



Social interaction shapes and boosts second language learning: virtual reality can show us how



Ana Zappa^{1,2}✉, Mel Slater³ & Antoni Rodriguez-Fornells^{1,2,4}

Social interaction can play a crucial role in how a second language (L2) is learned. In the current review, we examine theoretical frameworks and empirical studies demonstrating how social factors influence L2 learning, but we also identify gaps in the current literature. We propose using virtual reality (VR) as a methodology to fill these gaps with controlled, ecologically valid social simulations that can elucidate how social factors shape L2 learning.

The first time you heard someone say “Careful!” in a foreign language that you were trying to master (hereafter referred to as a second language (L2)), this utterance was likely framed by a series of multimodal cues, such as feeling a stranger grab your arm and seeing a bicycle racing towards you. These complex environments enriched by social contexts and physical cues are a far cry from the flat-screen paradigms that are generally used to examine L2 learning and processing. This discrepancy mirrors the diversity of real-life L2 learning contexts, which range from interactive and immersive to rigid and textbook-based. Learners also differ in how they approach L2 learning: some are prone to gather knowledge and learn from social interactions and personal feedback (social or communication-oriented learners), while others favor more analytical and non-social forms of learning^{1–3}. Importantly, practicing an L2 in social contexts is thought to be particularly beneficial for reaching a high level of proficiency^{4–6}. Indeed, no matter how much learners are exposed to an L2, if they do not engage in social interaction, attaining proficiency is nearly impossible^{7,8}. And, undeniably, most language learners aim to use the target language in meaningful exchanges⁵. Across disciplines, momentum is building to integrate social reality into models of language processing and L2 learning^{6,8,9}. With all this in mind, we believe it is crucial to consider how social processes affect L2 learning.

We begin by reviewing in-depth multidisciplinary theoretical approaches and studies that have investigated how social interaction can support L2 learning. We then examine four key components - multimodality, embodiment, interaction, and feedback - and consider how each supports L2 development. Next, we turn to recent neuroimaging studies that shed light on the neural mechanisms underlying socially interactive L2 learning, while highlighting the persistent trade-off between ecological validity and experimental control. To illustrate how these mechanisms might operate in practice, we present word learning as a case study, focusing

on how socially grounded contexts may enhance lexical encoding and integration. Finally, we review the potential of using virtual reality (VR) simulations as a promising methodological tool for advancing research on how social factors shape L2 learning.

Socially interactive language learning: theoretical frameworks

Studies examining the cognitive processes underlying language acquisition have traditionally focused on infants and children^{10–12}. Even before children can utter their first words, language is learned not only through exposure but through active participation as facilitated by social interaction¹³. As such, language acquisition theories and models emphasize the importance of social cues in developmental language learning. From a social-pragmatic perspective, communication motivates language acquisition, and interaction results in learning¹⁴. In particular, social interaction is thought to expedite referential learning by engaging sensorimotor and attentional mechanisms^{15,16}. Indeed, joint attention, or shared attentional focus between two individuals (e.g., a caregiver and a child), such as when playing with a common object¹⁷, has been associated with facilitated language learning (i.e., word segmentation¹⁸, semantic integration¹⁹). Language development is also tightly linked to parental responsiveness, or well-timed, relevant responses to infants' attempts to explore and communicate²⁰. A social behavior²¹, parental responsiveness is both temporally and conceptually connected to the learning environment, which is thought to support word-to-referent association²⁰. For instance, communicative feedback - informing children whether their utterance has been understood - is a particularly effective element of communicative coordination during language development¹³.

Compared to children, adults are thought to learn a second language in a more conscious and deliberate manner, with outcomes shaped by factors such as linguistic background, motivation, cognitive abilities, and other individual differences. Still, since the foundations of language learning are

¹Cognition and Brain Plasticity Group, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Barcelona, 08097, Spain. ²Department of Cognition, Development and Educational Science, Institute of Neurosciences, Campus Bellvitge, University of Barcelona, L'Hospitalet de Llobregat, Barcelona, 08097, Spain. ³Event Lab, Department of Clinical Psychology and Psychobiology, Institute of Neurosciences, University of Barcelona, Casanova, 143, 08036 Barcelona, Spain. ⁴Catalan Institution for Research and Advanced Studies, ICREA, Barcelona, Spain. ✉e-mail: anazappa@gmail.com

established early in life, it is reasonable to think that some of the same socially grounded mechanisms may continue to influence learning well into adulthood. Drawing parallels between first language (L1) acquisition and socially interactive L2 learning enables the exploration of how key factors in L1 acquisition (i.e., joint attention and communicative feedback) may also support L2 learning, thereby enriching our understanding of language acquisition across different contexts. As has been suggested for children¹⁴, successful adult L2 learning could be grounded in social interaction¹⁵. However, translating socially grounded frameworks into controlled experimental paradigms remains challenging. Traditional lab tasks struggle to capture the dynamic quality of real interaction, making it difficult to isolate the specific contributions of mechanisms like joint attention or communicative feedback.

While insights from L1 acquisition provide a valuable foundation for understanding the social basis of L2 learning, the field of Second Language Acquisition (SLA) has developed its own rich theoretical landscape. Drawing from a number of disciplines, including linguistics and psychology, SLA has long emphasized the critical role of socio-cultural factors in L2 learning, particularly how its underlying processes can be elucidated by language production and comprehension⁵. In the 1960s, anthropological linguistics and sociolinguistics contested the then-popular Chomskyan view, which framed language knowledge as a competence/performance dichotomy, and instead viewed language as a form of communicative action to be learned in situ^{22,23}. Another important influence for SLA was Vygotsky's sociocultural theory²⁴, later extended to L2 learning²⁵⁻²⁷, which claimed that meaningful interaction is at the root of how we learn. These models generally converge on the idea that consistent and meaningful language practice results in L2 proficiency⁵, emphasizing the critical roles of both the quality and quantity of language input and learner output in the L2 learning process²⁸. For example, corpus analyses reveal strong links between the amount and type of input and learners' L2 development and performance^{5,29}. Further, longitudinal case studies using conversational analysis revealed how social contexts shape L2 learning trajectories³⁰, often observing changes in how learners use social cues (i.e., gaze, gesture, or posture) as learning progresses^{17,18,31,32}. Indeed, social interaction is both a medium for practice and a catalyst for the eventual automatization of language skills³³⁻³⁵ in contexts ranging from study abroad³⁶ to task-based classroom interaction³⁷.

Building on these theoretical foundations, a key question is: what specific mechanisms make social interaction so powerful for L2 learning? Li and Jeong³⁸ recently proposed the Social L2 Learning (SL2) model, based on L2 learning with social interaction, defined as "real-life or simulated real-life environments where learners can interact with objects and people, perform actions, receive, use, and integrate perceptual, visuospatial, and other sensorimotor information, which enables learning and communication to become embodied"³⁹. In line with previous models^{39,40}, this approach claims that L2 learning processes only differ from L1 acquisition because L2 learning contexts are generally poorer. Furthermore, the Social L2 Learning model bases its premises on memory research, pointing to the interdependence between encoding and retrieval processes. For instance, according to the encoding-specificity principle, retrieval of learned content is facilitated when learners are in the context in which they first encoded semantic information⁴¹. Interestingly, learning conditions could have an important influence on which memory processes learners use, as individual differences in declarative and procedural memory capacity predict L2 learning^{42,43}. Menks and colleagues⁴⁴ posit that classroom instruction calls more upon the declarative memory system, whereas immersion-like learning is likely to involve the procedural memory system. This could, in part, explain the social advantage for L2 learning and aligns with the view that implicit learning facilitates automatization³³. Nonetheless, as noted by Bowden & Faretta-Stutenberg⁴⁵, opportunities for interaction vary considerably across cultural and classroom contexts, and many modern L2 classrooms now promote meaningful social interaction.

These interdisciplinary contributions converge on the view that language learning emerges through active, socially embedded communication.

Each framework brings valuable insight into how social interaction shapes L2 learning, yet they leave open questions about the specific mechanisms involved. We argue that advancing the field of second language learning calls for creating a unified account of how social interaction shapes L2 learning. For this, future research must more rigorously isolate, manipulate, and empirically evaluate the processes that each framework identifies as central to successful L2 learning.

Second language learning in social contexts: important factors

The previous section outlined the theoretical foundations of socially interactive L2 learning. We now turn to key mechanisms through which these interactions support language development. In particular, we highlight four core factors - **multimodality, embodiment, interaction, and social feedback** - that recur across frameworks and could play a central role in L2 learning.

Multimodal learning engages more than one sensory modality, such as receiving both auditory and visual input while listening to and watching someone speak. Here, multiple channels of information can enhance comprehension and solidify knowledge by tying learning to physical and social cues. Multimodal cues are intrinsically involved in a real-world language comprehension and production, and should hence be examined in L2 learning, including in naturalistic social contexts^{46,47}. Visual cues, often present in learning environments, can facilitate learning by enhancing the complexity of formed representations⁴⁸. Encoding new words with images, for instance, improves learning compared to verbal learning alone or learning with translations^{49,50}. On the premise that language is a "a system of communication in face-to-face interaction"⁵¹, engaging in social language use involves exposure to rich visual and auditory human-produced cues that provide critical contextual information, which enhances comprehension and supports more effective communication. Indeed, multimodality facilitates language comprehension^{52,53}, with auditory cues such as prosody^{54,55} and visual cues such as gestures⁵⁶ helping us to contextualize spoken language.

Two recent studies presented behavioral and electrophysiological evidence that mouth movements, gestures and prosody operate dynamically to facilitate word predictability in the L1⁵³, and, to a lesser extent, in the L2⁵⁷. This approach increases ecological validity by observing how co-occurring multimodal cues are used interactively during language processing, as opposed to only manipulating cues in isolation. Together, this evidence suggests that multimodally enriched learning strategies boost language comprehension and retention. In line with Zhang and colleagues^{53,57} and De Felice and colleagues⁵⁸ studies, L2 learning research would benefit from further examining whether and, if so, how different multimodal social cues affect processing and encoding in closer-to-life, social situations.

Embodied theories of language emphasize that in addition to engaging multiple sensory channels, socially interactive language learning is also grounded in physical and sensorimotor experiences. From this view, linguistic representations are shaped by our interactions with the surrounding environment and grounded in real-world, experience^{59,60}. Embodied or situated learning implies being physically present or involved in a learning context⁶¹. Physical action, whether it be manipulating objects⁶² or performing gestures^{63,64} has been shown to support word encoding via the activation of the motor system, forming "motor traces" that strengthen semantic representations. There is a substantial literature showing that learning new words while performing gestures leads to improved learning, compared to learning with pictures or using standard rote verbal learning^{49,64,65}. Furthermore, performing spontaneous gestures not only helps speakers to communicate⁶⁶ but is thought to support L2 learning⁶⁷. Interestingly, simply observing gestures or actions that are congruent with the meaning of words being learned enhances L2 learning⁶⁸⁻⁷¹. This indicates that embodied processes, as induced by action observation, support learning even in the absence of physical interaction. Pertaining to embodied language encoding facilitations, it has been suggested that, on a neural level, the association of spoken words to images or gestures links semantic concepts to visual⁷² and motor areas⁷³.

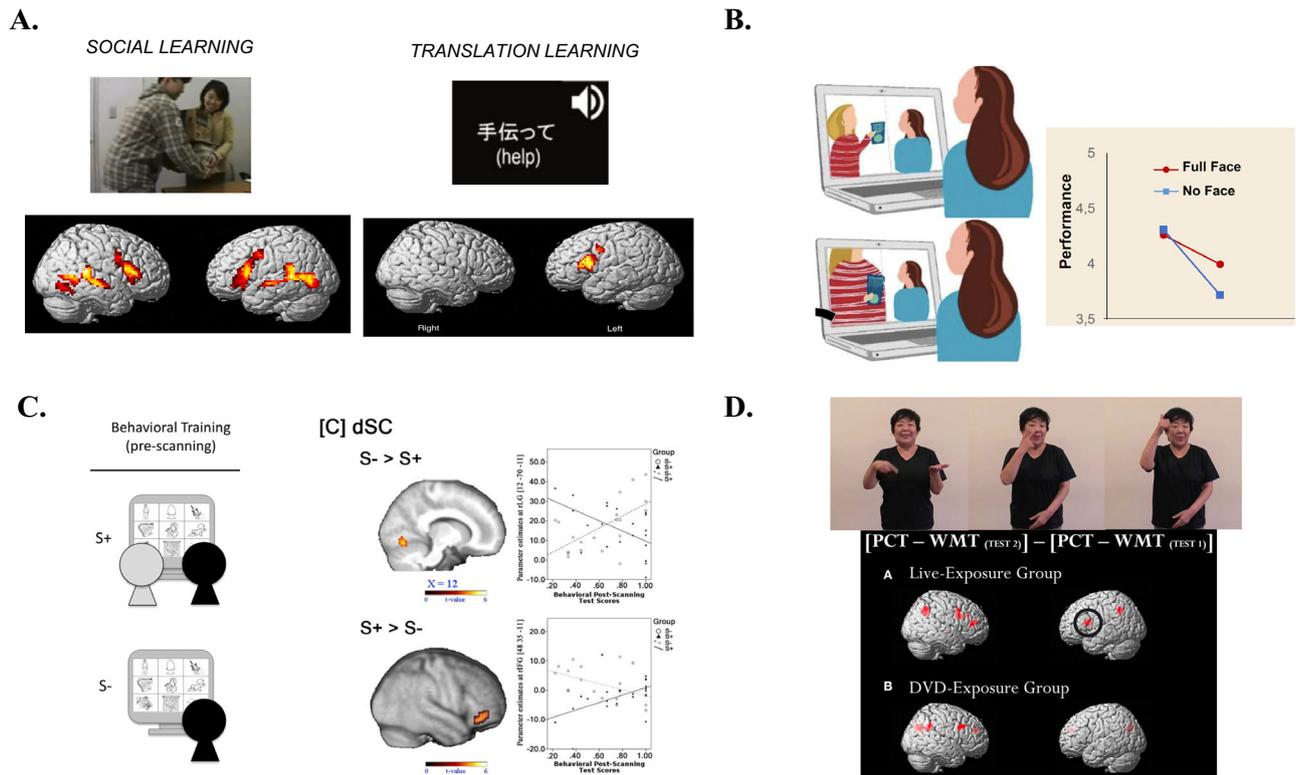


Fig. 1 | Examples of studies that examined the role of social interaction during semantic or language learning. **A** fMRI study that compared learning L2 words by viewing videos of social situations vs. translation. Here we see the brain areas activated in each type of learning (contrast between initial and learning and the following session). Note that Social learning shows greater activation in the right middle temporal gyrus, the superior temporal sulcus, and the inferior frontal gyrus, and both conditions show greater activation in the left inferior frontal gyrus, and supplementary motor area (Jeong et al., 2021; License: CC BY 4.0) **B** Behavioral study comparing learning conceptual information about new words by observing (Recorded) vs. engaging in (Live) social interaction. The Live condition resulted in greater behavioral accuracy post-training overall. Viewing the teacher’s full face

improved learning in the Recorded condition but not in the Live condition (De Felice et al., 2021; copyright permission from Elsevier, license number: 5812910454464) **C** fMRI study comparing L2 learning with an imagined social partner vs. a computer. A positive correlation emerged between accuracy post-training and activity in the right inferior frontal gyrus for the interactive group and in the fusiform gyrus for the non-interactive group. (Verga & Kotz, 2019; copyright permission from Elsevier, license number: 5812620564760) **D** fMRI study comparing sign language syntax learning with live interaction vs. through viewing videos. Only the Live interaction group showed activation in the left inferior frontal gyrus post-training. (Yusa et al., 2017; License: CC BY 4.0).

Interactive Learning occurs when learners actively engage with their environment, whether by manipulating objects or interacting with a partner, which plays a crucial role in language learning. Whereas multimodality and embodiment can be considered separately from interaction, the opposite seems nearly impossible. For example, at the crossroads between embodied, multimodal and interactive word learning, simply holding objects has been shown to aid in L2 learning⁷⁴. As for social interaction, one study had participants learn conceptual information about novel lexical items while observing social content compared to engaging in online social interaction⁵⁸. Seeing the teacher’s face resulted in improved learning, but only during interactive learning, suggesting that different mechanisms subservise interactional vs. observational learning processes (Fig. 1B). Therefore, while added visual and motor information appears to contribute to L2 learning in a number of experimental tasks, this multimodal advantage may be modulated by sociability. It would be worthwhile to explore whether this conceptual learning effect extends to L2 learning, where learners typically acquire new labels for familiar concepts.

Social feedback – broadly defined as contingent social information that learners receive in response to their language use – and consequential reinforcement can help shape and refine learners’ developing linguistic system. Corrective feedback draws learners closer to their goals by bringing attention to their errors, building declarative knowledge, and allowing them to find better alternatives via self-correction, which can impact procedural knowledge and eventually lead to automatization^{35,75}. Affective feedback,

such as praise or encouragement, could likewise play an important role in motivating learners. L2 learning conceived as a social activity might be governed by the complex rewarding dynamics involved in socialization⁷⁶ and the value we assign to these interactions^{77,78}. Not only do we prioritize social information^{79,80} but we treat it along with emotional content, associated with processing advantages^{81,82}. Being liked by others activates the dopaminergic system⁸³ and both avoiding social disapproval and anticipating social reward activate the ventral striatum/nucleus accumbens. Socially interactive L2 learning can engage multiple forms of social rewards (e.g., face attraction, gestures, subjective assigned value, positive feedback) and is tightly linked to social motivation⁷⁹. During its initial stages, L2 learning is largely controlled by external feedback – a form of social reward or punishment – whereby native or advanced speakers provide performance feedback and corrections to the initial production of new utterances. The quality of this feedback (including the positive or negative emotional content) and its impact in reward-motivational systems might also affect the encoding process of new linguistic information and speakers’ motivation to pursue their goals⁸⁴. Importantly, learning new words is associated with an increase in the activation of core reward brain regions (e.g., the ventral striatum^{85,86}) and these effects are modulated by dopaminergic pathways⁸⁷. An increase in dopamine has been shown to facilitate the encoding of new information in the hippocampal cortex⁸⁸. In a recent event-related potential (ERP) and pupillometry study, we showed that reliable social feedback in the form of dynamic videos, compared to symbolic feedback (a static image of

✓ or ✕), led to more widespread cortical activation and better differentiation between positive and negative feedback. Furthermore, social feedback had a greater overall impact on word learning, as shown by correlations between the ERP and pupillometry measures and post-training performance scores (Zappa et al., in press).

Given the above evidence, examining how social feedback can shape L2 learning as supported by reward processes could prove a very fruitful avenue. We propose that social feedback be conceptualized as both in its informative content as well as its emotional-rewarding value. Furthermore, future work should examine how individual learner profiles respond to different types of feedback (e.g., social, symbolic, encouraging, or critical). In doing so, we may gain a more nuanced understanding of how feedback shapes L2 development over time, and how its social dimension can contribute to optimizing both learning outcomes and learner motivation.

It is worth noting that although the above reviewed factors (multimodality, embodiment, interaction and social feedback) often overlap, they do not necessarily come as a package in real life, and even less so in controlled experimental studies. Face-to-face conversation involves all four dimensions, combining multimodal information (auditory and visual) with the interactive (including feedback), back-and-forth nature of dialog and embodied physical presence in a linguistic setting. Conversely, phone conversations typically lack these multimodal aspects; passively listening to a teacher is not interactive, and online interactions generally fall short of providing an embodied experience. While the effects described are compelling, most studies to date have been constrained by methodological limitations and, indeed, few have tested how these elements operate in truly interactive settings, highlighting the need for tools that can simulate authentic, multifactor, responsive environments while allowing for precise experimental control. To fully understand the contributions of these factors in second language learning, future research should test them both independently (i.e., isolate specific modalities or interaction types) and in combination (i.e., examine how they interact within contextually rich learning environments). This dual approach will clarify the mechanisms through which each supports learning and identify the conditions under which their integration produces additive or even multiplicative effects.

Neural evidence of socially interactive second language learning

Despite the growing number of studies on the role of social interaction in L2 learning^{6,8,9,45}, little is known about the core cognitive and neuronal processes underlying L2 acquisition involving social interaction. Examining brain activity during socially interactive L2 learning and processing can give us important clues as to the efficacy of different types of learning contexts, or even learning styles. Thus, L2 research stands to gain a great deal from neurocognitive studies that unveil the processes underlying socially interactive L2 learning, including L2 processing differences that occur during or as a result of different learning strategies.

While this area remains underexplored, a handful of functional magnetic resonance imagining (fMRI) studies on adults have examined the relationship between how words are encoded (i.e., socially vs. not) and the neural activations during the subsequent retrieval of these words. For instance, words learned by viewing videos depicting social situations activated the right supramarginal gyrus during retrieval, similarly to what is observed during L1 processing, whereas those learned via textual translations activated the left frontal middle gyrus³⁹. Furthermore, when focusing on neural processes during L2 word encoding in social vs. non-social situations as depicted by videos, the former also led to improved behavioral performance and greater activation of brain regions previously associated with social, affective, and perception-action processes during learning³⁰ (Fig. 1A). Overall, these results indicate that encoding words by viewing social situations could lead to more native-like processing. In line with the Social L2 Learning model, it is possible that processing complex social, emotional, and action-related information in a social vs. non-social context during encoding leads to the formation of richer representations of the newly learned L2 words and facilitates retrieval during the post-training tasks. These results also corroborate the idea that multimodality, as provided

by the multiple auditory and visual cues present in social contexts, can support learning. On the other hand, given that in both these studies the “social” condition consisted of participants passively viewing pre-recorded social scenes, this paradigm is limited as regards testing how embodiment or social interaction shapes L2 learning.

Another strategy used in fMRI studies to engage participants in social interaction is to scan them as they perform activities that they believe to be a live interaction¹⁵. Verga & Kotz⁹¹ used a simulated interactive context (i.e., playing with an imagined social partner vs. with a computer) to examine the effect of social interaction on L2 learning (Fig. 1C). During learning, social learners showed greater activation in the right supramarginal gyrus, associated with social cognition, similar to what Jeong and colleagues⁸⁹ found during retrieval. Activity in the right inferior frontal gyrus, associated with attention, correlated with post-training accuracy for words learned in non-informative contexts, but only for social learners. From a social pragmatic perspective, it is likely that social learning helped learners to integrate new meanings via attentional processes. However, although game-playing is naturally interactive, this imaginary social partner paradigm does not provide the embodied and multisensory richness of interacting with a social partner in real time, and it remains uncertain to what extent participants believed they were interacting with a real partner. In fact, both video viewing and simulated social interaction are limited in that they do not allow for the manipulation of multimodal social cues, or the observation of how these affect L2 learning. Furthermore, the Social L2 Learning model, which argues that L1 and L2 learning differences are due to learning context, cannot convincingly be tested via these L2 teaching methods that bear little resemblance to naturalistic L1 acquisition.

Yusa and colleagues⁹² used direct social interaction to compare post-training brain activity in participants who learned Japanese sign language syntax with a native deaf signer to that of those who viewed instructive videos (Fig. 1D). Both groups learned through physically signing, as opposed to passively viewing the teacher or videos. These two conditions directly tested the effect of embodiment, as induced by participants in the live exposure group being physically present with the teacher, and multimodal social cues, on learning. Syntactic processing post-training led to activations in the left inferior frontal gyrus, similar to that displayed by native listeners in the live exposure group but not in the video viewing group. These results further support the Social L2 Learning model by indicating that learning with a partner could lead to more native-like processing of an L2 compared to passive observation. Future studies should extend this type of paradigm, preferably adding actual back-and-forth interaction, to examine the neural correlates of how different aspects of sign and verbal L2 learning (e.g., word learning) as well as different types of social interaction (e.g., teacher-learner, learner-learner) and learning environments (e.g., classroom, café) impact L2 learning.

The above studies all use a single-brain approach, meaning that only one participant’s brain activity was measured during the experiments, which is sometimes thought to be limited in measuring social learning (for a review, see ref. 93). The second-person neuroscience approach seeks to increase ecological validity by observing neural processes while participants engage with a social partner in a close-to-life, dynamic, back-and-forth manner⁹⁴. Hyperscanning has recently been widely proposed to investigate the brain correlates of social interactions. Here, interbrain neural activity is explored via electroencephalography (EEG), fMRI or near-infrared spectroscopy (fNIRS) recordings of activity generated – often simultaneously – by two or more brains⁹⁵. Social interaction is associated with inter-brain synchrony between participants in face-to-face vs. back-to-back dialog and monologs⁶, during interactive vs. non-interactive object-naming and verbal descriptions⁹⁶, and in successful educational dialogs⁹⁷. Dikker and colleagues examined group interbrain synchrony between a group of students using portable EEG in a longitudinal, real-world classroom study. Watching videos or participating in a common discussion led to greater synchrony compared to listening to a teacher lecturing or reading out loud, which was interpreted as reflecting social interaction and underlying shared attention⁹⁸. We should note that this study conflated the effects of interaction with those

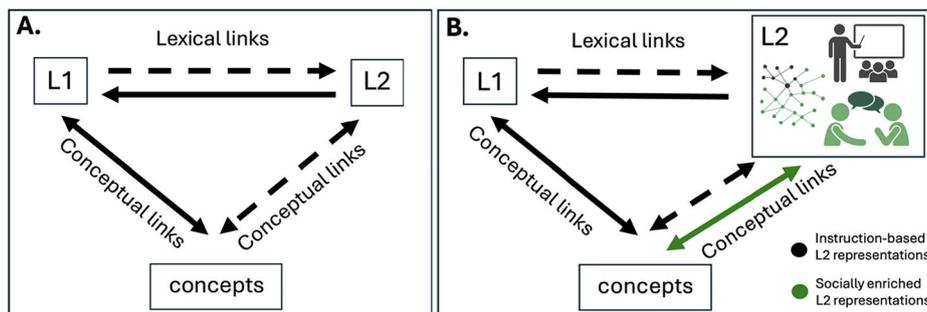


Fig. 2 | The Revised Hierarchical Model of bilingual language processing and our proposed extension, accounting for differences between L2 learning in social vs classical settings. **A** The Revised Hierarchical Model (adapted from Kroll & Stewart, 1994, copyright permission from Elsevier, license number: 5956110535190). L1 and L2 words are interconnected via lexical and conceptual links. The lexical links are

stronger from the L2 to the L1 than the other way around. Conceptual links are stronger for the L1 than for the L2. **B** Our proposed extension to the model: L2 learning in multimodal and social settings could form stronger links between L2 wordforms and concepts.

of multimodality. A similar “group neuroscience” study compared individual and teammate performance in problem-solving tasks. Overall, team members outperformed individuals, and a relationship emerged between performance and inter-brain synchrony within a group⁹⁹. However, although dual-brain synchronizations have been interpreted as reflecting the neural correlates of social interaction, it is important to consider that this phenomenon could simply reflect low-level perceptual mechanisms involved in perceiving a common environment (e.g., auditory processing)¹⁰⁰. Nevertheless, hyperscanning merits being contemplated as a potentially important tool for socially interactive L2 learning studies.

Overall, neuroimaging studies provide encouraging first steps toward understanding how social interaction shapes the brain mechanisms involved in L2 learning. Recent findings suggest that socially grounded encoding can lead to more native-like neural responses during word retrieval and may engage brain regions associated with social and attentional processes. However, stimuli are typically presented on flat screens, which clearly restricts our ability to capture how we learn languages in real life¹⁰¹. Furthermore, most have relied on the passive viewing of pre-recorded social scenarios or participants imagining social partners, thus limiting our perception of real-time, embodied, and multimodal social interactions. Finally, by dichotomizing conditions as “social” versus “non-social”, prior work has not isolated the specific components of social interaction that potentially drive advantages in learning. Future research should integrate neural measurements into more immersive, life-like settings that allow for real-time, responsive interaction with a social partner. Along these lines, Titone & Tiv⁶ recently proposed the “socially situated” Systems Framework of Bilingualism, whereby psycholinguists and cognitive neuroscientists are asked to “rethink experience”, going beyond individual learner abilities to thoroughly consider sociolinguistic and sociocultural experiences in understanding language and multilingualism as a whole. This is crucial for capturing the full complexity of socially interactive L2 learning as it unfolds in natural contexts, and for uncovering how such interactions shape underlying neurocognitive processes.

Word learning: a case study in socially interactive learning

Given the challenges and opportunities outlined above, it is important to explore new ways of advancing socially interactive L2 learning, both in theory and through innovative experimental set-ups. One illustrative sub-domain of L2 learning is word learning. While L2 language learning extends far beyond individual words, words are the natural starting point: they are the basic building blocks of comprehension and production and crucial contributors to linguistic fluency¹⁰². Social, embodied and multimodal learning might allow learners to more effectively integrate new words due to the richness of the semantic context, compared to word-meaning translation-type learning. Memory traces are reactivated after we encode and integrate new information that is semantically congruent with our existing

knowledge, also known as conceptual schemas¹⁰³. Considering that social contexts are a daily, familiar reality for most of us, they could represent “existing” knowledge within which a novel linguistic element is inserted during learning. Furthermore, using language for interacting with others binds processes related to perception, attention, memory, and emotion together¹⁰⁴. The complexity of social environments and interactions could lead to deeper encoding of new linguistic content and hence long-lasting memory traces⁹. This effect pertains to the levels-of-processing hypothesis¹⁰⁵, according to which memory encoding depends on the depth of processing required.

Following this reasoning, socially interactive L2 learning would ultimately allow for faster and more robust consolidation of associated information, such as the integration of new words into existing lexical-semantic networks¹⁰⁶. The mental lexicon is generally conceptualized as consisting of two interconnected areas of representation: word form, encompassing a word’s phonological and orthographic characteristics, and word meaning, which refers to a word’s semantic associations with other words and its relationship to the real world¹⁰⁷. A major question in the L2 literature is whether L2 lexicosemantic knowledge relies on the L1 lexicon to access meaning, or whether there is a direct connection between the L2 lexicon and meaning. The influential Revised Hierarchical Model¹⁰⁸ (Fig. 2A) suggests that the L2 lexicon initially relies on L1 translations during early learning and later develops direct connections to conceptual meanings as proficiency increases. Initial supporting evidence came from translation tasks showing that L2 learners whose L1 was still dominant had greater difficulty translating from the L1 to the L2 than the other way around^{108,109}.

However, L2 words may be represented as both lexical and conceptual entries, like L1 words, if acquired in environments where both form and meaning are emphasized¹¹⁰. By enriching the learning context, multimodal and socially interactive L2 learning environments could help L2 learners establish more direct connections between L2 words and concepts from the very beginning of L2 learning and bypass reliance on L1 translations. We propose an extension of the Revised Hierarchical Model, in which socially enriched contexts would facilitate the establishment of more direct links between L2 lexical items and conceptual representations compared to learning in classical non-social settings (Fig. 2B). Our model allows for the examination of how different aspects of social interaction (multimodal, embodied, interactive and social feedback factors) supports L2 word learning.

VR as a promising avenue for advancing research on socially interactive second language learning

Taken together, the studies and theoretical frameworks reviewed here underscore the central importance of social interaction in L2 learning, while also revealing a significant gap in our understanding of the mechanisms that support it. This gap is largely due to the challenges of

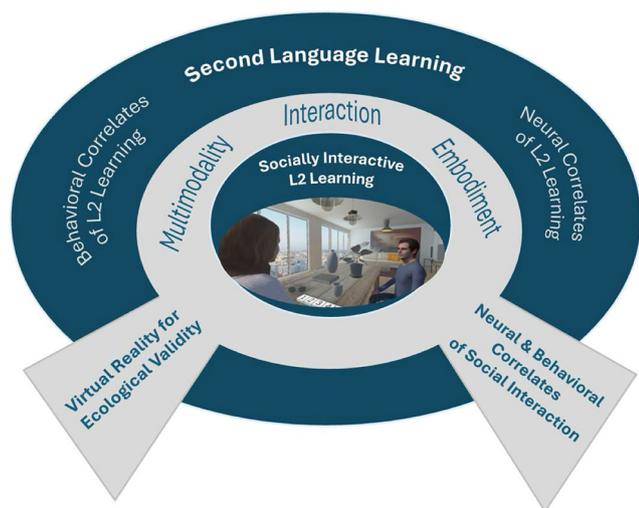


Fig. 3 | Versatility of socially interactive L2 learning studies that use virtual reality. The blue, outer circle of this figure describes different measures that can be used to examine L2 learning studies: behavioral and neural. The grey, inner circle lists three components of L2 learning in a social environment: multimodality, interaction, and embodiment. The left-hand side grey ray represents the benefit of virtual reality for examining socially interactive L2 learning, and the right-hand side ray, that of taking into account neural and behavioral correlates of social interaction in such studies. Finally, the inner blue circle shows a socially interactive L2 learning VR environment with two social agents – one represents the participant and the other an interactive virtual partner. The human characters in this figure are AI-generated.

creating appropriate experimental settings¹⁵. Truly grasping how social interaction shapes L2 learning calls for experimental paradigms that bridge social interactions and appropriate, reliable behavioral and brain measures in highly controlled yet ecologically valid environments. We argue here that virtual reality (VR) offers exactly this methodological bridge. VR provides a much sought-after means of sufficiently engaging the sensorimotor system to elicit meaningful responses¹¹¹, thus leading to findings that can generalize to real life⁴⁷ (Fig. 3). In what follows, we review how VR can help us to make significant strides in understanding the role of social interaction in L2 learning.

Although VR cannot replace real human interactions, it is a powerful tool that can facilitate insights into social dynamics and interactions, even without perfect realism^{112,113}. VR strives to replicate reality through three-dimensional environments with visual and auditory stimuli that often evoke a sense of presence in an immersive virtual world. Here, the concept of “presence” refers to the feeling of being in the virtual space and that events that are occurring there are truly occurring^{113,114}. These environments afford interactions with both objects and social agents (i.e., virtual characters that interact with participants), which can bring experimental studies on language processing and learning much closer to a real-life experience. According to Peeters, VR “shifts the theoretical focus toward the interplay between different modalities [...] in dynamic and communicative environments, complementing studies that focus on one modality in isolation”⁹⁷. Not only are users perceptively immersed, but they can perform naturalistic actions and receive real-time feedback for these actions, whether it be manipulating objects or conversing with a social agent. Crucially, VR allows for the full control of multimodal sensory stimulation (i.e., the onset, location, and duration of presentation of the stimuli) to observe how specific visual and auditory input influences linguistic processes.

Though limited in number, studies that used virtual environments (VE) to investigate L2 learning have found an advantage for “embodied” learning in comparison with less contextual learning^{115–117}. However, studies using VEs are limited in terms of visual immersion, physical interaction, and embodiment compared to VR studies that use head-mounted displays

(HMDs)¹¹⁸. HMDs and Cave virtual environments (CAVEs) afford a heightened sense of embodiment by surrounding participants visually and enabling direct interactions via hand-held controllers, finger trackers, and hand-tracking. In the case of HMDs – now widely regarded as the immersive, affordable, and portable gold standard in VR research –, when participants are embodied in a virtual body, this provides them with a sense of body ownership and agency⁶¹. Recently, interactive VR strategies involving either an HMD or a CAVE have been combined with brain measures in a growing range of language studies^{71,119–123}. Overall, these more immersive studies represent an important advance in observing the behavioral and neurocognitive substrates of situated and interactive language processing and learning. However, although these studies provide initial evidence in support of the idea that multimodal contextual cues and embodied processes can boost language processing and L2 learning, they lack a genuine social component. Indeed, most VR studies on L2 learning so far have focused on physical interaction with objects or environments but have not integrated meaningful social interactions. The absence of social agents capable of engaging in responsive communication within these environments leaves a significant gap in understanding how social factors shape L2 learning. Immersive VR studies offer a way forward: by surrounding participants visually and enabling naturalistic interaction, they provide a powerful means to investigate situated and socially grounded language learning under controlled conditions. In the remainder of this review, we use “virtual reality” (VR) to refer specifically to immersive experiences delivered via head-mounted displays (HMDs).

It is important to point out that there are clear limitations to using VR social agents as substitutes for real humans, particularly in the realm of language and multimodal communication. For instance, delivering high-quality, real-time language input, especially in the form of discourse, can be challenging. However, rapidly advancing AI technology is poised to make this possible in the near future. For experiments requiring complex behaviors such as negotiating meaning or providing corrective feedback, actors, confederates, or naïve participants should (continue to) be used. Indeed, humans provide cues during these activities that are extremely difficult to reproduce. On the other hand, numerous social cognition studies using VR have shown that fully replicating reality is not necessary for us to experience “real” feelings and display real-world behavior¹²⁴. VR can convince us that something is happening (plausibility) and that we are in a particular place (place illusion)^{112,113,125,126}.

Social paradigms that engage and manipulate simpler social, multimodal cues, along with environmental cues, could nevertheless provide the illusion of a real social interaction, representing significant potential to advance L2 learning studies. Again, these paradigms offer greater experimental control compared to using human counterparts. For instance, in VR, multimodal factors such as the speaking pace, accent, and prosody of a virtual agent can be altered while maintaining the same physical appearance (or vice versa). Additionally, the learning environment or the interlocutors’ physical appearance and behavior (i.e., social feedback such as smiling and nodding) can be finely altered to the desired degree. Moreover, VR design is rapidly advancing, creating interactive social agents modeled on humans who produce increasingly convincing facial and physical movement¹²⁷. During L2 learning, key variables, such as agent gaze, feedback valence, and physical characteristics (e.g., age, gender), can be manipulated to examine their effects on learning outcomes. Moreover, virtual agents provide consistency that is difficult to achieve with human counterparts. As mentioned earlier, the second-person neuroscience approach calls for the examination of the neural correlates of two-person social interactions, which are shown to differ from those seen during social observation and to engage the ‘mentalizing network’⁹⁴. Although VR social agents are not real humans, they can be perceived as social partners and engage similar processes as human-to-human interactions, contingent on visual, behavioral, and social factors^{126,128–131}. For instance, direct gaze from a virtual agent was shown to engage reward networks¹³².

Beyond creating immersive and interactive environments, VR also enables the precise measurement of behavioral and psychophysiological

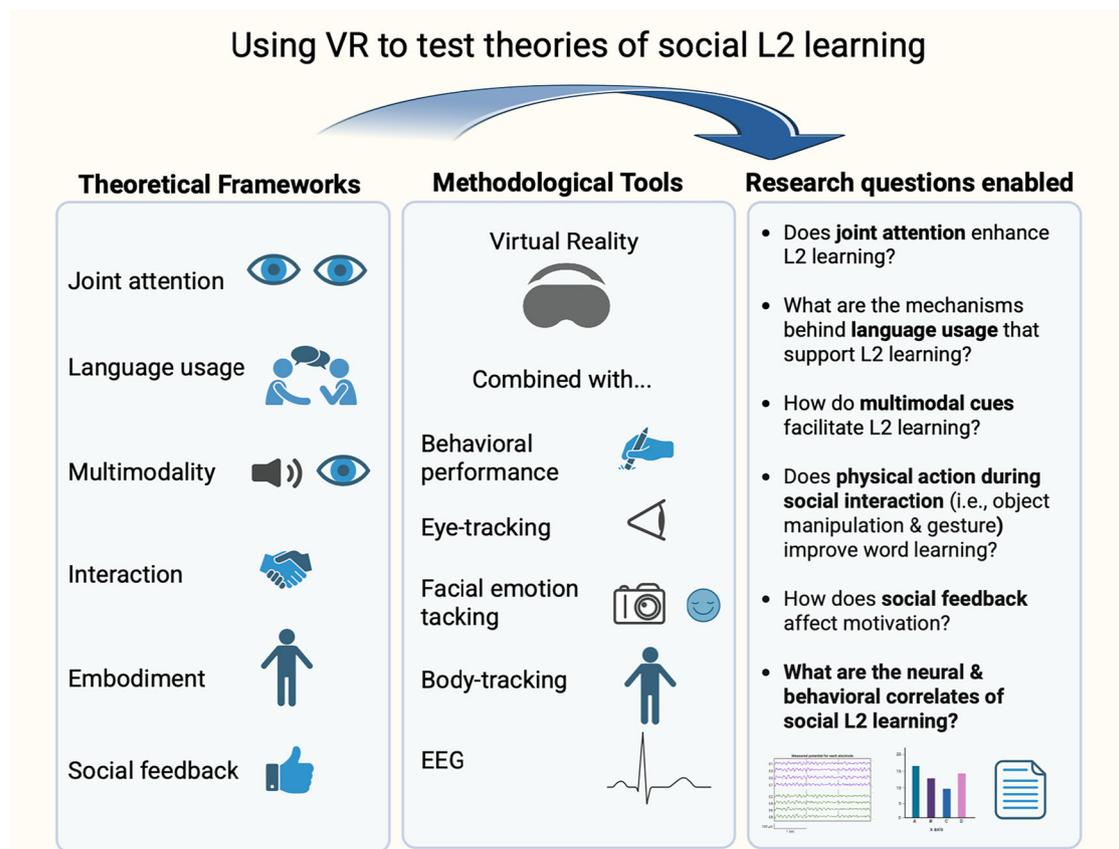


Fig. 4 | Illustration of how immersive VR, combined with behavioral and neurophysiological tools, enables targeted testing of core mechanisms proposed in socially grounded L2 learning theories. Created in BioRender. Zappa, A. (2025) <https://BioRender.com/n8mobrf>.

responses during language learning. Performance-based metrics, such as response times and accuracy, along with self-report measures of engagement and motivation, offer insight into the effectiveness of different social teaching strategies. But VR goes further by allowing for the continuous tracking of learners' movements and attentional dynamics. Through HMDs and hand controllers or hand-tracking systems, eye, head, hand, and body movements are recorded in real time, offering rich data on embodied and social behavior. Integrated eye-tracking systems, for instance, allow for the analysis of visual attention and joint attention, key mechanisms in social learning¹⁷, (e.g., by capturing when and how learners follow a virtual agent's gaze). Subtle changes in learners' facial muscle activity can also be measured via integrated cameras or sensors, enabling real-time emotion recognition during interactive learning tasks^{133,134}. Social engagement can be quantified via body movement synchrony between learners and virtual teachers, a phenomenon linked to communicative success in L1 contexts^{135,136}. Finally, communicative production can be recorded through metrics like hand movements and vocalizations⁴⁷. Additional measures, such as heart rate, electrodermal activity, and speech or gesture patterns, provide further insight into the emotional and interactive dimensions of the learning experience. The combination of immersive social scenarios with high-resolution behavioral and physiological tracking renders VR an unparalleled methodological platform for investigating how social cues and interactions shape L2 learning processes.

How VR can be used to test socially grounded frameworks of second language learning

VR is particularly valuable as a tool for investigating the socially interactive L2 learning frameworks discussed in the first half of this review in that it allows us to manipulate a specific aspect of social interaction (e.g., social gaze, social feedback) while controlling other complex factors (e.g., the

physical appearance and behavior of a social agent), facilitating replicability¹³⁷ (Fig. 4). In line with joint attention dynamics widely studied in L1 acquisition, we recently combined VR and eye-tracking to examine whether gaze following supported L2 word encoding in adults. As an example, we taught participants novel L2 words with three different virtual teachers: a Good teacher (who gazed at the correct object while producing a word), a Bad teacher (who gazed at the incorrect object), and a Random teacher (who gazed at the correct or incorrect object, randomly) (Zappa et al., under review). All teachers provided reliable social feedback. Words learned with the Good teacher showed improved performance post-training, particularly when participants followed his gaze towards the correct object during learning. According to self-reports, participants were unaware of the manipulation. Thus, reliable referential gaze seems to boost the mapping between words and objects unconsciously, beyond feedback alone, mirroring joint attention mechanisms in L1 acquisition. These results demonstrate how eye-tracking in VR can reveal non-conscious learning mechanisms that would remain undetected in traditional behavioral studies. This type of manipulation would be very difficult to achieve in a controlled manner using actors or confederates who were physically present. Furthermore, using a flatscreen would take away from participants' feeling of being in the same physical space as the agent (copresence), as well as that of looking at the same objects, in a shared physical space, essential for joined attention and life-like gaze following. Although immersive VR studies highlight the benefits of embodied interaction for L2 learning, few have tested contingent social cues like gaze or feedback. Our recent study is, to our knowledge, the first to do so, pointing to a promising direction for theory-driven research on social processes in language learning.

Second language acquisition (SLA) approaches align well with VR environments that could enforce iterative practice loops through contingent responses from virtual agents. VR also offers a unique opportunity to

examine multimodal components both separately and in combination by manipulating visual (gestures, gaze), and auditory (prosody) cues produced by virtual agents. Regarding embodied processes during social interaction, learners can manipulate objects (i.e., hand an object to a social agent), see and perform gestures, and move through social environments as they process and acquire new words. VR also provides an ideal context to test social feedback as informational and affective reinforcement. Virtual agents can deliver precisely timed verbal and nonverbal feedback, such as praise, corrections, or facial expressions, in response to learner performance. These cues can be systematically varied in valence, modality, and contingency to investigate differences in learning outcomes and motivation. Further, our extension of the Revised Hierarchical Model¹⁰⁸ (Fig. 2B) can be empirically tested using VR to compare word learning outcomes across socially enriched and more traditional learning contexts. Virtual agent referential gaze or gestures, and both agents and learners' interaction with objects can be manipulated while linguistic input remains strictly controlled.

A potentially crucial yet under-studied aspect of socially interactive second language learning are individual differences in willingness to communicate (WTC), or how likely a speaker is to engage in oral communication¹³⁸. Conceived as a dynamic construct shaped by both personality traits and context-sensitive variables^{139–141}, WTC is clearly influenced by established social dynamics¹⁴² and inherent social rewards. It therefore represents an important factor when examining how motivation impacts L2 learning^{38,143}. While it can be viewed as an individual trait (proneness to speak after learning), it can also vary over time, or according to learning situations¹⁴⁴. Ties between social interaction, learning motivation, and WTC make it an excellent candidate to study in conjunction with L2 learning in socially interactive environments. Using virtual social agents, future studies could measure the impact of different social learning scenarios, instructors, and feedback on WTC, L2 learning motivation, and learning success. Importantly, within such environments, we can assess how learners' communicative intent is modulated by specific features of interaction, for example, whether contingent feedback or perceived agent cooperativeness increases or decreases WTC over time. In addition, VR enables the collection of fine-grained behavioral and physiological data (e.g., gaze patterns, facial emotions, speech onset latencies, body orientation), offering converging measures of learners' social engagement and readiness to initiate communication (Fig. 4).

In all of the above examples, behavioral, psychophysiological (i.e., eye-tracking, facial emotion recognition), and neural responses can reveal how different aspects of social interaction shape L2 learning. VR is particularly well-suited to investigating the brain mechanisms underlying socially grounded L2 learning. As mentioned earlier, neural measures allow us to observe L2 learning phenomena that might not be measurable behaviorally, including semantic processing, syntactic processing, reward processes, feedback expectation and processing, and inter-subject correlation (when more than one learner is involved). One straightforward way to test the effect of socially interactive learning is to measure brain activity using EEG or fMRI during language processing before and after immersive VR learning sessions (compared to a control condition), to measure changes induced by learning^{71,119}. Although more challenging to set up, EEG can be directly measured while participants wear an HMD headset, as long as they remain seated and movement is limited during critical time windows, to avoid artifacts. It is crucial to measure the neural correlates of socially interactive L2 learning during the learning tasks themselves. EEG is especially valuable for studying how language learning unfolds in real time, as brain activity is captured with excellent temporal resolution that matches incredibly rapid language computations. Also, social reward, defined as positive verbal or non-verbal experience such as praise, smiling, or encouraging¹⁴⁵ has been linked to attentional bias¹⁴⁶ and improved learning but has not yet been investigated in connection with L2 learning with social partners. Importantly, although EEG-derived measures associated with processes related to socially interactive L2 learning have been studied, such as beta synchronization as a marker of reward¹⁴⁷ or inter-brain synchrony as a marker of shared attention and engagement^{98,99,148}, to our knowledge no studies have

employed these measures during socially interactive L2 learning in ecologically valid environments. Using VR to present learners with realistic, immersive, social experiences while acquiring brain and behavioral measures of social interaction and L2 learning yields a novel avenue of ecologically valid, yet controlled, L2 learning research.

Applications

While this review has emphasized VR as a tool for fundamental research, experimental advances in how social interaction impacts L2 learning are likely to inform innovations in instruction and rehabilitation. Knowledge about how learners respond to social cues and dynamics can inform the design of instructional strategies that promote active engagement with responsive interlocutors. It can also enrich language rehabilitation strategies. For example, people with aphasia show considerable variability in their ability to relearn words^{68,149,150}. While this is often attributed to lesion differences, interaction itself may serve as a crucial motivational and contextual support. It has been shown that incorporating meaningful communication into therapy is more effective than traditional production-based methods alone¹⁵¹. Interaction-based paradigms in VR, such as those involving multimodal cues and real-time feedback, could therefore offer a promising avenue for the development of more engaging and responsive rehabilitation protocols. Similar opportunities arise in autism research. A recent study by De Felice et al.⁹³ found that both autistic and neurotypical adults learned concepts most effectively during live interactive sessions, compared to observing recorded interactions or watching a monologue. Enjoyment was also highest in the interactive condition and strongly predicted learning outcomes. These findings suggest that social engagement can support learning even in populations typically considered less responsive to social cues. Transferring such paradigms to immersive VR would allow researchers to systematically manipulate degrees of co-presence, interactivity, and responsiveness, offering new insights into the role of social dynamics in learning across neurodiverse populations. In sum, research on word learning with social interaction can significantly advance both our theoretical understanding of L2 learning and the development of personalized, socially responsive teaching and rehabilitation strategies.

Conclusion

Given the central function that social interaction plays in language, this aspect could, in turn, significantly affect how an L2 is learned. Although the literature shows an increasing interest in this facet of L2 learning, researchers are often met with the challenge of providing learners with ecologically valid, socially interactive environments in which to examine the social component of L2 learning. Beyond offering a new methodological tool, VR provides a means to empirically test and refine existing theories of L2 learning. By isolating and manipulating factors such as joint attention, social feedback, embodiment, and multimodal input, VR can help to reveal which social mechanisms underlie comprehension and learning, thereby sharpening current theoretical accounts. Teaching participants an L2 via highly engaging virtual social agents makes it possible to measure when and how particular aspects of social environments and interactions modify or enhance L2 learning. This integration is not merely hypothetical – recent studies have already begun to demonstrate the feasibility and scientific value of using VR to examine these learning frameworks. Looking ahead, the combination of VR with neurophysiological and psychophysiological measures opens exciting possibilities for examining how social engagement influences language encoding and consolidation in real time. As adaptive, AI-driven virtual agents become increasingly sophisticated and capable of contingent dialog, they will offer unprecedented opportunities to investigate how motivation, social reward, and communicative alignment unfold during L2 learning. VR is not a mere technological upgrade; it represents a paradigm shift. Ultimately, this shift may enable a deeper understanding of the cognitive and neurocognitive mechanisms that underlie L2 learning in social contexts, refining and empirically grounding our proposed extension of the Revised Hierarchical Model (Fig. 2B). By enabling

experimental control without sacrificing ecological validity, this line of research opens the door to a more complete science of L2 learning: one that finally captures the social, dynamic nature of how languages are truly learned.

Data availability

No datasets were generated or analyzed during the current study.

Received: 27 October 2025; Accepted: 6 November 2025;

Published online: 19 December 2025

References

- Dörnyei, Z. *The Psychology of the Language Learner: Individual Differences in Second Language Acquisition*. xii, 270 (Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, 2005).
- Brown, H. D. *Principles of Language Learning And Teaching*. TESOL Q. **14**, 188–217 (2000).
- Richards, J. C. & Lockhart, C. *Reflective Teaching in Second Language Classrooms*. 52–72 (Cambridge University Press, 1994).
- Bardovi-Harlig, K. & Bastos, M.-T. Proficiency, length of stay, and intensity of interaction and the acquisition of conventional expressions in L2 pragmatics. **8**, 347–384 (2011).
- Usage in Second Language Acquisition: Critical Reflections and Future Directions. (Routledge, New York, 10.4324/9781032668475.(2024).
- Titone, D. A. & Tiv, M. Rethinking multilingual experience through a Systems Framework of Bilingualism. *Biling. Lang. Cogn.* **26**, 1–16 (2023).
- Duff, P. & Surtees, V. Chapter 11 Learning Through Social Interaction. in 101–109 <https://doi.org/10.1017/9781009024761.015> (2018).
- Group, T. D. F. A transdisciplinary framework for SLA in a multilingual world. *Mod. Lang. J.* **100**, 19–47 (2016).
- Li, P. & Jeong, H. The social brain of language: grounding second language learning in social interaction. *Npj Sci. Learn.* **5**, 1–9 (2020).
- Friedrich, M. & Friederici, A. D. Neurophysiological correlates of online word learning in 14-month-old infants. *NeuroReport* **19**, 1757 (2008).
- Childers, J. B. & Tomasello, M. Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Dev. Psychol.* **38**, 967–978 (2002).
- Friedrich, M., Wilhelm, I., Born, J. & Friederici, A. D. Generalization of word meanings during infant sleep. *Nat. Commun.* **6**, 6004 (2015).
- Nikolaus, M. & Fourtassi, A. Modeling the interaction between perception-based and production-based learning in children's early acquisition of semantic knowledge. In Proceedings of the 25th Conference on Computational Natural Language Learning (CoNLL) (Punta Cana, Dominican Republic, 2021).
- Tomasello, M. First steps toward a usage-based theory of language acquisition. *Cogn. Linguist.* **11**, 61–82 (2000).
- Verga, L. & Kotz, S. A. How relevant is social interaction in second language learning? *Front. Hum. Neurosci.* **7**, 550 (2013).
- Kuhl, P. K. Brain mechanisms in early language acquisition. *Neuron* **67**, 713–727 (2010).
- Tomasello, M. & Farrar, M. J. Joint attention and early language. *Child Dev.* **57**, 1454–1463 (1986).
- Kuhl, P. K., Tsao, F.-M. & Liu, H.-M. Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proc. Natl. Acad. Sci.* **100**, 9096–9101 (2003).
- Hirofani, M., Stets, M., Striano, T. & Friederici, A. Joint attention helps infants learn new words: Event-related potential evidence. *Neuroreport* **20**, 600–605 (2009).
- Tamis-LeMonda, C., Kuchirko, Y. & Song, L. Why is infant language learning facilitated by parental responsiveness?. *Curr. Dir. Psychol. Sci.* **23**, 121–126 (2014).
- Yurovsky, D. A communicative approach to early word learning. *N. Ideas Psychol.* **50**, 73–79 (2018).
- Eskildsen, S. W. *The Long Haul: Qualitative Usage-Based Research in Second Language Development*. (Routledge, 2022).
- Hymes, D. H. The ethnography of speaking. In *Anthropology and Human Behavior*. Gladwin, T. & Sturtevant, W. (Eds) 13–53 (Anthropological Society of Washington, Washington, DC, 1962).
- Vygotsky, L. S. & Cole, M. *Mind in Society: Development of Higher Psychological Processes*. (Harvard University Press, 1978).
- Lantolf, J. P. Second language learning as a mediated process. *Lang. Teach.* **33**, 79–96 (2000).
- Pavlenko, A. & Lantolf, J. Second language learning as participation: Reconstructing a self. In *Sociocultural theory and second language learning: Recent advances* 155–178 (2000).
- Thorne, S., Lantolf, J. & Thorne, S. L. (2007). Sociocultural Theory and Second Language Learning. In B. van Patten & J. Williams (eds.), *Theories in Second Language Acquisition* (pp. 201–224). Mahwah, NJ: Lawrence Erlbaum (2007).
- Ellis, N., Römer-Barron, U. & O'Donnell, M. Usage-Based Approaches to Language Acquisition and Processing: Cognitive and Corpus Investigations of Construction Grammar. (2016).
- Ellis, N. & Ferreira-Junior, F. Constructions and their acquisition Islands and the distinctiveness of their occupancy. *Annu. Rev. Cogn. Linguist.* **7**, 187–220 (2009).
- Pekarek Doehler, S. & Berger, E. L2 Interactional Competence as Increased Ability for Context-sensitive Conduct: A Longitudinal Study of Story-openings. *Appl. Linguist.* **39**, 555–578 (2018).
- Pekarek Doehler, S. & Skogmyr Marian, K. Functional diversification and progressive routinization of a multiword expression in and for social interaction: a longitudinal L2 Study. *Mod. Lang. J.* **106**, 23–45 (2022).
- Eskildsen, S. W. & Wagner, J. From Trouble in the Talk to New Resources: The Interplay of Bodily and Linguistic Resources in the Talk of a Speaker of English as a Second Language. In *Longitudinal Studies on the Organization of Social Interaction* (eds. Pekarek Doehler, S., Wagner, J. & González-Martínez, E.) 143–171 (Palgrave Macmillan UK, London, 2018).
- DeKeyser, R. Skill acquisition theory. In *Theories in second language acquisition*. 83–104 (Routledge, 2020).
- Segalowitz, N. *Cognitive Bases of Second Language Fluency*. Routledge, New York, <https://doi.org/10.4324/9780203851357> (2010).
- Suzuki, Y. Introduction: Practice and Automatization in a Second Language. In *Practice and Automatization in Second Language Research*, (Routledge, 2023).
- DeKeyser, R. M. Introduction: Situating the concept of practice. In *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (ed. DeKeyser, R.) 1–18 (Cambridge University Press, Cambridge, 2007).
- Ortega, L. Meaningful L2 practice in foreign language classrooms: A cognitive-interactionist SLA perspective. In *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (ed. DeKeyser, R.) 180–207 (Cambridge University Press, Cambridge, 2007).
- Elahi Shirvan, M., Khajavy, G. H., MacIntyre, P. D. & Taherian, T. A meta-analysis of L2 willingness to communicate and its three high-evidence correlates. *J. Psycholinguist. Res.* **48**, 1241–1267 (2019).
- Bates, E. & MacWhinney, B. *Competition, Variation, and Language Learning*. In *Mechanisms of Language Acquisition*, Psychology Press, (1987).
- Bates, E. & Macwhinney, B. Functionalism and the competition model. *Macwhinney Bates* **3**, 76 (1989).
- Tulving, E. & Thomson, D. M. Encoding specificity and retrieval processes in episodic memory. *Psychol. Rev.* **80**, 352–373 (1973).
- Hamrick, P. Declarative and procedural memory abilities as individual differences in incidental language learning. *Learn. Individ. Differ.* **44**, 9–15 (2015).

43. Morgan-Short, K., Faretta-Stutenberg, M., Brill-Schuetz, K. A., Carpenter, H. & Wong, P. C. M. Declarative and procedural memory as individual differences in second language acquisition. *Biling. Lang. Cogn.* **17**, 56–72 (2014).
44. Menks, W. M. et al. Study protocol: a comprehensive multi-method neuroimaging approach to disentangle developmental effects and individual differences in second language learning. *BMC Psychol.* **10**, 169 (2022).
45. Bowden, H. W. & Faretta-Stutenberg, M. Context of learning in second language neurocognition. In *The Routledge Handbook of Second Language Acquisition and Neurolinguistics*, Routledge, (2023).
46. Atkinson, D. Toward a sociocognitive approach to second language acquisition. *Mod. Lang. J.* **86**, 525–545 (2002).
47. Peeters, D. Virtual reality: A game-changing method for the language sciences. *Psychon. Bull. Rev.* **26**, 894–900 (2019).
48. Öttl, B., Dudschig, C. & Kaup, B. Forming associations between language and sensorimotor traces during novel word learning. *Lang. Cogn.* **9**, 156–171 (2017).
49. Mayer, K. M., Yildiz, I. B., Macedonia, M. & von Kriegstein, K. Visual and motor cortices differentially support the translation of foreign language words. *Curr. Biol.* **25**, 530–535 (2015).
50. Morett, L. M. The power of an image: images, not glosses, enhance learning of concrete L2 words in beginning learners. *J. Psycholinguist. Res.* **48**, 643–664 (2019).
51. Vigliocco, G., Perniss, P. & Vinson, D. Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philos. Trans. R. Soc. B Biol. Sci.* **369**, 20130292 (2014).
52. Drijvers, L. & Holler, J. The multimodal facilitation effect in human communication. *Psychon. Bull. Rev.* <https://doi.org/10.3758/s13423-022-02178-x> (2022)
53. Zhang, Y., Frassinelli, D., Tuomainen, J., Skipper, J. I. & Vigliocco, G. More than words: word predictability, prosody, gesture and mouth movements in natural language comprehension. *Proc. R. Soc. B Biol. Sci.* **288**, 20210500 (2021).
54. Cutler, A., Dahan, D. & van Donselaar, W. Prosody in the comprehension of spoken language: a literature review. *Lang. Speech* **40**, 141–201 (1997).
55. Dahan, D. Prosody and language comprehension. *Wiley Interdiscip. Rev. Cogn. Sci.* **6**, 441–452 (2015).
56. Dargue, N., Sweller, N. & Jones, M. P. When our hands help us understand: A meta-analysis into the effects of gesture on comprehension. *Psychol. Bull.* **145**, 765–784 (2019).
57. Zhang, Y. et al. The role of multimodal cues in second language comprehension. *Sci. Rep.* **13**, 20824 (2023).
58. De Felice, S., Vigliocco, G., Hamilton, A. F. & de, C. Social interaction is a catalyst for adult human learning in online contexts. *Curr. Biol.* **31**, 4853–4859.e3 (2021).
59. Barsalou, L. W. On staying grounded and avoiding quixotic dead ends. *Psychon. Bull. Rev.* **23**, 1122–1142 (2016).
60. Pulvermüller, F. Words in the brain's language. *Behav. Brain Sci.* **22**, 253–336 (1999).
61. Johnson-Glenberg, M. C. Immersive VR and Education: Embodied Design Principles That Include Gesture and Hand Controls. *Front. Robot. AI* **5**, 81 (2018).
62. James, K. H. & Swain, S. N. Only self-generated actions create sensori-motor systems in the developing brain. *Dev. Sci.* **14**, 673–678 (2011).
63. Krönke, K.-M., Mueller, K., Friederici, A. D. & Obrig, H. Learning by doing? The effect of gestures on implicit retrieval of newly acquired words. *Cortex* **49**, 2553–2568 (2013).
64. Macedonia, M. & Knösche, T. R. Body in mind: how gestures empower foreign language learning. *Mind Brain Educ.* **5**, 196–211 (2011).
65. Tellier, M. The effect of gestures on second language memorisation by young children. *Gesture* **8**, 219–235 (2008).
66. Goldin-Meadow, S. & Alibali, M. W. Gesture's role in speaking, learning, and creating language. *Annu. Rev. Psychol.* **64**, 257–283 (2013).
67. Morett, L. M. In hand and in mind: Effects of gesture production and viewing on second language word learning. *Appl. Psycholinguist.* **39**, 355–381 (2018).
68. Kelly, H. & Armstrong, L. New word learning in people with aphasia. *Aphasiology* **23**, 1398–1417 (2009).
69. Macedonia, M., Repetto, C., Ischebeck, A. & Mueller, K. Depth of Encoding Through Observed Gestures in Foreign Language Word Learning. *Front. Psychol.* **10**, 33 (2019).
70. Sweller, N., Shinooka-Phelan, A. & Austin, E. The effects of observing and producing gestures on Japanese word learning. *Acta Psychol.* **207**, 1–11 (2020).
71. Zappa, A. et al. The neural correlates of embodied L2 learning: Does embodied L2 verb learning affect representation and retention? *Neurobiol. Lang.* **5**, 360–384 (2024)..
72. Garagnani, M., Kirilina, E. & Pulvermüller, F. Semantic grounding of novel spoken words in the primary visual cortex. *Front. Hum. Neurosci.* **15**, 66 (2021).
73. Repetto, C., Mathias, B., Weichselbaum, O. & Macedonia, M. Visual recognition of words learned with gestures induces motor resonance in the forearm muscles. *Sci. Rep.* **11**, 17278 (2021).
74. Bara, F. & Kaminski, G. Holding a real object during encoding helps the learning of foreign vocabulary. *Acta Psychol.* **196**, 26–32 (2019).
75. Sato, M. Beliefs about peer interaction and peer corrective feedback: Efficacy of classroom intervention. *Mod. Lang. J.* **97**, 611–633 (2013).
76. Berridge, K. C. & Kringelbach, M. L. Affective neuroscience of pleasure: reward in humans and animals. *Psychopharmacology* **199**, 457–480 (2008).
77. Kohls, G. et al. The nucleus accumbens is involved in both the pursuit of social reward and the avoidance of social punishment. *Neuropsychologia* **51**, 2062–2069 (2013).
78. Williams, A. V. et al. Social approach and social vigilance are differentially regulated by oxytocin receptors in the nucleus accumbens. *Neuropsychopharmacology* **45**, 1423–1430 (2020).
79. Chevallier, C., Kohls, G., Troiani, V., Brodwin, E. S. & Schultz, R. T. The social motivation theory of autism. *Trends Cogn. Sci.* **16**, 231–239 (2012).
80. Williams, E. H., Cristino, F. & Cross, E. S. Human body motion captures visual attention and elicits pupillary dilation. *Cognition* **193**, 104029 (2019).
81. Hari, A. R., Tessoro, A., Mattay, V. S., Fera, F. & Weinberger, D. R. The Amygdala response to emotional stimuli: a comparison of faces and scenes. *NeuroImage* **17**, 317–323 (2002).
82. Norris, C. J., Chen, E. E., Zhu, D. C., Small, S. L. & Cacioppo, J. T. The interaction of social and emotional processes in the brain. *J. Cogn. Neurosci.* **16**, 1818–1829 (2004).
83. Davey, C. G., Allen, N. B., Harrison, B. J., Dwyer, D. B. & Yücel, M. Being liked activates primary reward and midline self-related brain regions. *Hum. Brain Mapp.* **31**, 660–668 (2010).
84. Mani, N. & Ackermann, L. Why do children learn the words they do?. *Child Dev. Perspect.* **12**, 253–257 (2018).
85. Ripollés, P. et al. The role of reward in word learning and its implications for language acquisition. *Curr. Biol. CB* **24**, 2606–2611 (2014).
86. Ripollés, P. et al. Intrinsic monitoring of learning success facilitates memory encoding via the activation of the SN/VTA-Hippocampal loop. *eLife* **5**, e17441 (2016).
87. Ripollés, P. et al. Intrinsically regulated learning is modulated by synaptic dopamine signaling. *eLife* **7**, e38113 (2018).

88. Lisman, J. E. & Grace, A. A. The Hippocampal-VTA Loop: Controlling the Entry of Information into Long-Term Memory. *Neuron* **46**, 703–713 (2005).
89. Jeong, H. et al. Learning second language vocabulary: Neural dissociation of situation-based learning and text-based learning. *NeuroImage* **50**, 802–809 (2010).
90. Jeong, H., Li, P., Suzuki, W., Sugiura, M. & Kawashima, R. Neural mechanisms of language learning from social contexts. *Brain Lang.* **212**, 104874 (2021).
91. Verga, L. & Kotz, S. A. Spatial attention underpins social word learning in the right fronto-parietal network. *NeuroImage* **195**, 165–173 (2019).
92. Yusa, N., Kim, J., Koizumi, M., Sugiura, M. & Kawashima, R. Social interaction affects neural outcomes of sign language learning as a foreign language in adults. *Front. Hum. Neurosci.* **11**, 115 (2017).
93. De Felice, S., Hamilton, A. F., de, C., Ponari, M. & Vigliocco, G. Learning from others is good, with others is better: the role of social interaction in human acquisition of new knowledge. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **378**, 20210357 (2023).
94. Redcay, E. & Schilbach, L. Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nat. Rev. Neurosci.* **20**, 495–505 (2019).
95. Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S. & Keysers, C. Brain-to-Brain coupling: A mechanism for creating and sharing a social world. *Trends Cogn. Sci.* **16**, 114–121 (2012).
96. Hirsch, J., Adam Noah, J., Zhang, X., Dravida, S. & Ono, Y. A cross-brain neural mechanism for human-to-human verbal communication. *Soc. Cogn. Affect. Neurosci.* **13**, 907–920 (2018).
97. Holper, L. et al. The teaching and the learning brain: A cortical hemodynamic marker of teacher-student interactions in the Socratic dialog. *Int. J. Educ. Res.* **59**, 1–10 (2013).
98. Dikker, S. et al. Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Curr. Biol. CB* **27**, 1375–1380 (2017).
99. Reiner, D. A., Dikker, S. & Van Bavel, J. J. Inter-brain synchrony in teams predicts collective performance. *Soc. Cogn. Affect. Neurosci.* **16**, 43–57 (2021).
100. Holroyd, C. B. Interbrain synchrony: on wavy ground. *Trends Neurosci.* **45**, 346–357 (2022).
101. Sobczak, A. & Bunzeck, N. Effects of positive and negative social feedback on motivation, evaluative learning, and socio-emotional processing. *Npj Sci. Learn.* **8**, 1–16 (2023).
102. Levelt, W. J. M. Speaking: From Intention to Articulation. The MIT Press. <https://doi.org/10.7551/mitpress/6393.001.0001> (1989).
103. van Kesteren, M., Ruiter, D., Fernández, G. & Henson, R. How schema and novelty augment memory formation. *Trends Neurosci.* **35**, 211–219 (2012).
104. Lantolf, J. P. & Xi, J. Digital language learning: a sociocultural theory perspective. *TESOL Q* **57**, 702–715 (2023).
105. Craik, F. I. M. & Lockhart, R. S. Levels of processing: A framework for memory research. *J. Verbal Learn. Verbal Behav.* **11**, 671–684 (1972).
106. Rodríguez-Fornells, A., Cunillera, T., Mestres-Missé, A. & de Diego-Balaguer, R. Neurophysiological mechanisms involved in language learning in adults. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **364**, 3711–3735 (2009).
107. Marslen-Wilson, W. D. Functional parallelism in spoken word-recognition. *Cognition* **25**, 71–102 (1987).
108. Kroll, J. F. & Stewart, E. Category interference in translation and picture naming: evidence for asymmetric connections between bilingual memory representations. *J. Mem. Lang.* **33**, 149–174 (1994).
109. Degroot, A. M. B., Dannenburg, L. & Vanhell, J. G. Forward and backward word translation by bilinguals. *J. Mem. Lang.* **33**, 600–629 (1994).
110. Kroll, J. F. & Tokowicz, N. The Development of Conceptual Representation for Words in a Second Language. in *One mind, two languages: Bilingual language processing*, 49–71. Blackwell Publishing, Malden, (2001).
111. Bohil, C. J., Alicea, B. & Biocca, F. A. Virtual reality in neuroscience research and therapy. *Nat. Rev. Neurosci.* **12**, 752–762 (2011).
112. Slater, M. et al. An experimental study of a virtual reality counselling paradigm using embodied self-dialogue. *Sci. Rep.* **9**, 10903 (2019).
113. Slater, M. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philos. Trans. R. Soc. B Biol. Sci.* **364**, 3549–3557 (2009).
114. Sanchez-Vives, M. & Slater, M. From presence to consciousness through virtual reality. *Nat. Rev. Neurosci.* **6**, 332–339 (2005).
115. Legault, J., Fang, S.-Y., Lan, Y.-J. & Li, P. Structural brain changes as a function of second language vocabulary training: Effects of learning context. *Brain Cogn.* **134**, 90–102 (2019).
116. Essoe, J. K.-Y. et al. Enhancing learning and retention with distinctive virtual reality environments and mental context reinstatement. *Npj Sci. Learn.* **7**, 31 (2022).
117. Vukovic, N. & Shtyrov, Y. Learning with the wave of the hand: Kinematic and TMS evidence of primary motor cortex role in category-specific encoding of word meaning. *NeuroImage* **202**, 116179 (2019).
118. Slater, M. & Sanchez-Vives, M. V. Enhancing our lives with immersive virtual reality. *Front. Robot. AI* **3**, 287 (2016).
119. Jiao, L. et al. An ERP study on novel word learning in an immersive virtual reality context. *Biling. Lang. Cogn.* **27**, 128–136 (2024).
120. Legault, J. et al. Immersive virtual reality as an effective tool for second language vocabulary learning. *Languages* **4**, 13 (2019).
121. Peeters, D. & Dijkstra, T. Sustained inhibition of the native language in bilingual language production: A virtual reality approach. *Biling. Lang. Cogn.* **21**, 1035–1061 (2018).
122. Tromp, J., Peeters, D., Meyer, A. S. & Hagoort, P. The combined use of virtual reality and EEG to study language processing in naturalistic environments. *Behav. Res. Methods* **50**, 862–869 (2018).
123. Zappa, A. et al. Motor resonance during linguistic processing as shown by EEG in a naturalistic VR environment. *Brain Cogn.* **134**, 44–57 (2019).
124. Hasler, B., Spanlang, B. & Slater, M. Virtual race transformation reverses racial in-group bias. *PLOS ONE* **12**, e0174965 (2017).
125. Martens, M. et al. It feels real: physiological responses to a stressful virtual reality environment and its impact on working memory. *J. Psychopharmacol.* **33**, 1264–1273 (2019).
126. Slater, M. et al. A separate reality: an update on place illusion and plausibility in virtual reality. *Front. Virtual Real.* **3**, 1–16 (2022).
127. Rekik, R., Wuhler, S., Hoyet, L., Zibrek, K. & Olivier, A. H. A survey on realistic virtual human animations: definitions, features and evaluations. *Comput. Graph. Forum* **43**, e15064 (2024).
128. Bailenson, J. N. & Yee, N. Digital Chameleons: automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychol. Sci.* **16**, 814–819 (2005).
129. Peeters, D. Bilingual switching between languages and listeners: Insights from immersive virtual reality. *Cognition* **195**, 104107 (2020).
130. Reeves, B. & Nass, C. I. *The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places*. xiv, 305 Cambridge University Press, New York, NY, US, (1996).
131. Slater, M. et al. Analysis of physiological responses to a social situation in an immersive virtual environment. *Presence* **15**, 553–569 (2006).
132. Schilbach, L. et al. Being with virtual others: Neural correlates of social interaction. *Neuropsychologia* **44**, 718–730 (2006).
133. Gjoreski, M. et al. Facial EMG sensing for monitoring affect using a wearable device. *Sci. Rep.* **12**, 16876 (2022).

134. Gong, Q., Liu, X. & Ma, Y. Real-time facial expression recognition based on image processing in virtual reality. *Int. J. Comput. Intell. Syst.* **18**, 8 (2025).
135. Shockley, K., Richardson, D. C. & Dale, R. Conversation and coordinative structures. *Top. Cogn. Sci.* **1**, 305–319 (2009).
136. Tsuchiya, A. et al. Body movement synchrony predicts degrees of information exchange in a natural conversation. *Front. Psychol.* **11**, 817 (2020).
137. Pan, X., Hamilton, A. F. & de, C. Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *Br. J. Psychol. Lond. Engl.* **109**, 395–417 (2018).
138. McCroskey, J. C. & Baer, J. E. Willingness to Communicate: The Construct and Its Measurement. <https://eric.ed.gov/?id=ED265604> (1985).
139. Katsaris, A. Willingness to Communicate (WTC): Origins, significance, and propositions for the L2/FL classroom. (2019).
140. Macintyre, P. D. Willingness to communicate in the second language: understanding the decision to speak as a volitional process. *Mod. Lang. J.* **91**, 564–576 (2007).
141. MacIntyre, P. Expanding the theoretical base for the dynamics of willingness to communicate. *Stud. Second Lang. Learn. Teach.* **10**, 111–131 (2020).
142. Dörnyei, Z. *Motivational Strategies in the Language Classroom*. (Cambridge University Press, (2001) <https://doi.org/10.1017/CBO9780511667343>).
143. Gardner, R. C., Lalonde, R. N. & Moorcroft, R. The role of attitudes and motivation in second language learning: correlational and experimental considerations. *Lang. Learn.* **35**, 207–227 (1985).
144. Zhang, J., Beckmann, N. & Beckmann, J. F. To talk or not to talk: A review of situational antecedents of willingness to communicate in the second language classroom. *System* **72**, 226–239 (2018).
145. Matyjek, M., Meliss, S., Dziobek, I. & Murayama, K. A Multidimensional View on Social and Non-Social Rewards. *Front. Psychiatry* **11**, 818 (2020).
146. Anderson, B. A. Social reward shapes attentional biases. *Cogn. Neurosci.* **7**, 30–36 (2016).
147. Marco-Pallarés, J., Münte, T. F. & Rodríguez-Fornells, A. The role of high-frequency oscillatory activity in reward processing and learning. *Neurosci. Biobehav. Rev.* **49**, 1–7 (2015).
148. Poulsen, A. T., Kamronn, S., Dmochowski, J., Parra, L. C. & Hansen, L. K. EEG in the classroom: Synchronised neural recordings during video presentation. *Sci. Rep.* **7**, 43916 (2017).
149. Coran, M., Rodríguez-Fornells, A., Ramos-Escobar, N., Laine, M. & Martin, N. Word learning in Aphasia: treatment implications and structural connectivity analyses. *Top. Lang. Disord.* **40**, 81–109 (2020).
150. Peñalosa, C., Martin, N., Laine, M. & Rodríguez-Fornells, A. Language learning in aphasia: A narrative review and critical analysis of the literature with implications for language therapy. *Neurosci. Biobehav. Rev.* **141**, 104825 (2022).
151. Stahl, B., Mohr, B., Dreyer, F. R., Lucchese, G. & Pulvermüller, F. Using language for social interaction: Communication mechanisms promote recovery from chronic non-fluent aphasia. *Cortex J. Devot. Study Nerv. Syst. Behav.* **85**, 90–99 (2016).

Acknowledgements

This research was funded by the European Union's H2021-MSCA-IF-2021 (project BraSILL No. 101062671). MS is supported by the European Research Council (ERC) Advanced Grant MoTIVE (#742989). This study has also received funding from the European Union's Horizon Europe research and innovation program under grant agreement No 101135025, PRESENCE project. We are very grateful to the anonymous reviewers for their thoughtful feedback and constructive suggestions.

Author contributions

A.Z. reviewed the literature and wrote the article. A.R.F. and M.L. contributed to the writing and revision of the article. All authors have read and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Ana Zappa.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025