

ARTICLE

Mental imagery effects in bilingual phrasal verb comprehension

Yi Hu and Helen Zhao 

School of Languages and Linguistics, Faculty of Arts, The University of Melbourne, Parkville, VIC, Australia

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

(Received 25 July 2025; Revised 26 November 2025; Accepted 29 December 2025)

Abstract

This study investigated how mental imagery is engaged during first language (L1) and second language (L2) speakers' incremental sentence processing of English phrasal verbs, using a self-paced sensibility judgment task interleaved with schematic diagrams. L1 speakers showed selective compatibility effects modulated by abstractness, semantic transparency of phrasal verbs, event plausibility and the timing of visual input. In contrast, L2 learners relied more generally on visual support, reflecting weaker integration of semantic and perceptual cues. Learner-internal factors such as L2 proficiency and language dominance modulated learners' sensitivity to integrate and resolve competing cues between semantic coherence and perceptual input. These findings support a simulation-based model of L2 comprehension, highlighting the developmental nature of sensorimotor activation in bilingual processing.

Keywords: English phrasal verb comprehension; individual learner differences; lexical-semantic properties; mental imagery

1. Introduction

Embodied cognition research emphasizes that mental simulation constitutes an internal mechanism in language processing, engaging modality-specific brain systems to reactivate memory-stored sensorimotor experiences (Bergen & Chang, 2013). Within this broader framework, mental imagery refers to the conscious reconstruction of perceptual and motor experiences without external sensory stimuli (Barsalou, 1999, 2008). Most investigations of mental imagery have focused on first language (L1) processing, demonstrating both compatibility and interference effects, with inconsistencies often attributed to language abstractness and task demands (e.g., Hauf et al., 2020; Liu & Bergen, 2016; Ostarek & Huettig, 2019).

In contrast, research on mental imagery in second language (L2) comprehension remains limited, with findings ranging from native-like to reduced or absent imagery effects (e.g., Chen et al., 2024; Shiang et al., 2024; Zeelenberg et al., 2025). A recent



development in this area is Zhao et al.'s (2025) simulation-based L2 comprehension model, which posits that crosslinguistic (dis)similarity, construction-specific semantic properties, learner-specific factors (e.g., proficiency, onset age of L2 learning and language dominance) and contextual factors (e.g., immersive environment, multimodal input) significantly influence L2 mental imagery effects. However, few studies have directly examined how construction-specific lexical-semantic properties affect L2 mental imagery or adopted a multidimensional view of learner variability.

The present study investigates the role of mental imagery in L1 and L2 speakers' processing of English phrasal verbs (PVs), a linguistic construction characterized by polysemy and semantic complexity. It examines how differences between abstract and concrete PV meanings, as well as semantic transparency, affect L1 and L2 mental representations during incremental sentence processing. To capture the temporal dynamics of mental imagery, this study used a self-paced sensibility judgment task interleaved with schematic diagrams.

1.1. *Mental imagery in L1 processing*

Rooted in embodied cognition, mental imagery is a fundamental framework in cognitive linguistics (CL) that emphasizes the sensorimotor basis of language and cognition. This perspective assumes that language processing reactivates neural patterns formed by prior sensorimotor experiences, enabling comprehenders to mentally reconstruct sensory and motor representations to interpret language input (Barsalou, 1999, 2008). This mechanism allows comprehenders to visualize objects, simulate related actions and recall associated sensory modalities (e.g., vision, sound and smell) (Bergen & Chang, 2005).

Research on L1 mental imagery has predominantly examined perceptual and motor simulation using sentence–picture verification tasks (SPVTs), where participants determine whether a picture matches the implied meaning of a preceding sentence. These studies manipulate perceptual dimensions (e.g., orientation, shape, size, color) and motor features, with most findings revealing compatibility effects where congruent perceptual-sensory input enhances processing speed and accuracy (e.g., Hauf et al., 2020; Norman & Peleg, 2022).

In contrast, some studies have reported interference effects, where language processing slows when paired with compatible images (Connell, 2007; Richardson et al., 2003). The underlying mechanism remains debated. One account attributes it to competition for shared cognitive resources, as simultaneous activation of perceptual-motor resources for language-related mental imagery and visual input may cause mutual inhibition and increased processing demands (Bergen et al., 2007). Others argue that certain perceptual properties, such as color, are inherently unstable and less conducive for simulation (Connell, 2007).

Variability in findings may reflect differences in task design, the sensorimotor features manipulated and language abstractness. For instance, Richardson et al. (2003) found that both abstract and concrete language facilitated performance in a visual memory task but interfered performance in a visual discrimination task. These contrasting results have contributed to an ongoing debate about whether abstract concepts engage sensorimotor systems similarly to concrete concepts. While some studies report comparable simulation effects across abstract and concrete language (e.g., Guan et al., 2013; Harpaintner et al., 2020), others argue that simulation is

restricted to concrete language (e.g., Bergen et al., 2007; Bergen & Wheeler, 2010). Further evidence of this distinction comes from Liu and Bergen (2016), who employed a sensibility judgment task to examine location–sentence compatibility effects. They found facilitation for concrete sentences (e.g., ‘Ben is feeding his child’) but interference for abstract sentences (e.g., ‘Calvin is submitting the request to the committee’). This pattern suggests that abstract language may involve distinct and more temporally extended simulation processes compared to concrete language, possibly due to weaker or more diffuse connections with sensorimotor experience.

1.2. L2 mental imagery: Mixed evidence and theoretical challenges

Despite extensive L1 research, evidence on L2 mental imagery remains limited and mixed. Some behavioral studies report native-like compatibility effects for features such as orientation, shape, color, motion and negation (e.g., Shiang et al., 2024; Zeelenberg et al., 2025), suggesting similar cognitive mechanisms for L1 and L2 mental simulation. Neuroimaging studies further indicate shared motor activation for both L1 and L2 processing (e.g., Birba et al., 2020; Zhang et al., 2020), though L2 learners may recruit motor cortex more extensively to compensate for weaker linguistic representations (Tian et al., 2020, 2023). Conversely, others report reduced or absent mental simulation effects in L2 processing, possibly due to weaker sensorimotor grounding and L1 dominance (e.g., Ahlberg et al., 2018; Chen et al., 2024; Xue et al., 2024). These inconsistencies raise questions about the extent of L1–L2 divergence in mental representations and the factors shaping such variation.

Abstractness is a critical factor influencing L2 mental imagery, as abstract language comprehension tends to pose greater challenges for L2 learners than concrete language (Shi et al., 2023). However, evidence for sensorimotor activation during L2 abstract language processing is still sparse and inconclusive. Some studies report compatibility effects regardless of abstractness (Wang & Zhao, 2023; Yang, 2016), while others report task-dependent interference effects influenced by conceptual abstraction (Wang & Zhao, 2024a, 2024b).

Considering abstractness as a continuum – ranging from literal (e.g., *seize the arm*), metaphorical (e.g., *seize the chance*) and abstract (e.g., *cherish the chance*) – Tian et al. (2020) found graded motor engagement (literal > metaphorical > abstract) in both L1 and L2 processing. The strongest motor activation for literal phrases likely reflects their direct connection with physical actions. However, Tian et al. (2023) failed to replicate these effects, possibly due to methodological differences (fMRI passive reading vs. MEG semantic judgment). These inconsistencies suggest that abstractness effects on L2 simulation are both context-sensitive and task-dependent.

More importantly, L1 and L2 mental simulations related to abstractness differ significantly. Embodiment effects for abstract concepts appear to be language-dependent, with stronger effects in L1 than in L2 (Wang & Zhao, 2024a; Wei et al., 2024). For instance, Chinese-English bilinguals process abstract senses of English prepositions more slowly than spatial senses, whereas native speakers show no such difference (Wang & Zhao, 2024a). This asymmetry may arise from weaker perceptual-motor associations in L2, impeding learners’ ability to intuitively construct L1–L2 constructional mappings (Dudschig et al., 2014).

To address these inconsistencies, Zhao et al. (2025) proposed a simulation-based L2 comprehension model, which assumes that cognitive simulations grounded in

real-world experiences underlie L2 comprehension, mirroring L1 processes. The model outlines three stages. In *constructional analysis* (Stage 1), learners identify constructions from L2 input and map L1 and L2 linguistic forms onto their corresponding meanings. L2 speakers' ability to achieve native-like mental simulation is strongly influenced by their existing linguistic construction knowledge and indirectly shaped by individual learner differences. Both the complexity and frequency of the construction would affect learners' simulation outcomes. In *contextual resolution* (Stage 2), discourse cues and world knowledge are integrated to construct a coherent semantic representation. In *embodied simulation* (Stage 3), this semantic representation engages sensorimotor systems, triggering perceptual-motor simulations that ground the linguistic input in embodied experiences.

The model further proposes that L2 mental simulation is modulated by linguistic knowledge (i.e., L1–L2 construction mapping), learner-specific factors (i.e., proficiency, onset age of L2 learning, language dominance) and contextual factors (i.e., world knowledge, communicative context). Variability in native-like to non-native-like perceptual-motor simulations may result from the interaction of these factors. Recent evidence shows that learner-specific and contextual factors collectively modulate the magnitude of L2 embodiment effects (e.g., Lu & Yang, 2025; Nishide et al., 2025).

Among these, L2 proficiency has received particular attention in relation to mental simulation, though findings are mixed – some report significant effects (Ahlberg et al., 2018; Nishide et al., 2025), while others find no or minimal effects (Lu & Yang, 2025; Zeelenberg et al., 2025). However, there is a paucity of research exploring other learner factors beyond proficiency. Lu and Yang (2025) found that earlier L2 age of acquisition (AoA) and greater exposure predicted stronger embodiment for English action verbs, whereas proficiency and dominance had no effect. Given the multidimensional nature of bilingual experience, learner factors alone cannot fully account for variability in L2 mental simulation. A comprehensive understanding of L2 embodiment requires examining the simulation-based model in relation to a broader range of variables, including proficiency, dominance, AoA and exposure.

Importantly, the L2 simulation model emphasizes the dynamic, temporally unfolding nature of L2 mental simulation, paralleling L1 simulation (Bergen & Chang, 2005, 2013), yet the temporal dimensions of L2 simulation remain under-explored.

1.3. Temporal dynamics in mental imagery

Numerous studies have demonstrated simulation effects during L1 incremental sentence processing, but the precise timing of mental activation remains debated (Sato et al., 2013; Taylor & Zwaan, 2008; Zwaan & Taylor, 2006). Using a self-paced reading task with manual knob-turning, Zwaan and Taylor (2006) found compatibility effects at the verb region but no effects at other regions (before/after the verb or sentence-final regions). Taylor and Zwaan (2008) further showed that motor resonance was triggered by action-relevant elements (e.g., verbs, action-modifying adverbs like *'quickly'*), but not agent-modifying adverbs (e.g., *'happily'*). Complementing these findings, Sato et al. (2013) found shape–sentence compatibility effects across pre-verbal, verb and postverbal regions, which were strongest when complete spatial information was available, typically at later stages of sentence processing. These

findings indicate that comprehenders integrate linguistic cues incrementally, continuously updating mental representations as information unfolds (Pickering & Gambi, 2018). These findings reinforce the dynamic and context-dependent nature of mental simulation.

While most studies on the timing of mental imagery focus on L1 speakers, fewer studies examine its temporal dynamics in L2 contexts. Nishide et al. (2025) applied a self-paced sentence–video verification task with Dutch and Japanese L2 English learners, reporting compatibility effects at both sentence-medial and sentence-final positions (e.g., *The/ ball/ rolled/ under/ [video]/ the/ car* vs. ...*under/ the/ car/ [video]*), with stronger effects when the video appeared at the sentence-final position. Similarly, Wang and Zhao (under review) showed increasing compatibility effects as diagrams appeared later in sentences among Mandarin–English bilinguals. These findings suggest that L2 learners engage in simulation incrementally rather than postponing it until spatial information is complete and benefit from more robust integration of perceptual information at later stages. This aligns with previous L1 studies (e.g., Sato et al., 2013; Taylor & Zwaan, 2008) and supports the incremental nature of imagery, where L2 learners form and refine visual representations as spatial information unfolds.

Existing research has implied that earlier sentence segments may not provide sufficient information to trigger robust imagery effects for both L1 and L2 speakers. However, given the complex and dynamic nature of L2 simulation, more empirical evidence is needed to explore the temporal dynamics of mental representations in L2 processing, particularly considering L2 learners' greater reliance on lexical-semantic cues in predictive processing (Ito et al., 2017).

1.4. The processing of English phrasal verbs

PVs are formulaic expressions combining a verb and a particle to form a single semantic unit, often marked by polysemy and idiomaticity. Their discontinuous structure, complex semantics and high polysemy encompassing both literal and abstract senses often pose challenges for L2 learners (Garnier & Schmitt, 2015).

In response, CL-inspired research has examined image schemas (Johnson, 1987) – fundamental cognitive structures (e.g., UP-DOWN, CONTAINER) derived from sensorimotor experiences – to support PV comprehension through spatial and metaphorical mappings (Mahpeykar & Tyler, 2015). Abstract senses often involve conceptual motion, where spatial and force dynamics are metaphorically mapped onto non-physical domains (Gibbs & Matlock, 2008). For example, the VERTICALITY schema explains *look up* both literally (as in physically looking upward) and figuratively (as in seeking information, metaphorically conceptualized as being stacked vertically). From an embodied cognition perspective, such abstract PVs may evoke simulation effects via activation of conceptual motion schemas (Bergen & Chang, 2005), even without physical motion. Schematic diagrams have been applied to visually represent spatiotemporal structures underlying image schemas and their corresponding linguistic expressions (Zhao et al., 2020).

CL-inspired experimental studies have shown that image-schema-based instruction enhances L2 PV learning outcomes (e.g., Hwang, 2023; Sato et al., 2022). However, far fewer studies have investigated mental imagery in PV comprehension, particularly incorporating schematic diagrams. In a self-paced sensibility judgment

task, Yang (2016) found semantically aligned diagrams (e.g., an ‘*up*’ diagram paired with the particle ‘*up*’) facilitating processing of both concrete and abstract PVs in L1 speakers. However, the validity of these compatibility effects is questioned by the study’s use of only four two-dimensional diagrams (lines and arrows) that likely fail to capture PVs’ semantic nuances, and the lack of norming procedures to verify PV–diagram alignment.

Additionally, Lindstromberg (2022) proposed that PVs vary along two semantic continua: from literal (i.e., concrete) to figurative (i.e., abstract) meanings, and from semantically compositional to non-compositional structures. Among these, semantic transparency – the extent to which a PV’s meaning can be inferred from its components – is a key internal feature, reflecting a graded spectrum of compositionality (Zhao & Le, 2016).

Empirical research shows that both L1 and L2 speakers are sensitive to semantic transparency, with literal PVs (e.g., *cover up*) generally perceived as more transparent than opaque ones (e.g., *chew out*), though this perception varies across items and individuals (Blais & Gonnerman, 2013). Transparent PVs, which retain clearer links between their components and overall meaning, have been shown to facilitate L2 comprehension (Zhao & Le, 2016) and elicit robust priming effects in lexical judgment tasks (e.g., *cover up* | *cover*). These findings suggest that transparency enhances access to compositional meaning by reducing the inferential burden during processing. From the mental imagery perspective, this more direct form–meaning mapping may better support the activation of perceptual representations during real-time comprehension. Within the L2 simulation-based model (Zhao et al., 2025), such effects likely emerge during the constructional analysis stage, where semantic transparency supports the retrieval of schematic structure and eases the activation of embodied representations. This underscores the need to move beyond categorical distinctions between abstract and concrete PVs and to examine how gradient semantic transparency shapes mental imagery in L2 learners’ processing.

1.5. The present study

Despite the extensive research on L1 mental imagery, L2 mental simulation, particularly in PV processing, remains largely underexplored and contested, with mixed evidence regarding the extent of embodiment. The simulation-based L2 comprehension model (Zhao et al., 2025) proposes that cognitive simulations grounded in real-world experiences operate similarly in L1 and L2 comprehension, while emphasizing the iterative and dynamic nature of L2 mental imagery shaped by linguistic and learner-specific factors. However, the impact of lexical-semantic properties of target constructions (i.e., PVs) and multidimensional learner factors on the dynamic formation of mental imagery has yet to be thoroughly examined.

This study addresses these gaps by investigating how schematic visual cues influence sentence-level comprehension of English PVs in L1 and L2 speakers. We used a self-paced sensibility judgment task interleaved with image-schematic diagrams to track incremental sentence processing. Schematic diagrams were chosen instead of realistic pictures because they can depict both concrete and abstract PV meanings with consistent spatial schemas (Wang & Zhao, 2024a); abstract PV senses (e.g., *pick up the melody*) are not easily or reliably illustrated with concrete images. This design provides fine-grained reaction time (RT) measures of online comprehension

Table 1. Language background of L2 participants

Measurement	Range	Mean	SD
L1 age of acquisition (AoA)	0–5	1.83	1.22
L1 proficiency	0.71–1	0.94	0.07
L1 exposure	0.73–1	0.92	0.06
L1 dominance	0.47–0.77	0.63	0.06
L2 AoA	3–13	8.11	2.40
L2 proficiency	0.32–0.82	0.61	0.12
L2 exposure	0.4–0.87	0.65	0.10
L2 dominance	0.18–0.56	0.35	0.08

and allows real-time assessment of mental imagery effects. These are advantages that alternative tasks such as post hoc sensibility judgment or simple SPVT cannot offer (cf. Sato et al., 2013; Yang, 2016). The schematic diagrams representing the literal or metaphorical motion implied by the PV were presented either immediately after the PV or after the sentence-final word to probe the temporal dynamics of mental imagery. PVs vary in semantic transparency across concrete and abstract senses, allowing us to examine how the lexical-semantic property and conceptual abstraction modulate mental imagery effects.

The following research questions guide the study:

- (1) Does diagram congruency facilitate sensibility judgments during incremental sentence processing of concrete and abstract PVs, and does this effect differ between L1 and L2 speakers?
- (2) How does the lexical-semantic property of PVs (i.e., semantic transparency) affect participants' sensitivity to diagram congruency in sentence comprehension?
- (3) How do individual learner differences, including L2 proficiency, AoA, L2 exposure and language dominance, influence L2 learners' sensitivity to diagram congruency in sensibility judgments?

2. Methodology

2.1. Participants

Fifty-six adult native English speakers (28 males, 27 females, 1 non-binary; mean age = 43.49, $SD = 6.63$) participated in the main experiment. They were recruited via Prolific and reported being right-handed with no literacy-related difficulties. All participants reported English as their mother tongue and were born and residing in the United Kingdom ($n = 19$), the United States ($n = 33$) or Canada ($n = 4$).

Sixty-three adult Chinese learners of L2 English (12 males, 51 females; mean age = 22.3, $SD = 1.62$) were recruited via Rednote, a Chinese social networking platform. L2 participants' English proficiency was indexed by their College English Test Band 4 (CET-4) scores, a nationally administered, large-scale standardized test and one of the most influential high-stakes English proficiency assessments in China (Zheng & Cheng, 2008). The test comprises four sections, including listening, reading, cloze/error correction and writing/translation, with a maximum score of 710. All L2 participants were college students in China, with an average CET-4 score of 538.11 ($SD = 61.64$), corresponding approximately to the B1–B2 range on the CEFR scale.

Participants completed the Language History Questionnaire (LHQ3.0, Li et al., 2020). Following Li et al.’s (2020) computing methods, Table 1 summarizes learner-related variables, including proficiency, exposure, dominance (standardized to a 0–1 scale) and self-reported AoA, which were all treated as continuous variables.

2.2. Design and materials

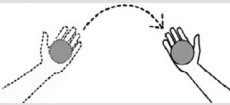
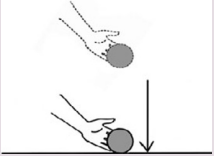
This study adopts a 2 × 2 × 2 × 2 (PV–diagram congruency × PV abstractness × sentence sensibility × temporal sequences) factorial design. PV–diagram congruency and PV abstractness (concrete vs. abstract senses) are categorical within-subject variables. Sequence (diagram presented after the PV vs. after the sentence) is a between-subject variable. PV semantic transparency is treated as a continuous within-subject variable. The experimental stimuli comprised 40 sentence pairs constructed from 20 PVs (Garnier & Schmitt, 2015), each used in both concrete and abstract senses, with sensible and nonsensical versions. Nonsensical sentences were syntactically well formed but semantically implausible, typically created by pairing PVs with incongruous or contextually inappropriate objects and adjuncts. All sentences were used in the simple past tense and active voice.

Table 2 shows sample sentence stimuli using ‘hand over’ and ‘take down’ in four conditions with corresponding diagrams. To reduce predictability and maintain task engagement, grammatical filler sentences (adapted from Ellis et al., 2009) with varied structures and plausibility were included. Full materials (PVs, sentence stimuli and fillers) are available in Supplementary Material 1, which details the strict multistep criteria used to select target PVs and reports their string frequency and collocational strength (Mutual Information, Church & Hanks, 1990).

2.2.1. Visual stimuli

Viewing verb–particle combinations as unified constructions, diagrams were designed and organized according to the spatial motion patterns underlying the

Table 2. Sample diagrams, senses and test sentences

Phrasal verb	Diagram	Abstractness	Sensibility	Sample sentence
hand over		Concrete	Sensible	The staff handed over the documents to the manager.
			Nonsensical	The staff handed over a mountain into the cupboard.
		Abstract	Sensible	The manager handed over the responsibilities to his staff.
			Nonsensical	The manager handed over social anxiety to the computer.
take down		Concrete	Sensible	The teacher took down a textbook from the bookcase.
			Nonsensical	The teacher took down a microwave from the curtain.
		Abstract	Sensible	The reporter took down some quotes in an interview.
			Nonsensical	The reporter took down the breeze in his notebook.

selected PVs (see [Supplementary Material 1](#) for selection rationale). Because many target PVs share a particle (e.g., *out*, *down*, *off*) and therefore a common image schema (OUT, DOWN, SEPARATION), several PVs draw on the same diagram, for instance, PVs instantiating the DOWN schema (*lay down*, *take down*, *put down*) and those instantiating the OUT schema (*lay out*, *give out*, *throw out*, *bring out*, *send out*). This overlap reflects the theoretical assumption that particle meaning is schematic and construction-based.

Accordingly, 13 schematic diagrams ([Appendix A](#)) were created to illustrate the literal and metaphorical motion configurations implied by the 20 target PVs. Within each diagram, dotted and solid lines distinguished prior and subsequent motion, and arrows indicated directionality. To ensure consistency, a small ball depicted the moving object, and a rectangle represented the CONTAINMENT schema. Body effectors (e.g., a hand or arm) were included only when needed to clarify the motion trajectories, and tools (e.g., a shovel for *break off*) were shown only when an external force was essential to convey the intended motion. Diagrams were primarily black-and-white line drawings, with limited accents of red, yellow and gray to enhance the perceptual salience of motion trajectories.

Seven additional filler diagrams ([Supplementary Material 1](#)) were included to illustrate semantic configurations of four modal verbs (*must*, *may*, *will*, *would*) (adapted from Tyler et al., 2010) and three tense variations (simple present, past and future). Each diagram was presented in either a congruent condition, visually matching the PV's conceptual motion, or an incongruent condition, visually conflicting with the PV's meaning. For incongruent trials, the diagram accompanying the sentence was randomly selected from either non-matching target diagrams or filler diagrams, which effectively minimized the likelihood that the specific diagram features would bias participants' judgments.

2.2.2. Norming study

A norming study was conducted online using Qualtrics (Qualtrics, 2025), including a sentence–diagram matching task and an event plausibility rating task. Twenty-two monolingual English speakers (11 males, 11 females; mean age = 42.73, $SD = 14.78$) were recruited via Prolific. Only sensible sentence stimuli were used in the norming study. To ensure enough critical stimuli and minimize plausibility bias, 20 PVs were initially used to generate 80 items (40 pairs crossing concrete/abstract senses), divided into two blocks of 40 sentences. Each participant completed one block, rating sentence–diagram compatibility (0 = *not at all matching*, 10 = *perfect match*) and event plausibility (0 = *extremely unlikely*, 10 = *extremely likely*) on ten-point Likert scales. Trials were randomized with no time limits or feedback. [Figure 1](#) presents a sample trial.

Based on mixed-effects modeling results, the study retained 40 high-quality sentence stimuli that are (i) well matched to their diagrams ($M_{\text{concrete}} = 7.11$, $SD = 2.78$; $M_{\text{abstract}} = 6.69$, $SD = 2.97$), with no significant difference across abstractness ($\beta = 0.45$, $SE = 0.32$, $t = 1.42$, 95 % CI $[-0.19, 1.09]$, $p = 0.17$); (ii) highly plausible ($M_{\text{concrete}} = 7.64$, $SD = 2.53$; $M_{\text{abstract}} = 7.21$, $SD = 2.55$), with comparable plausibility ratings for concrete and abstract items ($\beta = 0.43$, $SE = 0.22$, $t = 1.95$, 95 % CI $[-0.02, 0.87]$, $p = 0.06$); and (iii) evenly distributed over concrete and abstract categories. A sensitivity analysis in G*Power 3.1 (Faul et al., 2009) (t tests, difference between two dependent means, two-tailed, $\alpha = 0.05$, $N = 22$) indicated 80% power to detect a minimum effect of

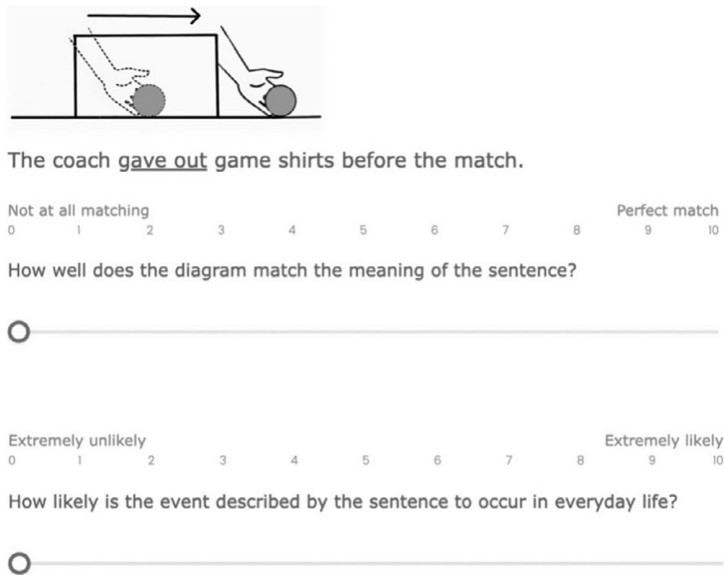


Figure 1. A sample trial in the norming study.

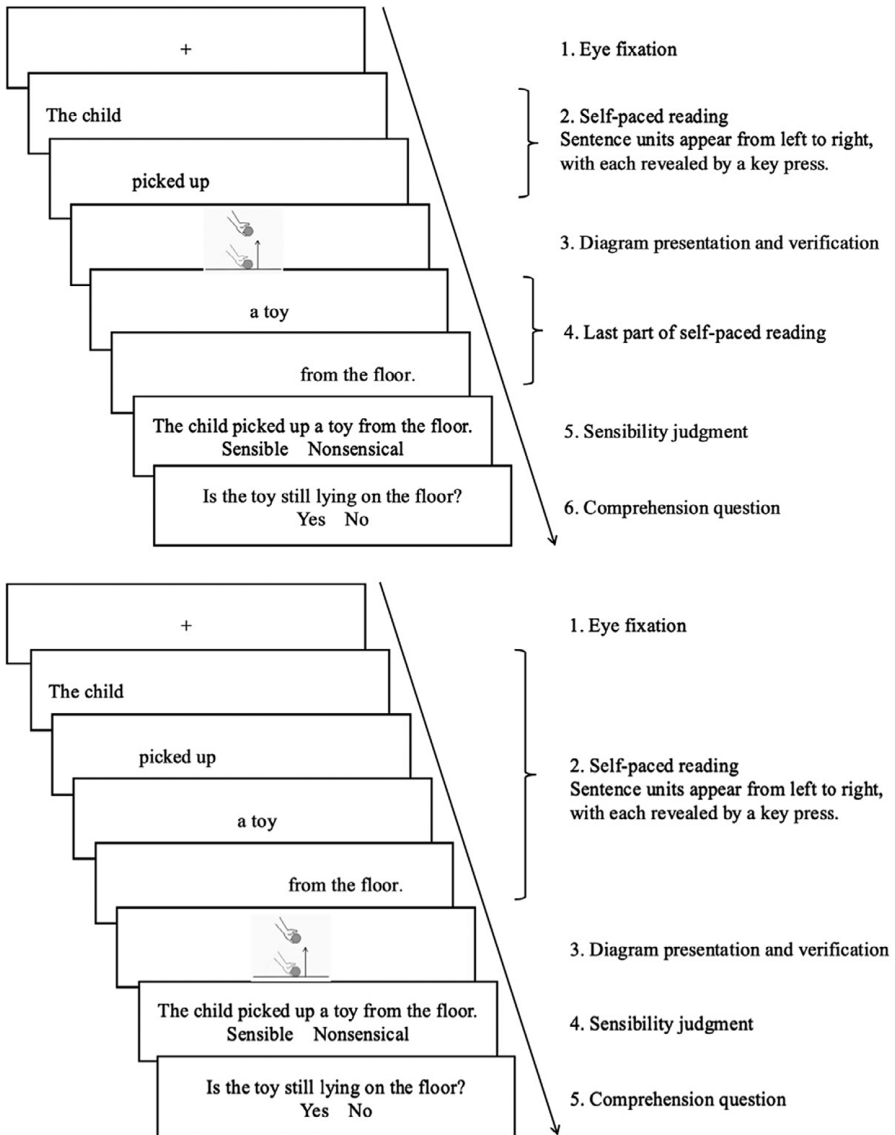
$d_z = 0.63$. These results confirm that the final stimuli set is both semantically plausible and visually congruent, supporting its validity for use in the subsequent experimental task.

2.2.3. Sensibility judgment task

The main experiment used a self-paced sensibility judgment task interleaved with diagram presentation, involving two blocks: diagrams appeared either after the PV or after the sentence-final word. Following Sato et al. (2013), each trial began with a central fixation cross, followed by a moving-window paradigm where participants pressed the spacebar to reveal sentence segments sequentially, each replacing the previous one to its right. Upon diagram presentation, participants made a diagram verification judgment to indicate whether the diagram matched the sentence meaning, allowing assessment of imagery accuracy. After completing each sentence reading, the full sentence reappeared for a sensibility judgment ('I' = sensible, 'F' = nonsensical) within 10,000 ms. For sensible sentences, a comprehension question followed, displayed until response. Accuracy was recorded for both sensibility judgments and comprehension responses. Participants completed seven practice trials with feedback before the formal task. Figure 2 illustrates a sample trial and two diagram presentation sequences.

2.2.4. Diagram familiarization

Before the sensibility judgment task, a diagram familiarization task was implemented to support participants' interpretation of abstract schematic diagrams. Thirteen animated GIFs were presented alongside their corresponding PV diagrams to strengthen intuitive form–meaning mappings and minimize visual ambiguity. Participants were required to review all pairings before proceeding the sensibility



judgment task. Figure 3 displays a sample diagram familiarization phase, with schematic diagrams on the left and motion animations on the right.

2.2.5. Semantic transparency rating task

Following the sensibility judgment task, all participants undertook an offline semantic transparency rating task. Materials were identical to the main task, with the target PVs underlined. Each participant was randomly assigned to one of two rating blocks

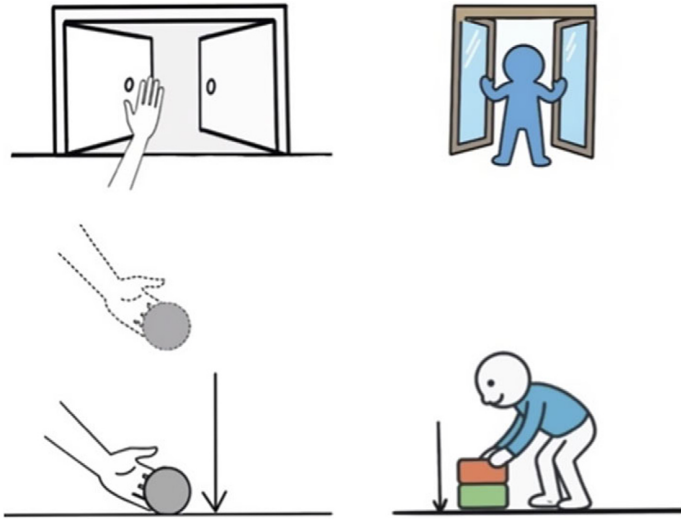


Figure 3. Example display from diagram familiarization task.

containing 20 sentences balanced for concrete and abstract items. Participants rated PV transparency on a 0–10 Likert scale, indicating how easily each PV’s meaning could be inferred from its verb and particle. Item-level transparency ratings were averaged and included as a continuous within-subject variable in the analysis.

2.3. Procedure

A digital informed consent form and a plain language statement were provided to participants prior to participation. The experiment began with a diagram familiarization phase using 13 animated GIFs, followed by the main self-paced sensibility judgment task implemented via PsyToolkit (version 3.6.2; Stoet, 2010, 2017). Participants were randomly assigned to one of two blocks: Sequence 1 (33 L1 and 31 L2 speakers), with the diagram presented after the PV, or Sequence 2 (23 L1 and 32 L2 speakers), with the diagram presented after the sentence. Within each block, participants were randomly assigned to one of eight counterbalanced lists, each containing 55 sentences (40 critical items and 15 fillers). After the main task, participants completed the PV transparency rating task via Qualtrics and the LHQ (Li et al., 2020). The entire session lasted approximately 20 minutes for L1 participants and 30 minutes for L2 participants.

2.4. Statistical analysis

Mental imagery effects were calculated as the RT difference between congruent and incongruent PV–diagram pairs. Data were analyzed in R software (version 4.5.0; R Core Team, 2025), using linear mixed-effects models with the *lme4* (version 1.1.37; Bates et al., 2025) and *lmerTest* (version 3.1.3; Kuznetsova et al., 2020) packages. Only correct sensibility judgments with RTs (200–8000 ms) were included and log-transformed. Trials falling outside this conventional range (< 200 ms or > 8000 ms)

for online self-paced sentence tasks and sensibility judgments were excluded as implausible responses, eliminating 2.68% of data overall ($L1 = 1.56\%$; $L2 = 3.69\%$). Exclusion rates were low and did not interact with congruency. Categorical variables (congruency, abstractness, sequence, sensibility) were treatment-coded.

The final models addressing RQ1 were fitted separately for L1 and L2 groups¹, including fixed effects of congruency, abstractness, sequence, sensibility and their interactions, with random intercepts for participants and items. For RQ2, models for each group included fixed effects of congruency and its interaction with covariates, including PVs' transparency, string frequency, collocational frequency, event plausibility and diagram–PV matching ratings. For RQ3, models tested each L2 learner factor (proficiency, AoA, dominance, exposure) interacting with congruency and sensibility. Random intercepts and participant-level slopes for congruency were retained when justified.

Fixed effects were evaluated using type III ANOVA with Satterthwaite's approximation for degrees of freedom and p values (Kuznetsova et al., 2017). Post hoc comparisons were conducted with *emmeans* (version 1.11.1; Lenth et al., 2025) and visualizations generated via *ggplot2* (version 3.5.2; Wickham et al., 2025). Effect sizes (Cohen's d) followed Plonsky and Oswald's (2014) interpretation: small (0.60), medium (1.00), large (1.40) for within-subjects comparisons; small (0.40), medium (0.70), large (1.00) for between-subjects comparisons. Full model formulas and outputs are reported in [Supplementary Material 2](#).

3. Results

3.1. Descriptive statistics of reaction times and accuracy rates

[Tables 3](#) and [4](#) present descriptive statistics for diagram verification and sensibility judgments across nativeness, abstractness, sequence and congruency conditions. The diagram verification served primarily as a manipulation check to confirm that participants attended to the diagrams and understood their intended mappings. In this procedure, L1 and L2 participants spent approximately 2000 ms to verify the diagram. Importantly, mean verification accuracy for both groups exceeded 70%, which is significantly above chance level and consistent with previous sentence–diagram verification studies (e.g., Wang & Zhao, 2024a, 2024b). However, no robust sentence–diagram compatibility effect emerged in the verification procedure. Accordingly, we interpret the verification as confirming adequate attention to and comprehension of the diagrams, but not as providing a reliable index of imagery-based compatibility².

¹To examine whether age differences influenced the results, we fit a model testing nativeness $\times$ age $\times$ congruency: $\log(\text{sensibilityRT}) \sim \text{nativeness} * \text{congruency} * \text{zAge} + \text{abstractness} * \text{sequence} * \text{sensibility} + (1 | \text{participant}) + (1 | \text{item})$. The model revealed no significant two-way or three-way interactions involving nativeness, age and congruency (all $ps > 0.14$), indicating that group differences in congruency effects are unlikely to be confounded by age difference. Similarly, to examine gender composition differences, we fit a model testing nativeness $\times$ gender $\times$ congruency: $\log(\text{sensibilityRT}) \sim \text{nativeness} * \text{congruency} * \text{gender} + \text{abstractness} * \text{sequence} * \text{sensibility} + (1 | \text{participant}) + (1 | \text{item})$. Results showed no significant two- or three-way interaction involving nativeness, gender and congruency (all $ps > 0.38$), indicating that L1–L2 congruency effects are not explained by gender.

²The diagram verification data did not yield a robust compatibility effect. As shown in [Table 3](#), L2 learners responded more slowly to congruent than to incongruent diagrams in Sequence 2. This pattern is consistent

Table 3. Diagram verification: Accuracy and response times by congruency × abstractness × sequence (L1 vs. L2)

Congruency	Abstractness	Sequence	L1 group		L2 group	
			Accuracy	RTs	Accuracy	RTs
Congruent	Abstract	1	0.77 (0.42)	2053.91 (1021.82)	0.73 (0.45)	2021.01 (1089.00)
		2	0.68 (0.47)	2149.51 (1029.93)	0.60 (0.49)	2237.85 (1157.56)
	Concrete	1	0.84 (0.37)	1975.54 (973.47)	0.79 (0.41)	2172.35 (1073.62)
		2	0.77 (0.42)	2170.77 (1053.48)	0.66 (0.47)	2168.94 (990.22)
Incongruent	Abstract	1	0.70 (0.46)	1991.32 (1045.63)	0.75 (0.43)	2010.54 (993.82)
		2	0.62 (0.49)	2136.13 (999.57)	0.74 (0.44)	2005.31 (964.56)
	Concrete	1	0.73 (0.45)	1925.89 (958.23)	0.71 (0.45)	2095.38 (1115.52)
		2	0.70 (0.46)	2101.17 (945.56)	0.72 (0.45)	1968.22 (960.22)

Table 4. Sensibility judgments: Accuracy and response times by congruency × abstractness × sequence (L1 vs. L2)

Congruency	Abstractness	Sequence	L1 group		L2 group	
			Accuracy	RTs	Accuracy	RTs
Congruent	Abstract	1	0.85 (0.36)	2255.22 (1280.68)	0.72 (0.45)	3056.06 (1812.40)
		2	0.82 (0.39)	2261.48 (1314.58)	0.76 (0.43)	2787.52 (1700.46)
	Concrete	1	0.91 (0.28)	2233.53 (1400.14)	0.81 (0.39)	2994.25 (1866.29)
		2	0.93 (0.26)	1955.81 (1000.82)	0.77 (0.42)	2578.18 (1569.38)
Incongruent	Abstract	1	0.86 (0.35)	2434.07 (1467.58)	0.69 (0.46)	3130.32 (1732.84)
		2	0.85 (0.36)	2329.40 (1345.17)	0.62 (0.49)	3044.02 (1754.13)
	Concrete	1	0.94 (0.23)	2001.11 (1213.24)	0.80 (0.40)	3002.84 (1667.12)
		2	0.89 (0.31)	2458.38 (1364.47)	0.70 (0.46)	2941.44 (1737.57)

For sensibility judgments, RTs of correct responses and accuracy rates were presented. Both groups responded faster in congruent trials. L2 learners showed higher accuracy in congruent trials compared to incongruent ones, whereas L1 speakers demonstrated similar accuracy across conditions. Overall, L1 speakers made faster and more accurate judgments than L2 learners. Additionally, both groups achieved high comprehension accuracy (L1: $M = 0.86$, $SD = 0.34$; L2: $M = 0.80$, $SD = 0.40$), confirming adequate task engagement and semantic understanding. Given the timed nature of the task, RTs serve as the primary dependent variable for assessing processing dynamics. Figure 4 provides violin plots depicting RTs and accuracy rates in diagram verification and sensibility judgment tasks, highlighting performance differences across language groups (L1 vs. L2) and congruency conditions.

with prior reports of interference effects in L2 mental imagery when schematic diagrams must be integrated with linguistic input during online processing (Wang & Zhao, 2024a, 2024b). Such interference has been attributed to mutual inhibition between simultaneously activated visual and linguistic representations, higher cognitive demands for multimodal integration in L2 and weaker L2 lexical-conceptual mappings. In the present study, variations in diagram distinctiveness (shared schemas, presence/absence of body effectors or tools, use of generic objects) may have further limited the verification procedure’s sensitivity to imagery-based compatibility. As diagram verification served only as an attention check rather than a diagnostic measure of simulation, its RT patterns are not used to infer imagery effects. Interpretive claims derive from the sensibility judgment data, where compatibility reliably emerged.

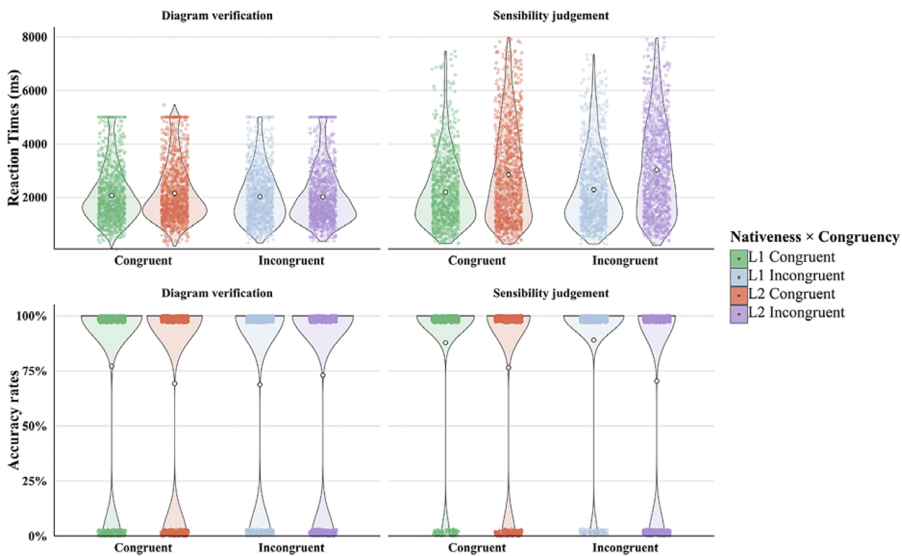


Figure 4. Violin plots of reaction times and accuracy rates in diagram verification and sensibility judgement tasks.

Furthermore, since accuracy rates were well above chance level and did not exhibit strong variability across congruency conditions, subsequent inferential analyses focus on RTs of correct responses to capture subtle processing differences.

3.2. Congruency facilitation effects on L1 and L2 sensibility judgments

Two linear mixed-effects models were performed to analyze the effects of congruency, sensibility, abstractness, sequence and their interactions on sensibility judgment RTs for L1 and L2 groups. Table 5 summarizes the ANOVA results for the L1 and L2 groups. In the L1 model³, results showed significant fixed effects of abstractness, with abstract PVs eliciting slower responses than concrete PVs ($\beta_{\text{abstract-concrete}} = 0.19$, $SE = 0.07$, 95% CI [0.059, 0.327], $t = 2.868$, $p = 0.005$; Cohen's $d = 0.40$, $SE = 0.141$, 95% CI [0.127, 0.678], corresponding to a small effect). Also, there is a significant two-way interaction between congruency and sequence. Pairwise comparisons showed that the congruency facilitation effect emerged only in Sequence 2 (see Figure 5): participants responded 199 ms significantly faster to the congruent trials ($M = 1801$, $SE = 130$, 95% CI [1559, 2081]) than incongruent trials ($M = 2000$, $SE = 145$, 95% CI [1731, 2311]; $t = -2.90$, $p = 0.004$, Cohen's $d = 0.46$, $SE = 0.16$, 95% CI [-0.76, -0.15], indicating a small compatibility effect). No significant congruency difference was observed in Sequence 1 ($p = 0.267$).

Additionally, the L1 model³ yielded a significant three-way interaction between congruency, abstractness and sequence. Post hoc tests revealed that the congruency facilitation effect occurred only for concrete PVs in both Sequence 1 ($t = 5.7$, $p = 0.008$, Cohen's $d = 0.49$, $SE = 0.17$, 95% CI [0.15, 0.83], indicating a small effect) and Sequence 2 ($t = -8.47$, $p = 0.004$, Cohen's $d = 0.91$, $SE = 0.22$, 95% CI [-1.34, -0.48],

³Model equation: $\log(\text{RT}) \sim \text{congruency} * \text{abstractness} * \text{sequence} * \text{sensibility} + (1 \mid \text{participant}) + (1 \mid \text{item})$.

Table 5. Summary of ANOVA results of congruency, abstractness, sensibility and sequence for L1 and L2 groups

Group	Effect	<i>F</i> (df1, df2)	<i>p</i>
L1	Congruency	2.44 (1, 1852)	0.12
	Abstractness	8.59 (1, 82)	0.004 **
	Sensibility	0.41 (1, 78)	0.53
	Sequence	0.29 (1, 56)	0.59
	Congruency × sequence	8.83 (1, 1814)	0.003 **
	Congruency × abstractness × sequence	15.91 (1, 1815)	< 0.001 ***
L2	Congruency	12.74 (1, 1681)	< 0.001 ***
	Abstractness	4.15 (1, 83)	0.045 *
	Sensibility	21.43 (1, 82)	< 0.001 ***
	Sequence	1.91 (1, 63)	0.17

Note. **p* < .05, ***p* < .01, ****p* < .001.

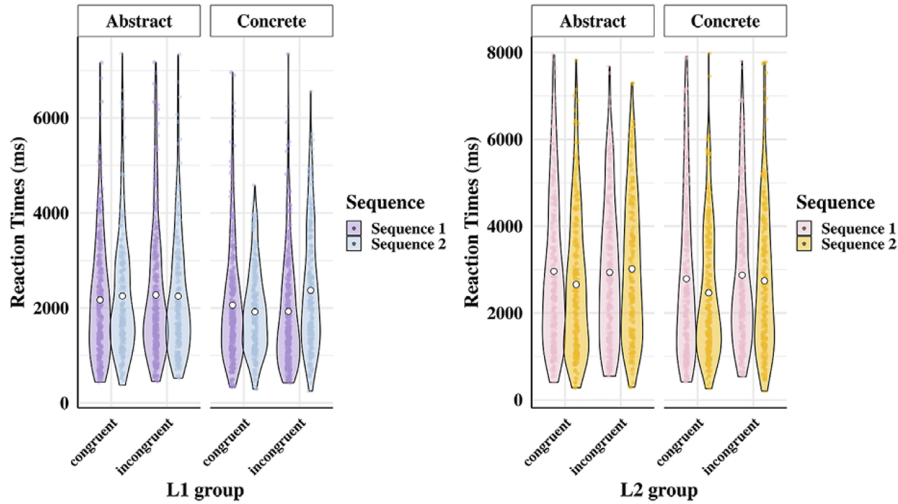


Figure 5. L1 and L2 reaction times across congruency, abstractness and sequences.

indicating a medium to large effect). No significant congruency effects were found for abstract PVs ($p > 0.23$).

In the L2 model⁴, significant main effects were found for congruency, abstractness and sensibility, but not for sequence. Post hoc analyses indicated that L2 learners judged 250 ms faster on congruent trials ($M = 2302$, $SE = 104$, 95% CI [2105, 2518]) than incongruent trials ($M = 2552$, $SE = 116$, 95% CI [2331, 2792]; $t = -3.56$, $p < 0.001$, Cohen's $d = 0.33$, $SE = 0.092$, 95% CI [-0.505, -0.146], indicating a very small effect). For the significant effect of abstractness, L2 learners judged 205 ms slower for abstract PVs ($M = 2528$, $SE = 122$, 95% CI [2298, 2782]) than concrete PVs ($M = 2323$, $SE = 111$, 95% CI [2114, 2553]; $t = 1.99$, $p = 0.049$, Cohen's $d = 0.27$, $SE = 0.134$, 95% CI [0.004, 0.53], corresponding to a very small effect). For the sensibility effects, post hoc results revealed that mean RTs in sensible sentences ($M = 2203$, $SE = 103$, 95% CI [2009, 2416]) were 464 ms shorter than in nonsensical sentences ($M = 2667$, $SE = 131$, 95% CI [2420, 2939]; $t = -4.53$, $p < 0.0001$, Cohen's $d = 0.61$,

⁴Model equation: $\log(RT) \sim \text{congruency} * \text{abstractness} * \text{sequence} * \text{sensibility} + (1 | \text{participant}) + (1 | \text{item})$.

$SE = 0.134$, 95% CI $[-0.867, -0.342]$, indicating a small to medium effect). No significant two-way or three-way interactions were found.

3.3. Lexical-semantic properties

According to the linear model⁵ for the L1 group, there were significant effects of PV string frequency and L1 transparency ratings. Table 6 presents a summary of lexical property effects for L1 and L2 groups. Follow-up post hoc analyses showed that higher string frequency ($\beta = -0.043$, $SE = 0.015$, $t = -2.86$, $p = 0.004$) and greater transparency ($\beta = -0.033$, $SE = 0.016$, $t = -2.14$, $p = 0.032$) were associated with shorter RTs. As illustrated in Figure 6, the model⁵ yielded a significant interaction

Table 6. Summary of ANOVA results for effects of lexical properties in L1 and L2 groups

Group	Effect	F (df1, df2)	p
L1	Congruency	0.36 (1, 1010)	0.548
	L1 PV transparency	4.60 (1, 1000)	0.032 *
	String frequency	8.18 (1, 994)	0.004 **
	Plausibility rating	1.48 (1, 1009)	0.224
	Congruency $\times$ L1 transparency	5.92 (1, 1027)	0.015 *
	Congruency $\times$ string frequency	0.63 (1, 991)	0.428
	Congruency $\times$ plausibility	3.31 (1, 991)	0.069
	Congruency	0.0001 (1, 1036)	0.991
L2	L1 transparency	1.93 (1, 38)	0.173
	Plausibility rating	4.50 (1, 37)	0.041 *
	PV–diagram matching rating	4.79 (1, 44)	0.034 *
	Collocational frequency	4.48 (1, 39)	0.041 *
	Congruency		

Note. * $p < .05$, ** $p < .01$, *** $p < .001$.

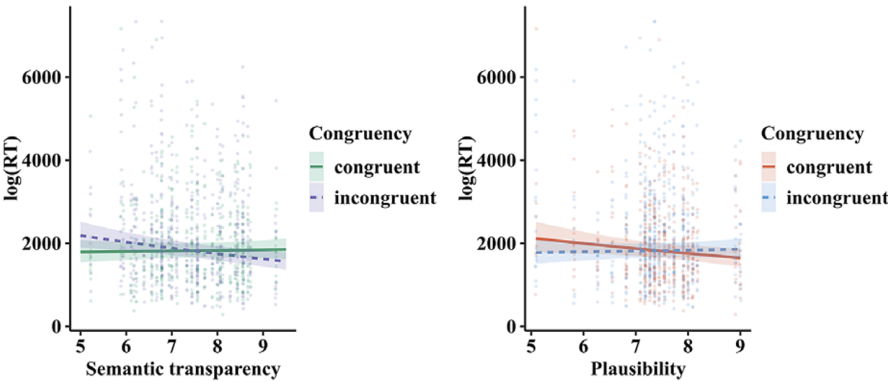


Figure 6. Interaction of congruency with PV transparency and event plausibility.

⁵Model equation: $\log(RT) \sim \text{plausibility_rating} * \text{congruency} + \text{L1_transparency} * \text{congruency} + \log(\text{string_frequency}) * \text{congruency} + (1 | \text{participant})$.

We compared model⁵ using L1 speakers' context-specific transparency ratings ($AIC = 1339.98$, $BIC = 1389.50$) with a model using non-specific ratings (termed as 'literality') adapted from Lindstromberg (2022) ($\text{transparency_L1non-specific}$, $AIC = 1347.78$, $BIC = 1397.31$). The context-specific ratings yielded better model fit.

between congruency and PV transparency. Higher transparency predicted significantly faster RTs in incongruent trials ($\beta = -0.075$, $SE = 0.023$, $t = -3.32$, $p = 0.001$), but no significant effects in congruent trials ($t = 0.32$, $p = 0.75$). An interaction between plausibility and congruency approached significance, with higher plausibility predicting faster RTs in congruent trials ($\beta = -0.064$, $SE = 0.031$, $t = -2.04$, $p = 0.042$), but not in incongruent trials ($p = 0.694$).

In the model⁶ investigating the effect of lexical properties in the L2 group, results showed significant main effects of event plausibility ratings, PV–diagram matching ratings and collocational frequency. Post hoc analyses revealed that higher-rated plausible sentences were marginally associated with faster RTs ($\beta = -0.08$, $SE = 0.04$, $t = -2.00$, $p = 0.052$). Similarly, there was a marginal trend indicating that L2 learners responded faster to sentences with higher PVs' collocational frequency ($\beta = -0.02$, $SE = 0.01$, $t = -2.00$, $p = 0.052$). Additionally, higher PV–diagram matching ratings predicted significantly faster RTs ($\beta = -0.06$, $SE = 0.03$, $t = -2.07$, $p = 0.04$). However, no significant interactions between lexical properties and congruency were found in the model⁶.

3.4. Individual difference factors

Table 7 provides a summary of ANOVA results examining bilingual experience effects. In the model⁷ concerning L2 proficiency, results showed a significant main effect of L2 proficiency and a three-way interaction between congruency, sensibility and proficiency. Follow-up analyses showed that higher L2 proficiency led to faster RTs ($\beta = -0.661$, $SE = 0.312$, $t = -2.117$, $p = 0.038$). Moreover, post hoc analyses of the three-way interaction indicated that high-proficiency learners responded significantly faster in incongruent-sensible trials ($\beta = -1.14$, $SE = 0.36$, $t = -3.16$, $p = 0.002$), but not in congruent-sensible trials ($\beta = -1.14$, $p = 0.38$). For nonsensical sentences, a significant proficiency effect was observed only in congruent trials ($\beta = -1.06$, $SE = 0.39$, $t = -2.73$, $p = 0.007$), indicating that higher proficiency was associated with faster responses. However, the effect was absent in incongruent-nonsensical trials ($\beta = -0.12$, $p = 0.79$).

In the model⁸ examining L2 dominance, results showed a significant main effect of L2 dominance, with higher L2 dominance predicting faster sensibility judgments ($\beta = -1.01$, $SE = 0.5$, $t = -2.017$, $p = 0.048$). Additionally, a significant three-way interaction emerged among dominance, congruency and sensibility. Pairwise comparisons revealed that greater L2 dominance led to significantly faster responses to incongruent-sensible sentences ($\beta = -1.69$, $SE = 0.58$, $t = 2.9$, $p = 0.005$). In contrast, for nonsensical sentences, higher dominance facilitated faster responses only in congruent trials ($\beta = -1.52$, $SE = 0.65$, $t = -2.34$, $p = 0.019$), with no significant effect in incongruent-nonsensical trials. Figure 7 presents the interaction effects between both learner-internal factors and congruency. The model⁹ concerning L2 AoA revealed no significant main effects or interaction effects.

⁶Model equation: $\log(\text{RT}) \sim \text{plausibility_rating} * \text{congruency} + \text{L1_transparency} * \text{congruency} + \text{match_rating} * \text{congruency} + \text{collo_freq} * \text{congruency} + (1 \mid \text{participant}) + (1 \mid \text{item})$.

⁷Model equation: $\log(\text{RT}) \sim \text{congruency} * \text{sensibility} * \text{L2proficiency} + (1 \mid \text{participant}) + (1 \mid \text{item})$.

⁸Model equation: $\log(\text{RT}) \sim \text{congruency} * \text{sensibility} * \text{L2dominance} + (1 \mid \text{participant}) + (1 \mid \text{item})$.

⁹Model equation: $\log(\text{RT}) \sim \text{congruency} * \text{sensibility} * \text{L2AoA} + (1 \mid \text{participant}) + (1 \mid \text{item})$.

Table 7. Summary of ANOVA results for effects of learner differences

Model	Fixed effect	<i>F</i> (df1, df2)	<i>p</i>
L2 proficiency	Congruency	0.10 (1, 1692)	0.756
	Sensibility	0.35 (1, 1603)	0.555
	L2 proficiency	4.64 (1, 64)	0.035 *
	Congruency × L2 proficiency	0.08 (1, 1690)	0.774
	Congruency × sensibility × L2 proficiency	7.36 (1, 1220)	0.007 **
L2 dominance	Congruency	0.12 (1, 1697)	0.725
	Sensibility	1.23 (1, 1555)	0.267
	L2 dominance	4.21 (1, 63)	0.044 *
	Congruency × L2 dominance	0.10 (1, 1697)	0.752
	Congruency × sensibility × L2 dominance	8.95 (1, 1508)	0.003 **
L2 age of acquisition (AoA)	Congruency	0.07 (1, 1685)	0.788
	Sensibility	2.47 (1, 1186)	0.117
	L2 AoA	0.23 (1, 64)	0.632
	Congruency × L2 AoA	0.37 (1, 1684)	0.545
	Congruency × sensibility × L2 AoA	2.21 (1, 1013)	0.138
L2 exposure	Congruency	3.94 (1, 1686)	0.047 *
	Sensibility	1.95 (1, 1709)	0.163
	L2 exposure	0.07 (1, 63)	0.787
	Congruency × L2 exposure	6.16 (1, 1686)	0.013 *

Note. **p* < .05, ***p* < .01, ****p* < .001.

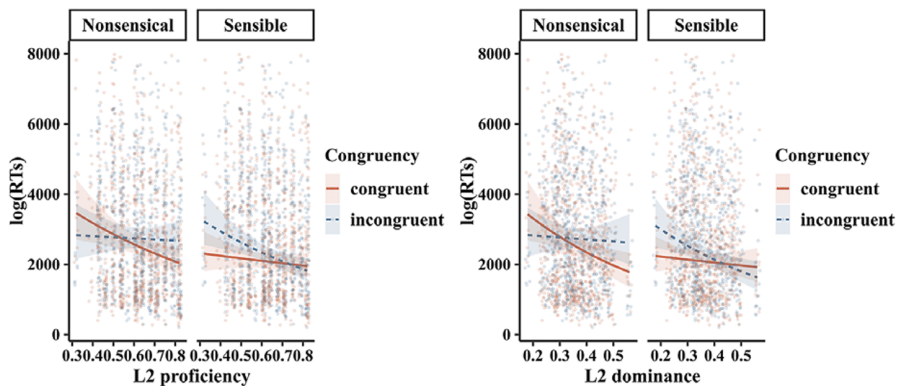


Figure 7. Interaction effects between learner-internal factors and congruency.

In the model¹⁰ examining L2 exposure, no significant main effect of exposure itself was found. However, a significant interaction between L2 exposure and congruency emerged. Post hoc comparisons revealed no significant RT differences between congruent and incongruent conditions across L2 exposure levels (*ps* > 0.27).

4. Discussion

4.1. RQ1: L1–L2 mental imagery effects

Both L1 and L2 speakers showed facilitation when a schematic diagram matched the PV during the sensibility judgment procedure, echoing earlier motion-imagery

¹⁰Model equation: $\log(\text{RT}) \sim \text{congruency} * \text{sensibility} * \text{L2exposure} + (1 | \text{participant}) + (1 | \text{item})$.

research (e.g., Dudschig et al., 2014; Hauf et al., 2020; Tomczak & Ewert, 2015), but our L1–L2 patterns differed markedly. L1 speakers showed facilitation only for concrete PVs and only when the diagram appeared after the sentence, whereas L2 speakers showed robust congruency effects across PV types and diagram presentation sequences. For L1 speakers, these findings are consistent with the recruitment of spatial-path imagery during sensibility-based judgment: mental representations of movement trajectories and spatial configurations implied by the verb–particle combination. Facilitation emerged only when diagrams followed the full sentence (Sequence 2), indicating that L1 comprehenders tend to first construct a complete semantic representation before integrating visual-spatial cues as confirmatory input. Presenting the diagram mid-sentence may interrupt or interfere with ongoing sentence integration, thereby reducing imagery-based facilitation. This sequence-sensitive pattern accords with prior work showing that full sentential context supports more robust simulation (e.g., Nishide et al., 2025; Sato et al., 2013).

The selective effect for concrete PVs aligns with embodied cognition literature, where spatial mental imagery is selectively recruited for spatially grounded expressions but less consistently for abstract language (Bergen & Chang, 2005; Bergen et al., 2007; Bergen & Wheeler, 2010). Importantly, the present effects differ from the motor-effector simulations reported in studies of action-verb processing, where participants reenact specific bodily movements (e.g., De Grauwe et al., 2014; Xue et al., 2015). Whereas those investigations targeted verbs like *kick* or *grasp* and documented activation in motor and sensorimotor brain areas, our schematic-diagram paradigm probed PVs whose concrete senses primarily encode spatial change-of-location (e.g., *throw out*, *take down*). The selective facilitation we observed thus reflects imagery of spatial configurations, not detailed motor-action reenactment. The absence of facilitation for abstract PVs further supports claims that abstract language evokes weaker or more temporally extended simulation processes requiring greater contextual integration (Liu & Bergen, 2016).

L2 learners, by contrast, showed generalized congruency effects for both concrete and abstract PVs and across diagram presentation sequences. This broader reliance on diagram congruency suggests task-constrained imagery that is more strongly anchored in external visual cues. This finding likely reflects reduced automaticity and weaker lexical-semantic entrenchment in L2 (Zhao et al., 2025), prompting learners to recruit visual input to support meaning construction within the demands of the sensibility judgment task. Comparable compensatory recruitment of sensorimotor resources during L2 comprehension has been observed in neuroimaging research: Tian et al. (2020) reported greater motor cortex activation during L2 processing, while Tian et al. (2023) found that language-related regions (e.g., insular gyri, superior temporal gyrus) were more active in L1 during early semantic processing (~300–500 ms), whereas L2 processing showed stronger motor activation (central sulcus) at a later stage (~600–800 ms). They attributed the increased motor activity during L2 processing to learners' compensatory recruitment of sensorimotor regions, arising from less efficient or incomplete activation of traditional language processing networks and reduced processing automaticity. The present finding that L2 speakers relied broadly on diagram congruency echoes this interpretation: in the context of sensibility judgment, the schematic diagrams, regardless of whether the accompanying PVs expressed a concrete or an abstract meaning or of when the diagram appeared, served as external visual-spatial scaffolding that L2 learners recruited to support task-driven imagery formation, complementing their less entrenched lexical-semantic representations.

4.2. RQ2: Lexical-semantic cue weighting in L1 and L2 imagery

Greater semantic transparency of PVs led to faster sensibility judgment in diagram-incongruent trials, suggesting that transparent PVs offer more decomposable form–function mappings and evoke more accessible mental imagery, which helped L1 speakers to detect incongruence efficiently. Within Zhao et al. (2025) simulation-based L2 comprehension model, this transparency advantage reflects more effective Stage 1 constructional analysis, which supports downstream embodied representation at Stage 3. This result aligns with embodied and usage-based models of language processing (Barsalou, 2008; Zwaan, 2008), which propose that meaning comprehension involves simulating sensorimotor and conceptual content. The ability to resolve conflict through lexical-semantic information also supports accounts of flexible cue integration in L1 processing (Altmann & Mirković, 2009; MacWhinney, 2005; Trueswell & Tanenhaus, 1994), whereby comprehenders adaptively weigh lexical, structural or contextual cues depending on the reliability of available information.

In contrast, higher plausibility ratings predicted faster sensibility judgment only in congruent trials, suggesting that L1 speakers draw more heavily on Stage 2 contextual integration when visual and verbal information align. Together, these findings demonstrate L1 speakers' flexible strategy adaptation based on visual-verbal congruency: prioritizing semantic plausibility when inputs align and reverting to lexical-semantic cues like PV transparency under conflict. This adaptability supports theories of adaptive simulation and predictive processing (Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018), where cue use is guided by input coherence and task demands.

L2 learners presented a different profile. They showed a strong diagram-congruency effect across all conditions, independent of PV transparency, PV–diagram matching ratings, event plausibility or collocational frequency. Their lack of sensitivity to transparency – a lexical property important for Stage 1 constructional processing – suggests difficulty in accurately identifying and mapping L2 PV constructions, especially given the lack of direct L1-Chinese equivalents. This finding echoes Tiv et al. (2019), who observed that real-time L2 sentence reading was largely unaffected by transparency, probably because learners depend on surface-level compositional strategies and underdeveloped lexicalized representations. In contrast, Blais and Gonerman (2013) found transparency effects in L2 decontextualized tasks such as semantic-similarity ratings and lexical priming (*finish up* – *finish*), tasks that tapped shallow lexical access with minimal contextual load, whereas our sentence-level comprehension task required deeper integration of visual and verbal information.

Similarly, the absence of any plausibility-by-congruency interaction suggests weaker Stage 2 contextual integration in L2 processing, likely due to limited discourse exposure or underdeveloped world knowledge. This constraint may hinder L2 learners' ability to construct coherent mental models that integrate linguistic and situational information, especially for semantically complex expressions like PVs. Their strong but unmodulated diagram congruency effect suggests that L2 representations in this task were largely perceptually anchored: external diagrams trigger representation building, while internal lexical-semantic and contextual integrations remain limited. This interpretation aligns with the simulation-based model's proposal that native-like Stage 3 embodied representations depend on robust constructional and contextual processing at earlier stages, which may be underdeveloped for many L2 learners.

Altogether, these findings show that in sensibility judgment, L1 speakers flexibly integrate lexical-semantic and contextual cues, drawing on PV transparency when visual

and verbal inputs conflict. L2 learners rely consistently on diagram congruency and exhibit weaker integration of semantic and contextual information – a pattern that aligns with bilingual embodiment research reporting weaker or more variable sensorimotor activation in L2 processing (e.g., Lu & Yang, 2025; Wei et al., 2024; Xue et al., 2024).

4.3. *RQ3: Influence of learner-specific and contextual factors*

L2 proficiency and language dominance enhanced learners' sensitivity to diagram incongruency, enabling faster sensibility judgments when visual and verbal information conflicted. Learners with higher proficiency or greater dominance integrated lexical-semantic and perceptual cues more efficiently under these demanding conditions, but this advantage disappeared in congruent trials where visual-verbal alignment already supported comprehension. Thus, while proficiency and dominance aid conflict resolution, their influence diminishes when inputs are fully coherent.

Overall, these findings align with prior research (e.g., Birba et al., 2020; Nishide et al., 2025), showing that higher L2 proficiency supports sensorimotor embodiment and efficient integration of linguistic knowledge with embodied mental models (Zhao et al., 2025). L2 proficiency, therefore, not only shapes linguistic and executive functioning but also modulates sensorimotor grounding (Kogan et al., 2020), explaining the observed proficiency-driven imagery effects. Greater L2 dominance, reflecting broader language usage, similarly fosters grounded representations that aid comprehension. However, no facilitation emerged for nonsensical sentences with incongruent diagrams, indicating that proficiency and dominance offer limited compensatory benefits under high semantic-perceptual conflict.

No significant effects of L2 AoA on mental imagery were found, possibly due to limited variability in our sample, as most participants were sequential bilinguals who began learning English around age eight. Although earlier AoA is typically linked to stronger L2 embodiment (Lu & Yang, 2025; Norman & Peleg, 2022), our results indicate that sensorimotor recruitment in L2 processing may not be strictly age-dependent, with certain lexicosemantic domains engaging embodied networks irrespective of acquisition timing.

Additionally, L2 exposure did not predict learners' sensitivity to diagram congruency, diverging from Lu and Yang (2025), who found both L2 exposure and AoA predictive of embodiment effects for action verbs. This discrepancy may reflect differences in focused constructions: action verbs are closely linked to sensorimotor experience, whereas interpreting semantically complex PVs may depend more on linguistic competence than on exposure alone. The null exposure effect may also reflect measurement limitations. According to Stage 2 of Zhao et al.'s (2025) simulation-based model, successful contextual integration requires immersive, multimodal L2 discourse engagement. While the LHQ offers standardized exposure metrics, it may overlook the depth, quality or context-sensitive nature of L2 engagement. Moreover, although our controlled PV set ensured experimental consistency, its limited scope may constrain ecological validity. Incorporating more diverse, contextually rich PVs could enhance generalizability.

4.4. *Future studies*

L2 learners in this study showed pronounced dependence on external visual support, suggesting that the self-paced sensibility judgment task interleaved with schematic

diagrams may impose substantial cognitive demands on lower-intermediate learners. Our L2 sample primarily comprised learners at lower-intermediate proficiency (B1–B2 on the CEFR) and did not span the full proficiency continuum. Accordingly, these findings may not extend to higher-proficiency learners, who may engage mental imagery in a more native-like manner when processing L2 input. Prior research indicates that higher L2 proficiency is associated with more advanced, native-like embodiment (Ahlberg et al., 2018; Nishide et al., 2025), likely reflecting richer lexical-semantic knowledge and more successful functional restructuring of form–function mappings that supports the development of L2-specific processing strategies (Zhao et al., 2025). Future studies should therefore recruit learners across a broader proficiency range and with varied ages of onset to test whether advanced or earlier-acquiring learners rely less on external visual support and show more native-like effects of diagram congruency.

Additionally, while our schematic diagrams were normed for PV–diagram congruency, subtle differences in visual details, such as the presence of tools, body effectors (e.g., a hand or arm) or other manipulable objects, may have provided varying degrees of visual affordance. Neuroimaging research shows that visually salient, action-related features can engage motor-related cortical regions (e.g., Buccino et al., 2001; Grèzes et al., 2003), suggesting that such features may influence how comprehenders integrate visual information when forming task-based mental imagery. In the present study, these variations did not produce systematic effects on sensibility judgment outcomes: the same set of schematic diagrams was used across concrete and abstract senses, and incongruent trials always paired sentences with non-matching diagrams or filler diagrams, reducing the likelihood that specific visual features could drive the observed compatibility patterns. Even so, because schematic diagrams necessarily simplify spatial relations and vary in their perceptual specificity, future research may benefit from further reducing diagrammatic variability or employing complementary temporal methods (e.g., eye-tracking, EEG) to more precisely trace how visual characteristics modulate imagery-based processing within sensibility judgment paradigms.

5. Conclusion

Overall, this study found that L2 learners' imagery-based simulation was not dynamically modulated by lexical-semantic or contextual factors, reflecting a broad reliance on external visual support. In contrast, L1 speakers strategically engaged visual cues, guided by their internalized semantic knowledge. Crucially, L2 imagery effects were shaped by L2 proficiency and dominance, whereas AoA and exposure showed no significant impact. These findings support the simulation-based model of L2 comprehension, suggesting that mental imagery reflects the gradual restructuring of conceptual and linguistic associations as they unfold in a visual-verbal integration context. Pedagogically, image-schematic diagrams representing spatiotemporal or metaphorical structures may facilitate learners' form–meaning mappings and promote deeper conceptual integration for complex expressions like PVs.

Supplementary material. The supplementary material for this article can be found at <http://doi.org/10.1017/langcog.2026.10062>.

Data availability statement. The supplementary materials and anonymized data are accessible from https://osf.io/qsfbn/?view_only=de18db648d0f4910839e5b20a2173827.

References

- Ahlberg, D. K., Bischoff, H., Kaup, B., Bryant, D., & Strozyk, J. V. (2018). Grounded cognition: Comparing language $\times$ space interactions in first language and second language. *Applied Psycholinguistics*, 39(2), 437–459. <https://doi.org/10.1017/S014271641700042X>.
- Altmann, G. T. M., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, 33(4), 583–609. <https://doi.org/10.1111/j.1551-6709.2009.01022.x>.
- 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>.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2025). *lme4: Linear mixed-effects models using “Eigen” and S4* (Version 1.1.37) [Computer software]. <https://doi.org/10.32614/CRAN.package.lme4>
- Bergen, B. K., & Chang, N. (2005). Embodied construction grammar in simulation-based language understanding. In J.-O. Östman & M. Fried (Eds.), *Constructional grammars: Cognitive grounding and theoretical extensions* (pp. 147–190). John Benjamins Publishing Company.
- Bergen, B., & Wheeler, K. (2010). Grammatical aspect and mental simulation. *Brain and Language*, 112(3), 150–158. <https://doi.org/10.1016/j.bandl.2009.07.002>.
- 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.
- Bergen, B., Lindsay, S., Matlock, T., & Narayanan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science*, 31(5), 733–764. <https://doi.org/10.1080/03640210701530748>.
- Birba, A., Beltrán, D., Martorell Caro, M., Trevisan, P., Kogan, B., Sedeño, L., Ibáñez, A., & García, A. M. (2020). Motor-system dynamics during naturalistic reading of action narratives in first and second language. *NeuroImage*, 216, 116820. <https://doi.org/10.1016/j.neuroimage.2020.116820>.
- Blais, M.-J., & Gonnerman, L. M. (2013). Explicit and implicit semantic processing of verb–particle constructions by French–English bilinguals. *Bilingualism: Language and Cognition*, 16(4), 829–846. <https://doi.org/10.1017/S1366728912000673>.
- Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R. J., Zilles, K., Rizzolatti, G., & Freund, H.-J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: An fMRI study. *European Journal of Neuroscience*, 13(2), 400–404. <https://doi.org/10.1111/j.1460-9568.2001.01385.x>.
- 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>.
- Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1), 22–29.
- Connell, L. (2007). Representing object colour in language comprehension. *Cognition*, 102(3), 476–485. <https://doi.org/10.1016/j.cognition.2006.02.009>.
- De Grauwe, S., Willems, R. M., Rueschemeyer, S.-A., Lemhöfer, K., & Schriefers, H. (2014). Embodied language in first- and second-language speakers: Neural correlates of processing motor verbs. *Neuropsychologia*, 56, 334–349. <https://doi.org/10.1016/j.neuropsychologia.2014.01.020>.
- Dudschig, C., De La Vega, I., De Filippis, M., & Kaup, B. (2014). Language and vertical space: On the automaticity of language action interconnections. *Cortex*, 58, 151–160. <https://doi.org/10.1016/j.cortex.2014.06.003>.
- Ellis, R., Loewen, S., Elder, C., Reinders, H., Erlam, R., & Philp, J. (2009). *Implicit and explicit knowledge in second language learning, testing and teaching*. Multilingual Matters.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>.
- Garnier, M., & Schmitt, N. (2015). The PHaVe list: A pedagogical list of phrasal verbs and their most frequent meaning senses. *Language Teaching Research*, 19(6), 645–666. <https://doi.org/10.1177/1362168814559798>.
- Gibbs, R. W., & Matlock, T. (2008). Metaphor, imagination, and simulation. In R. W. Gibbs, Jr. (Ed.), *The Cambridge handbook of metaphor and thought* (1st ed., pp. 161–176). Cambridge University Press. <https://doi.org/10.1017/CBO9780511816802.011>

- Grèzes, J., Tucker, M., Armony, J., Ellis, R., & Passingham, R. E. (2003). Objects automatically potentiate action: An fMRI study of implicit processing. *European Journal of Neuroscience*, 17(12), 2735–2740. <https://doi.org/10.1046/j.1460-9568.2003.02695.x>.
- Guan, C. Q., Meng, W., Yao, R., & Glenberg, A. M. (2013). The motor system contributes to comprehension of abstract language. *PLoS One*, 8(9), e75183. <https://doi.org/10.1371/journal.pone.0075183>.
- Harpaintner, M., Sim, E.-J., Trumpp, N. M., Ulrich, M., & Kiefer, M. (2020). The grounding of abstract concepts in the motor and visual system: An fMRI study. *Cortex*, 124, 1–22. <https://doi.org/10.1016/j.cortex.2019.10.014>.
- Hauf, J. E. K., Nieding, G., & Seger, B. T. (2020). The development of dynamic perceptual simulations during sentence comprehension. *Cognitive Processing*, 21(2), 197–208. <https://doi.org/10.1007/s10339-020-00959-7>.
- Hwang, H.-B. (2023). Image-schema-based-instruction enhanced L2 construction learning with the optimal balance between attention to form and meaning. *International Review of Applied Linguistics in Language Teaching*, 60(1), 1–25. <https://doi.org/10.1515/iral-2023-0015>.
- 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>.
- Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. University of Chicago Press.
- Kogan, B., Muñoz, E., Ibáñez, A., & García, A. M. (2020). Too late to be grounded? Motor resonance for action words acquired after middle childhood. *Brain and Cognition*, 138, 105509. <https://doi.org/10.1016/j.bandc.2019.105509>.
- Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1), 32–59. <https://doi.org/10.1080/23273798.2015.1102299>.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2020). *lmerTest: Tests in linear mixed effects models* (Version 3.1.3) [Computer software]. <https://doi.org/10.32614/CRAN.package.lmerTest>
- Lenth, R. V., Banfai, B., Bolker, B., Buerkner, P., Giné-Vázquez, I., Herve, M., Jung, M., Love, J., Miguez, F., Piaskowski, J., Riebl, H., & Singmann, H. (2025). *Emmeans: Estimated marginal means, aka least-squares means* Version 1.11.0) [Computer software]. R Foundation for Statistical Computing.
- Li, P., Zhang, F., Yu, A., & Zhao, X. (2020). Language history questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, 23(5), 938–944. <https://doi.org/10.1017/S1366728918001153>.
- Lindstromberg, S. (2022). The compositionality of English phrasal verbs in terms of imageability. *Lingua*, 275, 103373. <https://doi.org/10.1016/j.lingua.2022.103373>.
- Liu, N., & Bergen, B. (2016). When do language comprehenders mentally simulate locations? *Cognitive Linguistics*, 27(2), 181–203. <https://doi.org/10.1515/cog-2015-0123>.
- Lu, X., & Yang, J. (2025). Second language embodiment of action verbs: The impact of bilingual experience as a multidimensional spectrum. *Bilingualism: Language and Cognition*, 28(4), 1117–1133. <https://doi.org/10.1017/S1366728924000981>.
- MacWhinney, B. (2005). A unified model of language acquisition. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 49–67). Oxford University Press.
- Mahpeykar, N., & Tyler, A. (2015). A principled cognitive linguistics account of English phrasal verbs with up and out. *Language and Cognition*, 7(1), 1–35. <https://doi.org/10.1017/langcog.2014.15>.
- Nishide, T., Zhao, H., & 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>.
- 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>.
- Ostarek, M., & Huettig, F. (2019). Six challenges for embodiment research. *Current Directions in Psychological Science*, 28(6), 593–599. <https://doi.org/10.1177/0963721419866441>.
- Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, 144(10), 1002–1044. <https://doi.org/10.1037/bul0000158>.
- 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>.
- Qualtrics. (2025). *Qualtrics XM platform* [Computer software]. Qualtrics. <https://www.qualtrics.com>

- Richardson, D. C., Spivey, M. J., Barsalou, L. W., & McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science*, 27(5), 767–780. https://doi.org/10.1207/s15516709cog2705_4.
- R Core Team (2025). *R: A language and environment for statistical computing* (Version 4.5.0) [Computer software]. R Foundation for Statistical Computing. <https://www.rproject.org>
- Sato, T., Ogura, M., & Burden, T. (2022). What does image schema facilitate in English L2 vocabulary processing?: The effectiveness of schematic visual aids for the enhancement of metaphoric competence in learning phrasal verbs. *Cognitive Linguistic Studies*, 9(2), 202–222. <https://doi.org/10.1075/cogls.20022.sat>.
- 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>.
- Shi, J., Peng, G., & Li, D. (2023). Figurativeness matters in the second language processing of collocations: Evidence from a self-paced reading experiment. *Language Learning*, 73(1), 47–83. <https://doi.org/10.1111/lang.12516>.
- Shiang, R.-F., Chern, C.-L., & Chen, H.-C. (2024). Embodied cognition and L2 sentence comprehension: An eye-tracking study of motor representations. *Frontiers in Human Neuroscience*, 18, 1410242. <https://doi.org/10.3389/fnhum.2024.1410242>.
- Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, 42(4), 1096–1104. <https://doi.org/10.3758/BRM.42.4.1096>.
- Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1), 24–31. <https://doi.org/10.1177/0098628316677643>.
- 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>.
- Tian, L., Chen, H., Heikkinen, P. P., Liu, W., & Parviainen, T. (2023). Spatiotemporal dynamics of activation in motor and language areas suggest a compensatory role of the motor cortex in second language processing. *Neurobiology of Language*, 4(1), 178–197. https://doi.org/10.1162/nol_a_00093.
- Tian, L., Chen, H., Zhao, W., Wu, J., Zhang, Q., De, A., Leppänen, P., Cong, F., & Parviainen, T. (2020). The role of motor system in action-related language comprehension in L1 and L2: An fMRI study. *Brain and Language*, 201, 104714. <https://doi.org/10.1016/j.bandl.2019.104714>.
- Tiv, M., Gonnerman, L., Whitford, V., Friesen, D., Jared, D., & Titone, D. (2019). Figuring out how verb-particle constructions are understood during L1 and L2 reading. *Frontiers in Psychology*, 10, 1733. <https://doi.org/10.3389/fpsyg.2019.01733>.
- Tomczak, E., & Ewert, A. (2015). Real and fictive motion processing in Polish L2 users of English and monolinguals: Evidence for different conceptual representations. *The Modern Language Journal*, 99(S1), 49–65. <https://doi.org/10.1111/j.1540-4781.2015.12178.x>.
- Trueswell, J. C., & Tanenhaus, M. K. (1994). Toward a lexicalist framework of constraint-based syntactic ambiguity resolution. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 155–179). Lawrence Erlbaum Associates.
- Tyler, A., Mueller, C. M., & Ho, V. (2010). Applying cognitive linguistics to instructed L2 learning: The English modals. *AILA Review*, 23, 30–49. <https://doi.org/10.1075/aila.23.03tyl>.
- Wang, M., & Zhao, H. (2023). Mental simulation in L2 processing of English prepositional phrases. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 45, 595–602.
- Wang, M., & Zhao, H. (2024a). Are schematic diagrams valid visual representations of concepts? Evidence from mental imagery in online processing of English prepositions. *Language and Cognition*, 16(4), 1564–1587. <https://doi.org/10.1017/langcog.2024.18>.
- Wang, M., & Zhao, H. (2024b). Perceptual representations in L1 and L2 spatial and abstract language processing: Applying an innovative sentence-diagram verification paradigm. *Frontiers in Human Neuroscience*, 18, 1425576. <https://doi.org/10.3389/fnhum.2024.1425576>.
- Wang, M., & Zhao, H. (under review). *L2 learner factors in incremental English motion sentence processing* [Manuscript submitted for publication].
- Wei, Y. F., Yang, W. W., Oppenheim, G., Hu, J. H., & Thierry, G. (2024). Embodiment for spatial metaphors of abstract concepts differs across languages in Chinese-English bilinguals. *Language Learning*, 74(S1), 224–257. <https://doi.org/10.1111/lang.12632>.

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., & Van Den Brand, T. (2025). *ggplot2: Create elegant data visualisations using the grammar of graphics* (Version 3.5.2) [Computer software]. <https://doi.org/10.32614/CRAN.package.ggplot2>

Xue, J., Marmolejo-Ramos, F., & Pei, X. (2015). The linguistic context effects on the processing of body–object interaction words: An ERP study on second language learners. *Brain Research*, 1613, 37–48. <https://doi.org/10.1016/j.brainres.2015.03.024>.

Xue, J., Xie, D., Lu, X., Niu, Z., & Marmolejo-Ramos, F. (2024). The different effects of a sensorimotor grounding on AoA between bilingual concepts. *Frontiers in Human Neuroscience*, 18, 1387674. <https://doi.org/10.3389/fnhum.2024.1387674>.

Yang, T. (2016). Image schemas in verb–particle constructions: Evidence from a behavioral experiment. *Journal of Psycholinguistic Research*, 45(2), 379–393. <https://doi.org/10.1007/s10936-015-9354-6>.

Zeelenberg, R., Pecher, D., van der Meijden, M. E. M., Trott, S., & Bergen, B. (2025). Non-native language comprehenders encode implied shapes of objects in memory. *Cortex*, 182, 100–111. <https://doi.org/10.1016/j.cortex.2024.09.008>.

Zhang, X., Yang, J., Wang, R., & Li, P. (2020). A neuroimaging study of semantic representation in first and second languages. *Language, Cognition and Neuroscience*, 35(10), 1223–1238. <https://doi.org/10.1080/23273798.2020.1738509>.

Zhao, H., Huang, S., Zhou, Y., & Wang, R. (2020). Schematic diagrams in second language learning of English prepositions: A behavioral and event-related potential study. *Studies in Second Language Acquisition*, 42(4), 721–748. <https://doi.org/10.1017/S027226311900069X>.

Zhao, H., & Le, F. (2016). Measuring L2 explicit knowledge of English verb–particle constructions: Frequency and semantic transparency at two proficiency levels. In L. Ortega, A. Tyler, H. I. Park, & M. Uno (Eds.), *The usage-based study of language learning and multilingualism* (pp. 171–185). Georgetown University Press.

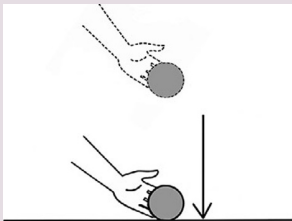
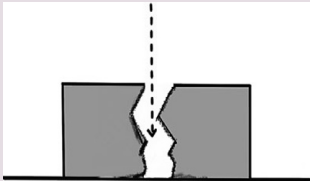
Zhao, H., Vanek, N., & MacWhinney, B. (2025). Mental simulation in bilingual and second language processing: New directions in the competition model. *Brain and Language*, 269, 105619. <https://doi.org/10.1016/j.bandl.2025.105619>.

Zheng, Y., & Cheng, L. (2008). Test review: College English test (CET) in China. *Language Testing*, 25(3), 408–417. <https://doi.org/10.1177/0265532208092433>.

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>.

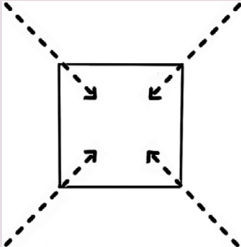
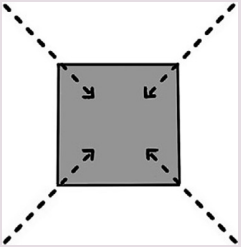
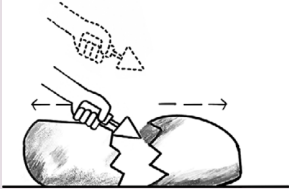
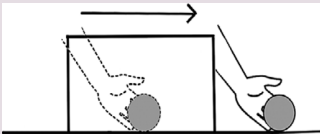
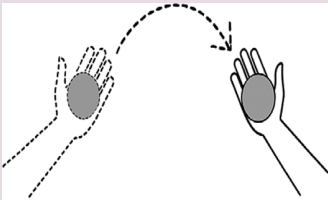
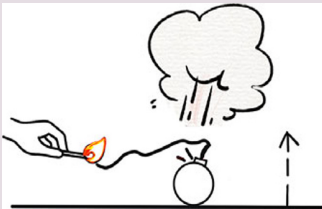
Zwaan, R. A. (2008). Experiential traces and mental simulations in language comprehension. In M. de Vega, A. Glenberg, & A. Graesser (Eds.) *Symbols, Embodiment, and Meaning* (pp. 165–180), Oxford University Press.

Appendix A. Target Phrasal Verbs and Corresponding Schematic Diagrams

Phrasal verb	Schematic diagram
lay down take down put down	
break down	

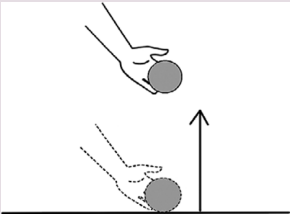
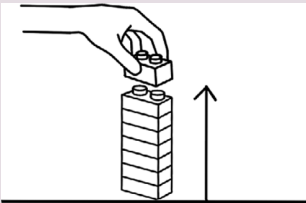
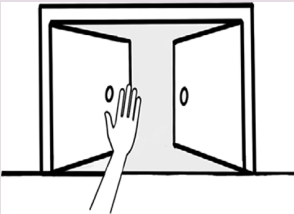
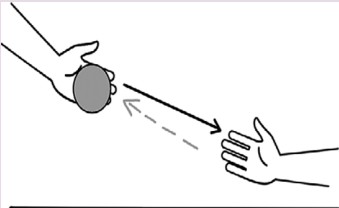
(Continued)

(Continued)

Phrasal verb	Schematic diagram
<i>take in</i>	
<i>fill in</i>	
<i>break off</i> <i>cut off</i>	
<i>lay out</i> <i>give out</i> <i>throw out</i> <i>bring out</i> <i>send out</i>	
<i>hand over</i>	
<i>blow up</i>	

(Continued)

(Continued)

Phrasal verb	Schematic diagram
<i>pick up</i>	
<i>build up</i>	
<i>open up</i>	
<i>break up</i>	
<i>give back</i>	

Cite this article: Hu, Y., & Zhao, H. (2026). Mental imagery effects in bilingual phrasal verb comprehension, *Language and Cognition*, 18, e13, 1–29. <https://doi.org/10.1017/langcog.2026.10062>