation that the bystander NP versions should be less plausible than the agentive versions.

Norming method and procedure

Stimuli included the 20 vignettes as described in Section "***Vignette design***", split into four lists of twenty stories each, with text condition counterbalanced across the lists. We designed a plausibility rating task using *jsPsych* (de Leeuw, 2015) and recruited 40 monolingual speakers of English living in the United States from the research crowdsourcing website Prolific Academic (Age: min = 18, max = 57, M = 33.5, SD = 10.6; 52.5% female, 45% male, 2.5% prefer not to answer). Participants were evenly and randomly assigned to one of the four counterbalanced lists of 20 stories. Each trial presented full text from a randomly selected story vignette, along

https://doi.org/10.1017/S0142716426100526 Published online by Cambridge University Press

with a 5-point rating scale (1 = not at all plausible, 5 = highly plausible) with the following instructions:

> *Plausibility is a measure of how likely you think something can occur. For example, being full after eating a meal is more plausible than being hungry after eating a meal. However, one could still feel hungry after eating a meal. It is just less plausible when compared to being full. Your task is to rate each story you read on a scale of plausibility. The scale ranges from 1 (not at all plausible) to 5 (highly plausible). When making your ratings, please consider all aspects of the story. Also, take your time – please do not rush your answers.*

Statistical analysis

We fit a Bayesian multilevel mixed effects regression model with plausibility ratings as the dependent variable and a two-way interaction between story condition (plausible, implausible, temporal filler, locative filler) and theme type (animate vs. inanimate) as the predictor variable. To fit our model, we used the *brms* package (Bürkner, 2017) in R (R Core Team, 2025) with a generic weakly informative prior normal distribution (a mean of 0 and a standard deviation of 1). The predictor variables were contrast-coded using the *contr.equalprior* function from *bayestestR* (Makowski et al., 2019), assigning each level of the variable to equalized marginal prior distributions. We included subjects and items as random intercepts, with a random slope of story condition fit on both subjects and items. After fitting the model, we obtained pairwise comparisons of posterior distributions using the *emmeans* package (Lenth, 2024).

We subjected these pairwise comparisons to a region of practical equivalence (ROPE) test using the *bayestestR* package. These analyses measure which proportion of a credible interval from the posterior distribution falls within a user-defined ROPE, centered around zero (i.e., representing both the absence of an effect as well as the practical absence of an effect). Essentially, if a high percentage of a credible interval from a posterior distribution related to any one effect falls within a ROPE, it can be thought to reflect the absence of a meaningful effect. Much like frequentist approaches, which use a 0.05 threshold for *p*-values, decisions about what counts as a significant effect based on the percentage of a distribution that falls within a ROPE are arbitrary. In this data, we opted to describe our effects in relation to one another and avoid designating any effect using binary classifications of significant or not significant. We use a 95% credible interval in all our analyses. For the plausibility ratings, we used the built-in functionality of the bayestestR package to define the ROPE.

Results

Means and standard deviations of plausibility ratings for the four different story versions are displayed in Table 2, with two noteworthy observations. First, the plausibility ratings are lower for the implausible text condition when compared to the other three conditions. Second, the plausibility ratings are lower for stories with animate themes (compared to inanimate) for all four text conditions. For animate

https://doi.org/10.1017/S0142716426100526 Published online by Cambridge University Press

Table 2. Descriptive statistics of plausibility ratings

	Animate		Inanimate	
Text condition	*M*	*SD*	*M*	*SD*
Implausible	0.93	0.91	1.98	1.28
Plausible	2.42	1.23	2.79	1.13
Locative filler	2.45	1.25	2.79	1.11
Temporal filler	2.20	1.22	2.51	1.08

stories, pairwise comparisons confirmed that ratings were reliably lower for the implausible versions when compared to the other three text conditions (all ROPE = 0%). For inanimate stories, reliable differences were found between implausible and the plausible or locative stories (ROPE = 0%); however, the comparison to temporal fillers had a ROPE overlap of 9.41%, indicating their difference was less credible. Additionally, the difference between ratings for the animate and inanimate versions of the implausible stories was also reliable (ROPE = 0%), whereas these differences were not reliable for the other three conditions. As such, these data provide evidence that the implausible stories were perceived to be less plausible than the plausible condition, but also suggest that differences in the animacy of the theme of the *by*-phrase moderated this effect, with animate themes resulting in lower perceptions of plausibility when compared to inanimate themes. For the sake of space, we provide the full statistical contrasts as an Appendix in the online Supplemental Materials.

Experimental study

The norming study confirmed that the stories in the implausible condition were seen as reliably less plausible than the plausible versions. We then conducted a reading comprehension experiment with participants from the three language backgrounds relevant to our research hypotheses.

Method

Text comprehension questions

Three comprehension questions with two possible answers were created for each story. The questions asked about the agent of the main verb (*who verb-ed THEME?*), the time of the main verb event (*when was THEME verb-ed?*), or the location of the main verb event (*where was THEME verb-ed?*). The agent question answer options included the two NPs in the vignette (bystander and semantically-biased agent), and the locative and temporal questions were included as fillers (so that participants would not learn to rely solely on the NP in the *by*-phrase when reading). The location question answer options included the location mentioned in the story or a different location that was not mentioned in any of the other stories. The temporal question answer options included the time mentioned at the start of the sentence or an incorrect time for all conditions aside from the temporal fillers. For those stories,

https://doi.org/10.1017/S0142716426100526 Published online by Cambridge University Press

Table 3. Example comprehension questions

S1. Last Monday, when Mary was planting a flower in her garden, John walked towards Mary while looking at his phone. S2. Mary's brother later heard that the flower was crushed . . .			
Story condition	Time question	Agent question	Location question
(one of four possible preposition + NP endings)	*When was the flower crushed?*	*Who crushed the flower?*	*Where was the flower crushed?*
. . . by John. (agentive)	Last Monday. Last Friday.	Mary. John.	A garden. A bar.
. . . by Mary. (bystander)			
. . . in the garden. (locative)			
. . . before breakfast. (temporal)	After breakfast. Before breakfast.		
S3. Mary told John to be more careful.			

the temporal question included options related to the time mentioned in the temporal preposition + NP after the verb. These possibilities are summarized in Table 3.

Reading task

We used the same 20 vignettes described in Section "***Materials***". The stimuli were separated into similar counterbalanced lists and presented using a non-cumulative self-paced reading (SPR) design. In this paradigm, a reader sees the first word on their screen, with underlines representing placeholder spots for the rest of the words in a text. The reader then pushes a key to advance to the next word, which subsequently masks the previous word, and thus only one word is presented at any one time. We used non-cumulative SPR because we wanted to prevent participants from being able to refer back to previous text during the task, forcing a heavy reliance on memory representations of the event participants in the text. We collected reading times during this task, but do not report them here, as our theoretical question concerned the memory representations formed during reading rather than the online processing difficulty reflected in reading times. Our dependent measure was therefore accuracy on the comprehension questions that followed each vignette.

LexTALE English proficiency

We incorporated a version of the LexTALE test of English proficiency (Lemhöfer & Broersma, 2012) into our *jsPsych* survey. The LexTALE test includes 40 real English words and 20 English nonwords. Participants see the words one at a time and indicate if they think the word is real or not by clicking on buttons labeled as *yes* or *no*. Only accuracy and not reaction time is used to calculate the scores. As such,

https://doi.org/10.1017/S0142716426100526 Published online by Cambridge University Press

LexTALE scores are primarily a measure of vocabulary knowledge, but validation studies have found associations between LexTALE scores and general measures of English proficiency for advanced users of English (Lemhöfer & Broersma, 2012). The primary benefit of LexTALE is that it is relatively quick (around five minutes), which means a consistent measure of English proficiency can be collected from participants at the time of the experiment.

Participants

Three groups of participants were recruited: English monolinguals living in the United States, Mandarin Chinese learners of English living in Taiwan, and Japanese learners of English living in Japan. An initial wave of 147 monolingual English speakers was recruited from the online research crowdsourcing platform Prolific Academic. To be eligible, participants needed to reside in the United States and indicate they were monolingual speakers of English (they also did not participate in the norming study). An additional 104 Mandarin Chinese L1 learners of English and 89 Japanese L1 learners of English were recruited from university student populations in Taiwan and Japan through the aid of colleagues who work at these universities. We also encouraged any of the participants in Taiwan or Japan who completed the study to pass the recruitment details of our study on to their classmates.

Some of the L2 participants were removed from the final data analysis due to low accuracy on the practice and/or filler items (see section ***Results***). As such, the final composition of the participant groups was 147 Monolingual English speakers (M_{age} = 26.02, SD_{age} = 8.79, 77.5% female, 21.1% male, 1.3% prefer not to answer), 73 Japanese (M_{age} = 22.05, SD_{age} = 7.25, 52.1% female, 43.8% male, 4.1% prefer not to answer), and 98 Taiwanese Mandarin (M_{age} = 22.28, SD_{age} = 4.24, 75.5% female, 21.4% male, 3.1% prefer not to answer). In terms of the proficiency scores, the LexTALE were Japanese M = 64.34, SD = 9.28, Min = 51.25, Max = 93.75; Taiwanese M = 61.43, SD = 11.89, Min = 40.00, Max = 96.25. Age of English Exposure was Japanese M = 9.26, SD = 3.47, Min = 0, Max = 13; Taiwanese M = 7.40, SD = 2.77, Min = 1, Max = 14.

Procedure

The experiment was conducted using a web browser. Instructions were provided in English and Mandarin Chinese or Japanese for L2 participants. Participants completed a demographic survey including age, sex, and language status. English L2 participants also reported their age of English acquisition, daily English usage percentage, and completed the LexTALE proficiency test. Participants were then introduced to the reading task, instructed to use the spacebar to advance through stories, and told they would answer three questions after each story. They completed a practice session with three similar texts that lacked the key prepositional constructs before proceeding to the main task.

During the reading task, participants read 20 stories in a random order from counterbalanced lists. Texts appeared in 20-point Trebuchet MS black font on a gray background. The first sentence (20 words) was presented as one line, with the

https://doi.org/10.1017/S0142716426100526 Published online by Cambridge University Press

second and third sentences below. A fixation cross over the first word was replaced by the actual word when participants pressed the spacebar, revealing placeholder lines for the remaining words. Participants read at their own pace, advancing words with the spacebar. The three comprehension questions appeared individually after each text, with two answer options as adjacent buttons. Question order and answer placement were randomly assigned per participant.

Statistical analysis

We again employed Bayesian multilevel mixed effects regression models to test our predictions. For these data, our outcome variable was the binary response to the *who* comprehension question asked after each story. The binary response was coded so that a score of 1 was assigned when participants chose the NP in the *by*-phrase, with other answers being scored as 0. We created logistic models following a Bernoulli distribution and logit link using the *brms* package (Bürkner, 2017) in R (R Core Team, 2025). Estimated marginal means and trends were calculated using *emmeans* (Lenth, 2024), and we used the same ROPE method from the *bayestestR* package (Makowski et al., 2019) to determine the significance of the effects in the models. All models were trained on 10,000 iterations. Because these models were logistic regressions, we present the posterior distributions in terms of probability.

Model 1: Agent selection model

The first model included data from all three language groups, with a three-way interaction among language group (English, Japanese, Mandarin), story condition (plausible vs. implausible; we did not analyse data from the filler story conditions), and animacy (animate vs. Inanimate; included to control for any differences in how animate vs. inanimate NPs were interpreted). All variables were contrast-coded to include equal marginal prior distributions for all levels of each factor. Random intercepts with a random slope of the interaction between story condition and animacy were fit for subjects, with a random slope of vignette condition fit for items. We again fit a generic weakly informative prior using a normal distribution with a mean of 0 and standard deviation of 1. Random effects were kept at the default priors (Student *t*-distribution with mean of 0 and standard deviation of 2.5).

Model 2: L2 English proficiency model

For the L2 participants, we fit a model that included our two measures of L2 English proficiency (LexTALE and Age of English Exposure) and responses for the implausible stories only. The model included three-way interactions among language group (Japanese or Mandarin), vignette animacy (animate or inanimate), and either LexTALE English proficiency scores or Age of English Exposure values (z-scored with a mean of 0 and standard deviation of 1). Random intercepts were fit for subjects and items, with a random slope of animacy fit for subjects. We used the same contrasts and prior specifications as the previous model.

https://doi.org/10.1017/S0142716426100526 Published online by Cambridge University Press

Predictions

H1. Our first hypothesis incorporates predictions of the SSH (i.e., that semantic information is more salient for L2 speakers) into the cue framework of the Competition Model. Because semantic information is thought to be a more reliable cue for L2 speakers, the Japanese and Mandarin L1 participants should have a lower frequency of choosing the *by*-phrase NP as the agent in the implausible vignette condition when compared to the L1 English speakers. However, this difference will be moderated by L2 English proficiency, with higher proficiency reducing the reliance on semantic cues and thus increasing the probability of choosing the *by*-phrase NP as the agent in the implausible vignette condition.

H2. Following the Competition Model, this hypothesis predicts that the strength of syntactic cues within the *by*-phrase will interact with corresponding L1 structures. Because English *by* and Japanese *ni* are multifunctional (allowing for agentive or locative interpretations), a testable prediction is that there will be a weaker agentive bias for English *by* (for Japanese, this is reinforced through similar L1–L2 cues). In contrast, Mandarin speakers have distinct L1 cues with specialized cue strengths: *bei* for agentive functions and *zai* for locative functions. When mapping onto English *by*, only the agentive cue strength from *bei* transfers, and thus, there will be a stronger agentive bias for these speakers. This means that the Mandarin speakers should choose the *by*-phrase NP as the agent at a higher rate when compared to the Japanese and English speakers for the implausible vignettes. However, this effect should be weaker for Mandarin participants with higher L2 English proficiency, as they should rely less on direct mapping of L1 to L2 cues.

Results

We first prepared the data for analysis by identifying any participants with extremely low accuracy scores on comprehension questions. We removed any participant with lower than 70% accuracy on the practice questions and/or the comprehension questions for the plausible and filler story conditions. In total, 16 participants from the Japanese group and six participants from the Taiwanese group were removed for low accuracy.

Descriptive statistics

Figure 1 presents the proportion of answers indicating whether the *by*-phrase NP was chosen as the agent, divided into vignettes with animate or inanimate themes in the *by*-phrase. All participants chose the *by*-phrase NP as the agent in plausible vignettes at a higher percentage when compared to implausible vignettes, and this difference was larger when comparing between animate versions of the vignettes. This interaction was largely driven by a difference within the plausible vignettes, where participants almost always chose the *by*-phrase NP as the agent for animate vignettes. In contrast, the percentages were relatively similar between animate and inanimate versions within the implausible condition.

https://doi.org/10.1017/S0142716426100526 Published online by Cambridge University Press

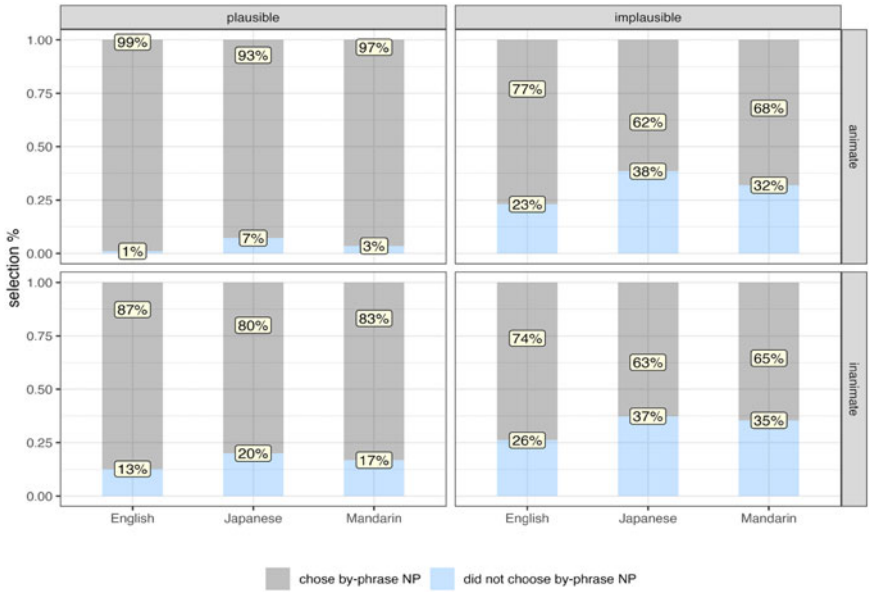


Figure 1. Proportion of answers choosing the by-phrase NP as agent when answering the who-questions (raw data).

Model 1: Agent selection

The relative differences among the percentages seen in Figure 1 were replicated by the multilevel model, which predicted a high probability for choosing the *by*-phrase NP as the agent in the animate (>96%) and inanimate (>89%) plausible vignettes for all three language groups (see Table 4). Within-group comparisons confirmed these differences were credible, in that the probability of choosing the *by*-phrase NP was lower for implausible vignettes, regardless of animacy (All ROPE = 0%; full contrasts provided in Supplemental Material).

Between-group comparisons, shown in Table 5, revealed comparable *by*-phrase NP selection among the groups in plausible vignettes across both animacy versions (all ROPE ≥ 32%), indicating baseline similarity in response patterns. For the implausible vignettes, language-specific differences emerged with varying degrees of certainty. For animate themed versions, English monolinguals showed higher *by*-phrase NP selection than the Japanese L1 group (Estimate = 0.19, 95% HDI [0.08, 0.31], ROPE = 0%), with the credible interval clearly excluding the ROPE. The difference between English monolinguals and Mandarin L1 showed a similar direction but with greater ROPE overlaps (Estimate = 0.10, 95% HDI [0.01, 0.19], ROPE = 10%). For inanimate themed versions, the English monolingual to Japanese L1 comparison again showed higher *by*-phrase NP selection for English monolinguals, though with slightly more uncertainty (Estimate = 0.13, 95% HDI [0.01, 0.25], ROPE = 6%), and the English monolingual to Mandarin L1 comparison showed a comparable pattern (Estimate = 0.10, 95% HDI [0.00, 0.21], ROPE = 11%). Japanese L1 and Mandarin L1 speakers did not differ meaningfully from each other in either animacy condition (ROPE ≥ 22%). In order to better

https://doi.org/10.1017/S0142716426100526 Published online by Cambridge University Press

Table 4. Model predicted probability of choosing by-phrase NP as agent

Implausible vignettes				
	Animate		Inanimate	
Language group	Estimate	95% HDI	Estimate	95% HDI
English L1	0.83	0.76, 0.89	0.79	0.69, 0.87
Japanese L1	0.64	0.51, 0.76	0.66	0.52, 0.79
Mandarin L1	0.73	0.63, 0.82	0.68	0.56, 0.80
Plausible vignettes				
	Animate		Inanimate	
Language group	Estimate	95% HDI	Estimate	95% HDI
English L1	1.00	0.99, 1.00	0.93	0.87, 0.97
Japanese L1	0.96	0.93, 0.99	0.86	0.77, 0.94
Mandarin L1	0.98	0.97, 1.00	0.89	0.81, 0.95

HDI = highest density interval. Estimate is median of posterior distribution. Positive estimates reflect higher probability of choosing the *by*-phrase NP.

understand these overlaps, we plot the posterior distributions and the ROPE in Figure 2 for these contrasts.

There are three main conclusions to draw from Model 1. First, both L1 and L2 participants were less likely to choose the *by*-phrase NP as the agent for implausible versus plausible vignettes. Yet, the likelihood of choosing the *by*-phrase NP was still higher than the other NP, even within the implausible stories. This finding suggests discourse-level information was being used by participants when answering the *who* question after the implausible stories, but only sometimes. Second, this tendency was stronger for the L2 participants, suggesting they drew more heavily from discourse than syntactic information when accessing their memory representations of the vignettes. Third, agent selection behavior in the implausible stories was more similar than different between the JP and TW participants, with a predicted difference of 2 to 9% (see Table 5). These results provide support for Hypothesis 1, with the strongest evidence found for animate implausible vignettes: L2 English speakers relied more heavily on semantic cues, whereas L1 English speakers had stronger reliance on syntactic cues. In contrast, Hypothesis 2 was not supported: Japanese and Mandarin L2 speakers showed similar patterns, and Japanese L1 speakers diverged more strongly from English monolinguals when compared to Mandarin L1 speakers, contrary to predictions of L1–L2 cue similarity.

Model 2: English proficiency

Model 2 examined agent selection within implausible stories for the Japanese L1 and Mandarin L1 participants in the context of two measures of English proficiency (LexTALE and Age of English Exposure). Results indicated that for Mandarin L1 speakers, higher LexTALE scores were associated with a higher probability of

https://doi.org/10.1017/S0142716426100526 Published online by Cambridge University Press

Table 5. Pairwise comparisons among language groups for *by*-phrase NP selection

	Implausible					
	Animate			Inanimate		
Comparison	Estimate	95% HDI	ROPE	Estimate	95% HDI	ROPE
English–Japanese	0.19	0.08, 0.31	0.00	0.13	0.01, 0.25	0.06
English–Mandarin	0.10	0.01, 0.19	0.10	0.10	0.00, 0.21	0.11
Japanese–Mandarin	−0.09	−0.21, 0.03	0.22	−0.02	−0.15, 0.11	0.50
	Plausible					
	Animate			Inanimate		
Comparison	Estimate	95% HDI	ROPE	Estimate	95% HDI	ROPE
English–Japanese	0.03	0.01, 0.06	0.82	0.06	−0.01, 0.15	0.32
English–Mandarin	0.01	0.00, 0.03	1.00	0.04	−0.02, 0.10	0.63
Japanese–Mandarin	−0.02	−0.05, 0.00	0.95	−0.03	−0.12, 0.05	0.67

ROPE range [−0.045, 0.045], HDI = highest density interval. Estimate is median of posterior distribution. Positive estimates reflect higher probability of choosing the *by*-phrase NP.

choosing the *by*-phrase NP in inanimate versions of the implausible vignettes (specifically, a one standard deviation increase in LexTALE was associated with a 12% probability increase; ROPE = 5%). All other effects were small and overlapped strongly with the ROPE, indicating no other credible effects. These contrasts are shown in Table 6 and plotted in Figure 3.

Discussion

To investigate how L1 and L2 English speakers interpret thematic roles when semantic and syntactic cues align or conflict, we collected data from monolingual English speakers, Taiwanese Mandarin speakers, and Japanese speakers of English. We measured comprehension of target sentences containing *by*-phrases embedded in short vignettes. The NP included in the *by*-phrase was either congruent (plausible condition) or not congruent (implausible condition) with the semantic information in the opening and closing sentences of the vignette. Participants read the vignettes one word at a time, answering three *wh*-questions after each vignette (*who*, *when*, and *where*).

We examined how participants reconciled discourse-level semantic cues with syntactic structure in implausible story contexts. We compared two possible hypotheses based on the Competition Model and its integration with the SSH. These hypotheses led to contrasting predictions: (H1) held that L2 speakers will attend more strongly to semantic cues when compared to L1 speakers, whereas the other (H2) held that L2 speakers would be influenced by L1–L2 cue similarity when making their decisions. Overall, our results provide stronger support for H1 when compared to H2.

https://doi.org/10.1017/S0142716426100526 Published online by Cambridge University Press

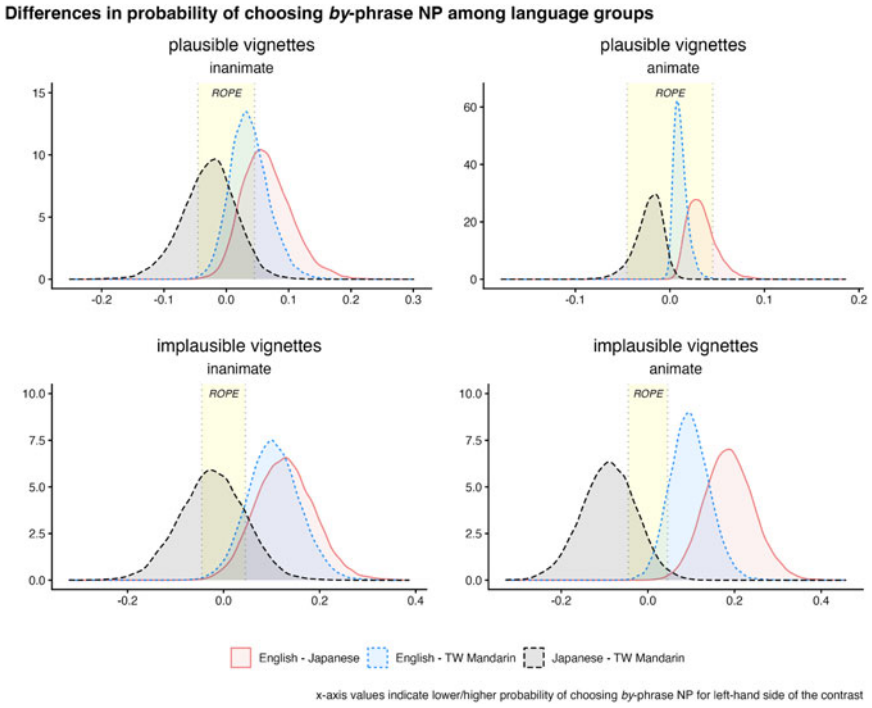


Figure 2. Posterior distributions of between-group contrasts comparing agent selection for implausible and plausible vignettes across animate and inanimate themes. For plausible vignettes, heavy overlap with the ROPE indicates comparability among the groups. For implausible vignettes, English speakers differed from both L2 groups, with the strongest effects between English and Japanese speakers for animate themes, while Japanese and Mandarin speakers remained comparable.

L1 and L2 comprehension of implausible vignettes

When reading plausible and implausible stories about past events, our L1 and L2 participants reliably interpreted the NP encoded by the *by*-phrase to be the agent of the event. This effect is unremarkable for the plausible vignettes, because the semantic and syntactic cues both supported interpreting the *by*-phrase NP as the agent. Yet, more often than not, participants made similar agent selection choices for implausible vignettes—choosing the *by*-phrase NP as the agent—even when such an interpretation conflicted with semantic information that came before and after reading the target sentence. This finding attests to a strong agentive cue of the English *by*-phrase when associated with an animate NP for both the L1 and L2 participants.

However, the weighting of this cue differed among our participant groups—and in the direction we had predicted—with L1 English speakers being more strongly influenced by the syntactic cue when faced with competition between syntax and semantics. In fact, while the probability difference between plausible and implausible stories for these participants was ∼12 to 17%, this reflected a probability decrease from 100% to 83% (animate) or 93% to 79% (inanimate; see Table 4). In other words, L1 English speakers would choose the *by*-phrase NP as the agent at

https://doi.org/10.1017/S0142716426100526 Published online by Cambridge University Press

Table 6. Effects of L2 English proficiency on probability of choosing *by*-phrase NP

LexTALE						
	Animate			Inanimate		
Language	Estimate	95% HDI	ROPE	Estimate	95% HDI	ROPE
Japanese	0.06	−0.10, 0.23	0.34	−0.07	−0.25, 0.08	0.30
Mandarin	0.04	−0.04, 0.13	0.55	0.12	0.02, 0.25	0.05
Age of English exposure						
	Animate			Inanimate		
Language	Estimate	95% HDI	ROPE	Estimate	95% HDI	ROPE
Japanese	0.11	−0.01, 0.26	0.13	−0.05	−0.18, 0.09	0.45
Mandarin	−0.03	−0.15, 0.07	0.56	0.04	−0.09, 0.17	0.49

ROPE range [−0.045, 0.045], HDI = highest density interval. Estimate is median of posterior distribution. Positive estimates reflect higher probability of choosing the *by*-phrase NP.

least ~80% of the time, even when information in the surrounding discourse conflicted with such a choice. The difference between plausible and implausible vignettes does suggest that discourse information can influence cue weighting for English speakers, but only to a relatively weak degree, mirroring previous research showing similar weak effects of prior discourse information on cue strength (Su, 2004).

In contrast to the L1 English speakers, results from Model 1 indicated the L2 participants made those same decisions when reading implausible stories about 68–73% (Mandarin) or 64–66% (Japanese) of the time (Table 4). This finding reflects the attested tendency for L2 speakers to rely more heavily on semantic and discourse-level information during L2 sentence processing (Clahsen & Felser, 2018; Hopp, 2022b). Recall that the SPR task we used was non-cumulative, meaning that words did not remain on the screen after participants read them. Such a task induces a greater reliance on memory representations, and it is under these conditions that L2 speakers are especially prone to rely more heavily on semantic versus syntactic information when recalling information (Cunnings, 2017). This does not mean that the L2 participants were unable to process the syntactic information, but rather that the syntactic cue was less influential during the mental construction of the stories for L2 participants.

Effects of L1–L2 cue mapping

Our second hypothesis predicted that differences between the Mandarin and Japanese participants' L1 structures would influence their agent selection behavior. If the unifunctional agentive nature of Mandarin *bei* is mapped to English *by* through cue transfer, as predicted by the Competition Model (Bates & MacWhinney, 1987), English *by* would become a valid and reliable cue for marking agentive roles in passive sentences for Mandarin speakers. When comparing the Mandarin speakers to the Japanese speakers, the descriptive results suggest that the

https://doi.org/10.1017/S0142716426100526 Published online by Cambridge University Press

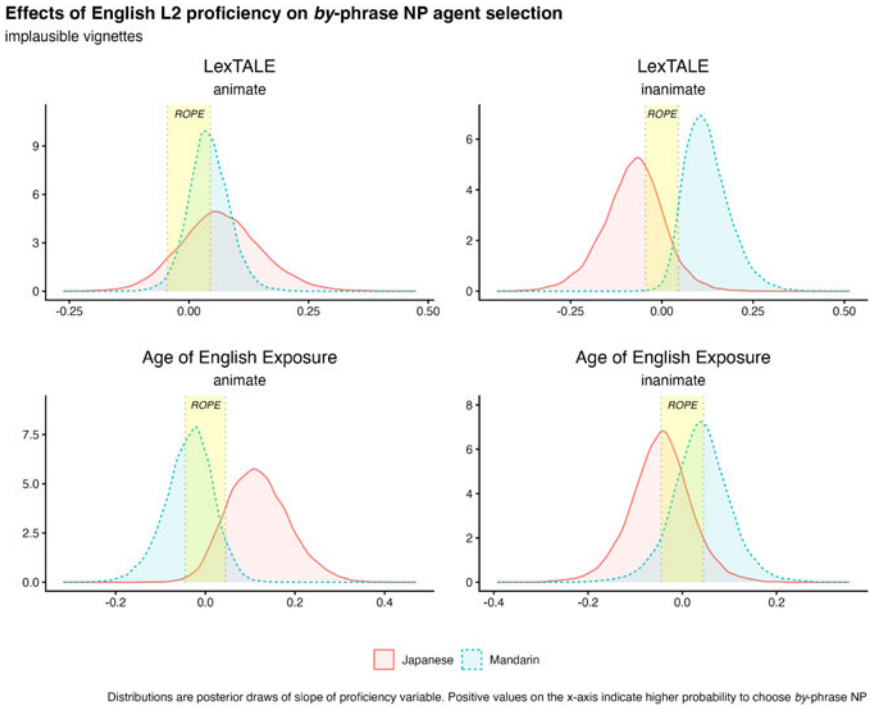


Figure 3. Posterior distributions of effects of L2 English proficiency measures on probability of choosing by-phrase NP as agent for the implausible vignettes. There was only one credible effect: higher LexTALE scores for Mandarin speakers were associated with higher probability of choosing the by-phrase NP for inanimate versions.

Mandarin participants did indeed interpret the *by*-phrase NP as the agent at a rate higher than the Japanese participants. Statistically, however, these differences were not robust, as there was a great deal of overlap between the posterior distributions of their differences (Figure 2). What this means is that the L2 speaker groups were more similar than different, and that predictions related to cue transfer were not supported.

The lack of any strong evidence for cue transfer effects in the current data may be attributed to our use of larger discourse contexts when compared to prior research focusing on the processing of single sentences. In such contexts, differences between L1 and L2 cues provide robust evidence of cue weighting strategies, L1 to L2 influence, and L2 to L1 influence (Bates & MacWhinney, 1987; Liu et al., 1992; Su, 2004). Any heuristics based on syntactic cues may therefore give way to semantic strategies when such information is available. For example, sentence stimuli used in prior studies, such as FOX EAT CAT and BOOK EAT FOX, are designed to amplify syntactic violations (Liu et al., 1992), making cue transfer more apparent and measurable. So it may be the case that our predictions related to cue transfer are less realistic when the L2 is working with larger and more cohesive discourse contexts, such as our story vignettes. This finding does not call into question the overall predictions of the Competition Model. Instead, we believe they support the main

https://doi.org/10.1017/S0142716426100526 Published online by Cambridge University Press

point—that in these contexts the L2 speakers relied on what was (to them) the most reliable cue to understanding meaning: the discourse semantics. That this also follows the predictions of the SSH showcases complementary rather than completely opposed predictions for our data.

However, it may also be the case that differences in L2 proficiency between the L2 groups may have affected the agent selection behavior and influence of L1 cues. Our second statistical model addressed this question and found a credible effect for the Mandarin speakers. Increases in LexTALE scores were associated with a higher chance of choosing the *by*-phrase NP as the agent, but only for the inanimate versions of the implausible vignettes. This means that higher proficiency Mandarin speakers were more similar to the L1 English speakers. That such a finding held only for the Mandarin group may, in turn, provide evidence of cue transfer, in that increased L2 proficiency strengthened the similarity between agentive *bei* and *by* (and moreover may have reduced the need to rely on semantic cues). In contrast, while the Japanese speakers may have been better equipped to incorporate the multifunctional nature of English *by* based on L1 cue similarity, the vignettes we used all made a locative interpretation very difficult, as the *by*-phrase NP was always an animate noun (not to be confused with the animacy contrast of the themes). Therefore, our stimuli may have enhanced the reliability and validity of agentive cues, which may have come more easily to the Mandarin speakers because of a perceived clear L1 to L2 mapping. However, this interpretation remains speculative and requires future research with contexts allowing locative interpretations in order to verify. And, in our study, a locative interpretation may not have been considered at all, meaning that the transfer related to any ambiguity of *by* was never made possible (this point is discussed further in the limitations).

Moreover, the fact that this proficiency effect only occurred for vignettes with inanimate themes is puzzling, as we did not initially expect the animacy of the theme would affect agent selection (to be clear, all possible agents were animate). Yet our plausibility norming study indicated animate-themed vignettes were taken as less plausible than the inanimate-themed vignettes (Table 2). This means people were more accepting of the events in our stories happening to inanimate objects versus people, and may, in turn, suggest a further interaction with the variety of verbs and event participants used in our stimuli. We accounted for some of this inherent variance among stimuli by using random intercepts and slopes in our multilevel models, yet the animacy effect suggests that when events involve human themes, additional constraints on typical event structures may create conditions under which proficiency differences became more apparent.

Limitations

This study is not without its limitations. Firstly, our use of a non-cumulative SPR design means that participants could not review prior information in the vignettes. This was intentional, but future work should incorporate such processing data into its analyses. For instance, future investigation with eye-tracking methods in a design that allows participants to review prior information in the vignettes while reading would provide further information about whether readers are returning to semantic cues in the stories when encountering conflicting syntactic cues.

https://doi.org/10.1017/S0142716426100526 Published online by Cambridge University Press

We are also aware that wide-scale use of the LexTALE test has recently come under scrutiny as a general measure of L2 proficiency (Puig-Mayenco et al., 2023). The main problem is that LexTALE is primarily designed to measure the proficiency of learners who are already at relatively advanced levels of L2 proficiency. Although our participants were all university students who had undertaken English language training throughout their schooling, our study would have further benefited from additional measures of L2 proficiency as a means to triangulate this effect. Moreover, LexTALE is largely a measure of vocabulary knowledge, which may not have been important for processing the vignettes in our study.

A final yet crucial limitation is the nature of ambiguity in our stimuli. Our desire to test a given cue (in this case, English *by*) was motivated by the possibility of agentive or locative interpretations. In practice, however, our use of agentive NPs in the *by*-phrase made the locative interpretation more difficult to obtain. While locative interpretation would allow for a pathway to reconcile any discrepancies between syntax and semantics, it is also the case that interpreting human bystanders as a location in a *by*-phrase is admittedly less common (though structurally allowed in syntax), particularly when compared to other prepositions such as *near* or *behind*. As such, although a locative interpretation can be made, this is only possible if the reader is willing to put in the requisite effort to construct such an interpretation. In our current design, we are unable to confidently determine whether participants made such an interpretation, which may further explain the lack of L1–L2 transfer effects.

Conclusion

We examined how L1 and L2 English speakers negotiate a competition between semantic and syntactic cues during discourse processing. Our study showed that the agentive bias of the English *by*-phrase is strong and robust, persisting in both L1 and L2 interpretations even in light of conflicts with discourse-level semantic information. Although we predicted that relative similarity between L1–L2 cues would influence L2 processing behavior, results from our Bayesian regression models indicate more similarity than difference among the L2 groups. This may be partially a function of recruiting relatively high-proficiency L2 speakers, as well as general similarity (rather than conflict) between L1–L2 cues. An important contribution of this work is our consideration of how global discourse-level cues interact with localized syntactic cues, and we encourage future research to continue investigating cue competition and other similar effects within discourse contexts.

Supplementary material. The supplementary material for this article can be found at https://osf.io/kbwun/

Replication package. Full data, code, and appendices are provided on the Open Science Framework: https://osf.io/kbwun/.

Author contributions. Skalicky: Conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, software, validation, visualization, writing—original draft.

Chen: Conceptualization, funding acquisition, investigation, methodology, writing—reviewing and editing.

https://doi.org/10.1017/S0142716426100526 Published online by Cambridge University Press

Competing interests. The author(s) has/have no competing interests to declare.

Ethical standard. This research was approved by the Victoria University of Wellington Human Ethics Committee #29346. Informed consent was obtained from all participants reported in this study.

Note

1. It is noteworthy, however, that in passive contexts, *ni* is compatible specifically with locative phrases that have a directional reading (e.g., go to X). Therefore, it is not entirely analogous to English *by* in passive environments, which can mark general locative phrases. The key point here, however, is that *ni* (in non-passive contexts) is indeed compatible with both agentive and locative NPs—a flexibility that is highly analogous to English *by*.

References

Bates, E., & MacWhinney, B. (1987). Competition, Variation, and Language Learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157–193). Lawrence Erlbaum.

Bürkner, P. C. (2017). brms: an *R* package for Bayesian multilevel models using *Stan*. *Journal of Statistical Software*, **80**(1), 1–28. https://doi.org/10.18637/jss.v080.i01.

Chu, C. (1973). The passive construction: Chinese and English. *Journal of Chinese Linguistics*, **1**(3), 437–470.

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

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

Crible, L., Wetzel, M., & Zufferey, S. (2021). Lexical and structural cues to discourse processing in first and second language. *Frontiers in Psychology*, **12**, 1–16. https://doi.org/10.3389/fpsyg.2021.685491.

Cunnings, I. (2017). Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition*, **20**(4), 659–678. https://doi.org/10.1017/S1366728916000675.

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, **47**(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Deniz, N. D. (2022). Processing syntactic and semantic information in the L2: evidence for differential cue-weighting in the L1 and L2. *Bilingualism: Language and Cognition*, **25**(5), 713–725. https://doi.org/10.1017/S1366728921001140.

Felser, C., & Drummer, J. D. (2022). Binding out of relative clauses in native and non-native sentence comprehension. *Journal of Psycholinguistic Research*, **51**(4), 763–788. https://doi.org/10.1007/s10936-022-09845-z.

Fillmore, C. J. (1968). The Case for Case. In E. Bach and R. T. Harms (Eds.), *Universals in linguistic theory* (pp. 1–25). Holt, Rinehart and Winston (Part Two).

Harrington, M. (1987). Processing transfer: language-specific processing strategies as a source of inter-language variation. *Applied Psycholinguistics*, **8**(4), 351–377. https://doi.org/10.1017/S0142716400000370.

Hashimoto, M. J. (1988). The Structure and Typology of the Chinese Passive Construction. In Shibatani (Ed.), *Passive and voice* (pp. 329–354). John Benjamins.

Hopp, H. (2022a). Second language sentence processing. *Annual Review of Linguistics*, **8**, 235–256.

Hopp, H. (2022b). Sentence Processing in a Second Language. In A. Godfroid & H. Hopp (Eds.), *The Routledge handbook of second language acquisition and psycholinguistics* (1st ed., pp. 216–228). Routledge. https://doi.org/10.4324/9781003018872-21.

Iwasaki, S. (2018). Passives. In Y. Hasegawa (Ed.), *The Cambridge handbook of Japanese linguistics* (pp. 530–556). Cambridge University Press.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological Review*, **87**(4), 329–354.

Kilborn, K. (1989). Sentence processing in a second language: the timing of transfer. *Language and Speech*, **32**(1), 1–23. https://doi.org/10.1177/002383098903200101.

https://doi.org/10.1017/S0142716426100526 Published online by Cambridge University Press

Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research*, **25**(2), 173–227. https://doi.org/10.1177/0267658308100283.

Larson, R. (1988). On the double object construction. *Linguistic Inquiry*, **19**(3), 335–391.

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

Lenth, R. (2024). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (Version 1.10.1) [Computer software]. https://CRAN.R-project.org/package = emmeans.

Li, P., Bates, E., & MacWhinney, B. (1993). Processing a language without inflections: a reaction time study of sentence interpretation in Chinese. *Journal of Memory and Language*, **32**, 169–192.

Li, P., & MacWhinney, B. (2012). Competition Model. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–5). John Wiley & Sons, Inc. https://doi.org/10.1002/9781405198431.wbeal0168.

Liang, L., Chondrogianni, V., & Chen, B. (2022). Online processing of the grammatical aspect marker by L2 Chinese learners. *Second Language Research*, **38**(4), 765–786. https://doi.org/10.1177/0267658321996423.

Liu, H., Bates, E., & Li, P. (1992). Sentence interpretation in bilingual speakers of English and Chinese. *Applied Psycholinguistics*, **13**(4), 451–484. https://doi.org/10.1017/S0142716400005762.

MacWhinney, B. (1987). The Competition Model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249–308). Lawrence Erlbaum.

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, **4**(40), 1541. https://doi.org/10.21105/joss.01541.

Mcdonald, J. L. (1987). Sentence interpretation in bilingual speakers of English and Dutch. *Applied Psycholinguistics*, **8**(4), 379–413. https://doi.org/10.1017/S0142716400000382.

McManus, K. (2021). *Crosslinguistic influence and second language learning*. Routledge.

O'Brien, E. J., & Cook, A. E. (2016). Coherence threshold and the continuity of processing: the RI-Val Model of Comprehension. *Discourse Processes*, **53**(5–6), 326–338. https://doi.org/10.1080/0163853X.2015.1123341.

O'Brien, E. J., Cook, A. E., & Lorch, R. F. (Eds.). (2015). *Inferences during reading*. Cambridge University Press.

Pan, H., & Hu, X. (2021). The Passive Construction in Chinese. In M. Aronoff (Ed.), *Oxford research encyclopedia of linguistics* (pp. 1–38). Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.884.

Pan, H.-Y., & Felser, C. (2011). Referential context effects in L2 ambiguity resolution: evidence from self-paced reading. *Lingua*, **121**(2), 221–236. https://doi.org/10.1016/j.lingua.2010.08.003.

Puig-Mayenco, E., Chaouch-Orozco, A., Liu, H., & Martín-Villena, F. (2023). The LexTALE as a measure of L2 global proficiency: a cautionary tale based on a partial replication of Lemhöfer and Broersma (2012). *Linguistic Approaches to Bilingualism*, **13**(3), 299–314. https://doi.org/10.1075/lab.22048.pui.

R Core Team. (2025). *R: A Language and Environment for Statistical Computing* [Computer software]. https://www.R-project.org/.

Su, I. R. (2001). Transfer of sentence processing strategies: a comparison of L2 learners of Chinese and English. *Applied Psycholinguistics*, **22**(1), 83–112. https://doi.org/10.1017/S0142716401001059.

Su, I. R. (2004). The effects of discourse processing with regard to syntactic and semantic cues: a competition model study. *Applied Psycholinguistics*, **25**(4), 587–601. https://doi.org/10.1017/s0142716404001286.

Tokowicz, N. (2014). *Lexical processing and second language acquisition*. Routledge.

Tokowicz, N., Warren, T., & Tolentino, L. (2024). The competition model and adult second language learning: cross-language similarity and L2 morphosyntax learning. *Language Teaching Research Quarterly*, **44**, 71–83. https://doi.org/10.32038/ltrq.2024.44.08.

van den Broek, P., & Helder, A. (2017). Cognitive processes in discourse comprehension: passive processes, reader-initiated processes, and evolving mental representations. *Discourse Processes*, **54**(5–6), 360–372. https://doi.org/10.1080/0163853X.2017.1306677.

Walter, D. R., Fischer, G., & Cai, J. (2024). The effect of cue length and position on noticing and learning of determiner agreement pairings: Evidence from a cue-balanced artificial vocabulary learning task. *Plos One*, **19**(7), e0302355. https://doi.org/10.1371/journal.pone.0302355.

https://doi.org/10.1017/S0142716426100526 Published online by Cambridge University Press

Wanner, A. (2009). *Deconstructing the English passive*. Mouton de Gruyter.

Washio, R. (1987). The Japanese passive. *The Linguistic Review*, **6**, 227–263.

Xiao, R., McEnery, T., & Qian, Y. (2006). Passive constructions in English and Chinese: a corpus-based contrastive study. *Languages in Contrast*, **6**(1), 109–149. https://doi.org/10.1075/lic.6.1.05xia

Zhao, H., & Fan, J. (2021). Modeling input factors in second language acquisition of the English article construction. *Frontiers in Psychology*, **12**, 1–15. https://doi.org/10.3389/fpsyg.2021.653258.

Zhao, H., Vanek, N., & MacWhinney, B. (2025). Mental simulation in bilingual and second language processing: new directions in the competition model. *Brain and Language*, **269**, 1–9. https://doi.org/10.2139/ssrn.5125974

Zufferey, S., Mak, W., Degand, L., & Sanders, T. (2015). Advanced learners' comprehension of discourse connectives: the role of L1 transfer across on-line and off-line tasks. *Second Language Research*, **31**(3), 389–411. https://doi.org/10.1177/0267658315573349

Cite this article: Skalicky, S. & Chen, V. (2026). First and second language speaker understanding of implausible English passives in discourse contexts. *Applied Psycholinguistics*. https://doi.org/10.1017/S0142716426100526

https://doi.org/10.1017/S0142716426100526 Published online by Cambridge University Press