EuroCALL 2025: Advancing CALL: New research agendas
27-30 August 2025, Università Cattolica del Sacro Cuore, Milan, Italy
DOI: https://doi.org/10.4995/EuroCALL2025.2025.21193

Exploring multimodal analysis in VR-assisted language learning: Insights and applications

Maria Christoforou
Cyprus University of Technology, , maria.christoforou@cut.ac.cy

How to cite: Christoforou, M. (2025). Exploring multimodal analysis in VR-assisted language learning: Insights and applications. In Y. Choubsaz, P. Díez-Arcón, A. Gimeno-Sanz, V. Morgana, A. C. Murphy & F. L. Seracini (Eds.), *Advancing CALL: New research agendas - EUROCALL 2025 Short Papers.* https://doi.org/10.4995/EuroCALL2025.2025.21193

Abstract

Multimodality recognizes that meaning making does not happen through language alone and examines the co-existence of various communicative modes, including sensory modes such as sight, hearing and touch (the haptic mode, concerned with tactile experience), as well as gestures, body movements, images, and both spoken and written language. As a theoretical and methodological framework, multimodality has been widely explored in education (Kress, 2012; Kress & van Leeuwen, 2001; Norris, 2004a, 2004b). However, its application within High-Immersion Virtual Reality (HiVR) environments has remained underexplored (Chen & Sevilla-Pavón, 2023; Jauregi-Ondarra et al., 2024). With its capacity for high immersion, embodiment and sensory engagement, HiVR enables contextualized, situated learning social interactions since learners can simulate authentic conditions that resemble real-life scenarios (Christoforou et al., 2019) and learn by doing. This paper explores how multimodal analysis can be methodologically applied in HiVR environments by examining a single learning episode from a larger study on immersive language learning (Thrasher et al., 2024b). Through a comparative application of three frameworks - Multimodal Discourse Analysis (MDA), Multimodal Interaction Analysis (MIA), and the Grammar of Transposition - the study aims to contribute to the underresearched field of multimodal analysis in VR-assisted language learning (VRALL) and to offer a practical entry point for future research in immersive language education.

Keywords: *Multimodal analysis; Virtual Reality-assisted language learning (VRALL); High-Immersion Virtual Reality (HiVR); intercultural communication.*

1. Introduction

Recent research has slowly begun to explore how multimodality can be investigated in High-Immersion Virtual Reality (HiVR) contexts. Kaplan-Rakowski and Gruber (2019) define HiVR as a computer-generated 360° virtual space that can be perceived as spatially realistic, due to the high immersion afforded by a head-mounted device, such as Meta Quest 2. Within such highly multimodal immersive environents, several possibilities for research can be identified such as the fostering of a deeper understanding of intercultural and linguistic dynamics (Jauregi-Ondarra et al., 2024). This study explored the potential of social VR environments to facilitate intercultural interaction and awareness, offering new opportunities for Dutch and Cypriot learners to negotiate meaning across linguistic and cultural boundaries in real-time. Moreover, Mills and Brown (2021) emphasized the role of transmediation of content across modes such as visual and haptic in digital designing within VR compared to more

BY-NC-SA 2025, Editorial Universitat Politècnica de València

conventional modes of drawing and writing, showing how learners fluidly shift meaning across visual, auditory, and haptic modes. Building on this, Mills et al. (2022) examined how embodiment in VR, through gestures, body movement and spatial positioning, deepens cognitive and communicative engagement in a meaning-making process and accentuates the role of bodily movement and sensory interaction in cognition. Finally, Christoforou and Efthimiou (2023) illustrated how the surrealistic experience in the VR application, *Dreams of Dali*, enhanced the students' sensory and linguistic engagement in art-related subjects, enabling them to engage with contextualized, discipline-relevant content through the VR affordances: immersion, simulation, and spatial navigation in the virtual painting. These studies collectively suggest that VR offers a fertile ground for multimodal meaning-making. However, to better understand how such meanings are constructed in HiVR environments, there is a need to apply robust multimodal analytical frameworks that can account for the diverse modes at play. When it comes to collecting and analyzing "real" data in HiVR settings and seeing how participants interact with the virtual space, Thrasher et al. (2024a) support that multimodality can be a pertinent methodological framework to capture the unique semiotic and embodied dimensions of interaction in these environments. Multimodal analysis is particularly appropriate for studying the rich layers of meaning that emerge through gesture, spatial navigation, interaction with the interface, and other embodied resources in immersive learning contexts.

Drawing on the VR affordances discussed above, this paper adopts a multimodal analytical perspective to explore meaning-making in VR-assisted language learning (VRALL). By analyzing a learning episode conducted in a HiVR environment, the study aims to examine how communicative processes unfold through the orchestration of multiple semiotic modes (gesture, spatial movement, and spoken language) particularly in the case of a low-proficiency language learner. The episode is examined through three distinct analytical frameworks: Multimodal Discourse Analysis (MDA), Multimodal Interaction Analysis (MIA), and the Grammar of Transposition, in order to compare their interpretive value and methodological demands. In doing so, this paper contributes to the limited but growing body of research on multimodal analysis in immersive environments and offers a practical foundation for researchers interested in analyzing language learning in VR.

2. Method

2.1. Context and participants

2.1.1. Immerse

This paper draws on data originally collected in the study by Thrasher et al. (2024b), which examined six adult learners' experiences in *Immerse* (https://www.immerse.online), a HiVR platform purpose-built for language learning. *Immerse* distinguishes itself through its emphasis on multimodal interaction, offering a highly embodied and sensory-rich environment that extends meaning-making beyond the verbal mode. Unlike many VR applications that limit users to solo experiences, *Immerse* is designed as a collaborative, real-time social space**,** allowing learners to engage with both peers and instructors through a range of interconnected semiotic modes, including gesture, spatial navigation, object manipulation, gaze, and voice. At the time of the study, the platform included over 40 interactive virtual settings (e.g., airports, restaurants, parks, a zoo), each supporting contextualized, embodied participation. Within these environments, learners could handle virtual objects, navigate spatially, and use built-in features like interactive whiteboards and object scanners that added layers of visual and auditory feedback. These affordances created dynamic opportunities for transmodal learning, where learners simultaneously processed and produced meaning through movement**,** touch (haptic), visuals, and sound, contributing to a deeper sense of presence.

BY-NC-SA

Figure 1. Multimodal VR environment in *Immerse* featuring avatars, interactive objects, and spatially rich context for real-time language interaction. (Image source: *Immerse*).

2.1.2. Participants

In the original case study, six adult learners employed by a Japanese education company participated voluntarily. Their ages ranged from 30 to 54 years. At the onset of the project, participants self-reported their English proficiency levels (A1-B2, CEFR). Since participation was voluntary and part of a pilot project, no formal assessment was administered; self-reporting was considered appropriate for establishing a general indication of their communicative proficiency and informing lesson planning. Over the course of eight weeks, the participants engaged in weekly thirty-minute tasks on the *Immerse* VR platform, with the aim of improving their English-speaking skills and building communicative confidence through immersive, interactive experiences. The lessons targeted general English for everyday communicative purposes, such as ordering at a restaurant, making plans with friends, or navigating an airport. Weekly tasks were situated in immersive VR scenes (e.g., a fast-food restaurant, a bar, an airport) and were designed to build both spoken fluency and communicative confidence in authentic, interactive contexts. All sessions were facilitated by a professional language instructor.

This paper focuses on a single learning episode from the final task (Week 8: Navigating the airport), in which only one participant, Nick (pseudonym), took part alongside the instructor. The task took place in a virtual airport setting, where Nick was instructed to go through airport security, board a plane, and select a destination to explore. Nick chose to travel with the instructor to Rantau Mountain in Nepal, a site of personal significance to him, while the instructor in turn guided him to Hamburg, Germany, where she lives. The instructor's role was to facilitate and prompt Nick to describe his chosen location, while also sharing her own. This reciprocal exchange foregrounded both personal storytelling and cultural sharing, illustrating how immersive VR tasks can create opportunities for language practice intertwined with intercultural meaning-making. The episode was selected for analysis as an illustrative case, following standard practices in qualitative multimodal research: it demonstrates how a learner with limited English proficiency (A1 level) effectively communicated and negotiated meaning using multimodal resources in the HiVR environment, developed intercultural awareness through contextualized interaction, and highlights methodological insights.

2.2. Multimodal analysis of a HiVR learning episode

In this section, three multimodal frameworks are presented for data analysis in a HiVR setting:

BY-NC-SA

Table 1. Multimodal frameworks

Multimodal Discourse Analysis (MDA) (Kress, 2012)	Multimodal Interaction Analysis (MIA) (Norris, 2004)	Grammar of Transposition (Cope & Kalantzis, 2020)
Analyzes mode orchestration and modal density in discourse	Examines embodied interaction, modal complexity over time	Traces how meanings are designed, redesigned, and recontextualized across forms (modes)

MDA focuses on the configuration of semiotic modes in a specific communicative event (e.g., how gesture, gaze, language co-occur) for analysis and it is rooted in social semiotics (Kress & van Leuwen, 2001). Methodologically, it can use a multimodal transcript or a software analysis tool (e,g. ELAN). The analysis of in-the-moment communication (e.g., how gesture, speech, and gaze create meaning) in the HiVR learning episode focused on how Nick and the instructor combine verbal speech, gesture, and spatial orientation in real time to co-construct meaning during their VR exchange. MIA is rooted in activity theory, interactional sociolinguistics, and social semiotics, and was developed by Norris (2004). The unit of analysis is the *mediated action* and is composed of *higher-level* and *lower-level actions* and methodologically it uses video data, and VR recordings with interaction-centered analysis with timestamped sequences of multimodal actions. In this paper, the higher-level action for the student participant was sharing his travel memory whereas some lower-level actions were using hand gestures, his body to orientate his avatar body, etc. MIA helps explain how these multimodal actions work together across time to construct meaning in the interaction. Like MDA, MIA also benefits from detailed multimodal transcripts, particularly time-stamped sequences of actions, to trace the unfolding of higher- and lower-level mediated actions over time. Norris (2004) also uses the terms *modal complexity* to refer to how tightly interwoven or synchronized the modes are in action and *modal density* to refer to the number of modes that are simulatenously involved in an interaction and how much communicative "weight" each carries. Finally, the Grammar of Transposition (Christoforou, 2025; Cope & Kalantzis, 2020) may not require a detailed moment-by-moment transcript like MDA or MIA, but it still draws on annotated data or transcript excerpts to trace meaning shift across forms (modes), such as body, speech, space, object, etc.. In addition to these form-based shifts, the framework also attends to how learners' meaning-making shifts in function, including changes in reference, agency, structure, context, and interest. In this paper, the Grammar of Transposition is used to trace how Nick's personal experience is re-articulated across forms, moving from body (i.e., simulating a helicopter with hand gestures), to speech (i.e., narrating his travel story), and space (i.e., spatial transposition through virtual movement to Rantau and later to Planten un Blomen park in Hamburg) in ways that reveal how the HiVR environment supports layered, agentive meaning-making.

2.2.1. Analysis of the HiVR learning episode following MDA

Table 2. Multimodal transcript in HiVR interaction following MDA

Time	**Verbal mode (speech / text)**	**Gestural mode (pointing, posture, etc.)**	**Spatial mode (movement, orientation, VR affordances)**	**Notes / intercultural significance**
3:14	Instructor: *"Can you tell me a touristic place in Nepal?"*	Faces Nick; head tilt	Avatars standing near mountain terrain	Initiates intercultural inquiry

BY-NC-SA

3:37	Instructor types: *"Temple? River?"*	Points at the surrounding mountains	VR landscape of mountains visible	Uses gesture to scaffold meaning
3:54	Instructor asks: *"Which mountain?"*	Types the question "What's the name of the mountain?"	The built-in text feature	Enables the student to express a place with a cultural / personal meaning
04:04	Nick: *"Mount Rantau"*	Points again to same mountain area	Avatar leans forward slightly	Refers to culturally specific place
04:15	Instrutor: *"I will look right now in Google Maps"*	Looks down; hands moving (typing)	Opens Google Maps inside VR	Uses external tool to bridge knowledge gap
5:08	Teleportation to Everest	—	Scene changes to snowy Everest peak	Shared virtual spatial shift
5:36	Nick: *"I, I, I I walked (.), I went to Rantau at helicopter. Mountain flight. No trekking"*	Points to a real mountain (not possible in a traditional classroom) when verbal explanation is difficult; simulates flying above the mountain with hands	Nick makes use of spatial positioning and hand tracking to physically enact his story	Shares personal travel experience
6:10	Teleportation to Rantau mountain	—	Scene changes to a different snowy slope	Shared virtual spatial shift
6:24	Instructor: *"Is it cold here?"*	Turns her body 180 degrees to inspect the terrain behind her	Rotates body to reorient spatial perspective	Initiates context-aware interaction based on environmental cues
6:34	Nick: *"Yes, a little cold and [switches to native tongue], air (.) very (.) very clean, a little oxy-, oxy-"*	Uses gesture space near the face to embody thinking / wondering	Proximity to the built-in text feature to observe what the instructor is writing	Uses external tool to help him find the right word
9:38	Teleport to Hamburg	—	Scene transitions to German city	Intercultural enrichment
9:58	Instructor: *"I live in Germany. Here. This is a very famous park in Germany. It's called Planten un Blomen"*	Points around environment (extended arm posture)	Uses VR pointer tool to spatially guide Nick through local landmarks	Cultural and spatial navigation / Sharing of local knowledge

2.2.2. Analysis of the HiVR learning episode following MIA

From an MIA perspective, the HiVR interaction between Nick and the instructor reveals a layered structure of mediated actions. The *higher-level action* involves Nick's participation in intercultural dialogue and navigation through virtual environments, composed of chained *lower-level actions* such as gesture (e.g., pointing, simulating flight), gaze orientation, and spatial repositioning relative to virtual landmarks. For instance, when Nick describes his visit to Rantau Mountain, he simulates walking and flying with his hands while narrating the experience,

BY-NC-SA

aligning embodied action with spoken language. This moment demonstrates both *high modal density*, as multiple modes operate simultaneously, and *modal complexity*, as the modes are closely coordinated to construct meaning. These embodied gestures are not isolated but function as integral semiotic resources in the co-construction of meaning. The instructor also adjusts her position and gaze dynamically to sustain engagement, reflecting sensitivity to the embodied and spatial aspects of interaction in HiVR. The sequence as a whole exemplifies how meaning unfolds in real time through layered, multimodal orchestration, consistent with Norris's (2004) emphasis on modal complexity and density within interactional moments.

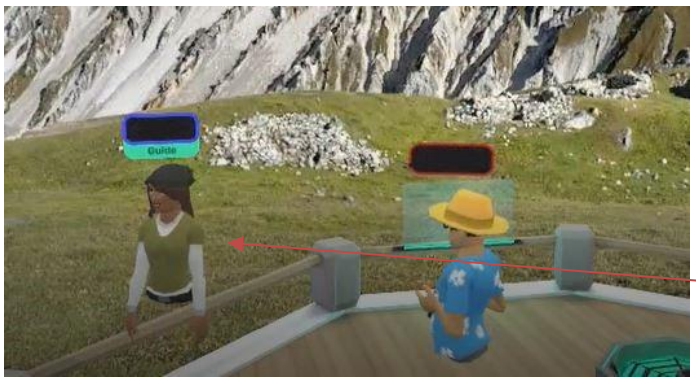
3:14: Nick's starting posture, hand slightly raised and avatar facing the instructor (forward gaze) set up the higher-level action (the start of intercultural sharing)

Figure 2. Initiation of interaction (Nick begins to explain his trip to Rantau)

3:37: Nick gestures flying while narrating his travel experience. His hands are in mid-air mimicking flight, body upright. The synchrony of gesture, speech, and spatial stance illustrates modal density and complexity within this moment.

Figure 3. Multimodal interaction in full

BY-NC-SA 2025, Editorial Universitat Politècnica de València



6:24: The instructor turns around, and repositions her avatar while asking, *"Is it cold here?"*. Her repositioning shows awareness of the spatial and embodied dimensions of communication in HiVR.

Figure 4. Instructor's embodied response

9:58: Both, Nick and the instructor, teleport to a new location. The spatial repositioning in HiVR is a mediated lower-level action that enables the higher-level action (Nick's participation in intercultural dialogue and navigation through virtual environments.

Figure 5. Spatial repositioning of Nick and his instructor to Hambug

2.2.3. Analysis of the HiVR learning episode following the Grammar of Transposition

Table 3. Multimodal transcript in HiVR interaction following Transpositional Grammar

Time	**Initial mode/ context**	**Transposition (shift)**	**New mode / context**	**Meaning-making outcome**
5:36	Gesture (body) (simulates flying, walking)	From gesture (body) ⇩ to speech	Verbal narration (helicopter, no trekking)	Embodied personal experience becomes a verbalized travel account
6:10	Verbal and embodied reference to Rantau	From speech/gesture (body) ⇩ to spatial movement (space)	Teleportation to virtual Rantau scene (space)	Reinforces personal memory through spatial immersion
9:58	Spoken explanation of home city	From speech ⇩	Guided virtual tour of Hamburg landmarks	Local cultural knowledge becomes

BY-NC-SA

		to spatial reference and object pointing (space / gesture)		spatially anchored in VR

This transcript illustrates the coordination of verbal, gestural, and spatial modes in the construction of intercultural meaning. The instructor and Nick co-navigate the virtual space while negotiating cultural knowledge, with gestures (e.g., pointing, simulating flight) and spatial transitions (e.g., teleportation) supporting language use and scaffolding comprehension. Built-in VR tools, such as the pointer, typing to see a real place, and Google Maps, facilitate rapid, embodied navigation and contextualized interaction, highlighting the unique affordances of HiVR environments.

3. Findings and Discussion

The multimodal transcripts revealed how learners in HiVR environments construct meaning using a dynamic combination of embodied, spatial, and linguistic resources. Across all three multimodal frameworks applied in this study, the orchestration of gesture, spatial repositioning, and speech played a central role in the co-construction of meaning within intercultural interaction. The intercultural dimension of this episode arises from the reciprocal sharing of culturally and personally meaningful locations: Nick described his experiences visiting Rantau Mountain in Nepal, and the instructor guided him to Hamburg, Germany, highlighting key landmarks. This exchange facilitated negotiation of meaning across cultural and linguistic boundaries.

In the selected episode, Nick, a learner with A1-level English proficiency, successfully participated in an intercultural exchange by drawing on the affordances of the HiVR environment. His embodied simulation of flying to Rantau mountain, combined with fragmented speech *("I went to Rantau at helicopter. No trekking"*), exemplifies a communicative moment of high modal density and complexity (Norris, 2004), where multiple semiotic modes operate in tight synchrony to convey meaning. These findings reflect what Mills et al. (2022) describe as the cognitive and communicative depth that emerges through embodied interaction in VR, where gesture, body orientation, and spatial positioning serve as meaning-making resources alongside language.

The immersive setting also supported intercultural understanding, a central goal of the original task. As shown in Jauregi-Ondarra et al. (2024), HiVR facilitates spontaneous, culturally situated communication in ways not possible through traditional tools. In this episode, Nick used both gesture and environmental reference to talk about culturally specific sites, while the instructor scaffolded the exchange by dynamically repositioning herself in space and pointing to elements in the virtual landscape, practices that reinforce the pedagogical value of spatial navigation in VR-based learning (Christoforou & Efthimiou, 2023).

Further insights emerged when comparing the frameworks. MDA allowed for an examination of modal orchestration at key moments, while MIA highlighted the temporal flow of actions, capturing how the higher-level intercultural exchange was supported by layered, embodied lower-level actions. The Grammar of Transposition revealed how Nick's personal narrative evolved through representational shifts (gesture, language, and spatial movement), demonstrating what Mills and Brown (2021) refer to as transmediation, or the re-articulation of meaning across modes.

From a methodological perspective, the study underscores that multimodal analysis in HiVR is both rich and time-consuming. Transcription alone required careful attention to gesture, speech, spatial positioning, and VR affordances. However, the depth of insight made visible through the frameworks justifies the analytical effort. Importantly, the choice of framework should align with the research aim, question, and unit of analysis. For example, studies interested in sequential interaction may benefit more from MIA, while those focusing on representational shifts across media may find the Grammar of Transposition more applicable.

BY-NC-SA

Overall, the findings highlight the potential of HiVR to support language learners, including those with limited proficiency, by enabling access to a broader repertoire of communicative resources. Multimodal analysis, when aligned with immersive affordances, provides a useful lens through which such interactions can be explored.

4. Conclusions

This paper contributes to the underexplored area of multimodal analysis in HiVR environments by offering a comparative perspective on three analytical frameworks applied to a single learning episode. By illustrating how meaning unfolds across gesture, space, and speech, the study provides an entry point for researchers interested in applying multimodal analysis in immersive language learning contexts.

Although the creation of a detailed multimodal transcript was time-intensive, it proved essential for uncovering the nuanced semiotic strategies used by the learner, particularly given his limited linguistic proficiency. The selected episode was intentionally chosen to demonstrate the communicative potential of HiVR for low-level learners and to foreground the methodological contribution of the analysis.

Future research could build on this approach by including multiple episodes, additional learner profiles, or by aligning framework selection more closely with specific learning goals. As HiVR environments gain momentum, especially in the post-COVID era, there is a growing need for robust, adaptable methodologies to analyze the multimodal richness of learner interaction. This study aims to serve as one such foundational tool, encouraging further experimentation and refinement in the field.

Acknowledgements

I would like to thank my colleagues, Tricia Thrasher (Immerse) and Amelia Ijiri (Kyoto Institute of Technology), for our collaboration in previous work based on this project.

References

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

Christoforou, M. (2025). Gen AI-assisted multimodal meaning Design: exercising a pedagogic metalanguage of transposition. Pedagogies: An International Journal, 1–25. https://doi.org/10.1080/1554480X.2025.2522884

Christoforou, M., & Efthimiou, F. (2023). Introducing Dreams of Dali in a tertiary education ESP course: Technological and pedagogical implementations. In P. Zaphiris & A. Ioannou (Eds.), *Learning and collaboration technologies. HCII 2023. Lecture notes in computer science* (Vol. 14041, pp. 53–65). Springer. https://doi.org/10.1007/978-3-031-34550-0_4

Christoforou, M., Xerou, E., & Papadima-Sophocleous, S. (2019). Integrating a virtual reality application to simulate situated learning experiences in a foreign language course. In F. Meunier, J. Van de Vyver, L. Bradley & S. Thouësny (Eds), *CALL and complexity – short papers from EUROCALL 2019* (pp. 82-87). Research-publishing.net. https://doi.org/10.14705/rpnet.2019.38.990

Cope, B., & Kalantzis, M. (2020). *Making sense: Reference, agency, and structure in a grammar of multimodal meaning*. Cambridge University Press. https://doi.org/10.1017/9781316459645

Jauregi-Ondarra, K., Meijerink, J. & Christoforou, M. (2024). Using high-immersion social virtual reality environments for researching interculturality. In K. Sadeghi (Ed.) *Routledge Handbook of Technological*

BY-NC-SA 2025, Editorial Universitat Politècnica de València

Advances in Researching Language Learning. Routledge (pp. 403-419). https://doi.org/10.4324/9781003459088-36

Kaplan-Rakowski, R., & Gruber, A. (2019). Low-immersion versus high-immersion virtual reality: Definitions, classification, and examples with a foreign language focus. In *Proceedings of the 12th International Conference Innovation in Language Learning* (pp. 552–555). Florence: Pixel

Kress, G. (2012). Multimodal discourse analysis. In J. P. Gee, & M. Handford (Eds.), The Routledge handbook of discourse analysis (pp. 35-50). Routledge. https://doi.org/l0.4324/9780203809068.ch3

Kress, G., & Van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication.* Arnold.

Mills, K. A., & Brown, A. (2021). Immersive virtual reality (VR) for digital media making: Transmediation is key. *Learning, Media and Technology, 47*(2), 179–200. https://doi.org/10.1080/17439884.2021.1952428

Mills, K. A., Scholes, L., & Brown, A. (2022). Virtual Reality and Embodiment in Multimodal Meaning Making. *Written Communication*, *39*(3), 335–369. https://doi.org/10.1177/07410883221083517

Norris, S. (2004a). *Analyzing multimodal interaction: A methodological framework*. Routledge.

Norris, S. (2004b). Multimodal discourse analysis: A conceptual framework. In P. LeVine & R. Scollon (Eds.), *Discourse and technology: Multimodal discourse analysis* (pp.101–115). Georgetown University Press.

Thrasher, T., Sadler, R., & Dooly, M. (2024a). Collecting 'real' data in virtual reality (VR) settings: Best practices. In K. Sadeghi (Ed.), *The Routledge handbook of technological advances in researching language learning* (pp. 27-39). Routledge. https://doi.org/10.4324/9781003459088-35

Thrasher, T., Christoforou, M. & Ijiri, A. (2024b). Virtual Reality-Mediated Language Learning: A Case Study of Immersive Learning in the Metaverse. In Vurdien, R. & Chambers, W. (Eds.). *Technology-Mediated Language Learning and Teaching*. IGI Global. https://doi.org/10.4018/979-8-3693-2687-9

BY-NC-SA 2025, Editorial Universitat Politècnica de València