Language Learning & Technology
ISSN 1094-3501 CC BY-NC-ND

February 2023, Volume 27, Issue 2
pp. 118–154

ARTICLE

Negotiation of meaning via virtual exchange in immersive virtual reality environments

Hsin-I Chen, *National Taipei University of Technology*

Ana Sevilla-Pavón, *Universiat de València*

Abstract

This study examines how English-as-lingua-franca (ELF) learners employ semiotic resources, including head movements, gestures, facial expression, body posture, and spatial juxtaposition, to negotiate for meaning in an immersive virtual reality (VR) environment. Ten ELF learners participated in a Taiwan-Spain VR virtual exchange project and completed two VR tasks on an immersive VR platform. Multiple datasets, including the recordings of VR sessions, pre- and post-task questionnaires, observation notes, and stimulated recall interviews, were analyzed quantitatively and qualitatively with triangulation. Built upon multimodal interaction analysis (Norris, 2004) and Varonis and Gass' (1985a) negotiation of meaning model, the findings indicate that ELF learners utilized different embodied semiotic resources in constructing and negotiating meaning at all primes to achieve effective communication in an immersive VR space. The avatar-mediated representations and semiotic modalities were shown to facilitate indication, comprehension, and explanation to signal and resolve non-understanding instances. The findings show that with space proxemics and object handling as the two distinct features of VR-supported environments, VR platforms transform learners' social interaction from plane to three-dimensional communication, and from verbal to embodied, which promotes embodied learning. VR thus serves as a powerful immersive interactive environment for ELF learners from distant locations to be engaged in situated languacultural practices that goes beyond physical space. Pedagogical implications are discussed.

Keywords: *Immersive Virtual Reality, Multimodality, Negotiation of Meaning, Virtual Exchange*

Language(s) Learned in This Study: *English*

APA Citation: Chen, H. I., & Sevilla-Pavón, A. (2023). Negotiation of meaning via virtual exchange in immersive virtual reality environments. *Language Learning & Technology*, *27*(2), 118–154. https://hdl.handle.net/10125/73506

Introduction

In the past years, the coronavirus pandemic has dramatically changed the way people live, act, work, and learn. Distant learning and remote collaboration have now become the 'new normal' of education. In language education, reforms of distance learning and remote collaboration can be witnessed with the prevalence of virtual reality (VR) technologies. Virtual reality is defined as "an immersive computer-enabled technology that replicates an environment and allows a simulation of the user to be present and interact in that environment" (Lloyd et al., 2017, p. 222). With the qualities of immersion, interaction, and imagination, VR not only addresses the limitations of traditional learning methods, but also facilitates the delivery of learning contents and mediated interaction through full immersion and interactivity (Burdea & Coiffet, 2003).

In second language learning, VR technologies have been shown to foster embodied learning and learner autonomy (Chen & Kent, 2020; Lan, 2014, 2020; Liaw, 2019). VR has the potential to revolutionize education based on its ability to "immerse students in their learning more than any other available medium" (Gadelha, 2018, p. 41). This could imply a connection to target language speakers in a way that closely resembles face-to-face communication with the immersive experience of VR environments (York et al.,

2021, p. 52). The immersion, active learner participation, social interaction, and authenticity of VR thus holds good potential for language learning (Lan, 2014).

While existing studies have examined the effectiveness of VR/AR (Augmented Reality) technologies to language learning, such as second language (L2) vocabulary learning (Alfadil, 2020; Legault et al., 2019), listening comprehension (Tai & Chen, 2021), writing motivation and performance (Lan et al., 2019), oral performance (Lan, 2014; Xie et al., 2019), and intercultural competence (Liaw, 2019), little research explores L2 learners' embodied interaction in immersive VR environments. More importantly, few examine learners' embodied interaction in a context of virtual exchanges where English-as-lingua-franca (ELF) learners from different geographical locations and diverse linguistic and cultural backgrounds interact with one another. Although learners' virtual exchanges have been explored in previous studies, most of them integrated asynchronous technological tools (e.g., emails, chats), synchronous voice or video conferencing platforms (e.g., Google Meet, Zoom, Skype), or two-dimensional or semi-immersive VR platforms (e.g., Second Life) to facilitate the collaboration with distant partners across the globe. The multimodal dimensions of immersive VR-supported virtual exchanges differ, in many important ways, from those supported by other common technologies such as videoconferencing. Immersive VR-mediated or VR-supported virtual exchanges, however, are still in its infancy and remain as yet largely unexplored. To address the gap in research and practice, this study examines the ways L2 learners engage in verbal and nonverbal interaction (e.g., gesture, eye gaze, facial expression, body posture) in an immersive VR environment during a Taiwan-Spain virtual exchange project. The purpose of the study is to explore how L2 learners utilize multimodal resources afforded in VR environments to fulfill specific communicative purposes and how different modalities can be used in a complementary, compensating, and competing manner to negotiate for meaning (Hampel & Sticker, 2012, p. 135). As *interaction* is considered the most salient and research-intensive aspect of language learning and teaching in virtual worlds (Wigham et al., 2018, p. 154), different modes of computer-mediated communication (CMC) directly influence how learners communicate their ideas and how they interact with each other (Stockwell, 2010).

Specifically, this study focuses on instances where L2 learners negotiate meaning via their L2 in VR space. Negotiation of meaning (NoM) is a collaborative attempt in a conversation between more or less fluent speakers in order to solve communication breakdowns and reach comprehension (Long, 1996). NoM, by asking for clarification, modifying utterances, improving message comprehensibility, or cooperating to solve a communicative breakdown or non-understanding, is widely recognized to be beneficial and essential for L2 learning (e.g., Long, 1983a, 1983b; Pica, 1991, 1992, 1994, 1996; Pica et al., 1987; Varonis & Gass, 1985a, 1985b). As Varonis and Gass (1985b) claim, discourse between two non-native speakers (NNS) allows greater opportunity than NS-NNS or NS-NS discourse for the negotiation of meaning when there has been an actual or potential communication breakdown. As such, this study examines NNS-NNS interaction. We frame NNS learners as ELF learners in this study because most of the learners in the study have multilingual and multicultural backgrounds. English is therefore used as a lingua franca in such multilingual interactions. Such explorations of ELF learners' NoM occurrences in immersive VR environments thus allow a comprehensive understanding of the ways immersive VR environments contribute to L2 development.

To explore the potentiality and practicality of immersive VR technologies for language learning, and ultimately, L2 development, this study addresses the following questions:

1. What are the characteristics of the negotiation of meaning by ELF learners in an immersive VR-supported multimodal environment?
2. To what extent does the multimodality of the immersive VR environment contribute to learners' negotiation of meaning and L2 learning?

Literature Review

VR, Social VR, and Language Learning

VR is reality simulated virtually—a digitally presented real-world-like environment for multiple users to see, play, and socially interact. Three important features distinguish VR from other CMC tools: (a) immersion, (b) interaction, and (c) imagination (Burdea & Coiffet, 2003). With the immersive dimension, the sensation of being there no longer necessitates a physical presence (Flower, 2015). Immersion allows L2 learners to combine learning an additional language with an intercultural experience beyond geographical limitations, with no need to step out of the classroom or leave their home countries (Wang et al., 2017, as cited in Lan, 2020, p. 1). Fully immersive VR experiences are enabled by wearing a VR headset or gear such as head-mounted devices (HMDs), haptic gloves, and treadmills. Immersive VR features strong spatial immersion, which means that users engage in VR tasks from the first-person view so that they perceive themselves as physically present in a non-physical world (Howard-Jones et al., 2014).

To create more immersive experiences for users, social VR emerged and was developed to facilitate situated social experiences so that users could feel that they are interacting with another person in a co-located virtual space. Social VR is defined as "3D virtual spaces where multiple users can interact with one another through VR head-mounted displays" (Maloney et al., 2020, p. 175). Social virtual worlds are defined as "3D, synchronous, immersive, persistent, graphical environments with generative capabilities in which participants are co-present through their avatars and interact with each other and the world's contents" (Wigham et al., 2018, p. 154). In the real-time and dynamic simulation environment, users can communicate in a multisensory way in social VR, which makes social interaction more embodied and closely resembles face-to-face communication.

The importance of language learners' embodied experience to language learning is stressed in *embodied cognition theory*. Embodied cognition theory emphasizes the role of the environment in one's cognitive process. It suggests that "the representation of knowledge is grounded in a person's experiences of interacting with and perceiving the environment, which involves whole-body involvement, including sensation, perceptions, and actions" (Lan, 2021, p. 3). Viewing language processing as an embodied process, embodied cognition theory advocates that one's bodily motions and actions will influence how one comprehends language and processes the information (Glenberg & Kaschak, 2002; Glenberg et al., 2013). As there is a connection between motor and visual processes, the more explicit the connection, the better the learning, suggesting that embodiment is important for language learning (Makransky & Petersen, 2021, p. 949). This has echoed Wei's (2018) observation that "language learning is a process of embodied participation and resemiotization" (p. 17). In immersive VR where learners can control the actions of an avatar through head-motion-tracking technologies that involve gestures and body motion, such environments provide learners with an embodied learning experience, which in turn could affect language learning. In addition to immersion and embodiment, the availability of diverse interactive modes is another key characteristic of social VR (Wang, 2020).

In real life, human communication can be divided into verbal and non-verbal communication, including gesture, eye gaze, body posture, and facial expressions. For example, videoconferencing, one of the dominant media for remote collaboration, provides users with a 2D, screen-based visual and audio connection with cameras. Due to the limit of screen frames, some drawbacks, such as partial loss of non-verbal cues, including gestures, eye contact, and body posture, which are shown to increase trust and collaboration, are reported in literature (Anton et al., 2018, p. 78). Another issue is the lack of spatial proximity and depth perception, which also plays an important role in collaboration (Salazar Miranda & Claudel, 2021). Immersive VR, an emerging alternative, uses motion tracking and VR headsets to place participants in a shared 3D environment, allowing participants to see the full range of gestures and facial expressions or proxemic adjustments, while videoconferencing remains screen based with a limited view. This immersive 3D experience is shown to provide a high level of social presence with conversational patterns that are very similar to face-to-face interaction (Smith & Neff, 2018). Jauregi et al. (2011)

compared VR-based SCMC (synchronous computer-mediated communication) to video-based SCMC. They reported that the participants preferred the VR modality due to high social presence. Some participants expressed that the anonymity of the avatar was more comfortable for them than the video modality. In a recent study, York et al. (2021) examined the effect of three different computer-mediated communication modalities (voice-based, video-based, VR-based oral interaction) onto EFL learners' foreign language anxiety. The post-study questionnaire indicated that Japanese EFL learners perceived VR as the easiest environment to communicate in and the most effective environment for language learning. They conclude that "the disembodiment of communicating with an avatar in VR environments may be perceived as a positive or negative depending on learner predispositions to technology, meaning that VR is not inherently natural or enjoyable for all learners" (York et al., 2021, p. 67). The affordances of embodiment and social presence are shown to facilitate learners' language learning.

In sum, it is important to explore how the affordances of VR impact conversational interactions that are unique compared with other media. Conversation or interaction is a collaborative process in which meaning is incrementally constructed together. It relies on both coordination and communication across verbal and nonverbal channels. The issue of how learners construct and negotiate meanings with semiotic resources afforded in VR environments thus deserves further exploration.

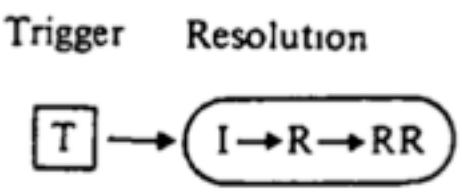
Negotiation of Meaning in Computer-mediated Communication

Negotiation of meaning (NoM) is a collaborative attempt in a conversation between more or less fluent speakers in order to solve communication breakdown and reach comprehension (Long, 1996). In the field of second language acquisition (SLA), NoM has been recognized as essential and beneficial for language acquisition, for it encourages learners to check, confirm, and clarify utterances to reach a shared understanding and to maintain conservational flow (Ellis, 2003; Gass & Mackey, 2007; Long, 1983a, 1983b; Mackey et al., 2012; Nakahama et al., 2001; Pica, 1991, 1992, 1994, 1996; Pica et al., 1987; Oliver, 2002; Van der Zwaard & Bannink, 2014, 2016, 2019, 2020; Varonis & Gass 1985a, 1985b). That being said, the more language learners engage in negotiated episodes, the better (i.e., the more they indicate non-understanding, the better; Van der Zwaard & Bannink, 2014, p. 139).

According to Varonis and Gass (1985a), negotiation episodes occur when non-understanding is explicitly acknowledged. Varonis and Gass' (1985a) NoM model has two phrases: the Trigger phase and the Resolution phase. Their model contains three components—trigger (T), indicator (I), and response (R)—and one optional phase—reaction to response (RR) (see Figure 1). In this model, the trigger (T) is the utterance that causes non-understanding, indicator (I) is the signal that shows the existence of a problem, response (R) is the utterance that aims to resolve the problem, and reaction to response (RR) is the acknowledgement that the problem is solved (Yin & Satar, 2020, p. 393).

Figure 1

Varonis and Gass' (1985a) Model of Negotiation of Meaning



Recent studies have examined NoM in computer-mediated communication (CMC) contexts, including text-based CMC (e.g., Akayoglu & Altun, 2009; Blake, 2000; Fernández-García & Martínez-Arbelaiz, 2002; Lee, 2001, 2009; O'Rourke, 2005; Pellettieri, 2000; Smith, 2003a, 2003b; Tudini, 2003, 2007; Yin & Satar, 2020) and audio or video-based CMC (e.g., Canals, 2021; Lee, 2020; Lee et al., 2019; Wang, 2006; Wang & Tian, 2013; Van der Zwaard & Bannink, 2014, 2016, 2019, 2020; Yanguas, 2010). Wang

(2006) examined Chinese learners' online videoconferencing sessions via an Internet-based videoconferencing tool, *NetMeeting*. She adopted Varonis and Gass' (1985a) NoM model in analyzing the interaction and found distinct trigger features in online synchronous multimodal environments. The researcher investigated the effects of the tutor's use of video in online class meetings. She indicated that gestures and facial expressions were used frequently by the tutor and tutees as semiotic tools for meaning-making to complete tasks in online video-based sessions. Similarly, in a study of online tutoring sessions, Hampel and Stickler (2012) analyzed the written and spoken communication in recorded videconferencing sessions in *FlashMeeting*. With a qualitative approach, they concluded that language teachers and learners used a number of multimodal strategies (e.g., switch to the text chats to comment on what the speaker said) to make meaning and maintain communication while not interrupting others.

Van der Zwaard and Bannink (2014) examined NoM between NS and NNS in two CMC channels: video calling and chat messaging. They found that distinct patterns of negotiated interaction can be identified between these two modes, indicating that learners' interactive patterns are dependent on the modality that mediates the interaction. Wang and Tian (2013) used Varonis and Gass' (1985a) NoM model in analyzing the interaction in an eTandem learning context between Mandarin and English students. They found that the dyads with different proficiency levels revealed different interactive patterns in negotiating for meaning with their partner, both quantitatively and qualitatively. They concluded that live video and Textchat are the two distinctive features of videoconferencing-supported environments that facilitate NoM in eTandem learning. Nonverbal cues such as laughing, nodding, and a puzzled look, among other facial expressions, are evident in all primes in the Indicator, Response, and Reaction to Response stage. The researchers also found that "visual cues constituted an integral part of the negotiation process as students made a deliberate and effective use of a variety of body gestures to generate comprehensible input and output via the live video" (Wang & Tian, 2013, p. 52).

In a recent study, Canals (2021) examined the interplay of multimodality and translanguaging in learners' meaning negotiation process in video-based CMC (Skype) during a tandem project between 18 college-level students from Spain and Canada. The study shows that the learners not only used translanguaging strategies involving English, Spanish, and other shared languages but also utilized multimodal resources such as postures, gestures, gaze, and digital and physical devices (e.g., computers, notes) during meaning negotiation. In exploring the role of gestures in meaning negotiation, Lee et al. (2019) examined the interaction among NNS learners in Skype videoconferencing sessions. They concluded that NNS learners use gestures to get their message across and understand their interlocutors in online videoconferencing contexts. Gestures were shown to aid comprehension and play a facilitating role in establishing joint attention and lexical retrievals in L2 oral interactions. As such, semiotic resources such as postures, gestures, gaze, and speech are shown to coordinate with each other in learners' negotiation occurrences in online video-based CMC environments.

Extending current research inquiries, this study examines the multimodality in immersive VR environments and how ELF learners attempt to deal with their communicative issues with semiotic resources afforded in an immersive virtual space, particularly in the context of cross-cultural virtual exchanges. To our knowledge, a focus on learners' NoM in virtual exchanges via immersive VR tools and the virtual contexts they enable remains unexplored. Thus, this study is designed to bridge such gaps in research.

Methodology

Participants

The participants in this study were five university students from a national university in Northern Taiwan and five university students from a public university in Spain. The participants were recruited by the authors respectively. The students were informed that the purpose of the project was to provide students with opportunities to interact with people from different cultural backgrounds and to engage in intercultural communication. The participants ranged in ages from 18 to 22 years, with three males and seven females.

The Taiwanese students were English majors, and the Spanish students majored in Translation and Interpreting. Both groups of students' English proficiency levels ranged from B2 (upper intermediate) to C1 (advanced) according to the Common European Framework for Languages (CEFL). Three students reported that they have experience studying abroad, while most of the students reported having experience traveling or visiting foreign countries. Half of the participants have participated in online interaction projects with people from other countries prior to this virtual exchange project, while the other half of the students have not. Almost all the participants have no prior experience with virtual reality or the Oculus headsets, except three Taiwanese students who had experience with VR prior to the study. The ten participants were paired up and communicated with each other from different geographical locations in Taiwan and Spain using VR technologies. Five dyads were formed. The five dyads completed two tasks in an immersive VR platform. All the participants' names were assigned with pseudonyms, with permission to use videos and images of them obtained within consent forms.

VR Equipment and Social VR Platform

In this study, *Oculus Quest 2* was used as the VR equipment, including a headset and two controllers. *Oculus Quest 2* was selected as the VR headset for the present study because of its compatibility with most of the AR and VR applications on the market. The tasks in this study took place on a social VR platform, *Spatial*. *Spatial* was selected in this study because it is user-friendly and provides free access to basic social functions. *Spatial* allows users to create a 3D-realistic avatar from a single selfie in seconds. It allows users to sit next to each other from across the world. The avatar comes to life as users talk, move, and interact with the lipsyncing and facial recognition functions, allowing multimodal representations in hand movements and gestures, facial expressions, body posture, and nonverbal expressions with hands.

Using the avatars they created in *Spatial*, the Taiwanese and Spanish participants completed two tasks—an information exchange task and a role-playing task—in the VR environment (see Appendix A for task descriptions). The participants were informed that the minimal time length was about 30 minutes per task; however, some dyads produced longer exchanges during the tasks. Task 1 is an information exchange task in which the participants introduced themselves and got to know each other. Task 2 is a role-playing task in which the participants took turns taking a role of an international student preparing for the transition to the new university and an international buddy who facilitated the international student's transition to the new environment. A one-hour training workshop on how to use the VR equipment and explorations on *Spatial* was provided. In the workshop, students on both ends received technical guidance and instructional support in utilizing the functions available in the VR platform from two research assistants, one in Spain and one in Taiwan, when completing the different tasks. The VR-based tasks took place at a virtual reality lab with necessary VR equipment at the university in Spain. Due to the school closure in Taiwan because of the pandemic situation during the time of the study, the tasks took place in the participants' households with remote guidance from the Taiwanese research assistant. These tasks were completed in collaboration with different Spanish dyads, who were in turn guided and supervised by the Spanish research assistant.

Data Collection

As part of a larger study, four datasets were collected and analyzed, including the participants' pre-task questionnaires, the video recordings of VR tasks, post-task questionnaires, and stimulated recall interviews. All the participants filled out a pre-task questionnaire in which they provided biographical information about past English learning experiences, VR experiences, international experiences, and cultural experiences. The two VR tasks of each dyad were video recorded from two angles: from the participant's view in the VR environment with the build-in recording function of the VR headsets and from the third-person view in the physical setting with a side camera being set up by the participant. The recordings from both the learner's VR view and side view allowed access to document the learner's use of semiotic resources in the virtual world (avatar representations in the virtual space) and the real world (real-person representations in the physical space), providing us a more comprehensive picture about how learners engaged in VR-mediated interaction. Both learner' avatar representations and real-person representations were documented and analyzed (e.g., see Figure 2). Approximately 30 hours of recordings were collected,

including the Taiwanese and Spanish participants' VR- and side-view recordings for both Tasks 1 and 2.

Figure 2

Screenshots of the Learner's Avatar Representation in VR and Real-person Representation in the Real World

Screenshot of the learner's VR avatar representation (from the partner's VR view)

Screenshot of the learner's real-person representation (from the learner's side camera)

After completion of the tasks, the participants completed a post-task questionnaire in which they were asked to reflect on their VR experience, cultural experience, and virtual exchange with their partner. After initial analyses of the data, stimulated recall interviews were conducted in which the participants were prompted to recall what they had done to manage occurrences of non-understanding and NoM with their partners in the VR tasks with the replay of the recorded videos. The stimulated recall interviews were audio recorded for later analyses. Given the scope of this paper, only the video data and stimulated recall interviews are reported.

Data Analysis

The video recordings of the VR sessions were transcribed and reviewed line-by-line by the first author and two research assistants. During the review, content analysis was adopted to identify utterances where Taiwanese and Spanish learners encountered communicative breakdowns and negotiated for meanings, which we identified as NoM episodes. The NoM episodes were further coded using Varonis and Gass' (1985a) model of NoM routines consisting of four primes (the Trigger, the Indicator, the Response, and the Reaction to Response, T-I-R-RR). After identifying the T-I-R-RR primes, this study analyzed learner-initiated signals of non-understanding and the categories of the four primes used respectively by ELF learners in immersive VR sessions. This study adopted a modified list of Wang and Tian's (2013) coding categories (see Table 1). The data were coded by three coders to establish interrater reliability. The amount of agreement reached 90% in the first round and was calculated by a simple percentage of agreement. The NoM episodes and categories which showed discrepancies among the three coders were discussed until a consensus was reached.

The quantitative and qualitative analyses of the characteristics of the NoM occurrences in the four phases during the two VR-mediated interaction tasks were documented and coded to understand the general characteristics and quality of the total numbers and types of the four primes and their subcategories. To address the research questions in this study, the identified NoM episodes then underwent a qualitative analysis and detailed multimodal transcription and coding using multimodal (inter)action analysis (Norris, 2004) by analyzing turns, including the spoken language and non-verbal qualities such as facial expressions, gestures, gaze, body posture, head movement, space proxemics, and object handling, and how those semiotic cues were used in context.

Table 1

Subcategories of the Four Primes (T-I-R-RR) (modified and adopted from Wang & Tian, 2013, p. 47)

Primes	Categories
Trigger	Lexical
	Syntactic
	Content
Indication	Explicit statement of non-understanding
	Non-verbal responses
	Echo
	Visual indicators
	Implicit statement of non-understanding
	Confirmation check
	Rephrasing
Response	Target language equivalent
	Comprehension check
	Repetition
	Request
	Overt explanation
	Expansion
	Rephrasing
	Reduction
Reaction to Response	Comprehension
	Modification of output
	Expansion
	Comprehension check
	Incomprehension
	Miscomprehension
	Repetition
	Confirmation

Multimodal (Inter)action Analysis

This study adopts *multimodal (inter)action analysis* to understand the degrees to which each mode is important to an interaction and the relationships between the modes in the interaction (Norris, 2004). The unit of analysis is action, as each action is mediated. Multimodal (inter)action analysis involves two phases

of analysis: *analysis of actions* and *analysis of modes*. The analysis begins with identifications of different types of mediated actions: *higher-*, *lower-*, and *frozen actions* (Norris, 2004, 2019). *Higher-level actions* are complex actions that have identifiable boundaries. For example, *response to non-understanding* is a higher-level action in communication between ELF learners in virtual exchange contexts. In many cases, communicative interactions consist of multiple higher-level actions. In virtual exchanges in the present study, ELF learners engage in multiple higher-level actions, such as *indicating non-understanding*, *response to non-understanding*, and *reaction to response*.

Higher-level actions consist of a chain or sum of *lower-level* actions. A lower-level action is the smallest interaction meaning unit (Norris, 2004). Each lower-level action is mediated by a system of representation, which, in other words, is one mode. Modes can be embodied during the interaction (e.g., a gesture or spoken utterance), but they may also be disembodied and produced prior to an interaction (e.g., a written assignment sheet). In the present study, in addition to their spoken language, learners' facial expressions and body language all serve as lower-level actions that contributed to their partner's understanding of the meanings learners intend to convey, which influenced the direction of further information exchanges and elaborations. *Frozen actions*, the third type of actions, are higher-level actions performed by social actors at an earlier time and are now entailed in disembodied modes (e.g., printed material and the layout of an environment) or more or less permanent, material objects (Norris, 2004). For example, in the present study, the higher-level actions of completing the VR-mediated tasks and negotiating meanings are mediated by frozen actions such as the layout of the VR environment and objects available (e.g., sofa, desk, and outdoor terrace), which also contribute meaning to the real-time interaction.

The second phase of analysis in multimodal (inter)action analysis is to explore how modes accomplish higher-level actions in hierarchical and non-hierarchical ways (Norris, 2019). This is determined by analyzing *modal intensity* (i.e., the weight that a mode carries in a higher-level action) and *modal complexity* (i.e., the relationships between modes that rely on each other for meaning) (Norris, 2017). In some NoM occurrences in the present study, for instance, the spoken mode (e.g., lexical or syntactic), often identified as a common trigger for a communicative breakdown, may take on high modal intensity because it is the dominant mode of communication, while augmented minimally with other semiotic sources (e.g., gestures, facial expressions, or body posture). Modal intensity thus analyzes the hierarchical orders of different modes. *Modal complexity* refers to "the interplay of numerous communicative modes that make the construction of a higher-level action possible" (Norris, 2004, p. 87) and analyzes multiple modes in non-hierarchical ways. The explorations of modal complexity thus examine how the learner uses different modes that are intricately intertwined to construct higher-level actions.

To sum up, adopting multimodal (inter)action analysis (Norris, 2004), this study examines how ELF learners engage in the use of multiple modes to negotiate meanings in a virtual exchange context within an immersive VR environment. The mediated actions, including higher-level, lower-level, and frozen actions, and the analytical lenses of modal intensity and multimodal complexity, allow us to analyze the VR-mediated ELF interactions in detail. Specifically, this study modified the coding schemes used in Wigham and Satar (2021) and added *space proxemics* to the coding schemes as a modality afforded in VR environments. This modality was identified when the participants presented proxemic behaviors and exploited the VR space by transportation to a different place. This study focused on the participants' use of verbal mode (linguistic and para-linguistic cues), visual mode (avatar representation and real-person representation), gestural mode (facial expression, gesture, gaze, body posture, head movement, and object handling), spatial mode (space proxemics), and other semiotic modes in constructing and negotiating meanings (see Table 2 for different color codes used in multimodal transcriptions).

The video recordings of each learner's avatar representation and real-person representation were imported into the ELAN software and reviewed side-by-side by the first researcher and two research assistants. Notes were taken and compared throughout the reviews of the video recordings, specifically on lower-, higher-, and frozen actions represented in the VR ELF interaction, and the verbal, gestural, visual, and spatial modes that carry out those mediated actions in ELF learners' NoM routines.

Table 2

Semiotic Modes and Coding Schemes in Multimodal (Inter)action Analysis

Semiotic mode	Coding schemes
Verbal mode	linguistic and para-linguistic cues
Visual mode	avatar representation, real-person representation
Gestural mode	facial expression, gesture, gaze, body posture, head movement
Spatial mode	space proxemics
Other semiotic modes	object handling

Findings

To address the research questions, we first take a quantitative approach to explore the characteristics and quality of the VR-supported virtual exchanges between the five dyads. Using the Varonis and Gass (1985a) model, a total of 56 occurrences of NoM routines that contain the four primes (Trigger, Indicator, Response, and Reactions to Response) were identified. Table 3 shows the proportions of different types of the four primes produced by the five dyads. In the Trigger phase, 71.5% of triggers were content-driven, while 28.5% were lexical. Among Indicators, an explicit statement of nonunderstanding (52.6%) was mostly used by the participants to indicate the non-understanding, followed by a confirmation check (33.3%) and an echo (14.1%). To resolve the problem indicated, overt explanation (30.1%) was utilized most frequently, followed by repetition (20.5%), rephrasing (20.5%), confirmation (16.4%), expansion (8.2%), and skip (4.3%). In the Reaction to Response phase, comprehension (64.8%), followed by comprehension check (11.4%), and confirmation (9.8%) were the three most frequently used strategy types by the participants to acknowledge that the non-understanding was resolved.

It is important to note that, unlike Wang and Tian's (2013) study, we coded 'visual indicators' and 'non-verbal responses' in the categorization for instances where there was no verbal response at all but visual indicators or non-verbal responses in a specific prime. In the case of this study, we did not observe any instances in our data that contained only non-verbal responses or visual indicators and were without verbal utterances at all primes. Instead, in addition to verbal speech, we witnessed ELF learners' constant usage of a wide range of non-verbal modalities during their NoM. As driven by multimodal (inter)action analysis (Norris, 2004), lower-level actions and the mode(s) that realize such actions were identified. Seven non-verbal modes could be identified during the NoM routines, including facial expression, gesture, gaze, body posture, head movement, space proxemics, and object handling. Table 4 indicates the frequency and percentage of non-verbal resources used by the five dyads. Non-verbal resources such as head movement (37.9%) and gesture (34.5%) were found to be most commonly used by the participants to solve communication breakdown and reach comprehension when negotiating meaning. Other non-verbal cues, including body posture (10.2%), facial expression (7.4%), object handling (5.9%), gaze (3.6%), and space proxemics (0.5%), were shown to play a role in the participants' process of meaning negotiation. Close examinations of the data indicated that non-verbal modalities were extensively employed by the ELF learners to construct the Trigger, Indicator, Response, and Reaction to Response primes.

Table 3

Frequency and Percentage of Negotiation of Meaning (NoM) Occurrence and Subcategories of the Four Primes Used by the Five Dyads

Primes	Subcategories	D1	D2	D3	D4	D5	Total
Trigger	Lexical	1(50%)	5(27.7%)	3(30%)	7(53.8%)	1(7.1%)	17(29.8%)
	Syntactic	0	0	0	0	0	0
	Content	1(50%)	13(72.3%)	7(70%)	6(46.2%)	13(92.9%)	40(70.2%)
	Total	100%	100%	100%	100%	100%	100.00%
Indication	Explicit statement of nonunderstanding	1(50%)	14(73.6%)	2(20%)	7(43.7%)	6(46.1%)	30(52.6%)
	Non-verbal Responses	0	0	0	0	0	0
	Echo	1(50%)	2(10.5%)	1(10%)	2(12.6%)	2(15.3%)	8(14.1%)
	Visual Indicators	0	0	0	0	0	0
	Implicit statement of nonunderstanding	0	0	0	0	0	0
	Confirmation check	0	3(15.9%)	7(70%)	7(43.7%)	5(38.6%)	19(33.3%)
	Rephrasing	0	0	0	0	0	0
	Total	100%	100.00%	100%	100%	100.00%	100.00%
Response	Target language equivalent	0	0	0	0	0	0
	Comprehension check	0	0	0	0	0	0
	Repetition	0	7(23.3%)	1(11.1%)	3(17.6%)	4(28.5%)	15(20.5%)
	Request	0	0	0	0	0	0
	Overt explanation	2(66.6%)	10(33.3%)	2(22.2%)	6(35.2%)	2(14.3%)	22(30.1%)
	Expansion	0	2(6.6%)	2(22.2%)	1(5.8%)	1(7.1%)	6(8.2%)
	Rephrasing	1(33.3%)	4(13.4%)	3(33.3%)	5(29.4%)	2(14.3%)	15(20.5%)
	Reduction	0	0	0	2(12%)	0	0
	Confirmation	**0**	**4(13.4%)**	**1(11.1%)**	**0**	**5(35.8%)**	**12(16.4%)**
	Skip	**0**	**3(10%)**	**0**	**0**	**0**	**3(4.3%)**
	Total	100%	100.00%	100%	100%	100%	100.00%

Reaction to Response	Comprehension	1(33.3%)	16(59.3%)	8(88.9%)	11(61.1%)	10(71.4%)	46(64.8%)
	Modification of output	0	0	0	1(5.5%)	1(7.1%)	2(2.8%)
	Expansion	0	0	0	0	0	0
	Comprehension check	0	5(18.5%)	0	2(11.2%)	1(7.1%)	8(11.4%)
	Incomprehension	0	0	0	1(5.5%)	0	1(1.3%)
	Miscomprehension	0	0	1(11.1%)	0	0	1(1.3%)
	Repetition	1(33.3%)	0	0	0	1(7.1%)	2(2.8%)
	Confirmation	1(33.3%)	5(18.5%)	0	1(5.5%)	0	7(9.8%)
	Overt explanation	**0**	**1(3.6%)**	**0**	**0**	**0**	**1(1.3%)**
	Explicit statement of non-understanding	**0**	**0**	**0**	**1(5.5%)**	**0**	**1(1.3%)**
	Explaining miscomprehension	**0**	**0**	**0**	**1(5.5%)**	**0**	**1(1.3%)**
	Echo	**0**	**0**	**0**	**0**	**1(7.1%)**	**1(1.3%)**
	Total	100%	100.00%	100%	100.00%	100.00%	100.00%

Note. The number indicates frequency count, and the number in the parentheses indicates the percentage. New categories emerging from the data are emphasized in **bold text**.

Table 4

Frequency and Percentage of Non-verbal Resources Used by the Five Dyads in NoM

Non-verbal resources	D1	D2	D3	D4	D5	Total
Facial expression	0	31(11.1%)	8(6.4%)	9(7.4%)	0	48(7.4%)
Gesture	12(46.1%)	90(32.1%)	46(36.8%)	52(42.3%)	25(25.3%)	225(34.5%)
Gaze	1(3.8%)	6(2.1%)	8(6.4%)	5(4.1%)	4(4%)	24(3.6%)
Body posture	5(19.3%)	19(6.7%)	17(13.6%)	6(4.9%)	20(20.4%)	67(10.2%)
Head movement	8(30.8%)	124(44.5%)	38(30.4%)	36(29.2%)	42(42.4%)	248(37.9%)
Space proxemics	0	2(0.7%)	0	0	0	2(0.5%)
Object handling	0	8(2.8%)	8(6.4%)	15(12.1%)	8(8%)	39(5.9%)
Total	100%	100%	100%	100%	100%	100%

Note. The number indicates frequency count, and the number in the parentheses indicates the percentage.

As shown in Table 5, in the Trigger prime, gestural modes such as gesture (38.6%), head movement (36.4%), and object handling (10.4%), a semiotic mode, were the three most common non-verbal cues used when a

non-understanding in the conversation occurred. To indicate a non-understanding, the participants employed head movement (28.9%), gestures (27.7%), body posture (20.4%), and other modes to either provide explicit statements of non-understanding or perform comprehension checks. With regard to the Response phase, head movement (45.1%), gestures (33.6%), and body posture (9.2%) were mostly generated by the ELF learners in an attempt to resolve the communication breakdown. Interestingly, spatial resource, such as space proxemics, was utilized by the learners to skip the communication breakdown. Similar behavior patterns were observed in the Reaction to Response phase, in which gestures (35.4%), head movement (34.3%), and body posture (12.3%) were used as the common non-verbal resources to provide answers, present confirmations, or perform comprehension checks.

To summarize, the quantitative analyses of the four primes and their subcategories and types of non-verbal resources have provided us with an understanding of the general characteristics and quality of the VR-supported virtual exchanges among ELF learners. To further uncover such quality and features, we take a qualitative and multimodal approach to examine how the ELF learners constructed multiple higher-level actions related to negotiation of meaning as mediated by a chain of lower-level actions and simultaneous engagement in multiple modes or non-verbal resources. Given the space limit, three excerpts were reported, representing different aspects of ELF learners' multimodal engagement in NoM within immersive VR space. The findings reported mainly focus on Taiwanese students, while minimal descriptions of Spanish students are also provided.

The interaction, as shown in Excerpt 1, took place in the first task, where the learners exchanged basic information about their interests, major, and college life with each other. Elena, a Spanish student, asked Jie, a Taiwanese student, about his major and areas of study (see Figure 3 for the spatial proxemics of Dyad 1). A communicative breakdown occurred when Jie could not understand Elena's question.

In Line 1, a *communicative breakdown* was triggered when Elena initiated a wh-question, "what were you study?" with downward gaze and object handling (adjusting the headset) (#1). Elena's adjustment of the headset was not represented in her VR avatar because she did not hold the controller with the hand adjusting the headset object in the physical world. In Line 2, Jie *indicated his non-understanding* of Elena's question in an attempt to confirm his understanding of the spoken language "Where am I studied?". Along with his question with a rising intonation, Jie also expressed a puzzled look on his face and moved his head towards the right with his body leaning against the wall. His eye gaze also shifted from the environment to the partner while verbalizing his non-understanding of his partner's question (#2). These demonstrate that Jie not only indicated his non-understanding with the verbal utterance but also expressed it with the coordinated use of a facial expression, body posture, head movement, and eye gaze, which are reflected in both his avatar and real-person representation. In Line 3, Elena perceived Jie's non-understanding and *responded to his confirmation question by rephrasing her question* while shifting her gaze towards Jie and adjusting her headset (#3). She further expanded her question by verbalizing "yes, what and where, yes" with her head nodding and eye gaze downwards (#4). She then shifted her gaze towards Jie, standing in front of him in the VR space, and touched her hair, uttering, "I… I don't remember" (#6).

In Line 4, Jie *reacted to Elena's response by expressing comprehension*, utilizing the spoken language "Oh, I'm an English major" as well as other semiotic modes, including smiling, gazing at the partner, stretching hands, and head leaning forward (#6). Elena acknowledged Jie's answer to her question by verbalizing "okay" in addition to nodding (#7). In Line 6, Jie continued the higher-level action-*reaction to response* by providing additional information about his study using four main lower-level actions, including the spoken language ("so basically just… some uh some reading… uh some literature about English or…"), facial expression (smiling), gesture (iconic gesture and touching hair), and gaze (gaze at the partner) (#8 and #9). The gestures were used by Jie as an aid to search for and retrieve appropriate lexical words, here "reading" and "literature" in his explanations to Elena on what English-major students do. In Line 7, although the verbal production was not perceived, Elena employed paralinguistic resources like head movement, here nodding to acknowledge Jie's sharing and to show her listener presence (#10). In Line 8, Jie further expanded his response with simultaneous use of head movement (head nodding), gesture (hand stretching),

gaze (looking up), and facial expression (smiling) (#11 and #12).

Table 5

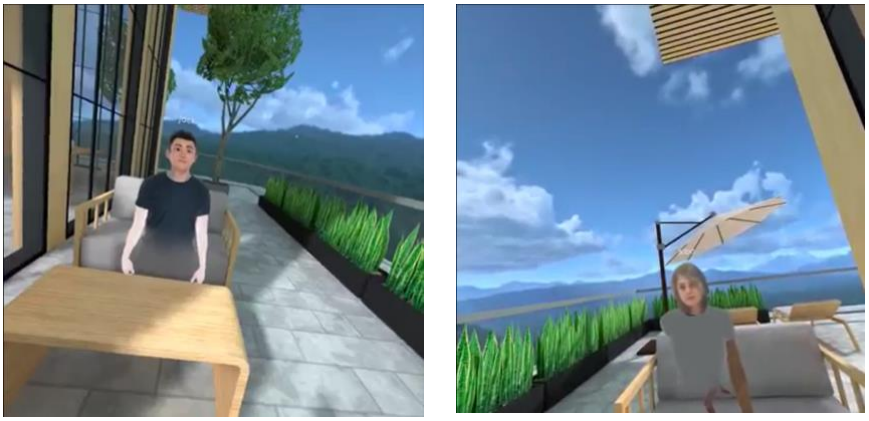
Frequency and Percentage of Non-verbal Resources Used in the Four Primes

Non-verbal resources / Prime	Trigger	Indication	Response	Reaction to Response
Facial expression	12(7.3%)	7(8.4%)	10(5.4%)	16(8.6%)
Gesture	63(38.6%)	23(27.7%)	62(33.6%)	67(35.4%)
Gaze	5(3.1%)	4(4.8%)	5(2.7%)	10(5.2%)
Body posture	7(4.2%)	17(20.4%)	17(9.2%)	23(12.3%)
Head movement	59(36.4%)	24(28.9%)	83(45.1%)	65(34.3%)
Space proxemics	0	0	2(1.1%)	0
Object handling	17(10.4%)	8(9.8%)	5(2.9%)	8(4.2%)
Total	100%	100%	100%	100%

Figure 3

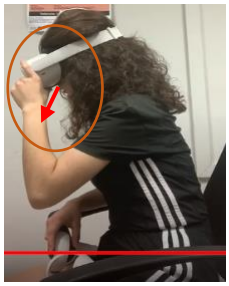
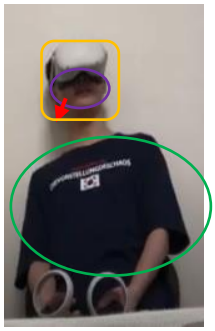
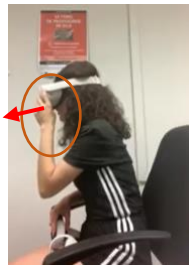
Spatial Proxemics of the Avatar Representations of Dyad 1 (Jie & Elena) in the VR Environment

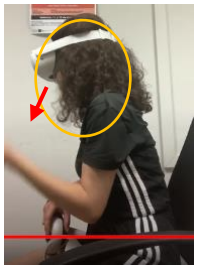
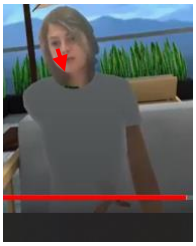
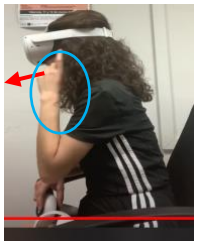
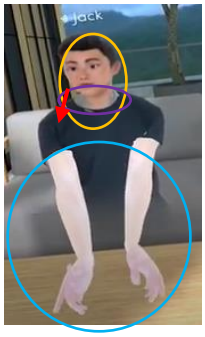
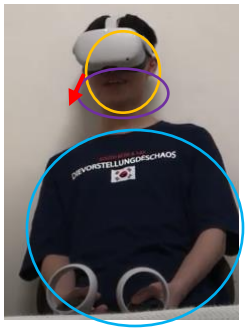
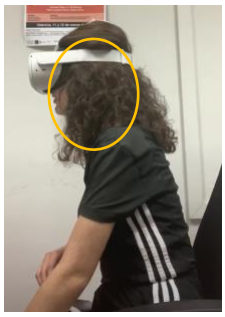
Jie (Taiwanese student) Elena (Spanish student)

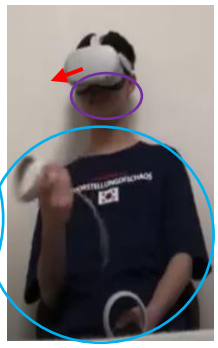
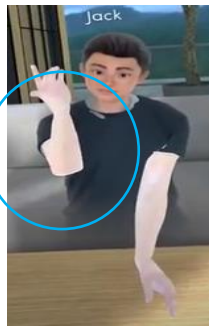
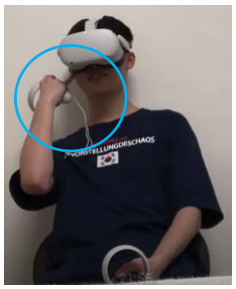
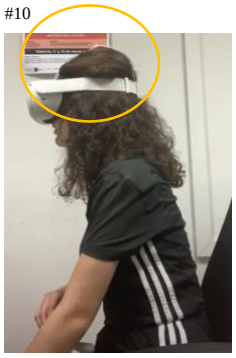


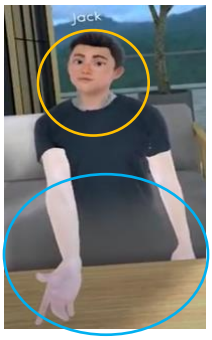
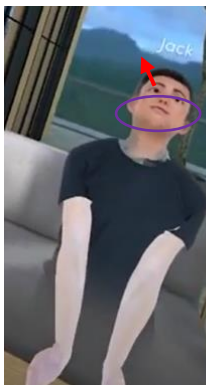
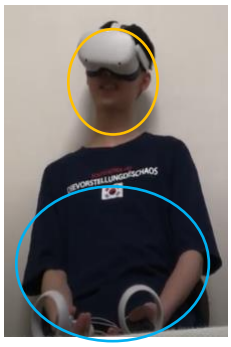
Excerpt 1

Negotiation of Meaning Episode from Dyad 1: Jie & Elena

Turn	**Speaker**	**Verbal**	**Gestural**	**Higher-level actions (NoM prime)**	**Category**	**Visual** Avatar representation (Virtual space)	Real-person representation (Physical space)
1	Elena	But well [#1]what were you study? What did you...	[#1] Gaze: downwards Object handling: adjusting headset	Trigger	Content	#1 	#1 
2	Jie	[#2]Where am I studied?	[#2] Facial expression: A puzzled look Gaze: gaze shifts towards the partner Head movement: head moving towards right Body posture: leaning against the wall	Indicator	Confirmation check	#2 	#2 
3	Elena	[#3]What are you studying…[#4] yes what and where, yes. [#5]I... I don't remember.	[#3] Object handling: adjusting headset Gaze: gaze shifts towards the partner	Response	Rephrasing & Expansion	#3 	#3 

			#4 Head movement: nodding Gaze: gaze downwards			#4 	#4 
			#5 Gaze: gaze shifts towards the partner Gesture: touching hair			#5 	#5 
4	Jie	Oh, I'm an Engli… #6I'm an English major…	#6 Head movement: head leaning forward Facial expression: smiling Gesture: stretching hands Gaze: gaze at the partner	Reaction to Response	Comprehension	#6 	#6 
5	Elena	#7Okay.	#7 Head movement: nodding			#7 	#7 

6	Jie	so basically just… #8 some uh #9 some reading…uh some literat-ure about English or …	#8 Facial expression: smiling Gesture: iconic gesture Gaze: gaze at the partner #9 Gesture: touching ear	Reaction to Response	Expansion	#8  #8  #9  #9 
7	Elena	∅	#10 Head movement: nodding			#10  #10 

8	Jie	[#11]some-thing like that. [#12] Or business English, yes.	[#11] Head movement: nodding Gesture: hand stretching [#12] Gaze: looking up Facial expression: smiling	Expansion	#11  #12 	#11  #12 

Note. #[n] refers to the image number within the excerpted transcript being discussed

This excerpt shows that both Jie and Elena employed various semiotic resources, including the spoken language mode, facial expression, gesture, gaze, body posture, head movement, and object handling afforded in the immersive VR environment to mitigate a communicative non-understanding and negotiate meaning in the interaction. For instance, Jie utilized various semiotic resources in initiating a resolution phase of NoM, particularly in the Reaction to Response stage. The higher-level actions, such as Trigger, Indicator, Response, and Reaction to Response, in meaning negotiation were shown to be mediated by the ELF learners' coordinated and combined use of lower-level actions with different modalities. Learners' employment of visual, gestural, and spatial modes all serve as lower-level actions that contribute to their partner's understanding of meanings learners intend to convey, which influence the direction of further information exchanges and elaborations (Norris, 2004). This NoM episode was replayed during the stimulated recall interview with Jie. When asked how he managed occurrences of non-understanding and negotiation of meaning with his partner in the VR tasks, Jie remarked,

> I wasn't sure whether she asked me 'what am I studying or where am I studying'… so I just wanted to clarify with her. I didn't realize that I'm gesturing… I use lots of gestures in my speech in daily life. I feel like having conversation through this VR environment is more vivid, just like normal life. Maybe that's why I use many gestures. I feel quite relax when talking in VR. Using gestures let me verbalize things better and explain things more clearly. (Interview with Jie)

It is worth noting that during the interaction, Jie and Elena exploited the spatial resource afforded in the VR environment. In the middle of the interaction, they played with the 'transporting' feature in the VR platform and decided to continue their conversation at the outdoor terrace due to the beautiful view from the terrace. In the interview, Jie remarked that having a conversation at the outdoor terrace is more casual and relaxing, which he enjoyed very much. The learners were also shown to utilize their space differently and reduced the proxemic space when interacting with each other in an outdoor setting on the VR platform.

In Excerpt 2, Teng (Taiwanese student) and Mandy (Spanish student) discussed the weather in Taiwan and Spain. The spatial juxtaposition of Teng's avatar and Mandy's avatar in the VR space is shown in Figure 4.

Figure 4

Spatial Proxemics of the Avatar Representations of Dyad 2 (Teng & Mandy) in the VR Environment

Teng (Taiwanese student) Mandy (Spanish student)

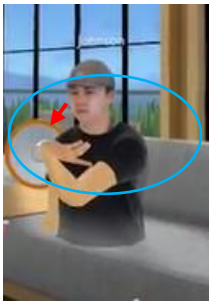
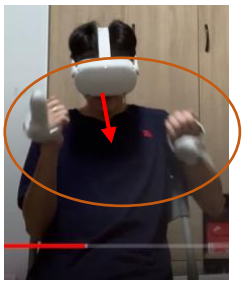
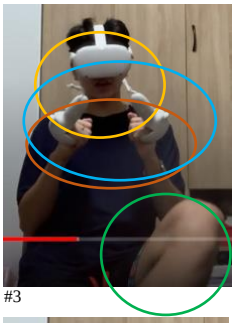
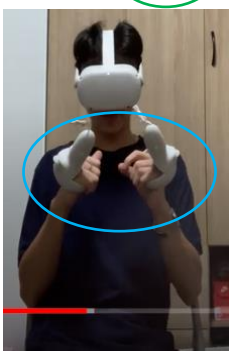
In Line 1, when asked about the weather in Taiwan, Teng replied with a verbal explanation, "it depends on the season that you're coming in…" along with other semiotic modes such as gaze (downwards), object handling (holding the controller upside down), head movement (head moves upwards), body posture (crossing the legs), and gestures (pointing) to deliver the messages (#1-8). Gestures were shown to play an important role in Teng's articulation of verbal utterances. Not only did Teng use gestures extensively as a resource to provide information to his Spanish partner, but he also employed iconic gestures which conveyed meaning semantically related to the content of the co-occurring speech when he uttered "typhoon" and "a thing" (McNeill, 1992).

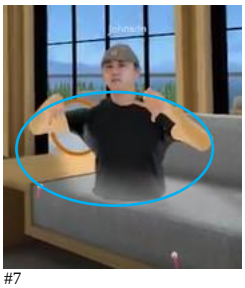
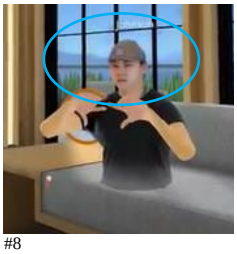
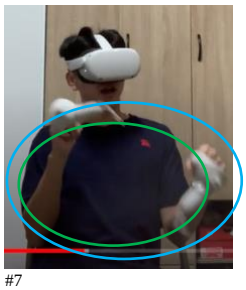
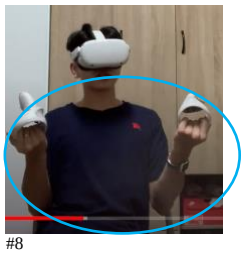
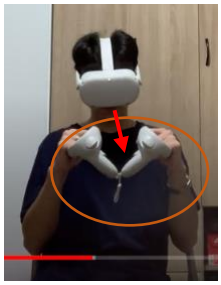
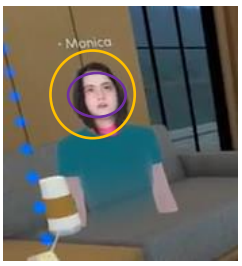
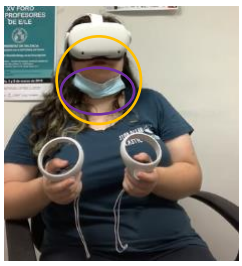
Although Teng provided gestural input along with verbal production with an aim to aid Mandy's comprehension of the weather situation in Taiwan, a communicative breakdown still occurred when Teng uttered, "in summer we have a lot of rain and we have typhoon." The word "typhoon" served as the trigger of a non-understanding in the negotiation episode. In Line 2, Mandy indicated a non-understanding with an explicit statement, "what do you mean… typhoons?". Concurrent with her spoken language mode, Mandy also showed a puzzled look and shook her head when expressing non-understanding (#9). To reduce face-threatening, Mandy presented a smiley face and lifted one hand when posing a direct question, "what do you mean…?" to Teng (#10).

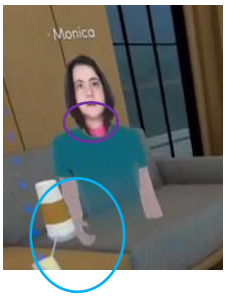
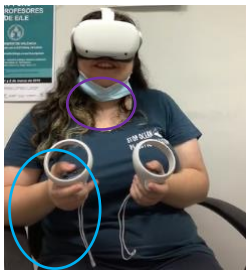
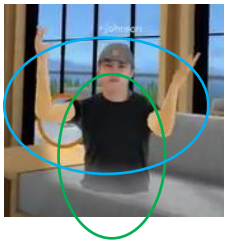
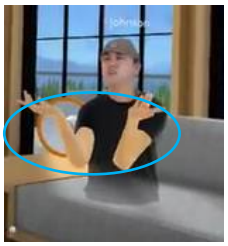
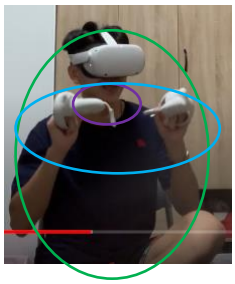
As shown in Line 3, Teng responded to Mandy's explicit non-understanding by providing overt explanations of what a typhoon is. To facilitate his explanations, Teng was again engaged in an extensive use of iconic gestures to represent the state of "thunder," "lightning," and "heavy rain" when a typhoon approaches (#11-#17). He also shifted his eye gaze from downwards to the partner while articulating the response. Teng's body posture also displayed shifting positions during this turn. In Line 4, Mandy reacted to Teng's response with an explicit comprehension statement, "Okay, I know" along with other visual and gestural modes like nodding, smiling, and lifting one hand (#18).

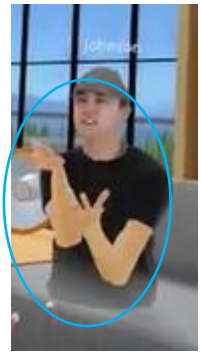
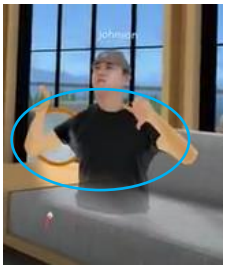
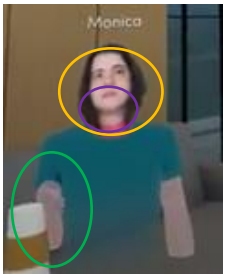
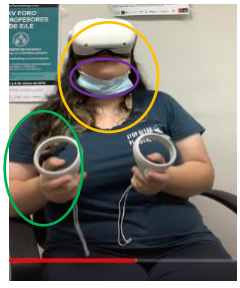
Excerpt 2

Negotiation of Meaning Episode from Dyad 2: Teng & Mandy

Turn	Speaker	Verbal	Gestural	Higher-level actions (NoM prime)	Category	Visual Avatar representation (Virtual space)	Visual Real-person representation (Physical space)
1	Teng	#1 …Um … it depends #2 on the season that you're coming in… #3 like when you're coming like #4 in summer we have a lot of rain and we have #5 typhoon. I think #6 there's no such kind of #7 a thing you call typhoon in Spain, #8 right?	#1 Gaze: looking down Object handling: holding controller upside down #2 Head movement : head up Body posture: crossing one leg Object handling: holding controller wrong #3 Gesture: touching fingers #4 Gesture: crossing fingers Gaze: gaze shifts towards the partner	Trigger	Content	#1  #2  #3  #4 	#1  #2  #3  #4 

			#5 Gesture: iconic gesture #6 Gesture: pointing in air Body posture: moving towards left #7 Gesture: iconic gesture #8 Gaze: looking down Object handling: putting controllers together			#5  #6  #7  #8 	#5  #6  #7  #8 
2	Mandy	#9Um… no. honestly you scared me like, what do you mean…#10 typhoons?	#9 Head movement: head shaking Facial expression: puzzled look	Indicator	Explicit statement of non-understanding	#9 	#9 

			#10 Gesture: lifting up one hand Facial expression : smiling			#10 	#10 
3	Teng	#11 Typhoon is like #12 huge rain that compose of #13 huge wind and #14 like thunder, #15 lightning #16 and heavy rain #17 like we are having it right now.	#11 Facial expression : smiling Body posture: leaning forward Gesture: pointing in air #12 Body posture: leaning backward Gesture: drawing shapes in air #13 Gesture: putting hands down	Response	Overt explanation	#11  #12  #13  #14 	#11  #12  #13  #14 
			#14 Gesture: pointing in air				

			#15 Gesture: beat gesture			#15 	#15 
			#16 Gesture: pointing in air			#16 	#16 
			#17 Gaze: looking up			#17 	#17 
4	Mandy	#18 Okay I know…	#18 Facial expression: smiling Body posture: lifting one hand Head movement: nodding	Reaction to Response	Comprehension	#18 	#18 

Note. #[n] refers to the image number within the excerpted transcript being discussed

What characterizes the interaction between this dyad is the greater variety and frequency of semiotic resources used in negotiating non-understanding. While Mandy was shown to often utilize gestural resources in the NoM routines, Teng was shown to extensively exploit gestural and body posture modes to explain and elaborate on additional information. This excerpt also indicates how visual cues such as gestures, body posture, and facial expressions in each prime facilitated the NoM. In the interview, Teng commented on the different modes of communication available in VR environments. He illustrated that "I think VR works as a medium where we can see some hand gestures and body languages which helps

me to learn what my partner's feelings are and her reaction to what I said."

In Excerpt 3, Shih (Taiwanese student) and Amanda (Spanish student) discussed the festivals and holidays in Taiwan and Spain (see Figure 5 for the spatial proxemics of Dyad 3). Shih shared some traditions and cultural practices (e.g., eating moon cake) Taiwanese people do during the Moon Festival, one of the traditional festivals in Taiwan. Shih remarked, "… and also we have moon festival on September … in September and we will eat moon cake" while shifting her gaze from upwards to Amanda and downwards as presented in Line 1 (#1-#3). In Line 2, to inquire more about the ingredients of a moon cake, Amanda posed the question, "And what is it made of?" with the hope of seeking more information about the moon cake. However, a communicative breakdown was triggered when Amanda initiated the question using the pronoun "it," Shih expressed uncertainty about what "it" refers to with gestural resources like smiling and eye gaze (#4).

Figure 5

Spatial Proxemics of the Avatar Representations of Dyad 3 (Shih & Amanda) in the VR Environment

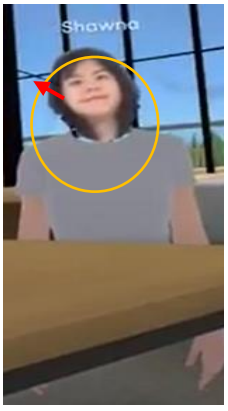
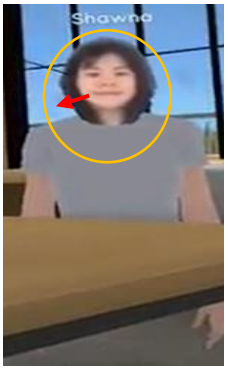
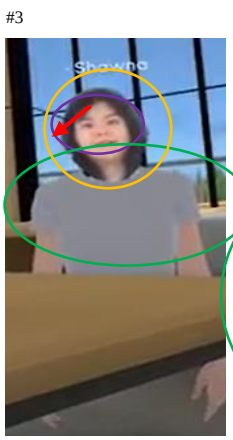
Shih (Taiwanese student) Amanda (Spanish student)

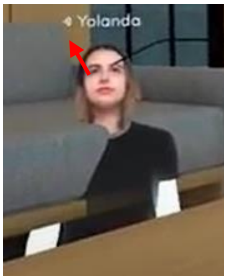
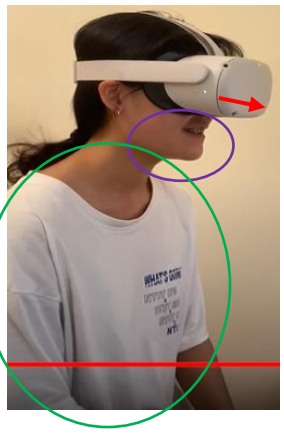
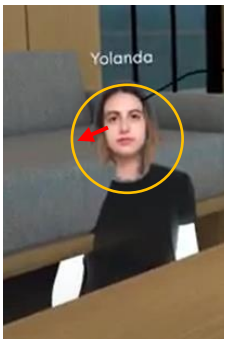
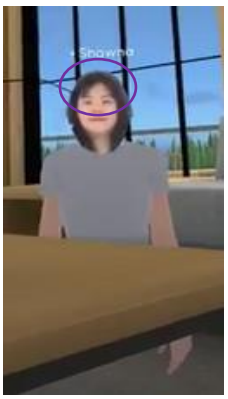
In Line 2, Shih indicated the non-understanding with a confirmation check in the spoken language mode "Moon cake? You say moon cake?". While confirming with Amanda what she meant, Shih also leaned her body forward and expressed her non-understanding with a polite smile on her face as captured in the side camera view. Shih was shown to use a 'leaning forward' body posture and a smiley face to display her attempt to resolve the communicative breakdown gently and politely (#5).

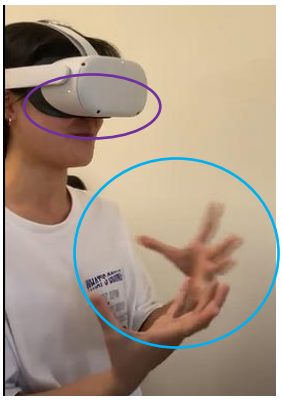
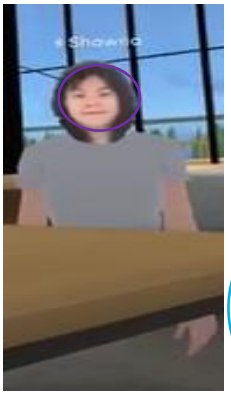
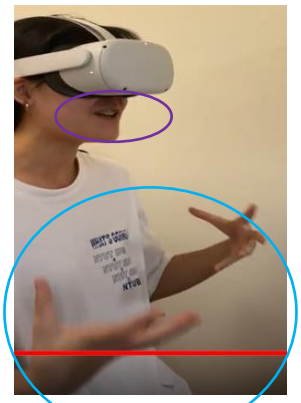
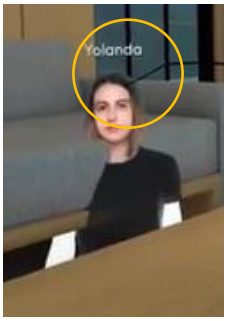
In Line 3, Amanda responded to Shih's confirmation check with a positive confirmation using both the spoken language mode ("yeah") and the head movement mode (nodding) (#6). As shown in Line 4, Shih then reacted to Amanda's response by explaining what a moon cake is. During this phrase, Shih positioned her body with a more relaxing posture by leaning the body backward and employing gestural resources to facilitate the thinking process and verbalization (#7-10). During the occurrence, Shih also removed the headset and looked at the computer, trying to look up some information about the moon cake on the Internet (#9). Such action indicates that Shih shuttled between the virtual space and the physical space, where the headset serves as a material object that divides the two spaces. The headset also functions as a spatial resource or gateway that enables learners like Shih to traverse back and forth between two spaces during the VR-mediated interaction with her Spanish counterpart. Shih viewed the computer and the Internet as material objects and semiotic resources to resolve the negotiation of meaning occurrence with Amanda.

Excerpt 3

Negotiation of Meaning Episode from Dyad 3 (Shih & Amanda)

Turn	Speaker	Verbal	Gestural	Higher-level actions (NoM prime)	Category	Visual Avatar representation (Virtual space)	Visual Real-person representation (Physical space)
1	Shih	… and also [#1] we have moon [#2] festival on September … in September and [#3] we will eat moon cake.	[#1] Head movement: tilting head Gaze: gaze upwards			#1 	#1 
			[#2] Head movement: nodding Gaze: gaze shifts towards the partner			#2 	#2 
			[#3] Head movement: nodding Gaze: gaze downwards Facial expression: smiling Body posture: sit still			#3 	#3 

2	Amanda	[#4] And what is it made of? (laughter)	[#4] Gaze: upwards	Trigger	Content	#4 	#4 
3	Shih	[#5] Moon cake? You…you say moon cake?	[#5] Body posture: leaning forward Gaze: gaze shifts towards the partner Facial expression: smiling	Indicator	Confirmation check	#5 	#5 
4	Amanda	[#6] Yeah.	[#6] Head movement: nodding Gaze: gaze shifts towards the partner	Response	Confirmation	#6 	#6 
5	Shih	Moon cake is [#7] like a desert. Yeah, [#8] I don't know what… I don't know [#9] how to made it … but yeah [#10] it's a kind of cake.	[#7] Body posture: leaning backward Facial expression: smiling Gesture: iconic gesture	Reaction to Response	Answer	#7 	#7 

			#8 Facial expression: smiling Gesture: beat gesture		#8 	#8 
			#9 Facial expression: smiling Object handling: removing headset and look at the computer Gaze: at the computer		#9 	#9 
			#10 Facial expression: smiling Gesture: beat gesture		#10 	#10 
6	Amanda	#11 Yeah.	#11 Head movement: nodding	Interaction resumes	#11 	#11 

Note. #[n] refers to the image number within the excerpted transcript being discussed

According to Norris & Pirini (2016), multimodal (inter)action analysis is "a holistic analytical framework that understands the multiple modes in (inter)action as all together building one system of communication" (p. 24). It examines interactions between social actors and other social actors, objects, or the environment. It comprises the social actors and the mediated actions in which the actors are engaged and the frozen actions entailed in material objects within the interactional setting. The observed frozen actions are entailed in electronic resources on the Internet as projected on the computer in front of Shih. It is important to note that among the five dyads in this project, Shih is the only learner who constantly shuttles between the visual and physical space during the immersive VR-mediated interaction. In the interview, Shih remarked, "I wanted to find some pictures about moon cakes on the Internet and share them with my partner, so she will understand what moon cake is better." Shih's use of objects, here computers, facilitates her process of reaction to response in NoM.

It is worth noting that in the middle of the interaction, as Shih put away the controllers, her use of gestures and body posture was not fully captured by her Spanish partner's VR view and not perceived by her partner either. However, the side camera did document Shih's employment of those gestural and visual resources in negotiating meaning throughout the virtual exchanges via the VR platform.

Discussion

Negotiation of Meaning Shifts from Verbal to Embodied in Immersive VR Environments

The quantitative findings indicate that the ELF participants also displayed negotiation of meaning behaviors in VR-supported virtual exchanges. While NoM behaviors were observed during the VR-mediated interaction, the ELF learners presented different NoM patterns in the immersive VR environment compared with those in videoconferencing, as reported in Wang and Tian's (2013) study. For example, in the Trigger stage, while lexical and syntactic triggers were shown to be the dominant ones that trigger non-understanding in Wang and Tian (2013), more content-driven triggers were identified in our study. With regard to the Indicator stage, similar findings were observed that an explicit statement of non-understanding was used most frequently by the participants to indicate non-understanding. Confirmation check was found to be the second most frequently used indicator among the ELF participants in the VR environment, whereas Echo was evident as the second most frequently used indicator among the participants in the videoconferencing context, as reported in Wang and Tian (2013). In the Response prime, we found an overt explanation as the most frequent way employed by the participants to respond to non-understanding; however, expansion was identified as the most common response in Wang and Tian (2013). In the Reaction to Response phase, in Wang and Tian's study (2013), incomprehension (42%) was identified as the most common type among the observed reaction to response behaviors. However, comprehension (64.8%) was found to be used most frequently by the ELF participants to react to responses during NoM routines in our study. Comprehension considerably outweighs other subcategories, such as comprehension check (11.4%) and confirmation (9.8%). This finding could imply that learner responses or explanations in embodied VR may potentially enhance learners' comprehension and reduce the chance of incomprehension when non-understanding occurs during the interaction. As Tai and Chen (2021) argue, the access to simulated and embodied interaction in immersive VR environments helps learners activate prior knowledge and make proper inferences, which in turn aids comprehension.

In addition to verbal utterance, multimodal forms and nonverbal cues of negotiations of meaning were found to be used coordinately by the ELF participants at all the stages when negotiating meaning. The findings indicate that, in addition to the spoken language mode, visual cues such as gestures, eye gaze, head movement, facial expression, and body posture are evident in all primes in the Trigger, Indicator, Response, and Reaction to Response stages of ELF learners' process of NoM in the VR environment. The learners utilized those modalities in interaction in order to fulfill higher-level actions such as *indicating non-understanding*, *response to non-understanding*, and *reaction to response*. Interestingly, the EFL learners also used VR space and built-in objects in the VR environment as semiotic resources to construct and negotiate meaning in the VR-mediated interaction. As such, space proxemics and object handling emerged

as the two distinct features and affordances of VR-supported environments that the ELF learners acted upon when negotiating. To construct those higher-level negotiation actions, both groups of students drew on a series of lower-level actions through the coordinated use of various multimodal semiotic resources, including verbal (e.g., speech), visual (e.g., avatar representation), gestural (e.g., eye gaze, gesture, facial expression, body posture), spatial (e.g., spatial layout and proxemics), and other semiotic resources (e.g., objects). The learners' combined use of various modes in negotiating meaning with each other demonstrates the modal complexity of L2 learners' NoM actions (Norris, 2004).

The three excerpts, as reported in this study, displayed similarities and differences in the ways and strategies of handling communicative breakdowns and the kinds of modalities the ELF learners employed to exercise those strategies in negotiating meaning via VR technologies. For example, Jie indicated his non-understanding of his Spanish partner's question using the confirmation check strategy. Jie constructed the higher-level action—confirmation check—not only by the spoken language mode but also by facial expression (a puzzled look), gaze (shifts towards the partner), head movement (head moving towards the right), and body posture (leaning against the wall). In the NoM interaction with the Spanish partner, Teng extensively utilized gestural resources and body posture in his overt explanation to respond to Mandy's non-understanding of the term "typhoons." Gestures and other visual modes were shown to play an important role in facilitating explanation and elaboration in L2, which in turn helped him to resolve the communicative breakdown. In the case of Excerpt 3, although more NoM episodes can be identified in the interaction between Shih and Amanda compared with other dyads, in addition to the verbal confirmation, Shih often utilized the visual device like smiling as a politeness strategy to initiate her non-understanding of Amanda's questions or response, which can be represented in her avatar and real-person representations. Shih was shown to constantly shuttle between the visual and physical space during the immersive VR-mediated interaction.

While the multimodal characteristics of NoM in the immersive VR space as identified in the present study have shared features with those observed in Wang and Tian's (2013) exploration of learners' NoM in the videoconferencing context (e.g., paralinguistic cues such as nodding, laughing, and a puzzled look, and visual cues such as gestures), one of the most distinctive features in the immersive VR environment compared with other CMC contexts is that learners' actions and use of the various semiotic resources were fully embodied in 3D. According to embodied cognition theory, one's bodily motions and actions will influence how they comprehend language and process information (Glenberg & Kaschak, 2002). In the context of videoconferencing, learners often gaze at the web camera to show social presence during the interaction. In immersive VR environments, however, learners are able to look into each other's eyes through representations of avatars and to engage in a stronger sense of social presence with embodied gaze and gaze shifts situated in the co-virtual space.

The affordance of embodied actions such as embodied gaze, embodied body movement, embodied gesture, embodied facial expression, and embodied engagement with objects through avatars in the VR environment thus provides learners with an embodied learning experience, which in turn creates a strong sense of social presence and facilitates learners' process of meaning creation. The sense of embodiment and affordances of spatial resources and object handling using immersive VR technologies enable learners to situate their meaning negotiation practice with cross-cultural partners in embodiment.

Embodiment and Semiotic Assemblages of VR for L2 Learning

Different modes of CMC directly influence how learners express and communicate their ideas and how they interact with each other (Stockwell, 2010). In this study, the affordances of verbal, spatial, visual, gestural, and other semiotic resources of the immersive VR environment allow ELF learners to simultaneously stimulate different sensations with a high level of realism. The built-in speakers in the VR headset enable learners to deliver and perceive 3D positional audio and verbal messages with a high quality of sound. The findings also indicate that space proxemics and objects handling are the two distinct features of VR-supported environments that facilitate NoM in virtual exchanges. The affordances of spatial proxemics in the VR setting create a virtual but shared social space for learners from geographically distant

locations to meet, interact, and collaborate, enabling a high degree of social interaction and spatial navigation during virtual exchanges. As shown in the data, Shih's engagement with different objects (i.e., Internet webpages and built-in gadgets) was shown to facilitate interaction and meaning creation with her Spanish partner.

With the availability of visual resources, an immersive VR environment allows learners to create their own 3D-realistic avatars and to be immersed in bodily representations during virtual exchanges. The head-mounted display affords learners a full vision of the spatial layout of the VR space, other social actors (e.g., cross-cultural partners), and material objects (e.g., sofa, desk, window, door), which promotes learners' participation and engagement with the VR space and the situated social interaction with their partners. With lipsyncing and facial recognition functions, learners' virtual representations become more vivid and resemble real-life interactions. Furthermore, the hand controller of the VR device serves as a mediator that transforms learners' gesture and body position performed in the physical space into the virtual space during interactions. VR technologies provide various gestural resources, allowing L2 learners to perform multimodal representations using hand and head movements, eye gaze, gestures, and body posture in order to experience embodied interaction with each other. The visual and gestural stimuli provided by the VR tools could "create an immersive sense of presence, so as to trigger the brain to activate schema to process information in multisensory representations" (Xie et al., 2019, p. 11). The access to a full range of gestures and non-verbal cues is shown to facilitate key processes for language development/learning, such as negotiation of meaning, comprehension, noticing, and ultimately, learning (York et al., 2021).

Overall, the various semiotic resources provided by VR technologies and VR platforms hold good potential for L2 social interactions and learning. While not devaluing the importance of linguistic resources to language learners' L2 development, it is essential to understand how semiotic resources in addition to linguistic resources can be utilized simultaneously to aid comprehension and facilitate the process of meaning making in the L2. This requires a conceptual lens to view interaction and see VR-mediated interactions as "semiotic assemblages" that allow "an understanding of how different trajectories of people, semiotic resources, and objects meet at particular moments and places and thus helps us to see the importance of things, the consequences of the body, and the significance of place alongside the meanings of linguistic resources" (Pennycook, 2017, p. 269). This notion of semiotic assemblages thus enables us to uncover the diverse pathways ELF learners from remote locations take in interacting with each other with a full embodiment using semiotic resources and material objects in the shared virtual space at a given time. Consequently, virtual exchanges in immersive VR environments can provoke our students' embodied learning and engage learners in a natural, embodied, and perception-action rich context.

Conclusion and Implications

This study explores how Taiwanese and Spanish students employ semiotic resources to negotiate for meaning in the immersive VR environment. Adopting multimodal (inter)action analysis (Norris, 2004) and Varonis and Gass' (1985a) negotiation of meaning model to analyze learners' meaning negotiation instances, the findings indicate that learners perform lower-actions by an appropriation of different modes to construct higher-level actions such as indicating non-understanding, confirming comprehension, and explaining in immersive VR space. Learners' use and engagement with semiotic resources, including verbal (e.g., spoken language), visual (e.g., avatar representation), gestural (e.g., gaze, gestures, facial expression, body posture), spatial (e.g., spatial layout), and other semiotic objects (e.g., coffee, wine, sofa) were shown to be fully embodied, which facilitates indication, comprehension, confirmation, and explanation to solve communicative breakdowns in the L2. This study calls for a multimodal and semiotic lens to look at VR-mediated interactions and the role of semiotic resources in L2 learners' process of meaning construction and negotiation.

The exploration of this study is timely as multimodal, immersive interaction supported by VR technologies is still under-explored in CALL research, both theoretically and in practice. Such research inquiries would contribute to the field of language education in two ways. First, VR immersive environments could hold

good potential for fostering meaning creation as it represents the closest form of interaction to face-to-face interaction. The findings would allow a more comprehensive understanding of ELF learners' use and process of negotiating for meaning in immersive VR-mediated virtual exchanges.

Pedagogically, language teachers could provide students with adequate instruction and facilitation to help them develop strategies for the effective negotiation of meaning. For example, conversational strategies such as paraphrasing, rephrasing, recasting, expansion, and reduction in their breakdowns could be integrated into language classrooms to promote students' quality L2 production and achieve effective communication. Other multimodal strategies, including the use of gestures, eye gaze, facial expressions, and body posture, and the affordances of immersive VR technologies for language learning could be introduced to students and discussed in classrooms to enhance their awareness of utilizing those multimodal resources in a creative way to mitigate non-understanding and to promote quality interaction in L2. With the affordances of verbal, visual, gestural, spatial, and other semiotic resources, immersive VR technologies break the boundaries across physical spaces and geographical borders by providing learners immersive and embodied virtual experiences with diverse cultures and communities. Telecollaboration projects using immersive VR technologies can be integrated into L2 curricula by connecting the L2 classroom to the multifaceted contexts and challenges of language use beyond the classroom.

To conclude, VR environments allow language learners a more flexible approach to time, space, and body by offering learners an immersive experience in situated environments where communicating in a second language feels like being in a flow state. The use of VR technologies for learning has resulted in "a change from the notion of a place to that of space or spatiality, which includes a shift from physical institutions to distributed communities which can act as communities of practice or be experienced as affinity spaces" (Hampel, 2019, p. 289). As such, virtual technologies promote a new perspective, which moved from individual agency to distributed practice and from competence to emplacement (Canagarajah, 2018). VR technologies also entail a turn from verbal to embodied communication and online re-embodiment in online virtual worlds. While the virtual spaces open up new opportunities and communities for language learning, there are also challenges when interacting in VR environments. Learners need to develop an awareness of multimodality and new literacy skills so they can use different modalities available to participate in both the physical and virtual world.

Acknowledgements

We would like to express gratitude for the editors and reviewers for their valuable comments. We also wish to extend thanks to the students who participated in the study. This project was supported by the National Science and Technology Council in Taiwan (Grant Number: NSTC 110-2410-H-027-011) and the Ministry of Science, Innovation and Universities in Spain (Ref: RTI2018-094601-B-100).

References

Abdullah, A., Kolkmeier, J., Lo, V., & Neff, M. (2021). Videoconference and embodied VR: Communication patterns across task and medium. *Proceedings of the ACM on Human-Computer Interaction, 5*, 1–29.

Akayoglu, S., & Altun, A. (2009). The functions of negotiation of meaning in text-based CMC. In R. de Cassia Veiga Marriott & P. Lupion Torres (Eds.), *Handbook of research on e-learning methodologies for language acquisition* (pp. 291–306). Information Science Reference.

Alfadil, M. (2020). Effectiveness of virtual reality game in foreign language vocabulary acquisition. *Computers & Education, 153*, 103893.

Anton, D., Kurillo, G., & Bajcsy, R. (2018). User experience and interaction performance in 2D/3D telecollaboration. *Future Generation Computer System, 82*, 77–88.

Blake, R. (2000). Computer-mediated communication: A window on L2 Spanish interlanguage. *Language Learning & Technology, 4*(1), 120–136. http://doi.org/10125/25089

Burdea, G. C., & Coiffet, P. (2003). *Virtual reality technology*. John Wiley & Sons.

Canagarajah, S. (2018). Translingual practice as spatial repertoires: Expanding the paradigm beyond structuralist orientations. *Applied Linguistics, 39*(1), 31–54. https://doi.org/10.1093/applin/amx041

Canals, L. (2021). Multimodality and translanguaging in negotiation of meaning. *Foreign Language Annals, 54*(3), 1–24. https://doi.org/10.1111/flan.12547

Chen, J., & Kent, S. (2020). Task engagement, learner motivation and avatar identities of struggling English language learners in the 3D virtual world. *System, 88*, 102168.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Fernández-García, M., & Martínez-Arbelaiz, A. (2002). Negotiation of meaning in nonnative speaker-nonnative speaker synchronous discussions. *CALICO, 19*(2), 279–294.

Flower, C. (2015). Virtual reality and learning: Where is the pedagogy? *British Journal of Educational Technology, 46*(2), 412–422.

Gadelha, R. (2018). Revolutionizing education: The promise of virtual reality. *Childhood Education, 94*(1), 40–43. https://doi.org/10.1080/00094056.2018.1420362

Gass, S. M., & Mackey, A. (2007). Input, interaction, and output in second language acquisition. In B. VanPatten & J. Williams (Eds.), Theories in second language acquisition (pp. 175–200). LEA.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychnomic Bulletin & Review, 9*(3), 558–565.

Glenberg, A. M., Witt, J. K., & Metcalfe, J. (2013). From the revolution to embodiment: 25 years of cognitive psychology. *Perspectives on Psychological Science, 8(5)*, 573–585. https://doi.org/10.1177/1745691613498098

Hampel, R. (2019). The conceptualization of time, space, and the body in virtual sites and the impact on language learner identities. In S. Bagga-Gupta, G. Messina Dahlberg, & Y. Lindberg (Eds.), *Virtual sites as learning spaces* (pp. 269–294). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-26929-6_10

Hampel, R., & Stickler, U. (2012). The use of videoconferencing to support multimodal interaction in an online language classroom. *ReCALL, 24*(2), 116–137.

Howard-Jones, P., Ott, M., van Leeuwen, T., & De Smedt, B. (2014). The potential relevance of cognitive neuroscience for the development and use of technology-enhanced learning. *Learning, Media and Technology, 40*(2), 131–151. https://doi.org/10.1080/17439884.2014.919321

Jauregi, K., Canto, S., Graaff, R., Koenraad, T., & Moonen, M. (2011). Verbal interaction in Second Life: Towards a pedagogic framework for task design. *Computer Assisted Language Learning Journal, 24*(1), 77–101.

Lan, Y. J. (2014). Does second life improve Mandarin learning by overseas Chinese students? *Language Learning & Technology, 18*(2), 36–56. http://dx.doi.org/10125/44365

Lan, Y. J. (2020). Immersion, interaction, and experience-oriented learning: Bringing virtual reality into FL learning. *Language Learning & Technology, 24*(1), 1–15. http://hdl.handle.net/10125/44704

Lan, Y. J. (2021). Language learning in virtual reality: Theoretical foundations and empirical practices. In Y. J. Lan & S. Grant (Eds.), *Contextual language learning* (pp. 1–22). Springer.

Lan, Y. J., Lyu, B. N., & Chin, C. K. (2019). Does a 3D immersive experience enhance Mandarin writing by CSL students? *Language Learning & Technology, 23*(2), 125–144. https://doi.org/10125/44686

Lee, H. (2020). *Gesture in multimodal language learner interaction via videoconferencing on mobile devices* [Unpublished doctoral dissertation]. The Open University.

Lee, H., Hampel, R., & Kukulska-Hulme, A. (2019). Gesture in speaking tasks beyond the classroom: An exploration of the multimodal negotiation of meaning via Skype videoconferencing on mobile devices. *System, 81*, 26–38.

Lee, L. (2001). Online interaction: Negotiation of meaning and strategies used among learners of Spanish. *ReCALL, 13*(2), 232–244.

Lee, L. (2009). Exploring native and nonnative interactive discourse in text-based chat beyond classroom settings. In L. B. Abraham & L. Williams (Eds.), *Electronic discourse in language learning and language teaching* (pp. 127–150). John Benjamins.

Legault, J., Zhao, J., Chi, Y., Chen, W., Klippel, A., & Li, P. (2019). Immersive virtual reality as an effective tool for second language vocabulary learning. *Languages, 4*(1), 1–13. https://doi.org/10.3390/languages4010013

Liaw, M. L. (2019). EFL learners' intercultural communication in an open social virtual environment. *Educational Technology & Society, 22*(2), 38–55.

Lloyd, A., Rogerson, S., & Stead, G. (2017). Imagining the potential for using virtual reality technologies in language learning. In M. Carrier, R. M. Damerow, & K. M. Bailey (Eds.), *Digital language learning and teaching: Research, theory and practice* (pp. 222–234). Routledge.

Long, M. (1983a). Native speaker/non-native speaker conversation in the second language classroom. In M. Clarke & J. Handscombe (Eds.), *On TESOL'82: Pacific perspectives on language and teaching* (pp. 207–228). TESOL.

Long, M. (1983b). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics, 4(2)*, 126–141.

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.

Mackey, A., Abbuhl, R., & Gass, S. (2012). Interactionist approach. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 7–23). Routledge.

Makransky, G., & Petersen, G. B. (2021). The cognitive affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review, 33*, 937–958.

Maloney, D., Freeman, G., & Wohn, D. (2020). "Talking without a voice": Understanding non-verbal communication in social virtual reality. Proceedings of the ACM on Human-Computer Interaction, *4*(CSCW2), Article 175, 1–25. https://doi.org/10.1145/3415246

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

Nakahama, Y., Tyler, A., & Van Lier, L. (2001). Negotiation of meaning in conversational and information gap activities: A comparative discourse analysis. *TESOL Quarterly, 35*(3), 377–405.

Norris, S. (2004). *Analyzing multimodal interaction*. Taylor & Francis.

Norris, S. (2017). Modal density and modal configurations. In C. Jewitt (Ed.), *The Routledge handbook of multimodal analysis* (pp. 86–99). Routledge.

Norris, S. (2019). *Systematically working with multimodal data: Research methods in multimodal discourse analysis*. Wiley.

Norris, S., & Pirini, J. (2016). Communicating knowledge, getting attention, and negotiating disagreement via videoconferencing technology: A multimodal analysis. *Journal of Organizational Knowledge Communication*, *3*(1), 23–48.

O'Rourke, B. (2005). Form-focused interaction in online Tandem learning. *CALICO Journal*, *22*(3), 433–466.

Oliver, R. (2002). The patterns of negotiation for meaning in child interactions. *The Modern Language Journal*, *86*(1), 97–111.

Pellettieri, J. (2000). Negotiation in cyberspace: The role of chatting in the development of grammatical competence. In M. Warschauer & R. Kern (Eds.), *Network-based language teaching: Concepts and practice* (pp. 59–86). Cambridge University Press.

Pennycook, A. (2017). Translanguaging and semiotic assemblages. *International Journal of Multilingualism*, *14*(3), 269–282.

Pica, T. (1991). Classroom interaction, participation, and comprehension: Redefining relationships. *System*, *19*, 437–452.

Pica, T. (1992). The textual outcomes of native-speaker-non-native speaker negotiation: What do they reveal about second language learning. In C. Kramsch & S. McConnell-Ginet (Eds.), *Text and context: Cross-disciplinary perspectives on language study* (pp. 198–237). D C Heath & Co.

Pica, T. (1994). Research on negotiation: What does it reveal about second-language learning conditions, processes, and outcomes? *Language Learning*, *44*(3), 493–527.

Pica, T. (1996). The essential role of negotiation in the communicative classroom. *JALT Journal*, *78*, 241–268.

Pica, T., Young, R., & Doughty, C. (1987). The impact of interaction. *TESOL Quarterly*, *21*(4), 737–758.

Salazar Miranda, A., & Claudel, M. (2021). Spatial proximity matters: A study on collaboration. *PLoS ONE*, *16*(12), e0259965.

Smith, B. (2003a). The use of communication strategies in computer-mediated communication. *System*, *31*, 29–53.

Smith, B. (2003b). Computer-mediated negotiated interaction: An expanded model. *The Modern Language Journal*, *87*(1), 38–57.

Smith, H., & Neff, M. (2018). Communication behavior in embodied virtual reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Article 289, pp. 1–12). Association for Computing Machinery.

Stockwell, G. (2010). Using mobile phones for vocabulary activities: Examining the effect of the platform. *Language Learning & Technology*, *14*(2), 95–110. http://dx.doi.org/10125/44216

Tai, T., & Chen, H. (2021). The impact of immersive virtual reality on EFL learners' listening comprehension. *Journal of Educational Computing Research*, *59*(7), 1272–1293. https://doi.org/10.1177/0735633121994291

Tudini, V. (2003). Using native speakers in chat. *Language Learning & Technology*, *7*(3), 141–159. http://dx.doi.org/10125/25218

Tudini, V. (2007). Negotiation and intercultural learning in Italian native speaker chat rooms. *The Modern Language Journal*, *91*(4), 577–601.

Van der Zwaard, R., & Bannink, A. (2014). Video call or chat? Negotiation of meaning and issues of face in telecollaboration. *System*, *44*, 137–148.

Van der Zwaard, R., & Bannink, A. (2016). Nonoccurrence of negotiation of meaning in task-based synchronous computer-mediated communication. *The Modern Language Journal, 100*(3), 625–640. https://doi.org/10.1111/modl.12341

Van der Zwaard, R., & Bannink, A. (2019). Toward a comprehensive model of negotiated interaction in computer-mediated communication. *Language Learning & Technology, 23*(3), 116–135. http://hdl.handle.net/10125/44699

Van der Zwaard, R., & Bannink, A. (2020). Negotiation of meaning in digital task-based language teaching: Task design versus task performance. *TESOL Quarterly*, *54*(1), 56–89. https://doi.org/10.1002/tesq.537

Varonis, E., & Gass, S. (1985a). Non-native/non-native conversations: A model for negotiation of meaning. *Applied Linguistics, 6(1)*, 71–90.

Varonis, E., & Gass, S. (1985b). Miscommunication in native/non-native conversation. *Language in Society, 14*(3), 327–343.

Wang, M. (2020). Social VR: A new form of social communication in the future or a beautiful illusion? *Journal of Physics: Conference Series*, *1518*, 1–8.

Wang, Y. (2006). Negotiation of meaning in desktop videoconferencing-supported distance language learning. *ReCALL, 18*(1), 122–146.

Wang, Y. F., Petrina, S., & Feng, F. (2017). VILLAGE–Virtual immersive language learning and gaming environment: Immersion and presence. *British Journal of Educational Technology, 48*(2), 431–450.

Wang, Y., & Tian, J. (2013). Negotiation of meaning in multimodal tandem learning via desktop videoconferencing. *International Journal of Computer-Assisted Language Learning and Teaching, 3*(2), 41–55.

Wei, L. (2018). Translanguaging as a practical theory of language. *Applied Linguistics, 39(1)*, 9–30. https://doi.org/10.1093/applin/amx039

Wigham, C. R., & Satar, M. (2021). Multimodal (inter)action analysis of task instructions in language teaching via videoconferencing: A case study. *ReCALL, 33*(3), 195–213. https://doi.org/10.1017/S0958344021000070

Wigham, C. R., Panichi, L., Nocchi, S., & Sadler, R. (2018). Interactions for language learning in and around virtual worlds. *ReCALL, 30*(2), 153–160.

Xie, Y., Ryder, L., & Chen, Y. (2019). Using interactive virtual reality tools in an advanced Chinese language class: A case study. *TechTrends*, *63*(3), 251–259.

Yanguas, I. (2010). Oral computer-mediated interaction between L2 learners: It's about time! *Language Learning & Technology, 14*(3), 72–93. http://dx.doi.org/10125/44227

Yin, Q., & Satar, M. (2020). English as a foreign language learner interaction with chatbots: Negotiation for meaning. *IOJET, 7*(2), 390–410. http://iojet.org/index.php/IOJET/article/view/707

York, J., Shibata, K., Tokutake, H. & Nakayama, H. (2021). Effect of SCMC on foreign language anxiety and learning experience: A comparison of voice, video, and VR-based oral interaction. *ReCALL, 33*(1), 49–70.

Appendix A. Description of the VR Tasks

Scenario: You have been admitted to the university in Taiwan/Spain for the fall semester in 2022. The university has assigned you with a local student as your international buddy who will assist you in preparation for the transition to the new environment. You will arrange **2 tasks** with your international buddy.

Task #1 (30 minutes)

Introducing yourselves and getting to know each other.

Below is a list of questions for your partner (not limited to this list):

1. Could you tell me about yourself?
2. What challenges do you encounter at school?
3. What is your experience learning English?
4. What is COVID-19 situation in Spain/Taiwan?
5. What do you plan to do after graduation?
6. Do you have part-time jobs? What kinds of jobs do you have? Why?
7. What is the college life like at universities in Spain/Taiwan?
8. How do you celebrate the holidays in your country?
9. What is your impression about Asians/Spanish?
10. What do you know about Spanish/Taiwanese culture?
11. What kinds of activities do Spanish/Taiwanese students usually do at school?

The list of questions is for your reference. Please feel free to ask other questions if you would like.

Task #2 (45 to 60 minutes)

Taking turns playing these two roles.

Role 1, International student: You are a student participating in a study abroad program, and you are preparing for your transition to the new university in a foreign country. You have been assigned an international buddy who is going to mentor you and provide you with some useful information in order to help you adjust to the academic and social life in the new country. Ask your international buddy as many questions as possible so as to better adapt to the new country and its culture (e.g., what is the academic culture at National Taipei University of Technology (NTUT) in Taiwan/University of Valencia (UV) and University Jaume I (UJI) in Spain, what are the fun things to do in Taipei/Valencia and Castellón, what places should you visit, student life on campus, social events, student organizations/activities/clubs, how to find accommodation, transportation and how to move around the city, interesting cultural events or festivals, traditions and customs, where to eat, where to buy your groceries, weather, current COVID19 situation, etc.).

Role 2, International buddy from the local university: You are a local student who has volunteered to become an international buddy. Your role is to welcome an international student participating in a study abroad program and to become their mentor and counselor. In other words, you need to help them prepare for their transition to the new university in the foreign country by providing them with some useful information regarding cultural aspects as well as academic and social life in the new country. The international student will ask you questions about your university, your country, its culture, and more. Answer all their questions and provide any additional information they might find useful.

About the Authors

Hsin-I Chen is an Associate Professor in the Department of English at the National Taipei University of Technology, Taiwan. Her research focuses on technology-enhanced language learning and teaching, telecollaboration, multimodality, and teacher education.

E-mail: hichen@mail.ntut.edu.tw

Ana Sevilla-Pavón is an Associate Professor at the Universitat de València, Spain. Her research interests include computer-assisted language learning, English for specific purposes, teacher training and the design of resources for educational innovation, and intercultural communication through virtual exchange.

E-mail: ana.m.sevilla@uv.es