Bilingualism: Language and Cognition

cambridge.org/bil

Cross-linguistic L1–L2 dis/similarity effect on mental imagery in incremental motion event processing

Taketo Nishide[1], Helen Zhao[1] and Simon De Deyne[2]

[1]School of Languages and Linguistics, Faculty of Arts, The University of Melbourne, Parkville, VIC, Australia and [2]Melbourne School of Psychological Sciences, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Parkville, VIC, Australia

Research Article

Cite this article: Nishide, T., Zhao, H. and De Deyne, S. (2025). Cross-linguistic L1–L2 dis/similarity effect on mental imagery in incremental motion event processing. *Bilingualism: Language and Cognition* 1–17. https://doi.org/10.1017/S1366728925100801

Received: 28 February 2025
Revised: 8 September 2025
Accepted: 24 October 2025

Keywords:
mental imagery; second language processing; crosslinguistic influence; incremental processing; spatial language

Corresponding author:
Helen Zhao;
Email: helen.zhao@unimelb.edu.au

This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

Abstract

Despite abundant studies on motion events and mental simulation in first languages (L1s), research on how cross-linguistic dis/similarity – whether an L1 shares constructions with a second language (L2) – affects mental simulation during incremental L2 processing remains limited. This study used a novel self-paced reading task with video verification to investigate L1 influence on mental imagery of the dual (directional/locational) interpretation of locative prepositions. Participants included native English speakers and advanced L2 English learners whose L1s were either similar (Dutch) or dissimilar (Japanese) to English. Results revealed an L1 dis/similarity effect on the reaction times for the directional interpretation, but not for the locational interpretation, which was readily accessible across all L1 groups. Factors such as L2 proficiency and onset age of L2 acquisition were found to be constrained by L1, suggesting that L1–L2 constructional correspondence limits the influence of learner factors. These findings support the simulation-based model of L2 sentence processing.

1. Introduction

Theories of embodied cognition emphasise the central role of mental simulation: the reactivation of the perceptual-motor experience accumulated through the interaction between one's body, mind and the world (Barsalou, 1999, 2008; Barsalou et al., 2008). Within this theoretical framework, the experiential traces account (Zwaan & Madden, 2005) argues that linguistic representations are intrinsically linked to co-occurring referents and events in the real world. These interconnected experiences are stored in memory as experiential traces. From this perspective, mental simulation is a language-induced phenomenon, where language comprehension involves reactivating these stored experiential traces of language.

While substantial behavioural evidence demonstrates that native speakers (NSs) perform mental simulation during language comprehension (e.g., Zhang et al., 2022; Zwaan et al., 2002), relatively little is known about how this process operates during L2 sentence comprehension (Ahn & Jiang, 2018; Wang & Zhao, 2024). Several issues related to mental simulation remain unresolved. First, although extensive research has explored mental simulation in relation to perceptual (e.g., shape, colour, distance) and sensory-motor features (e.g., Ahlberg et al., 2018; Glenberg & Kaschak, 2002), studies investigating mental imagery of motion events are scarce, particularly in the context of L2 processing. Second, while the role of mental simulation in *incremental* sentence processing has been studied in L1 comprehension (Sato et al., 2013; Wang & Zhao, under review), almost no research has extended this to L2 contexts. To date, no studies have investigated mental imagery effects on motion event interpretations during incremental L2 sentence processing. Third, there has been limited research on the potential influence of cross-linguistic differences in L1s on the mental simulation in L2 processing (Ahlberg et al., 2018). To address these gaps, the present study employed a self-paced reading task embedded with video verification to examine English motion event processing. We compared English native speakers (NSs) with two groups of advanced L2 English learners: L1-Dutch speakers (whose L1 is similar to English) and L1-Japanese speakers (whose L1 is dissimilar to English).

2. Literature review

2.1. Mental simulation in L1 processing

Research has consistently demonstrated that humans perform mental simulation and create mental representations of linguistically described events in both behavioural experiments (e.g., Stanfield & Zwaan, 2001) and neuroimaging studies (e.g., Zhang et al., 2022). Within behavioural paradigms, embodied cognition in language processing has been extensively investigated using

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

the sentence-picture verification task (SPVT), in which participants receive a linguistic prompt followed by a related image and judge whether the image matches the sentence. Studies using the SPVT have shown that participants simulate object properties such as shape (e.g., Hoeben Mannaert et al., 2019; Kang et al., 2020; Sato et al., 2013; Schütt et al., 2023; Zwaan et al., 2002), colour (Hoeben Mannaert et al., 2017, 2021) and orientation (e.g., Engelen et al., 2011; Stanfield & Zwaan, 2001), as well as the dynamic aspects of motion (Wang & Zhao, under review; Zwaan et al., 2004). For example, Stanfield and Zwaan (2001) found that participants responded faster to picture stimuli that matched the orientation implied by a sentence than those that did not.

However, the SPVT typically presents images at the end of a sentence, which may not adequately capture how mental simulation unfolds over time (e.g., Sato et al., 2013; Taylor & Zwaan, 2008; Zwaan & Taylor, 2006). For instance, Sato et al. (2013) investigated mental representations of object shape in Japanese, a prototypical subject-object-verb (SOV) language in which the final verb can alter the mental representation of the object shape. They conducted two experiments with different picture presentation points in a SPVT: before the final verb (Experiment 1) and after the entire sentence (Experiment 2). They found compatibility effects in both experiments in which matched images were responded to faster than the mismatched ones. More importantly, they found that mental representations shifted dynamically as participants integrated sentence cues, demonstrating the incremental nature of simulation. A recent study by Wang and Zhao (under review) investigated mental imagery during incremental motion sentence comprehension in Chinese, a subject-verb-object (SVO) language, using a self-paced reading task with schematic-diagram verification at multiple points in the sentence. Participants were faster at judging congruent diagrams over incongruent ones at the different presentation points, suggesting that the formation of mental imagery is not delayed but immediate during the incremental processing of incoming information in the sentence. Interestingly, while Sato et al. (2013) observed increasing compatibility effects as sentences progressed, Wang and Zhao (under review) reported decreasing effects, potentially highlighting linguistic or task-dependent differences in mental simulation processes.

2.2. Mental simulation in L2 processing

Despite a substantial body of research on L1 speakers' mental simulation, fewer studies have focused on L2 processing. Some studies showed reduced or absence of mental simulation in L2 speakers for features such as orientation (Koster et al., 2018), shape (Chen et al., 2020, 2024; Norman & Peleg, 2022), motion directionality (Wang & Zhao, 2024) and emotion (Foroni, 2015). These findings contrast with other studies observing that L2 learners can create native-like mental representations (e.g., Ahn & Jiang, 2018; Dudschig et al., 2014; Vukovic & Williams, 2014). For example, Ahn and Jiang (2018) demonstrated that both NSs and L2 learners of Korean exhibited sentence-picture compatibility effects for object shape and orientation, suggesting native-like mental simulation abilities among L2 learners.

Zhao et al. (2025) proposed a simulation-based model of L2 processing, building on Bergen and Chang's (2005, 2013) L1 framework. Their L1 model emphasises the role of linguistic constructions in linking phonological forms with conceptual meaning. Language understanding proceeds through three stages: constructional analysis (identifying constructions and associated meaning schemas), contextual resolution (mapping meaning to communicative context) and embodied simulation (generating inferences based on simulated scenarios). This framework emphasises how language users draw on both linguistic and extralinguistic resources to construct meaning.

Extending this model to L2 processing, Zhao et al's model posits that L2 comprehension similarly involves cognitive simulations grounded in real-world experiences, with mental models central to mastering L2. Embodied simulation is constrained by how L2 speakers comprehend and produce language. The model assumes that bilinguals activate representations from both languages at lexical, syntactic and conceptual levels. While L2-based simulation resembles L1 simulation, it is moderated by language-internal factors (i.e., L1–L2 constructions), learner-internal factors (e.g., proficiency) and learning context (e.g., immersion in the L2 environment).

The model assumes that the influence from L1 entrenchment is inevitable in L2-based simulations. By performing embodied simulation, learners analyse L2 constructions by linking forms and functions while simultaneously activating form-function mappings from their L1. As L1 is typically the dominant language in sequential bilinguals, processing will inevitably involve transfer from their L1 mental models. Cross-linguistic similarity between L1–L2 mental models facilitates nativelike L2 simulation, whereas significant dissimilarities lead to non-target-like simulation and errors in semantic or conceptual interpretation. The degree of L1 transfer is further moderated by learner factors such as L2 proficiency.

Despite the importance of language-internal factors, research on the cross-linguistic effects of L1 dis/similarity on mental simulation remains limited. One exception is Ahlberg et al. (2018), who examined how German spatial prepositions – *auf* (on), *über* (above) and *unter* (under) – were mentally simulated among L1 and L2 speakers. The prepositions *auf* and *über* exhibit a spatial distinction in the upper space, differentiating between contact (*auf*) and lack of contact (*über*), similar to the English distinction between *on* and *above*. The study compared German native speakers (NSs) with L2 groups from different L1 backgrounds: those whose L1 exhibited a similar split usage and the others whose L1 used a single form for both types of upperspace relations. The findings showed an action compatibility effect: participants responded faster when the spatial meaning of a preposition (e.g., *über* "above" or *unter* "under") matched the direction of a required motor response (e.g., upward for *über*, downward for *unter*). Importantly, these effects varied by L1 background, suggestive of L1 dis/similarity effects.

In addition to L1 influence, learner-internal and contextual variables are also important to L2 simulation (Zhao et al., 2025). While L2 proficiency has received the most attention, findings remain mixed. Some studies report significant effects of proficiency and age of onset on L2 mental simulation (Ahlberg et al., 2018; Lu & Yang, 2025), whereas others suggest that L1 entrenchment or limited access to authentic, multimodal input may constrain simulation despite high L2 proficiency (Chen et al., 2020, 2024; Norman & Peleg, 2022). Recent research highlights the importance of language exposure and immersive interaction, with greater exposure correlating positively with the magnitude of L2 embodiment (Lu & Yang, 2025). However, few studies have examined how these learner and contextual factors jointly interact with crosslinguistic influence during real-time L2 sentence processing.

Finally, the temporal dynamics of mental simulation in L2 comprehension remain underexplored. Studies on predictive language processing suggest that L2 learners often struggle to integrate linguistic cues for anticipation (e.g., Ito et al., 2016; Ito et al., 2017; Martin et al., 2013), and that predictive processing in L2 is subjected

https://doi.org/10.1017/S1366728925100801 Published online by Cambridge University Press

to the influence of L1–L2 dis/similarity (van Bergen & Flecken, 2017). Therefore, L2 learners, especially those from dissimilar L1 backgrounds, may face challenges in integrating linguistic cues necessary for mental simulation in incremental sentence processing.

2.3. Cross-linguistic variation in motion event description and processing

Talmy (1985, 2000) proposed two typological categories based on the lexicalisation pattern of motion events: satellite-framed and verb-framed languages. Satellite-framed languages, including English and Dutch, tend to express the manner of motion in the main verb, while the path is encoded in accompanying elements called satellites, such as prepositional phrases in English. Conversely, in verb-framed languages such as Japanese, the main verbs express the path of the motion, and the manner is often described using an adverbial or gerundive phrase. The differences between English (1a), Dutch (1b) and Japanese (1c) are illustrated below:

(1a) The boy [runs]$_{Manner}$ [to]$_{Path}$ school.

(1b)
De	jongen	[rende]$_{Manner}$	[naar]$_{Path}$	school.
The	boy-NOM	run-PRESENT	to	school

(1c)
Sono	shonen-wa	gakko-ni	[hasit-te]$_{Manner}$
The	boy-NOM	school-GOAL	running

[iku]$_{Path}$.
go-PRESENT

Another characteristic that distinguishes satellite-framed and verb-framed languages is the potential dual interpretation (directional or locational) of locative prepositions (e.g., *in, under*, etc.) when combined with the manner-of-motion verbs (e.g., Beavers et al., 2010; Inagaki, 2002; Levin & Rappaport Hovav, 1995; Nikitina, 2008), as illustrated in the following examples: 2a for English, 2b–2c for Dutch and 2d–2f for Japanese.

(2a) The bird flew in the cage.

(2b)
De	vogel	vloog	in	de	kooi.
The	bird	fly-PAST	in	the	cage

(2c)
De	vogel	vloog	de	kooi	in.
The	bird	fly-PAST	the	cage	in

(2d)
Sono	tori-wa	kago-no	naka-wo	ton-da
The	bird-NOM	cage-GEN	in(side)-LOC	fly-PAST

(2e)
*Sono	tori-wa	kago-no	naka-ni	ton-da
The	bird-NOM	cage-GEN	in(side)-GOAL	fly-PAST

(2f)
Sono	tori-wa	kago-no	naka-ni	ton-de	hait-ta
The	bird-NOM	cage-GEN	in-GOAL	flying	enter-PAST

The ambiguity in (2a) arises from the locative preposition *in the cage*, which can imply either a directional interpretation (the bird flies towards the cage) or a locational interpretation (the bird's motion occurs entirely within the cage) (Levin & Hovav, 1995). This ambiguity of the dual interpretation is not absolute; rather, the interpretation of directionality is context-dependent and influenced by verb choices. For example, punctual verbs (e.g., *jump*) often facilitate a directional reading (e.g., Beavers et al., 2010; Nikitina, 2008; Tutton, 2009).

The dual interpretation is observed in many satellite-framed languages (Beavers et al., 2010). Dutch shares this dual interpretation feature with English, but its syntactic characteristics are partially different. In Dutch, spatial prepositions, such as *op* (on (to)), *in* (in(to)), *achter* (behind) and *onder* (under), can be used with manner-of-motion verbs to express the locational and directional motion. The dual interpretation depends on the verb choice and varies among speakers (Den Dikken, 2010). Importantly, Dutch allows the dual interpretation of the same construction as English, and directionality can be explicitly expressed by other means, such as postposition (*vloog de kooi in*: flew the cage in), circumpositional phrase (*loopt onder de brug door*: walks under the bridge through) and morphologically complex postpositional phrase (*loopt de brug onderdoor*: walks the bridge under-through) (Den Dikken, 2010).

In contrast, Japanese restricts such constructions to locational interpretations, as in (2d), where the event of the bird flying occurs entirely within the cage. The use of the accusative case marker *-wo* licenses an interpretation of movement within a bounded space. In (2e), by contrast, the use of the dative case marker *-ni* with the locative noun phrase *kago-no naka* ("inside the cage") results in an unnatural expression under the intended directional interpretation. While *-ni* can mark either a goal or a location, its default interpretation with stative or locative noun phrases tends to be locational, not directional (Matsumoto, 2018). Therefore, when paired with a manner-of-motion verb like *tonda* ("flew"), the sentence fails to clearly convey movement into the space. Japanese typically avoids conflating manner and path in a single verb as English does (e.g., *flew into*). Instead, as shown in (2f), a directional interpretation requires an explicit path verb like *haitta* ("entered") alongside the manner verb *tonde* ("flying"), yielding a natural expression of goal-directed motion.

While these crosslinguistic patterns are well described, several issues remain underexplored in the current literature. First, although both directional and locational interpretations are licensed in satellite-framed languages, it is unclear whether native speakers exhibit a bias towards one interpretation, that is, whether a prototypical reading exists in context-neutral environments. Second, few studies have examined whether L2 learners from typologically distinct backgrounds are equally sensitive to this ambiguity or whether their interpretation is shaped by L1-specific constraints. Third, prior research has primarily relied on offline tasks, leaving it unclear how both NSs and L2 learners compute spatial meanings incrementally during real-time sentence processing. These open questions form the basis for the current study.

The typological categorisation of motion events (Slobin, 2004; Talmy, 2000) has been extensively studied under Slobin's (1996, 2000, 2003) *Thinking for Speaking* framework. According to this framework, language-specific constraints influence thought processes during speaking, writing, translating and remembering. Empirical studies have demonstrated typological differences between satellite-framed and verb-framed languages in manner/path usage in verbal production (e.g., Allen et al., 2007; Hendriks et al., 2022; Özçalışkan, 2015; Slobin, 2003), attention allocation (Gennari et al., 2002; Papafragou et al., 2008) and visual processing (e.g., Fu et al., 2024).

In L2 research, these typological differences are often explored to understand how linguistic knowledge in one language affects

https://doi.org/10.1017/S1366728925100801 Published online by Cambridge University Press

the processing of another (e.g., Jarvis, 2011; Jarvis & Pavlenko, 2008). Much of this work has investigated how L2 learners of two typologically different languages align their verbal and non-verbal behaviours with either their L1 or L2 (e.g., Brown & Gullberg, 2008, 2011, 2013; Engemann, 2022; Filipović, 2011, 2022; Hohenstein et al., 2006; Kamenetski et al., 2022; Konishi et al., 2014; Lai et al., 2014; Park, 2020). While some studies have distinguished between inter-typological influences (transfer effects between typologically different languages, for example, verb-framed versus satellite-framed) and intra-typological effects (difference within the same typological group) (e.g., Lewandowski & Özçalışkan, 2019), the current study focuses specifically on inter-typological influences, particularly how L1–L2 dis/similarity may shape motion event construal in an L2. This shift in focus allows us to extend previous findings on crosslinguistic differences in manner/path encoding to the domain of locative prepositions, where L2 learners must contend with potential ambiguities in spatial interpretation.

Despite this progress, only a handful of studies have specifically addressed whether L2 learners from verb-framed L1s (e.g., Japanese) interpret English locative prepositions with manner-of-motion verbs in a native-like fashion. Existing studies were conducted using corpus analyses (Nikitina, 2008; Tutton, 2009) and offline measures such as sentence-picture matching tasks (Inagaki, 2002) and translation tasks (Kong, 2021). For example, Inagaki (2002) compared 35 intermediate-level L1 Japanese learners of English with English NSs. Using an offline sentence-picture matching task, Inagaki found that Japanese learners persistently favoured locational interpretations, whereas NSs accepted both directional and locational readings. These findings highlight the challenges L2 learners face when their L1 lacks the dual mental representation found in satellite-framed languages. However, these studies have not systematically examined whether one interpretation is more prototypical than the other, nor have they explored the real-time processing mechanisms underpinning these interpretations in L2 speakers. These gaps motivate the current study.

3. The present study

As seen above, the formation of mental representations in incremental sentence processing in L2 contexts, particularly the influence of cross-linguistic L1–L2 dis/similarity, is a complex and underexplored phenomenon. Previous studies on L2 processing have observed cross-linguistic effects on motor-related simulation (Ahlberg et al., 2018) and predictive eye movement (van Bergen & Flecken, 2017). Despite these limited findings, L1–L2 cross-linguistic effects on mental simulation remain largely unexplored.

The simulation-based model of L2 comprehension (Zhao et al., 2025) highlights the significant role of various learner factors in achieving nativelike mental simulation. Yet, existing studies have primarily focused on L2 proficiency, overlooking other critical factors such as the onset age of L2 acquisition and the length of immersion in the target language environment. These variables are well-documented predictors of L2 success in the second language acquisition (SLA) literature (e.g., Foster et al., 2013; Kotz & Elston-Güttler, 2007; Park, 2020). The present study aims to address these important gaps by investigating how L1–L2 dis/similarity affects the dynamic creation of semantic interpretations of English locative prepositions during incremental L2 sentence processing. Specifically, we addressed the following research questions (RQs):

RQ1: Are the locational and directional interpretations of English locative prepositions equally acceptable to native English speakers?

RQ2: How does cross-linguistic variation in the interpretation of locative prepositions influence the formation of mental representations during incremental sentence processing among adult L1 and L2 English speakers?

RQ3: How are L2 learners' mental representations influenced by learner-specific factors, including L2 proficiency, onset age of L2 acquisition and length of immersion?

To address RQ1, we conducted a norming study to determine whether English NSs exhibit a preferential bias towards locational or directional interpretations. The norming study also ensured that the stimuli used in the main experiment were capable of eliciting both locational and directional interpretations from English NSs. Additionally, the norming study assessed the event plausibility of the sentence stimuli to control for potential (im)plausibility effects on reaction times in the main experiment. In the following sections, we first report the method and results of the norming study. Then, we present the methodology and findings of the main experiment on L1 and L2 speakers of English (Dutch and Japanese L1s), which involved a self-paced reading task interleaved with video verification.

4. The norming study

4.1. Participants

We recruited monolingual English speakers via Prolific, an online participant recruitment platform. Eligibility criteria specified that participants must be native speakers (NSs) of English with limited or no knowledge of a second language. To ensure linguistic and cultural homogeneity, the participants' countries of residence were restricted to major English-speaking regions, including the United States, the United Kingdom, Canada and Australia. Twenty-nine English NSs initially participated. Two participants were removed from the analysis: one reported knowledge of a second language beyond the eligibility criteria, and the other participant took an exceptionally long time to complete the task. The final sample included 27 monolingual English speakers (12 female, 14 male and 1 non-binary), aged between 21 and 60 years (*mean* = 38.44, *SD* = 11.55).

4.2. Materials

Five prepositions (*in*, *on*, *under*, *behind* and *above*) were used to create sentence stimuli. Each preposition was used to create eight sentences featuring the English locative construction and manner-of-motion verbs (e.g., *run*, *walk* and *swim*), resulting in a total of 40 sentences, all of which could potentially invoke dual interpretations. The full list of the sentence stimuli is presented in Appendix A of the Supplementary Material.

Each sentence was paired with two videos: one depicting the locational interpretation and the other the directional interpretation. This resulted in a total of 80 videos. The videos consisted of black-and-white line drawings, with some videos incorporating

https://doi.org/10.1017/S1366728925100801 Published online by Cambridge University Press

grey and light blue to enhance the visibility of objects for participants. Each video has a duration of 5 seconds and a resolution of 750 px (width) × 450 px (height). Following Talmy's (2000) terminology, in the locational videos, the entire motion of the *Figure* (a moving object) occurs within the property of the *Ground* (a reference object) as indicated by the prepositional phrase. In the directional videos, the motion of the Figure begins near the Ground and then reaches it. The Figure arrives at the Ground 3500 ms after the onset of the video. The relevant components of motion (Figure, Ground and Manner) are all visible from the first frame of the video. For the prepositions *behind* and *under*, semi-transparent or relatively smaller objects, such as a fence or bush, were used so that the scene could be represented in a two-dimensional video. Example snapshots of directional and locational videos corresponding to the sentence stimulus *The ball rolled under the car* are given in Figure 1 (A and B).

4.3. Procedure

In this norming study, a sentence rating task was employed, in which the participants rated the sentence-video matching on a 6-point Likert scale (1 = completely mismatching; 2 = very mismatching; 3 = somewhat mismatching; 4 = somewhat matching; 5 = very matching; 6 = completely matching). For each video, participants were also asked to judge event plausibility on a 6-point Likert scale (1 = completely implausible; 2 = very implausible; 3 = somewhat implausible; 4 = somewhat plausible; 5 = very plausible; 6 = completely plausible). The participants were presented with one video and one sentence on each screen and rated all 80 sentence-video pairs (locational/directional videos × 40 sentences). The trial sequence was randomised for each participant.

The task was administered online using Qualtrics (Provo, UT, USA, version 5.2024). The participants first filled out their demographic information (e.g., age, gender and additional languages). Then they proceeded to the sentence-video matching norming task, which took approximately 25 minutes to complete.

4.4. Data analysis

Data were analysed using R (version 4.3.3, R Core Team, 2024). First, to ensure the reliability of the rating data of sentence-video matching and event plausibility, the split-half reliability was calculated using the Spearman–Brown formula with correction for length.

The participants' event plausibility ratings were *z*-transformed. The mean scores of the *z*-transformed event plausibility score were calculated for each item in each video condition for data screening to control for potential plausibility effects in the main experiment. Items with mean plausibility scores less than −1.0 were removed from the data analysis of the experiment to exclude semantically implausible stimuli that might otherwise compromise the reliability of participants' judgment.

Participants' sentence-video matching rating scores were analysed using the *lme4* package (version 1.1–35.3; Bates et al., 2024) and *lmerTest* package (version 3.1–3; Kuznetsova et al., 2017). A linear fixed effects model was constructed to examine the influence of the fixed effect of Interpretation (Directional/Locational) and the random effects of Sentence and Participant. The *emmeans* package (version 1.10.1; Lenth et al., 2024) was used for a post-hoc analysis. Data visualisation was performed using the *ggplot2* package (version 3.5.0, Wickham, 2016).

4.5. Results

The results of the split-half reliability showed high reliability for both the directional plausibility ratings (r_{sb} = .90) and the directional sentence-video matching ratings (r_{sb} = .84). Similarly, the locational plausibility ratings (r_{sb} = .90) and locational sentence-video matching ratings (r_{sb} = .84) demonstrated comparable reliability. The descriptive statistics of the mean event plausibility score of each item in each video condition are presented in Appendix A of the Supplementary Material.

The results of the linear mixed effects model revealed a significant effect of Interpretation on sentence-video matching ratings, $t(2155) = 2.70$, $p = 0.007$. A post-hoc analysis revealed that locational videos (M = 5.4, SE = 0.1, 95% CI [5.13, 5.63]) received significantly higher ratings than directional videos (M = 5.3, SE = 0.1, 95% CI [5.04, 5.54]) (see Figure A in the Supplementary Material). These findings suggest that when context cues, such as accompanying sentence structures and matching visual stimuli, are explicitly provided, both directional and locational interpretations are highly acceptable. More importantly, NSs strongly favoured the locational interpretation over the directional interpretation, suggesting that the locational interpretation represents the prototypical meaning for locative prepositions. The findings align with and support the argument that dual interpretation is possible when context cues for the directional interpretation are available (e.g., Beavers et al., 2010; Nikitina, 2008; Tutton, 2009).

5. The main experiment

The aim of this experiment was to investigate how mental representations of the directional/locational interpretations are formed incrementally by NSs and L2 learners from similar (Dutch) and

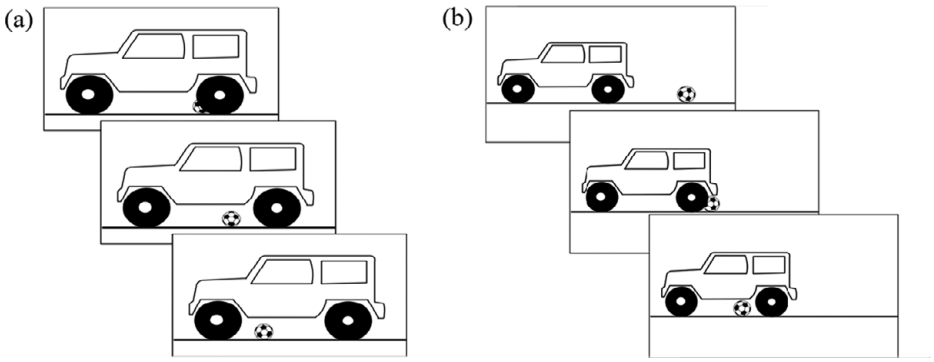


Figure 1. (A) Snapshot of a locational video. (B) Snapshot of a directional video.

https://doi.org/10.1017/S1366728925100801 Published online by Cambridge University Press

dissimilar (Japanese) L1 backgrounds. Based on the norming study results, it was hypothesised that the locational interpretation, being more frequently selected and highly acceptable, would be formed first, leading to faster reaction times (RTs) for locational videos than directional videos at both video presentation points. Based on previous studies on ambiguity resolution and mental representation (e.g., Roberts & Liszka, 2019; Sato et al., 2013), the directional interpretation was expected to require reanalysis, resulting in delayed verification.

Regarding the locational interpretation, the three L1 groups were predicted to show no difference due to the well-established experiential traces of this interpretation in their respective L1s. However, cross-linguistic L1–L2 dis/similarity was expected to influence RTs for the directional interpretation. Specifically, L1-Japanese participants were predicted to take longer than both NSs and L1-Dutch participants due to the limited experiential association of locative prepositions with the directional interpretation in their L1. No significant differences were anticipated between NSs and L1-Dutch participants, given their typological similarity. Likewise, for acceptance rates (ARs), no differences were expected for locational interpretations across the groups. However, L1-Japanese participants were predicted to show lower acceptance rates for the directional interpretation due to the restrictions in Japanese that limit this interpretation. In addition to these cross-linguistic predictions, we hypothesised that learners with higher L2 proficiency, earlier age of L2 acquisition (AoA) and longer immersion in L2-speaking environments (LoR) would exhibit more native-like representations. They were expected to show higher acceptance rates and shorter RTs for the less prototypical directional interpretation, reflecting enhanced restructuring of L2-specific representations.

5.1. Participants

A total of 144 participants participated in this experiment from three L1 groups: English, Dutch and Japanese. Two NSs and four L1 Japanese learners were removed from the analysis based on the data screening criteria of 70% accuracy rate of the comprehension questions. The resulting sample included data from 138 participants. The baseline group consisted of 44 NSs of English (19 males and 25 females) with an average age of 38.2 (*SD* = 11.3). The L1 Dutch group consisted of 44 participants (25 males and 19 females) with an average age of 33.1 (*SD* = 10.8). They reported that they started learning English at the age (AoA) of 1 to 14 (*mean* = 8.8, *SD* = 3.7). Their length of residence in English-speaking countries (LoR) ranged from 0 to 21 years (*mean* = 2.31, *SD* = 5.1). The L1 Japanese group consisted of 50 participants (16 males, 33 females and 1 other) with an average age of 32.1 (*SD* = 9.8). Their AoA ranged from 1 to 18 (*mean* = 9.9, *SD* = 4.0; three participants did not report this), and their LoR were from 0 to 36 years (*mean* = 8.06, *SD* = 11.04).

Participants were recruited via different platforms depending on availability. English NSs and Dutch learners of English were recruited through Prolific using eligibility criteria to ensure native proficiency and residence in relevant language communities. Due to the limited availability of Japanese participants on Prolific, most were recruited via social media and invited directly to the experiment. The same eligibility criteria were applied across all recruitment platforms. No systematic differences were observed in comprehension accuracy rates or dropout rates across groups recruited from different sources. The study adhered to APA ethical standards and received approval from the University of Melbourne Human Research Ethics Committee (ID: 2024-29010-52416-3). All participants received monetary compensation for their participation.

For L2 learners, English proficiency was measured using a cloze test adopted from a practice set of the Michigan Examination for the Certificate of Proficiency in English (ECPE). This cloze test was selected because it taps into learners' grammatical and vocabulary knowledge and has been shown in prior studies to effectively differentiate English proficiency levels among L1 speakers, ESL and EFL learners (Domazetoska & Zhao, 2025). The material of the cloze test is provided in Appendix B of the Supplementary Material. Each item was scored dichotomously, with 1 point for a correct response and 0 point for an incorrect response, yielding a total score ranging from 0 to 20. For the L1 Japanese group, the mean score was 12.6 (*SD* = 3.9), with scores ranging from 4 to 20. For the L1 Dutch group, the mean score was 16.5 (*SD* = 1.8), with scores ranging from 10 to 20.

5.2. Material and design

The present study employed a factorial design with response times (RTs) and acceptance rates in video verification as the dependent variables. The independent variables of interest were L1 (English/Japanese/Dutch), Interpretation (Locational/Directional) and Sequence (Post-preposition/Post-sentence video presentation). L1 was a between-subjects variables, whereas Interpretation and Sequence were within-subjects variables. Each participant completed two counterbalanced blocks of trials. In the Post-preposition block, the video appeared after the fourth word (the preposition in target trials). In the Post-sentence block, the video was presented after the participant read the entire sentence.

Each block comprised 40 trials: 20 target trials and 20 filler trials. The 20 target trials (10 directional and 10 locational videos) tested the interpretation of locative prepositions by presenting ambiguous motion sentences (e.g., *The bird flew in the cage*) followed by a video showing either a locational or directional interpretation. These target stimuli were identical to those used in the norming study and directly targeted the dual interpretation under investigation.

The 20 filler trials served to reduce participants' strategic attention to the target manipulation and to balance the number of "yes" and "no" responses as much as possible. Of the 20 fillers, 8 were sentence-video matching trials and 12 were sentence-video mismatching trials. All filler stimuli depicted unambiguous motion or action events that did not rely on prepositional ambiguity. The matching fillers described events such as *The pen fell from the desk* or *The airplane is flying higher than the helicopter*, and the accompanying video accurately depicted the described scenario. In the mismatching fillers, the video diverged from the sentence in some specific respect, such as subject/object noun phrase, manner, direction, or outcome (e.g., *The bird landed on the ground* paired with a video of a bird flying in the sky). A full list of these filler sentences and corresponding video descriptions is provided in Appendix C of the Supplementary Material. Therefore, participants verified a total of 80 videos (40 trials × 2 sequences) over the course of the experiment.

To mitigate potential bias towards habitual "yes" responses, particularly among English and Dutch participants who are likely to accept both locational and directional interpretations due to the dual interpretations in their L1s, we included more mismatching filler trials than matching ones. This yielded a response pattern in which English and Dutch participants were expected to produce more "yes" responses overall, while Japanese participants whose L1 lacks directional interpretations of locative prepositions were anticipated to reject directional targets, resulting in a more balanced or even reversed "yes/no" ratio. Rather than artificially balancing "yes" and "no" trials across language groups, we opted for a consistent design that preserved the interpretive contrast central to our

https://doi.org/10.1017/S1366728925100801 Published online by Cambridge University Press

research question. This approach allowed us to maintain experimental control while encouraging all participants to engage in careful semantic evaluation of the sentence-video pairs.

5.3. Self-paced reading embedded with video verification

A self-paced reading task interleaved with video verification was created using jsPsych (de Leeuw, 2023). For the self-paced reading part, a moving window paradigm was employed, in which participants pressed the spacebar repeatedly to read a sentence one word at a time. Participants were instructed to judge as quickly and accurately as possible whether a video, inserted at either the post-preposition or post-sentence position, matched the meaning of the sentence. Responses were recorded by pressing "f" (left hand) for sentence-video matches and "j" (right hand) for mismatches. This key mapping was applied uniformly across all participants and is consistent with common practice in psycholinguistic experiments (e.g., Hammerly, Staub & Dillon, 2019).

Sample illustrations of the task procedure are shown in Figure 2 (A and B). Each trial began with a fixation cross indicating the

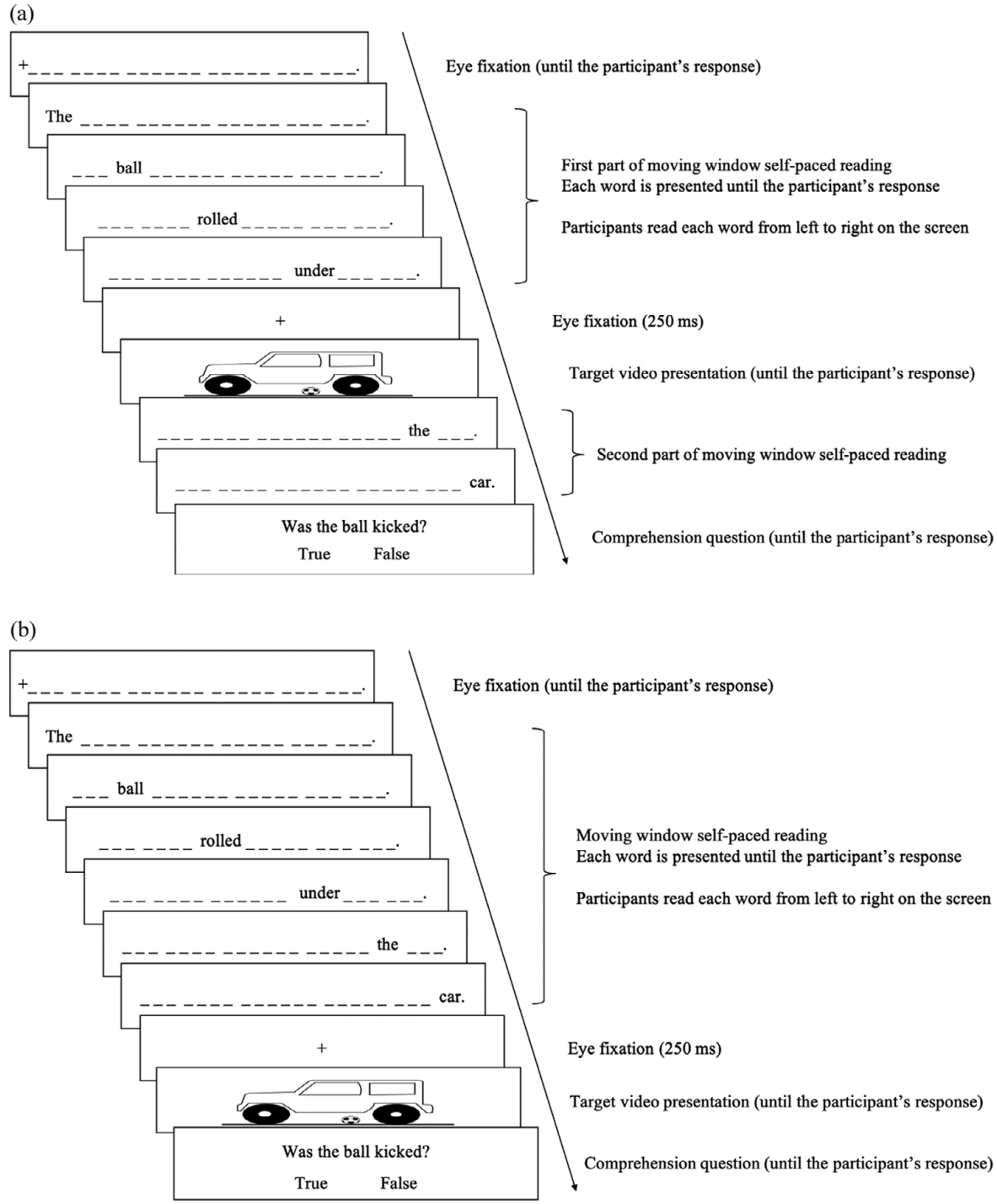


Figure 2. (A) Task procedure of post-preposition video verification. (B) Task procedure of post-sentence video verification.

https://doi.org/10.1017/S1366728925100801 Published online by Cambridge University Press

starting point of the first word. Sentential components were presented word by word, revealed by spacebar pressing. A video appeared after the fourth word (Post-preposition) or at the end of the sentence (Post-sentence). Before the video display, a fixation cross was shown at the centre of the screen for 250 ms to direct participants' attention. At the end of each trial (after the video verification in the Post-sentence video presentation), participants completed a comprehension check to ensure they processed the meaning of the sentence. The participants responded by clicking the "true" or "false" button on the screen. The RTs of video verification were recorded from the onset of the video display to the time the participant made the initial judgment response.

5.4. Procedure

In the experiment, participants first filled out the demographic survey. Upon completing the questionnaire, they proceeded to the English proficiency test and then the self-paced video verification task. At the instruction phase, the participants were instructed to read the sentence word by word and judge the video as quickly and accurately as possible. Before the target trials, they were first given six practice trials to familiarise themselves with the task in each block. After each practice trial, they were given correctness feedback on their video verification. The experiment took approximately 25 minutes for NSs and 30 minutes for L2 learners.

5.5. Data analysis

The software and packages used for data analysis were identical to those reported in the norming study. After removing the practice trials, data screening was conducted based on the accuracy of the comprehension questions. Six participants with comprehension accuracy below 70% were removed. In addition, RTs shorter than 300 ms and longer than 9000 ms were excluded from the analysis. For items, based on the results of event plausibility ratings in the norming study, sentence-video stimuli with z-transformed mean scores lower than −1.0 were excluded from the data analysis, resulting in the removal of six sentence-video stimuli. The excluded items are indicated by an asterisk in Appendix A of the Supplementary Material.

A treatment-coding scheme was applied to the categorical variables: L1 (English/Dutch/Japanese), Interpretation (Directional/Locational) and Sequence (Post-Preposition/Post-Sentence), with the first level of each variable as the reference. Since all sentence-video trials were designed to be interpretable as matching (either locationally or directionally), we analysed participants' acceptance rates and RTs, rather than accuracy. Trials in which participants judged the sentence and video as a "match" were coded as 1 (acceptance), and those judged as a "mismatch" were coded as 0 (rejection). Acceptance rates were analysed separately based on all video verification responses. For the RT analysis, only log-transformed RTs from trials in which participants accepted the video as matching the sentence were included[1].

The full models included the fixed effects of L1, Interpretation, Sequence and log-transformed Cloze test scores, and the interaction between them. Participant and Sentence were included as random effects. A generalised linear mixed-effects model (GLMM) with the logit link function was constructed for acceptance rate analysis, and a linear mixed-effects model was used for RT analysis. The full generalised linear mixed-effects model[2] failed to converge. Therefore, a model comparison was carried out with a cloze score as a covariate (Model 1)[3] and without a cloze score (Model 2)[4] using the ANOVA function. Since Model 1 showed a significantly better fit than Model 2 (Model 1: $\chi^2 = 8.58$, $p = 0.003$, Cohen's $d = 0.12$, $SE = 0.04$, 95% CI [0.02, 0.20], corresponding to a small effect size), subsequent analyses were based on Model 1. For the RT analysis, event plausibility ratings of the sentence stimuli obtained from the norming study were added as an additional covariate, which increased model fit and helped control for potential confounding effects of stimulus plausibility. Including cloze score in both models also served to control for the influence of L2 proficiency. All predictors of theoretical relevance were retained in the model to preserve model hierarchy, even if their main effects or lower-order interactions were not significant. This approach ensures that higher-order interactions are properly estimated and interpretable. The *emmeans* package (version 1.10.1; Lenth et al., 2024) was used for post hoc analyses.

In addition, for L2 learners, separate models for log-transformed Length of Residence in English-speaking countries (LoR) and log-transformed Age of L2 Acquisition (AoA) were constructed. These models included interactions with L1, Interpretation and Sequence to investigate the effects of learner factors on RTs and acceptance rates. Since some participants reported 0 for LoR, it was augmented by 1 before log-transformation to avoid undefined values.

To assess the sensitivity of the design, post hoc power analyses were conducted in G*Power (Faul et al., 2007). The analysis indicates that the study had 80% power to detect small within–between interactions across all participants ($f \approx 0.11$; $\eta^2p \approx .012$) and small-to-moderate incremental effects of the learner-factor three-way among L2 learners ($f^2 \approx .033$; partial $R^2 \approx .031$). Finally, effect sizes were calculated using Cohen's d (1977), with interpretation following Plonsky and Oswald's (2014) guidelines: for between-group contrasts, small (.40), medium (.70) and large (1.00); for within-group contrasts, small (.60), medium (1.00) and large (1.40).

6. Results

6.1. Descriptive statistics of reaction times and acceptance rates

The descriptive statistics of reaction times (RTs) of acceptance responses and of acceptance rates are presented in Table 1. Overall, the RTs for the locational videos were faster than the directional videos among the three L1 groups. Both the locational and directional videos were accepted at high rates, with English and Dutch speakers showing higher acceptance for the directional videos.

[1]For completeness, an additional RT analysis was conducted on rejection trials using the same model structure as the analysis on acceptance trials (log (RT) ~ L1*Interpretation*Sequence*log(Cloze)). However, no significant main or interaction effects were found. Given that rejection responses comprised less than 10% of the data and were unevenly distributed across participants and conditions, these results should be interpreted with caution. The main RT analysis therefore focused on acceptance trials, which provided a more reliable and representative dataset for evaluating processing differences.

[2]Model equation: response ~ L1*Interpretation*Sequence*log(Cloze) + (1|Participant) + (1|Sentence)

[3]Model equation: response ~ L1*Interpretation*Sequence + log(Cloze) + (1|Participant) + (1|Sentence)

[4]Model equation: response ~ L1*Interpretation*Sequence + (1|Participant) + (1|Sentence)

https://doi.org/10.1017/S1366728925100801 Published online by Cambridge University Press

Table 1. Descriptive statistics of RTs and acceptance rates in the sentence-video verification task

L1 group	Interpretation	Mean RTs (SD)	Mean acceptance rate (SD)
Dutch	Directional	3140 (1471)	0.97 (0.18)
	Locational	2250 (1284)	0.94 (0.24)
English	Directional	3286 (1528)	0.98 (0.19)
	Locational	2432 (1317)	0.95 (0.22)
Japanese	Directional	3743 (1954)	0.93 (0.26)
	Locational	2809 (3234)	0.93 (0.25)

Table 2. Results of reaction time analysis

Model term	df1	df2	F-ratio	χ^2	*p*-value
L1	2	∞	2.975	5.950	0.051
Interpretation	1	∞	735.095	735.095	<0.001
Sequence	1	∞	3.483	3.483	0.062
Cloze	1	∞	0.040	0.040	0.841
plausibility	1	∞	3.688	3.688	0.055
L1 × interpretation	2	∞	6.359	12.718	0.002
L1 × sequence	2	∞	1.492	2.984	0.225
L1 × cloze	2	∞	1.047	2.094	0.351
Interpretation × sequence	1	∞	35.230	35.230	<0.001
Interpretation × cloze	1	∞	0.087	0.087	0.768
Sequence × cloze	1	∞	0.056	0.056	0.814
L1 × interpretation × sequence	2	∞	0.072	0.144	0.931
L1 × interpretation × cloze	2	∞	8.926	17.852	<0.001
L1 × sequence × cloze	2	∞	0.957	1.914	0.384
Interpretation × sequence × cloze	1	∞	1.271	1.271	0.260
L1 × interpretation × sequence × cloze	2	∞	0.259	0.518	0.772

6.2. Results of reaction times of video verifications

According to the linear model[5], there was a significant main effect of Interpretation ($F = 735.095$, $p < 0.001$), such that the directional videos were verified more slowly than the locational videos ($\beta_{\text{Directional-Locational}} = 0.36$, $SE = 0.01$, 95% CI [0.33, 0.38], $z = 27.24$, $p < 0.001$; Cohen's $d = 0.98$, $SE = 0.04$, 95% CI [0.91, 1.05], corresponding to a small to medium effect size). Also, there was a significant interaction between L1 and Interpretation ($F = 6.359$, $p = 0.002$). Table 2 presents the statistical outputs of the linear model on the RT analysis. Figure 3 shows the RTs for the directional and locational videos by three L1 groups.

Specifically, the post hoc analysis revealed that the mean RTs of English speakers' verification of the locational videos ($M = 2058$, $SE = 130$, 95% CI [1787, 2371]) were estimated to be 726 ms shorter than the directional videos ($M = 2784$, $SE = 176$, 95% CI [2416, 3207]). Dutch speakers' mean verification RTs for locational videos ($M = 1956$, $SE = 110$, 95% CI [1725, 2219]) were 948 ms shorter than those for the directional videos ($M = 2904$, $SE = 164$, 95% CI [560, 3294]). Japanese speakers responded to the locational videos ($M = 2327$, $SE = 137$, 95% CI [2039, 2655]) 1004 ms faster than the directional videos ($M = 3331$, $SE = 197$, 95% CI [2919, 3801]). The pairwise comparison between L1-English and L2 speakers showed no statistically significant difference in the RTs for the locational videos ($\beta_{\text{English-Dutch}} = 0.07$, $p = 0.39$; $\beta_{\text{English-Japanese}} = -0.16$, $p = 0.14$). However, there was a significant difference in the RTs for the directional videos between L1-English and L1-Japanese participants ($\beta_{\text{English-Japanese}} = -0.19$, $SE = 0.08$, $z = -2.35$, $p = 0.02$, Cohen's $d = 0.50$, $SE = 0.22$, 95% CI [0.08, 0.93], corresponding to a small to medium effect size), but not between L1-English and L1-Dutch speakers ($\beta_{\text{English-Dutch}} = -0.05$, $SE = 0.08$, $z = -0.60$, $p = 0.55$), indicating the cross-linguistic influence of L1 dis/similarity effect on the mental imagery. For the Dutch-Japanese comparisons, the post hoc analysis shows a marginally significant difference for the directional interpretation ($\beta_{\text{Dutch-Japanese}} = -0.14$, $SE = 0.07$, $z = -1.90$, $p = 0.06$) and a significant difference in locational interpretation ($\beta_{\text{Dutch-Japanese}} = -0.18$, $SE = 0.07$, $z = -2.48$, $p = 0.01$).

In addition, a significant interaction between Interpretation and Sequence ($F = 35.230$, $p < 0.001$) revealed that the RT difference in the directional video and locational video verifications was significantly larger in the post-sentence sequence ($\beta_{\text{directional-locational}} = 0.44$, $SE = 0.02$, $z = 23.52$, $p < 0.0001$) than the post-preposition sequence ($\beta_{\text{directional-locational}} = 0.28$, $SE = 0.02$, $z = 15.14$, $p < 0.0001$).

Furthermore, the model yielded a significant three-way interaction between L1, Interpretation and Cloze ($F = 8.926$, $p < 0.001$). But the post hoc test did not yield any significant results. The three-way interaction between L1, Interpretation and Sequence was not statistically significant ($F = 0.072$, $p = 0.931$).

In the model built for investigating the effect of AoA[6], results (see Table 3) showed a significant two-way interaction between Interpretation and AoA ($F = 5.76$, $p = 0.016$). The post hoc analysis on the two-way interaction showed a significant effect of AoA on the RTs for the locational videos ($\beta = -0.01$, $SE = 0.005$, $z = 2.36$, $p = 0.018$), but not the directional videos ($\beta = 0.007$, $SE = 0.005$, $z = 1.35$, $p = 0.178$). Furthermore, there was a three-way interaction between L1, Interpretation and AoA ($F = 9.30$, $p = 0.002$). Post hoc analysis revealed a marginal (non-significant) trend suggesting that AoA may influence L1-Dutch learners' directional video verification ($z = 1.780$, $p = 0.07$), such that later AoA was associated with longer RTs (see Figure 4). No such trend was observed in the L1-Japanese group. Although this pattern was not statistically significant, it may indicate a potential modulation of simulation-based processing by AoA in a way that is constrained by L1-specific representational affordances. The model for examining the effect of L2 learners' LoR[7] did not yield any significant results.

6.3. Results of acceptance rates

The generalised linear mixed effects model[8] revealed a significant influence of Interpretation ($F = 4.1$, $p = 0.043$). Table 4 presents the model outputs. The post hoc analysis revealed that the directional videos had a significantly higher acceptance rate than that of the

[5]Model equation: log(RT) ~ L1*Interpretation*Sequence*log(Cloze) + plausibility + (1|Participant) + (1|Sentence)

[6]Model equation: log(RT) ~ L1*Interpretation*Sequence*log(AoA) + (1|Participant) + (1|Sentence)

[7]Model equation: log(RT) ~ L1*Interpretation*Sequence*log(LoR + 1) + (1|Participant) + (1|Sentence)

[8]Model equation: response ~ L1*Interpretation*Sequence + log(Cloze) + (1|Participant) + (1|Sentence)

https://doi.org/10.1017/S1366728925100801 Published online by Cambridge University Press

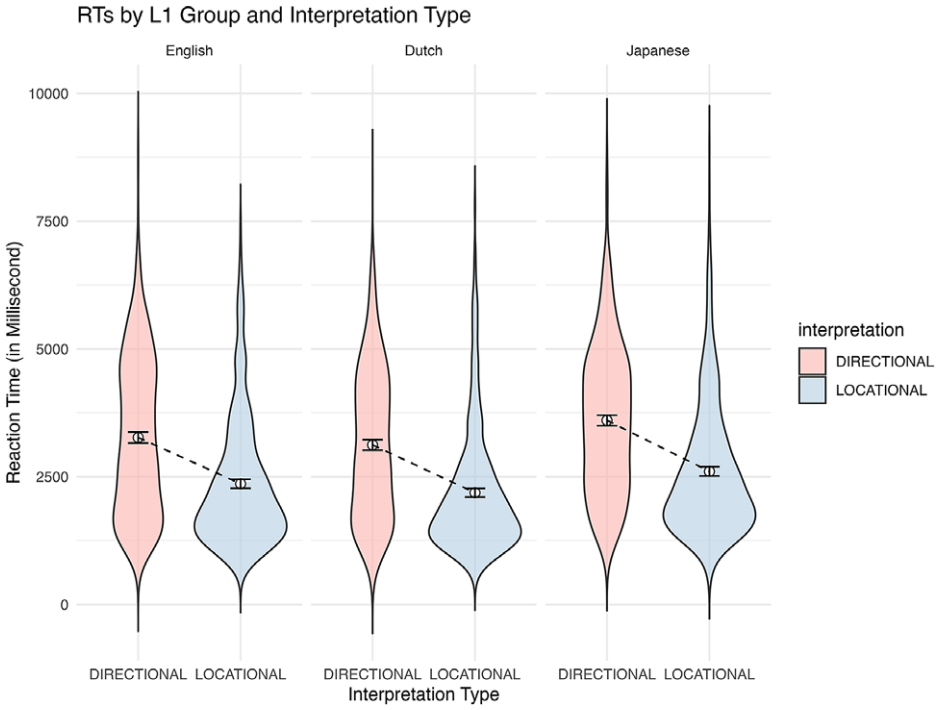


Figure 3. Violin plot of RTs by L1 groups and interpretation types.

Table 3. Results of L2 AoA effects on reaction times

Model term	df1	df2	F-ratio	χ^2	*p*-value
L1	1	∞	6.727	6.727	0.009
Interpretation	1	∞	675.42	675.42	<0.001
Sequence	1	∞	10.259	10.259	0.001
AoA	1	∞	3.586	3.586	0.058
L1 × interpretation	1	∞	0.799	0.799	0.371
L1 × sequence	1	∞	3.623	3.623	0.057
L1 × AoA	1	∞	0.359	0.359	0.549
Interpretation × sequence	1	∞	32.848	32.848	<0.001
Interpretation × AoA	1	∞	5.763	5.763	0.016
Sequence × AoA	1	∞	0.843	0.843	0.359
L1 × interpretation × sequence	1	∞	0.421	0.421	0.517
L1 × interpretation × AoA	1	∞	9.298	9.298	0.002
L1 × sequence × AoA	1	∞	1.206	1.206	0.272
Interpretation × sequence × AoA	1	∞	0.069	0.069	0.793
L1 × interpretation × sequence × AoA	1	∞	0.211	0.211	0.646

locational videos ($\beta_{\text{directional-locational}} = 0.31$, $SE = 0.16$, 95% CI [0.01, 0.62], $z = 2.03$, $p = 0.04$). Sequence was not a significant predictor.

There were significant interactions between L1 and Interpretation ($F = 6.2$, $p = 0.002$) (illustrated by Figure 5). The pairwise comparison between L1 groups revealed a significant difference in the acceptance of the directional videos between English and Japanese speakers ($\beta_{\text{English-Japanese}} = 1.01$, $SE = 0.34$, $z = 2.95$, $p = 0.003$; Cohen's $d = 1.01$, $SE = 0.34$, 95% CI [0.34, 1.68], corresponding to a large effect size), but not between English and Dutch speakers ($\beta_{\text{English-Dutch}} = 0.38$, $SE = 0.36$, $z = 1.07$, $p = 0.29$), suggesting the difference deriving from the L1 dis/similarity. In support of this, there was no such significant difference regarding the locational videos, which three L1 groups have in common, between English and Japanese speakers ($\beta_{\text{English-Japanese}} = -0.04$, $SE = 0.29$, $z = -0.14$, $p = 0.89$) and between English and Dutch speakers ($\beta_{\text{English-Dutch}} = 0.31$, $SE = 0.26$, $z = 1.16$, $p = 0.25$). The interaction between L1 and Sequence is statistically significant, but it is neither theoretically relevant nor meaningful without considering their relationship with the Interpretation variable.

Also, Cloze was found to be a statistically significant covariate ($F = 9.29$, $p = 0.002$). The post hoc analysis suggested that participants with a higher cloze test score showed a higher acceptance in the video verification ($z = 3.05$, $p = 0.002$).

In the model for investigating the L2 learner effect of AoA[9], the results (Table 5) showed a significant main effect of AoA ($F = 5.78$, $p = 0.016$). The post hoc analysis suggested that participants with a later AoA scored lower in terms of acceptance ($z = -2.40$, $p = 0.02$). Furthermore, AoA showed a marginally significant interaction with Interpretation and Sequence ($F = 2.78$, $p = 0.09$). Pairwise comparisons showed a significant AoA influence on acceptance only in the directional interpretation in the post-sentence sequence ($z = -2.254$, $p = 0.02$), and not in the other interpretation and sequence.

The model on the effect of LoR[10] showed a significant three-way interaction between L1, Interpretation and LoR ($F = 10.47$, $p = 0.001$). The model outputs are presented in Supplementary Table A. The post hoc analysis showed significant effects of LoR on the acceptance of directional videos by both Dutch learners ($z = 2.73$,

[9]Model equation: response ~ L1*Interpretation*Sequence *log(AoA) + (1|Participant) + (1|Sentence)

[10]Model equation: response ~ L1*Interpretation*Sequence*log(LoR) + (1|Participant) + (1|Sentence)

https://doi.org/10.1017/S1366728925100801 Published online by Cambridge University Press

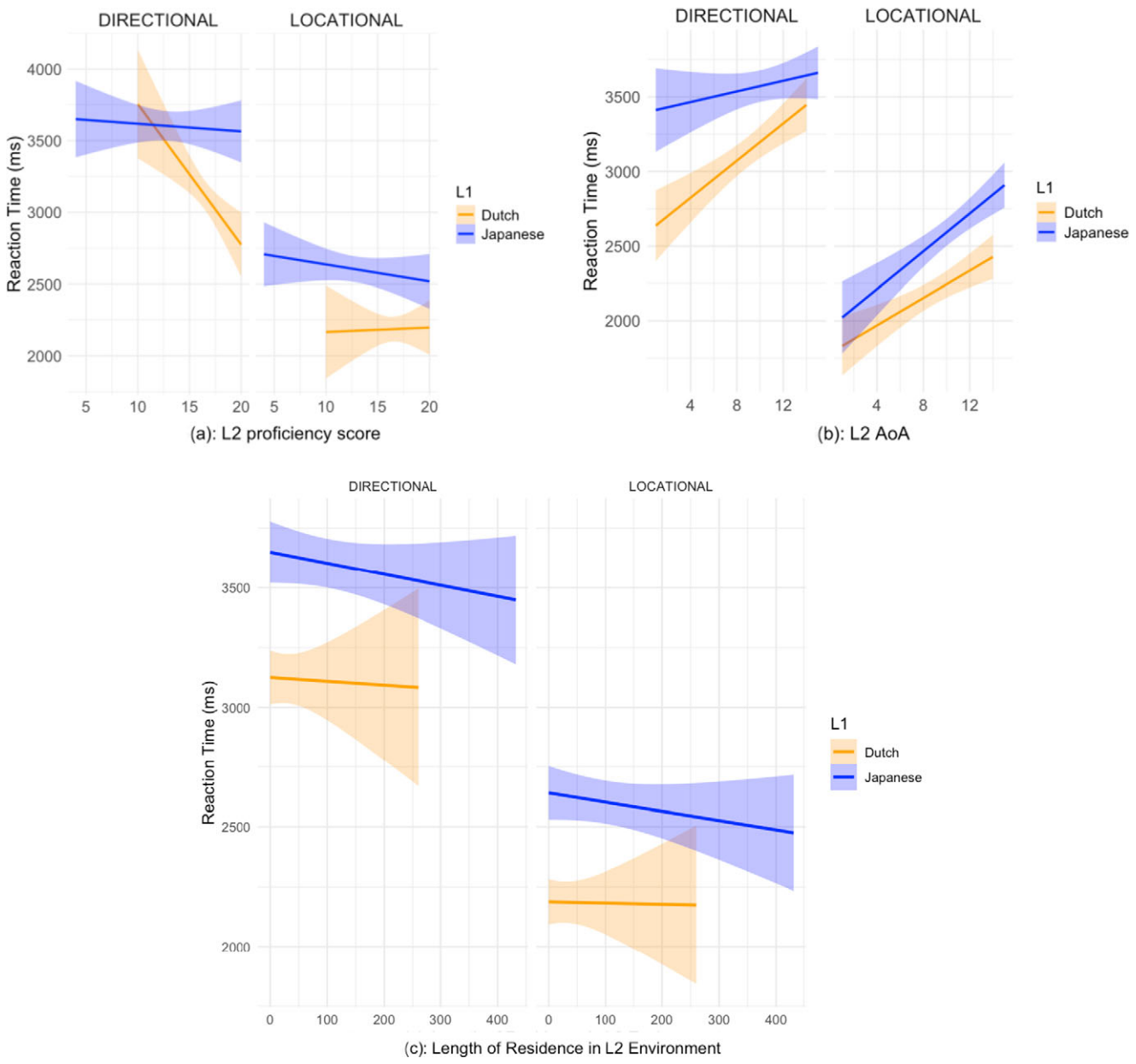


Figure 4. Relationship between reaction time and individual difference variables for the Dutch and Japanese speakers: (A) L2 proficiency, (B) L2 AoA and (C) LoR, by L1 group and interpretation type.

Table 4. Results of acceptance rate analysis

Model term	df1	df2	F-ratio	χ^2	*p*-value
L1	2	∞	1.857	3.714	0.156
Interpretation	1	∞	4.102	4.102	0.043
Sequence	1	∞	0.396	0.396	0.529
Cloze	1	∞	9.291	9.291	0.002
L1 × interpretation	2	∞	6.2	12.4	0.002
L1 × sequence	2	∞	5.165	10.33	0.006
Interpretation × sequence	1	∞	0.002	0.002	0.964
L1 × interpretation × sequence	2	∞	0.994	1.988	0.37

$p = 0.006$) and Japanese learners ($z = 4.07$, $p < 0.0001$), and no effects on the locational video judgments. Furthermore, an Interpretation × Sequence × LoR three-way interaction was observed, though this effect approached but did not reach conventional significance ($p = 0.056$). Given our a priori interest in LoR effects on directional interpretations, we conducted follow-up comparisons to clarify the pattern. Exploratory contrasts indicated that the LoR effect on the directional interpretation was restricted to the post-sentence sequence (Dutch $z = 2.62$, $p = 0.009$; Japanese $z = 3.10$, $p = 0.002$). Thus, while Sequence did not exert a significant main effect, its marginal interaction with LoR and Interpretation helped identify the contextual conditions under which LoR influenced participants' judgments. These findings suggest that longer LoR or L2 immersion experience enhanced learners' ability to recognise the post-sentential directional interpretation.

https://doi.org/10.1017/S1366728925100801 Published online by Cambridge University Press

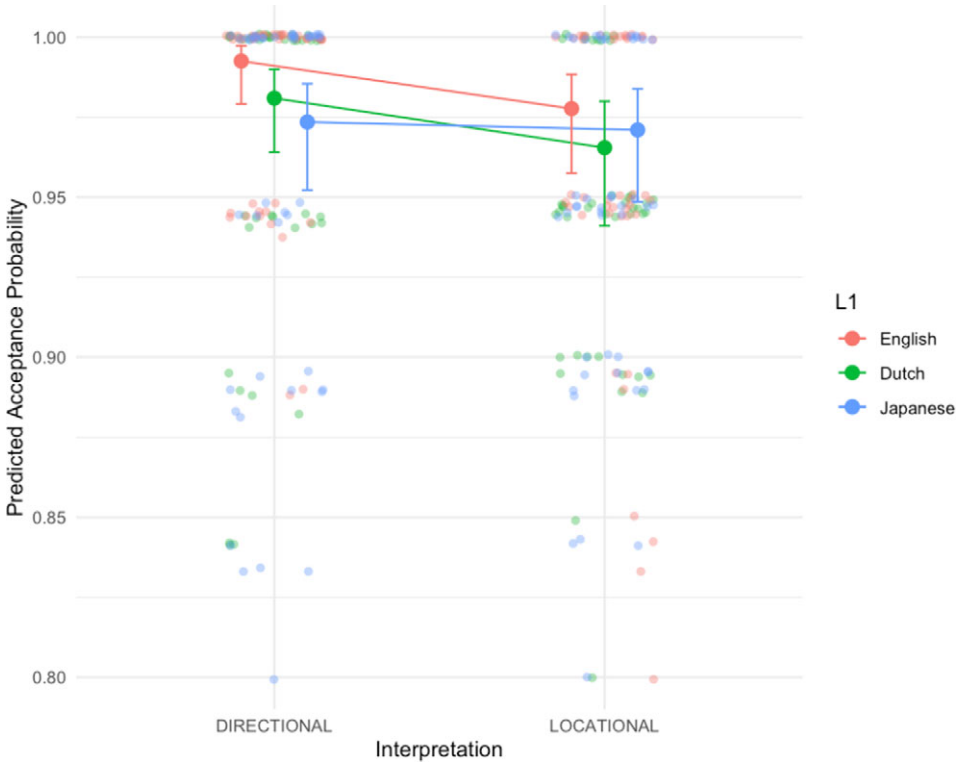


Figure 5. Predicted acceptance probabilities by L1 group and interpretation type. Error bars show 95% confidence intervals. Jittered dots represent individual participants' raw acceptance proportions. Y-axis truncated at 0.8 to enhance visibility of group differences.

Table 5. Results of L2 AoA effects on acceptance rates

Model term	df1	df2	F-ratio	χ^2	*p*-value
L1	1	∞	2.452	2.452	0.117
Interpretation	1	∞	1.041	1.041	0.308
Sequence	1	∞	1.474	1.474	0.225
AoA	1	∞	5.779	5.779	0.016
L1 × interpretation	1	∞	3.947	3.947	0.047
L1 × sequence	1	∞	8.877	8.877	0.003
L1 × AoA	1	∞	1.478	1.478	0.224
Interpretation × sequence	1	∞	0.08	0.08	0.778
Interpretation × AoA	1	∞	1.239	1.239	0.266
Sequence × AoA	1	∞	2.177	2.177	0.14
L1 × interpretation × sequence	1	∞	0.354	0.354	0.552
L1 × interpretation × AoA	1	∞	0.918	0.918	0.338
L1 × sequence × AoA	1	∞	0.772	0.772	0.38
Interpretation × sequence × AoA	1	∞	2.798	2.798	0.094
L1 × interpretation × sequence × AoA	1	∞	1.482	1.482	0.223

7. Discussion

The current study investigated the cross-linguistic L1–L2 dis/similarity effect on the formation of mental imagery for dual interpretations during incremental motion event processing. We hypothesised no differences in reaction times (RTs) and acceptance rates (ARs) for judging the locational videos across groups, whereas directional videos would result in group differences in the ARs and RTs between English and Japanese (dissimilar L1s), but not between English and Dutch (similar L1s). The results supported these predictions. The analysis of the RTs revealed a cross-linguistic L1 dis/similarity effect: while no significant differences were observed for the locational videos, the difference emerged in the directional videos between the dissimilar L1 groups (English and Japanese), but not between the similar L1 groups (English and Dutch). Likewise, the results of ARs revealed the L1 effect: L1-Japanese learners of English made more rejections of the directional videos compared to L1-English and L1-Dutch speakers, while no differences were observed for locational videos.

7.1. L1–L2 dis/similarity effect on the formation and reanalysis of mental imagery

The absence of significant differences in mental imagery for locational interpretations across groups, contrasted with the reduced embodiment effect for directional interpretations among L1-Japanese learners, indicates a clear L1 transfer effect in L2 mental model formation. This supports the simulation-based model of L2 comprehension (Zhao et al., 2025), which posits that constructional knowledge provides the foundation for mental simulation by specifying the semantic aspects to be mentally represented during language processing. As learners co-activate both L1 and L2 constructional knowledge, transfer is inevitable. Dominant L1 representations influence the formation of L2 representations. Locational interpretations, being the shared meaning of locative prepositions across all three languages, resulted in positive transfer and similarly high acceptance among both learner groups.

https://doi.org/10.1017/S1366728925100801 Published online by Cambridge University Press

In contrast, directional interpretations, where cross-linguistic differences arise, led to positive transfer for Dutch learners but negative transfer for Japanese learners, whose L1 lacks equivalent constructions.

The Japanese learners exhibited delays and higher rejection rates for directional interpretations due to the absence of equivalent dual interpretations in Japanese. This finding aligns with previous research on Japanese learners' acquisition of dual interpretations in English (Inagaki, 2002). For these learners, L2 forms may initially connect to conceptual meanings parasitically and indirectly via L1 mental representations (Kroll & Sholl, 1992; Kroll & Stewart, 1994). Developing a more autonomous linguistic system requires functional restructuring (Zhao et al., 2025; MacWhinney, 1992, 2012), a process in which bilinguals adapt and reconfigure cognitive systems to accommodate the demands of both languages. The absence of the directional interpretation in Japanese required them to establish a new category for this semantic meaning within their L2 mental models. To consolidate this category, learners needed to reinforce L2-specific neural pathways through repeated use in communicative contexts. Although the advanced Japanese participants in our study demonstrated good awareness of directional meanings – indicated by their high acceptance rates in verifying directional videos (above 90%) – their significant delays in reaction times (RTs) for directional verification compared to locational verification suggested greater cognitive effort in processing the directional interpretation. Their knowledge of the new L2-specific category was likely not fully proceduralised, limiting their fluency in applying it. Additionally, Japanese learners face further challenges due to structural differences between the two languages, such as divergent word orders (e.g., Fender, 2003) and the typological transition from a verb-framed language to a satellite-framed language (e.g., Brown & Gullberg, 2008, 2011; Pavlenko, 2014). These substantial cross-linguistic differences impede the development of autonomous and embodied L2 proficiency, requiring additional cognitive resources and practice to overcome.

These findings align with the L2 simulation model's (Zhao et al., 2025), which predicts that cross-linguistic constructional correspondences facilitate simulation in L2 processing, whereas mismatches necessitate a more effortful reanalysis, increasing processing difficulty and the likelihood of non-target-like interpretations. Prior findings that indicate L1 constraints on mental simulation in L2 contexts have either focused on constructions completely absent in the learner's L1 (Ahlberg et al., 2018) or have not included direct comparisons between learners with cross-linguistically similar and different L1 backgrounds (Foroni, 2015; Norman & Peleg, 2022). By explicitly considering cross-linguistic (dis)similarities, the current findings provide valuable insights into the nuanced ways in which L1 constructional knowledge influences L2 processing. They highlight the importance of constructional correspondences in enabling native-like mental imagery, particularly when interpreting semantically ambiguous constructions.

Interestingly, both English and Dutch participants showed higher acceptance of directional interpretations than locational ones in sentence-video verification – a finding that appears inconsistent with the results of the norming task, where locational interpretations were rated as more prototypical. However, the difference in mean ratings between the two interpretations was minimal (5.3 versus 5.4) despite being statistically significant, indicating a reliable but small effect size (d = 0.12). This discrepancy may reflect task-related differences in processing demands. In the norming task, participants had ample time to reflect and rate sentence-video pairings on a Likert scale, likely relying on entrenched semantic representations and metalinguistic judgment. In contrast, the sentence-video verification task required rapid, intuitive decisions under time pressure, potentially engaging more perceptually driven processes. Insights from event perception research help shed light on this discrepancy. According to Radvansky and Zacks (2014), event representations are structured and often organised around goal-directed actions and causal transitions. Building on Gibson's (2014) framework, visually salient events typically involve changes in surface layout and the appearance or disappearance of objects. In this context, the directional videos in our study, which depicted dynamic, goal-directed motion into a landmark (e.g., an object/person entering a location), entailed a clear transition from one spatial configuration to another. These scenes align more closely with Gibson's criteria for what constitutes a perceptually salient visual event. As such, participants may have been more inclined to accept these videos as matching the sentence, even if the directional interpretation is conceptually less prototypical. In contrast, the locational videos, though also dynamic, depicted motion occurring entirely within the boundaries of a landmark and lacked a clear containment transition, possibly making them appear less structured or goal-directed. While this account is speculative, it provides a plausible explanation for the observed discrepancy. Testing it is something that will be considered in future work.

At the same time, all three L1 groups responded faster to the locational interpretation than to the directional interpretation. This suggests that locational mental imagery was formed first, with the directional interpretation emerging through a reanalysis of the imagery (e.g., Roberts & Liszka, 2019; Sato et al., 2013). The faster formation of the locational interpretation aligns with the norming study findings, confirming the semantic prototypicality of locational meanings. Locational interpretations are more entrenched in the mental lexicon, leading to quicker processing times. Taken together, these findings suggest a dissociation between speed of mental imagery formation and final verification decisions: while locational interpretations were accessed more quickly due to their conceptual dominance, directional videos, possibly by virtue of their perceptual saliency, were more readily accepted under time pressure. In addition, the faster response to the locational interpretation may have been influenced by the lack of rich linguistic context in the sentence stimuli (Nikitina, 2008; Tutton, 2009). Tutton (2009) argues that the directional interpretation is more easily inferred in contexts with specific serial events. For example, in the sequence, "*The boy ran in the room and found the house key*," the first event (running in the room) implies directionality as a prerequisite for the second event (finding the house key). Such linguistic contexts provide critical cues for interpreting locative phrases directionally. In contrast, the simple sentences used in the present study may have reinforced the dominance of locational meanings over directional ones. To further explore this phenomenon, future research could incorporate serial event contexts to examine whether enriched linguistic cues can override the dominant locational representation in online processing.

It is worth acknowledging that while the study adopts a simulation-based account, we cannot definitively determine whether participants formed conscious mental images. This limitation is consistent with prior studies using similar sentence-picture verification paradigms (e.g., Sato et al., 2013; Zwaan et al., 2002, 2004), where mental simulation is inferred from task performance. Drawing on the perspective of Paivio's Dual Coding Theory (1986), it remains possible that automatic verbal associations may have contributed to the observed effects. However, because the sentence

https://doi.org/10.1017/S1366728925100801 Published online by Cambridge University Press

stimuli were identical across the visual conditions, differences in responses must arise from how participants internally represented the sentence content in relation to the visual scene. This lends more support to the view that perceptual simulation underlies the effects found in the verification task. Future studies employing neuroimaging methods may help to further disentangle the contributions of verbal and imagery-based systems in L2 simulation.

7.2. Mental simulation in incremental sentence processing

While cross-linguistic variation in L1s influenced L2 imagery, an interaction between Interpretation and Sequence was observed. Participants from all three L1 backgrounds verified faster at the post-sentence point than at the post-preposition point, suggesting that they could integrate all sentential cues and more easily form a mental representation after reading the entire sentence, compared to the incomplete sentence sequence. This finding supports the view that mental simulation unfolds incrementally as linguistic input accumulates over time (Altmann & Kamide, 1999; Zwaan, 2004), with comprehenders dynamically updating mental representations in response to semantically informative cues. The lack of cross-linguistic differences in this temporal pattern suggests that L2 learners, like native speakers, rely on similar cognitive mechanisms to build mental simulations in real time, supporting the simulation-based L2 comprehension model (Zhao et al., 2025) and prior findings that argue for functional similarity in the simulation processes of L1 and L2 users (Vukovic & Shtyrov, 2014).

A similar temporal trend was observed in Sato et al. (2013), who examined the formation of mental imagery of object shape during incremental processing of Japanese sentences. In their study, participants responded faster to pictures presented after the entire sentence (Experiment 2) than those presented prior to the final verbs (Experiment 1). Sato et al. (2013) attributed the delay in the incomplete sentence condition to differences in participant samples and the cognitive load involved in the task. However, the present study suggests that this delay reflects heightened cognitive load due to the additional need to match predictions with the presented video and revise representations if required. Unlike Sato et al.'s between-subjects design, the present study used a within-subjects design with counterbalanced sequences, further supporting the idea that delays arose from representation-level processing demands rather than participant-level factors.

In contrast, a recent study by Wang and Zhao (under review) found that the compatibility effect of mental imagery was strongest when diagrams were presented immediately after Chinese directional verbs (e.g., *jìn* "*enter*", *chū* "*exit*"). The effect weakened when diagrams appeared later in the sentence, such as after locational noun phrases. This pattern suggests that directional verbs may have provided sufficient schematic cues to evoke mental imagery related to CONTAINMENT and directionality. Consequently, the imagery effect was attenuated at the later-occurring locational noun phrases, which contributed comparatively less new information for refining the simulation.

By contrast, the final verbs in Sato et al. (2013) and the locational NPs in the present study appeared more critical for invoking concrete mental imagery. In Sato et al.'s study, the final verb in a SOV language like Japanese carried essential information determining the object's shape. Similarly, in the present study, sentence stimuli allowed for a wider range of possible interpretations for the final NPs, as in *The bird hooped behind* (*the log/the bench/the rock, etc.*). Participants required the information provided by the NPs to disentangle the ambiguous dual interpretations and select their preferred interpretation. This finding underscores the importance of complete sentential cues in shaping mental imagery but requires further investigation to deepen our understanding of mental simulation in incremental sentence processing.

7.3. L2 mental simulation and learner factors

Overall, L2 AoA influenced both reaction times and acceptance rates, but its effects varied depending on L1 background and interpretation type. L1-Dutch learners' directional simulation was more sensitive to the AoA effect, while L1-Japanese learners showed no such pattern, highlighting L1 constraints on L2 mental representation development. LoR did not affect RTs, but was associated with increased acceptance of post-sentential directional interpretation. Although this pattern emerged alongside a marginal omnibus interaction, the exploratory follow-up analyses point to an immersive learning advantage for endorsing non-prototypical meanings without necessarily accelerating simulation speed. These findings provide empirical support for the L2 simulation model (Zhao et al., 2025), particularly regarding the influence of learner-internal factors (AoA) and learning context (LoR) in influencing L2 mental simulations.

A key insight emerging from this study is that learner factors influenced processing only for the directional interpretation, not the prototypical locational one. Since the locational interpretation is the prototypical meaning of prepositional locatives and is shared across English, Dutch and Japanese, it is likely to be universally accessible and conceptually stable, reducing the impact of learner-internal factors. Conversely, the directional interpretation is non-prototypical and subject to cross-linguistic variation. The marginal effect of AoA on directional judgments only in L1-Dutch learners suggests a possible influence of cross-linguistic similarity in facilitating L2-specific restructuring. As our post hoc sensitivity analyses show, the study was adequately powered to detect small-to-moderate effects, but may have been less sensitive to very subtle higher-order interactions.

We acknowledge, however, that differences in learning context beyond formal immersion (indexed by LoR) may have contributed to the observed group differences. Dutch learners of English typically acquire English in a more communicative, ESL-like environment with rich exposure to authentic L2 input, whereas Japanese learners are often educated in EFL contexts where such input is limited. Repeated engagement with L2 constructions in authentic communicative settings likely fosters more flexible and context-sensitive representations. While our study included LoR as a proxy for contextual exposure, consistent with standard practice in bilingual research (e.g., Birdsong & Molis, 2001; Johnson & Newport, 1991), we did not directly assess the quality or quantity of participants' everyday L2 exposure in their home environments. This limitation should be addressed in future work using more nuanced indices of communicative experience and L2 usage.

Nonetheless, our data provide converging evidence that L1 constructional differences remain a major constraint on L2 mental simulation. Japanese learners' lack of sensitivity to proficiency, AoA or and LoR may reflect not only reduced input, but also the absence of equivalent constructions in the L1 that would support the acquisition of the directional interpretation. The infrequent mapping between form (manner-of-motion verb + locative prepositions) and meaning (directionality) in English (Tutton, 2009) poses additional challenges for L2 learners, particularly those from typologically distinct L1 backgrounds. A similar observation was made by Ahlberg et al. (2018), where L2 German from linguistically dissimilar L1 backgrounds showed an attenuated compatibility

https://doi.org/10.1017/S1366728925100801 Published online by Cambridge University Press

effect for *über* ("above"), which is more strongly associated with the upward spatial relation than *auf* ("on"). They attributed this finding to the more frequent usage and earlier acquisition of *auf* than *über* in L2 German. Since the directional interpretation is context-dependent and occurs less frequently than the locational interpretation (e.g., Nikitina, 2008; Tutton, 2009), Japanese learners – unlike Dutch speakers who can rely on their L1 constructional knowledge – did not encounter sufficient exemplars to make the directional interpretation as part of their experiential traces. This reinforces the idea that frequent exposure is crucial for integrating less prototypical meanings into L2 mental representations.

In summary, these findings collectively suggest that L2 simulation restructuring is a gradual and experience-sensitive process, which is significantly influenced by L1–L2 similarity constraints and semantic prototypicality, and may be modulated by early exposure and extended immersion. While prototypical meanings are processed more efficiently and remain stable across languages, non-prototypical meanings require greater cognitive effort and are more susceptible to both crosslinguistic constraints and learner-specific factors.

8. Conclusion

The present study provides empirical evidence that L1–L2 conceptual (dis)similarity significantly influences how L2 learners construct mental imagery during incremental sentence processing. Specifically, learners from dissimilar backgrounds may experience greater difficulty in shifting from entrenched L1 mental models to L2-specific simulations for non-prototypical meanings. These findings confirm and extend the simulation-based model of L2 sentence processing by demonstrating that the effects of proficiency, age of acquisition and immersion experience are moderated by L1–L2 constructional correspondence. Rather than developing mental simulations entirely anew, L2 learners build them through a process of functional restructuring shaped by both prior L1-based representations and the L2 learning context. These insights contribute to theoretical models of bilingual cognition by recognising that bilinguals develop distinct representational systems that are not simply approximations of monolingual norms. Pedagogically, these findings suggest the value of supporting learners' engagement with less prototypical, context-dependent meanings to enhance the flexibility of L2 mental representations for real-world communication.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S1366728925100801.

Data availability statement. The data that support the findings of this study are openly available in the Open Science Framework (OSF) at https://osf.io/u4x82/.

Acknowledgements. The authors would like to thank Dr. Ben Stone for programming the experimental task.

Competing interests. The authors declare none.

References

Ahlberg, D. K., **Bischoff, H.**, **Kaup, B.**, **Bryant, D.**, & **Strozyk, J. V.** (2018). Grounded cognition: Comparing language × space interactions in first language and second language. *Applied PsychoLinguistics*, **39**(2), 437–459. https://doi.org/10.1017/s014271641700042x.

Ahn, S., & **Jiang, N.** (2018). Automatic semantic integration during L2 sentential reading. *Bilingualism: Language and Cognition*, **21**(2), 375–383. https://doi.org/10.1017/s1366728917000256.

Allen, S., **Özyürek, A.**, **Kita, S.**, **Brown, A.**, **Furman, R.**, **Ishizuka, T.**, & **Fujii, M.** (2007). Language-specific and universal influences in children's syntactic packaging of manner and path: A comparison of English. *Japanese, and Turkish. Cognition*, **102**(1), 16–48. https://doi.org/10.1016/j.cognition.2005.12.006.

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

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, **22**(4), 577–660. https://doi.org/10.1017/S0140525X99002149.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, **59**(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639.

Barsalou, L. W., **Santos, A.**, **Simmons, W. K.**, & **Wilson, C. D.** (2008). Language and simulation in conceptual processing. In M. de Vega, A. Glenberg, & A. Graesser (Eds.), *Symbols and embodiment: Debates on meaning and cognition* (pp. 245–283). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199217274.001.0001.

Bates, D., **Maechler, M.**, **Bolker, B.**, **Walker, S.**, **Christensen, R. H. B.**, **Singmann, H.**, **Dai, B.**, **Scheipl, F.**, **Grothendieck, G.**, **Green, P.**, **Fox, J.**, **Bauer, A.**, **Krivitsky, P. N.**, **Tanaka, E.**, & **Jagan, M.** (2024). Package 'lme4' (Version 1.1–35.3) [Computer software]. Comprehensive R Archive Network. https://cran.r-project.org/web/packages/lme4/lme4.pdf

Beavers, J., **Levin, B.**, & **Wei Tham, S.** (2010). The typology of motion expressions revisited. *Journal of Linguistics*, **46**(2), 331–377. https://doi.org/10.1017/s0022226709990272.

Bergen, B. K., & **Chang, N.** (2005). Embodied construction grammar in simulation-based language understanding. In J. Östman & M. Fried (Eds.), *Constructional approaches to language* (pp. 147–190). John Benjamins Publishing.

Bergen, B. K., & **Chang, N.** (2013). Embodied constructional grammar. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford handbook of construction grammar* (pp. 168–190). Oxford University Press.

Birdsong, D., & **Molis, M.** (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language*, **44**(2), 235–249. https://doi.org/10.1006/jmla.2000.2750.

Brown, A., & **Gullberg, M.** (2008). Bidirectional crosslinguistic influence in L1–L2 encoding of manner in speech and gesture: A study of Japanese speakers of English. *Studies in Second Language Acquisition*, **30**(02), 225–251. https://doi.org/10.1017/s0272263108080327.

Brown, A., & **Gullberg, M.** (2011). Bidirectional cross-linguistic influence in event conceptualization? Expressions of path among Japanese learners of English. *Bilingualism: Language and Cognition*, **14**(1), 79–94. https://doi.org/10.1017/S1366728910000064.

Brown, A., & **Gullberg, M.** (2013). L1–L2 convergence in clausal packaging in Japanese and English. *Bilingualism: Language and Cognition*, **16**(3), 477–494. https://doi.org/10.1017/S1366728912000491.

Chen, D., **Su, J.**, & **Wang, R.** (2024). Differences in perceptual representations in multilinguals' first, second, and third language. *Frontiers in Human Neuroscience*, **18**(1408411). https://doi.org/10.3389/fnhum.2024.1408411.

Chen, D., **Wang, R.**, **Zhang, J.**, & **Liu, C.** (2020). Perceptual representations in L1, L2 and L3 comprehension: Delayed sentence–picture verification. *Journal of Psycholinguistic Research*, **49**(1), 41–57. https://doi.org/10.1007/s10936-019-09670-x.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. Routledge.

de Leeuw, J. R., **Gilbert, R. A.**, & **Luchterhandt, B.** (2023). jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software*, **8**(85), 5351. https://doi.org/10.21105/joss.05351.

Den Dikken, M. (2010). On the functional structure of locative and directional PPs. In G. Cinque & L. Rizzi (Eds.), *Mapping spatial PPs: The cartography of syntactic structures* (Vol. **6**, pp. 74–126). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195393675.001.0001.

Domazetoska, I., & **Zhao, H.** (2025). First and second language speakers' sensitivity to the distributional properties of *wh*-clauses: Effects of proficiency, acquisitional context and language experience. *Australian Review of Applied Linguistics*, **48**(1), 28–54. https://doi.org/10.1075/aral.23011.dom.

Dudschig, C., **de la Vega, I.**, & **Kaup, B.** (2014). Embodiment and second-language: Automatic activation of motor responses during processing spatially associated L2 words and emotion L2 words in a vertical Stroop

paradigm. *Brain and Language*, **132**, 14–21. https://doi.org/10.1016/j.bandl.2014.02.002.

Engelen, J. A., **Bouwmeester, S.**, **De Bruin, A. B.**, & **Zwaan, R. A.** (2011). Perceptual simulation in developing language comprehension. *Journal of Experimental Child Psychology*, **110**(4), 659–675. https://doi.org/10.1016/j.jecp.2011.06.009.

Engemann, H. (2022). How (not) to cross a boundary: Crosslinguistic influence in simultaneous bilingual children's event construal. *Bilingualism: Language and Cognition*, **25**(1), 42–54. https://doi.org/10.1017/s1366728921000298.

Faul, F., **Erdfelder, E.**, **Lang, A.-G.**, & **Buchner, A.** (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, **39**(2), 175–191. https://doi.org/10.3758/BF03193146.

Fender, M. (2003). English word recognition and word integration skills of native Arabic- and Japanese-speaking learners of English as a second language. *Applied PsychoLinguistics*, **24**(2), 289–315. https://doi.org/10.1017/s014271640300016x.

Filipović, L. (2011). Speaking and remembering in one or two languages: Bilingual vs. monolingual lexicalization and memory for motion events. *International Journal of Bilingualism*, **15**(4), 466–485. https://doi.org/10.1177/1367006911403062.

Filipović, L. (2022). First language versus second language effect on memory for motion events: The role of language type and proficiency. *International Journal of Bilingualism*, **26**(1), 65–81. https://doi.org/10.1177/13670069211022863.

Foroni, F. (2015). Do we embody second language? Evidence for 'partial' simulation during processing of a second language. *Brain and Cognition*, **99**, 8–16. https://doi.org/10.1016/j.bandc.2015.06.006.

Foster, P., **Bolibaugh, C.**, & **Kotula, A.** (2013). Knowledge of nativelike selections in a L2: The influence of exposure, memory, age of onset, and motivation in foreign language and immersion settings. *Studies in Second Language Acquisition*, **36**(1), 101–132. https://doi.org/10.1017/s0272263113000624.

Fu, X., **Vanek, N.**, & **Roberts, L.** (2024). Matched or moved? Asymmetry in high- and low-level visual processing of motion events. *Language and Cognition*, **16**(2), 283–306. https://doi.org/10.1017/langcog.2023.37.

Gennari, S., **Sloman, S. A.**, **Malt, B. C.**, & **Fitch, W. T.** (2002). Motion events in language and cognition. *Cognition*, **83**(1), 49–79. https://doi.org/10.1016/s0010-0277(01)00166-4.

Gibson, J. J. (2014). *The ecological approach to visual perception*. Psychology Press. https://doi.org/10.4324/9781315740218.

Glenberg, A. M., & **Kaschak, M. P.** (2002). Grounding language in action. *Psychonomic Bulletin & Review*, **9**(3), 558–565. https://doi.org/10.3758/BF03196313.

Hammerly, C., **Staub, A.**, & **Dillon, B.** (2019). The grammaticality asymmetry in agreement attraction reflects response bias: Experimental and modelling evidence. *Cognitive Psychology*, **110**, 70–104. https://doi.org/10.1016/j.cogpsych.2019.01.001.

Hendriks, H., **Hickmann, M.**, & **Pastorino-Campos, C.** (2022). Running or crossing? Children's expression of voluntary motion in English, German, and French. *Journal of Child Language*, **49**(3), 578–601. https://doi.org/10.1017/s0305000921000271.

Hoeben Mannaert, L. N., **Dijkstra, K.**, & **Zwaan, R. A.** (2017). Is color an integral part of a rich mental simulation? *Memory & Cognition*, **45**(6), 974–982. https://doi.org/10.3758/s13421-017-0708-1.

Hoeben Mannaert, L. N., **Dijkstra, K.**, & **Zwaan, R. A.** (2019). How are mental simulations updated across sentences? *Memory & Cognition*, **47**(6), 1201–1214. https://doi.org/10.3758/s13421-019-00928-2.

Hoeben Mannaert, L. N., **Dijkstra, K.**, & **Zwaan, R. A.** (2021). Is color continuously activated in mental simulations across a broader discourse context? *Memory & Cognition*, **49**(1), 127–147. https://doi.org/10.3758/s13421-020-01078-6.

Hohenstein, J., **Eisenberg, A.**, & **Naigles, L.** (2006). Is he floating across or crossing afloat? Cross-influence of L1 and L2 in Spanish–English bilingual adults. *Bilingualism: Language and Cognition*, **9**(3), 249–261. https://doi.org/10.1017/s1366728906002616.

Inagaki, S. (2002). Japanese learners' acquisition of English manner-of-motion verbs with locational/directional PPs. *Second Language Research*, **18**(1), 3–27. https://doi.org/10.1191/0267658302sr196oa.

Ito, A., **Corley, M.**, **Pickering, M. J.**, **Martin, A. E.**, & **Nieuwland, M. S.** (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language*, **86**, 157–171. https://doi.org/10.1016/j.jml.2015.10.007.

Ito, A., **Martin, A. E.**, & **Nieuwland, M. S.** (2017). On predicting form and meaning in a second language. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, **43**(4), 635–652. https://doi.org/10.1037/xlm0000315.

Jarvis, S. (2011). Conceptual transfer: Crosslinguistic effects in categorization and construal. *Bilingualism: Language and Cognition*, **14**(1), 1–8. https://doi.org/10.1017/s1366728910000155.

Jarvis, S., & **Pavlenko, A.** (2008). Crosslinguistic influence in language and cognition. Routledge. https://doi.org/10.4324/9780203935927

Johnson, J. S., & **Newport, E. L.** (1991). Critical period effects on universal properties of language: The status of subjacency in the acquisition of a second language. *Cognition*, **39**(3), 215–258. https://doi.org/10.1016/0010-0277(91)90054-8.

Kamenetski, A., **Lai, V. T.**, & **Flecken, M.** (2022). Minding the manner: Attention to motion events in Turkish–Dutch early bilinguals. *Language and Cognition*, **14**(3), 456–478. https://doi.org/10.1017/langcog.2022.10.

Kang, X., **Eerland, A.**, **Joergensen, G. H.**, **Zwaan, R. A.**, & **Altmann, G. T.** (2020). The influence of state change on object representations in language comprehension. *Memory & Cognition*, **48**(3), 390–399. https://doi.org/10.3758/s13421-019-00977-7.

Kong, S. (2021). Adult mandarin Chinese speakers' acquisition of locational and directional prepositional constructions in second language English. *Lingua*, **249**, 102993. https://doi.org/10.1016/j.lingua.2020.102993.

Konishi, H., **Wilson, F.**, **Golinkoff, R. M.**, **Maguire, M. J.**, & **Hirsh-Pasek, K.** (2014). Late Japanese bilinguals' novel verb construal. *Bilingualism: Language and Cognition*, **19**(4), 782–790. https://doi.org/10.1017/s136672891400073x.

Koster, D., **Cadierno, T.**, & **Chiarandini, M.** (2018). Mental simulation of object orientation and size: A conceptual replication with second language learners. *Journal of the European Second Language Association*, **2**(1), 38. https://doi.org/10.22599/jesla.39.

Kotz, S. A., & **Elston-Güttler, K. E.** (2007). Bilingual semantic memory revisited: ERP and fMRI evidence. In J. Hart & M. A. Kraut (Eds.), *Neural basis of semantic memory* (pp. 105–132). Cambridge University Press. https://doi.org/10.1017/CBO9780511544965.

Kroll, J., & **Sholl, A.** (1992). Lexical and conceptual memory in fluent and nonfluent bilinguals. In R. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 191–206). Elsevier.

Kroll, J., & **Stewart, E.** (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, **33**(2), 149–174. https://doi.org/10.1006/jmla.1994.1008.

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

Lai, V. T., **Rodriguez, G. G.**, & **Narasimhan, B.** (2014). Thinking-for-speaking in early and late bilinguals. *Bilingualism: Language and Cognition*, **17**(1), 139–152. https://doi.org/10.1017/s1366728913000151.

Lenth, R. V., **Buerkner, P.**, **Giné-Vázquez, I.**, **Herve, M.**, **Jung, M.**, **Love, J.**, **Miguez, F.**, **Piaskowski, J.**, **Riebl, H.**, & **Singmann, H.** (2024). Package 'emmeans' (Version 1.10.1) [Computer software]. https://cran.r-project.org/web/packages/emmeans/emmeans.pdf

Levin, B., & **Hovav, M. R.** (1995). *Unaccusativity: At the syntax-lexical semantics interface*. MIT Press.

Lewandowski, W., & **Özçalışkan, Ş.** (2019). How language type influences patterns of motion expression in bilingual speakers. *Second Language Research*, **37**(1), 27–49. https://doi.org/10.1177/0267658319877214.

Lu, X., & **Yang, J.** (2025). Second language embodiment of action verbs: The impact of bilingual experience as a multidimensional spectrum. *Bilingualism: Language and Cognition*, 1–17. https://doi.org/10.1017/S1366728924000981.

MacWhinney, B. (1992). Transfer and competition in second language learning. In R. J. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 371–390). Elsevier.

MacWhinney, B. (2012). The logic of the unified model. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 211–227). Routledge. https://doi.org/10.4324/9780203808184.

https://doi.org/10.1017/S1366728925100801 Published online by Cambridge University Press

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

Matsumoto, Y. (2018). Motion event descriptions in Japanese from typological perspectives. In P. Prashant & K. Taro (Eds.), *Handbook of Japanese contrastive linguistics* (pp. 273–290). De Gruyter Mouton. https://doi.org/10.1515/9781614514077-010.

Nikitina, T. (2008). Pragmatic factors and variation in the expression of spatial goals. In A. Asbury, J. Dotlačil, B. Gehrke, & R. Nouwen (Eds.), *Syntax and semantics of spatial P* (pp. 175–195). John Benjamins Publishing. https://doi.org/10.1075/la.120.09nik.

Norman, T., & **Peleg, O.** (2022). The reduced embodiment of a second language. *Bilingualism: Language and Cognition*, **25**(3), 406–416. https://doi.org/10.1017/S1366728921001115.

Özçalışkan, Ş. (2015). Ways of crossing a spatial boundary in typologically distinct languages. *Applied PsychoLinguistics*, **36**(2), 485–508. https://doi.org/10.1017/s0142716413000325.

Paivio, A. (1986). *Mental representation: A dual coding approach*. Oxford University Press.

Papafragou, A., **Hulbert, J.**, & **Trueswell, J.** (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, **108**(1), 155–184. https://doi.org/10.1016/j.cognition.2008.02.007.

Park, H. I. (2020). How do Korean–English bilinguals speak and think about motion events? Evidence from verbal and non-verbal tasks. *Bilingualism: Language and Cognition*, **23**(3), 483–499. https://doi.org/10.1017/s1366728918001074.

Pavlenko, A. (2014). *The bilingual mind: And what it tells us about language and thought*. Cambridge University Press. https://doi.org/10.1017/CBO9781139021456.

Plonsky, L., & **Oswald, F. L.** (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, **64**(4), 878–912. https://doi.org/10.1111/lang.12079.

R Core Team (2024). *R (4.3.3): A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

Radvansky, G. A., & **Zacks, J. M.** (2014). *Event cognition*. Oxford University Press.

Roberts, L., & **Liszka, S. A.** (2019). Grammatical aspect and L2 learners' online processing of temporarily ambiguous sentences in English: A self-paced reading study with German, Dutch and French L2 learners. *Second Language Research*, **37**(4), 619–647. https://doi.org/10.1177/0267658319895551.

Sato, M., **Schafer, A. J.**, & **Bergen, B. K.** (2013). One word at a time: Mental representations of object shape change incrementally during sentence processing. *Language and Cognition*, **5**(4), 345–373. https://doi.org/10.1515/langcog-2013-0022.

Schütt, E., **Dudschig, C.**, **Bergen, B. K.**, & **Kaup, B.** (2023). Sentence-based mental simulations: Evidence from behavioral experiments using garden-path sentences. *Memory & Cognition*, **51**(4), 952–965. https://doi.org/10.3758/s13421-022-01367-2.

Slobin, D. I. (1996). From "thought and language" to "thinking for speaking.". In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge University Press.

Slobin, D. I. (2000). Verbalized events: A dynamic approach to linguistic relativity and determinism. In S. Niemeier & R. Dirven (Eds.), *Evidence for linguistic relativity* (pp. 107–138). John Benjamins Publishing.

Slobin, D. I. (2003). Language and thought online: Cognitive consequences of linguistic relativity. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 157–191). MIT Press.

Slobin, D. I. (2004). The many ways to search for a frog: Linguistic typology and the expression of motion events. In L. Verhoeven & S. Stromqvist (Eds.), *Relating events in narrative, volume 2: Typological and contextual perspectives* (pp. 219–257). Taylor & Francis.

Stanfield, R. A., & **Zwaan, R. A.** (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, **12**(2), 153–156. https://doi.org/10.1111/1467-9280.00326.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description* (Vol. **3**, pp. 57–149). Cambridge University Press.

Talmy, L. (2000). *Toward a cognitive semantics: Typology and process in concept structuring*. MIT Press. https://doi.org/10.7551/mitpress/6848.001.0001.

Taylor, L. J., & **Zwaan, R. A.** (2008). Motor resonance and linguistic focus. *Quarterly Journal of Experimental Psychology*, **61**(6), 896–904. https://doi.org/10.1080/17470210701625519.

Tutton, M. (2009). When in means into: Towards an understanding of boundary-crossing in. *Journal of English Linguistics*, **37**(1), 5–27. https://doi.org/10.1177/0075424208329308.

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

Vukovic, N., & **Shtyrov, Y.** (2014). Cortical motor systems are involved in second-language comprehension: Evidence from rapid mu-rhythm desynchronisation. *NeuroImage*, **102**, 695–703. https://doi.org/10.1016/j.neuroimage.2014.08.045.

Vukovic, N., & **Williams, J. N.** (2014). Automatic perceptual simulation of first language meanings during second language sentence processing in bilinguals. *Acta Psychologica*, **145**, 98–103. https://doi.org/10.1016/j.actpsy.2013.11.002.

Wang, M., & **Zhao, H.** (2024). Perceptual representations in L1 and L2 spatial and abstract language processing: Applying an innovative sentence-diagram verification paradigm. *Frontiers in Human Neuroscience*, **18**, 18(1425576), 1. https://doi.org/10.3389/fnhum.2024.1425576.

Wang, M., & **Zhao, H.** (Under review). Mental simulation in the incremental processing of Mandarin motion sentences: Task dependence and temporal dynamics.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer. https://ggplot2.tidyverse.org

Zhang, Y., **Lemarchand, R.**, **Asyraff, A.**, & **Hoffman, P.** (2022). Representation of motion concepts in occipitotemporal cortex: fMRI activation, decoding and connectivity analyses. *NeuroImage*, **259**, 119450. https://doi.org/10.1016/j.neuroimage.2022.119450.

Zhao, H., **Vanek, N.**, & **MacWhinney, B.** (2025). Mental simulation in bilingual and second slanguage processing: New directions in the competition model. *Brain and Language*. http://doi.org/10.1016/j.bbandl.2025.105619.

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. **44**, pp. 35–62). Elsevier.

Zwaan, R. A., & **Madden, C. J.** (2005). Embodied sentence comprehension. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking* (pp. 224–245). Cambridge University Press. https://doi.org/10.1017/CBO9780511499968.

Zwaan, R. A., **Madden, C. J.**, **Yaxley, R. H.**, & **Aveyard, M. E.** (2004). Moving words: Dynamic representations in language comprehension. *Cognitive Science*, **28**(4), 611–619. https://doi.org/10.1207/s15516709cog2804_5.

Zwaan, R. A., **Stanfield, R. A.**, & **Yaxley, R. H.** (2002). Language Comprehenders mentally represent the shapes of objects. *Psychological Science*, **13**(2), 168–171. https://doi.org/10.1111/1467-9280.00430.

Zwaan, R. A., & **Taylor, L. J.** (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General*, **135**(1), 1–11. https://doi.org/10.1037/0096-3445.135.1.1.

https://doi.org/10.1017/S1366728925100801 Published online by Cambridge University Press