frontiers | Frontiers in Communication

TYPE Original Research
PUBLISHED 09 April 2025
DOI 10.3389/fcomm.2025.1519165

OPEN ACCESS

EDITED BY
Monika Messner,
University of Innsbruck, Austria

REVIEWED BY
Stephan Packard,
University of Cologne, Germany
Alexandra Gubina,
Leibniz Institute for the German Language (IDS), Germany

*CORRESPONDENCE
Milica Lazovic
milica.d.lazovic@gmail.com

RECEIVED 29 October 2024
ACCEPTED 05 March 2025
PUBLISHED 09 April 2025

CITATION
Lazovic M (2025) Spatial resources in pre-service teachers' instructional practices in VR tandems: co-constructing shared spaces and embodied spatial scaffolding. *Front. Commun.* 10:1519165. doi: 10.3389/fcomm.2025.1519165

Spatial resources in pre-service teachers' instructional practices in VR tandems: co-constructing shared spaces and embodied spatial scaffolding

Milica Lazovic*

Institut for German Linguistics, Department for German as a Second and Foreign Language, Philipps-University Marburg, Marburg, Germany

This study examines the use of spatial resources in instructional practices during virtual reality (VR) tandem interactions between pre-service teachers of German as a foreign language and learners with A2 language proficiency. These interactions take place within the highly immersive virtual environment *Wander*, designed to facilitate (inter)cultural learning. The linguistic, perceptual, epistemological, and technical asymmetries within this setting necessitate scaffolding the co-participant through the virtual environment, guiding them in spatial exploration, orienting them to usability cues, leveraging spatial resources to support interactive and learning processes, developing embodied practices, and fostering mutual alignment. The analysis focuses on pre-service teachers' use of spatial resources and their practices of embodied spatial scaffolding to support learning in three key areas: instructing on app functionality, developing new embodied action patterns, and fostering a functional understanding of virtual objects. Prior to this, the study examines the instructional grounding for upcoming actions, such as directing instructions, by focusing on two key aspects: the co-construction of shared focus and the alignment of perspectives. This is achieved through eliciting, monitoring, and adjusting according to the position of the co-participant's avatar in situations of dynamic spatial perception. These situations are characterized by the interplay between changing position, orientation, and floating attention in the context of exploratory spatial navigation, the presence of distractors, or transitions. Methodologically, this study is grounded in conversation analysis and interactional linguistics. Video recordings capture participants' perspectives in a split-screen format, documenting parallel perspectivization in action flow and revealing shifts in interactional coordination. The results indicate, among other things, that participants prefer using pointing gestures accompanied by local adverbs, which are semantically subsequently extended, specified, varied, or reduced in a scaffolded way. Over time, synchronized co-referencing practices involving joint and matched pointing become central to negotiating a blended origo. The sequential analysis identifies meta-regulatory practices for perspective alignment by eliciting the other's perspective and monitoring the co-participants' avatar orientation and spatial relation to align for goal-directed action before co-constructing the focus. This study contributes to the understanding of immersive instruction in virtual learning environments by highlighting key aspects such as pre-instructional spatial self-and other-monitoring activities designed to support spatial self-alignment. Embodied spatial scaffolding involves some of the following supporting aspects: the adjustment of the internal spatial interface, the transition from static to dynamic interaction within the virtual environment, the management of spatial relations (explorative vs. concrete references), the control of spatial

interaction and coherence, the orientation to calls for alignment, the bridging of spatial transformations in the action flow, and the monitoring of the co-participants' avatar. The interactions tend to emphasize spatial engagement, with participants sometimes "overdoing" spatial elements rather than using spatial cues to develop more complex interactions.

KEYWORDS

co-constructing space, immersive instruction, virtual reality, learning tandem, foreign language learning

1 Introduction

Foreign language learning in virtual reality (VR) tandems is increasingly gaining presence and popularity, supported by a growing body of empirical research that highlights its numerous benefits (Ahlers et al., 2020, 2021; Steinbock et al., 2022; Plötner and Nowotny, 2023; Senkbeil, 2024; Lazovic, 2025). These benefits include not only educational games that enhance language acquisition and problem-solving skills but also the creation of immersive spaces for cooperative exploration, construction, discussion, negotiation, and creative-esthetic experiences. Additionally, VR enhances intercultural competence, fosters diversity, encourages critical self-reflection and empathy, and develops interaction skills while creating new socialization opportunities that help establish shared patterns of interpretation and action. From the perspective of teacher education, VR proves to be an ideal environment for training future teachers, promoting, among other things, the development of interactive adaptability, micro-scaffolding, and practices that establish intersubjectivity. Instructions in VR contexts, designed to bridge different spatial environments and fractured ecologies, are also being increasingly explored, with a focus on explaining words or providing instructions during the initial phases of immersion (Olbertz-Siitonen and Piirainen-Marsh, 2023; Spets, 2023). Due to the unreliability of directional cues typically relied upon in co-presence, perceptual differences, asymmetrical access to the shared space, challenges related to distributed agency and relative spatiality, as well as the high engagement required for the alignment of virtual and physical gestures, virtual environments require more explicit and effortful engagement to establish a shared focus and perspective on an object as a basis for further action. These are not merely practices of focus alignment but rather intensive mutual instructing actions regarding where to look and how to position or orient oneself to spatially align and co-construct the interactional space, elevating them to central micro-instructional practices of co-orienting, navigating, and establishing mutual spatial alignment in both static and dynamic positions. These foundational actions relate to more elaborate instructional steps that prepare learners for upcoming tasks. Pointing has become a key resource for identifying and co-constructing interaction spaces. However, how participants orient it to establish a shared focus, co-construct it, and align their perspectives during exploratory or dynamic navigation—while instructing each other in the use of certain features or explaining the functionality of new objects in the virtual environment—remains largely unexplored.

Given the importance of this topic and the existing research gap in the context of foreign language tandems, the following study aims to explore instructional practices in the high-immersion context of the app *Wander*, specifically in the context of German as a foreign language, with pronounced asymmetry in the linguistic competence of the co-participants, since the tandems consist of learners at the A2 language proficiency level and L1 speakers. The specificity of the interactional setting also arises from the tandem's goal of promoting culturally reflective learning through visits and exchanges within cultural hotspots, which are designed as 2D, as Google Street View, but come to life in a 3D sense during the interaction. Furthermore, the setting is specifically shaped by the novice status of the participants in the VR setting and the tandem's integration into a project-based seminar for prospective teachers, which aims to increase awareness of virtual reality as a resource for teaching and learning. The current study focuses on pre-service teachers' practices in instructing L2 learners in the use of an app function and the functionality of a virtual object and spatially scaffolding them within the process. Furthermore, it examines their instructional grounding, scaffolding character, sequential organization, the role of spatial resources involved in co-constructing a shared focus, the use of perspective alignment during dynamic spatial navigation, and the repair of misalignments and distractions. The interaction was recorded as a screen capture from both perspectives and analyzed using methods from interactional linguistics. The data reveal an interesting interplay between environmental accommodation, the dynamics of interactional co-adaptation, co-constructive practices of mutual alignment, and interactive micro-learning processes. The following section provides theoretical insights into research on shared spaces in real and virtual environments and their relevance for instructional practices. Subsequently, the study design, methodology, and findings are presented and discussed in three analytical sections, with the instructional grounding related to the co-construction of shared focus and perspective alignment serving as preparatory actions for upcoming activities, presented first, followed by more elaborate instructions as more complex instructional practices.

2 Co-constructing shared spaces in real and virtual environments

Space and its dimensions have long been central to language and cognition research, even before the "spatial turn" (Döring and Thielmann, 2008). The analysis of situated interactions across different spatial contexts has contributed to the understanding of space as a dynamic process of *doing space* (Jucker and Hausendorf, 2022; Jucker et al., 2018), where interactional partners engage with spatial affordances, adapt them to their communicative needs, and simultaneously co-construct and adjust spatial configurations through discourse. Space serves both as a resource that shapes

interaction and as an interactive and performative achievement (Jucker and Hausendorf, 2022; Mondada, 2013). This is particularly evident in virtual realities, where interactants not only apply spatial knowledge from their real-life experiences and replicate physical space but also produce a real sense of space by *constructing interactional space* (Jucker et al., 2018, p. 16). Mondada (2013, p. 250) describes interactional space as a dynamic, social, and embodied construct that emerges through the situated, mutually adjusted changing arrangements of participants' bodies within it. It is constituted by their mutual attention, shared focus, the objects they manipulate, and the ways they coordinate in joint action. Physical and interactional space can be expanded, according to Gibson (1979), by individual actional space (with spatial self-monitoring) and the imaginary space that is central to certain contexts, such as storytelling (Heller, 2022). Space as a mental co-construction is a multifaceted concept that encompasses interrelated and shared mental representations and interpretations of actions within a given context, enabling the maintenance of intersubjectivity, even in cases of underspecified references and fluid referential transitions (Ono and Tompson, 2024).

Different settings allow the use of various spaces, enabling participants to combine spatially focused and unfocused interactions based on their communicative needs (Jucker et al., 2018). Deleuze and Guattari's concept of assemblage expands our understanding of space by framing it as a situated activity—a lived production formed by an arrangement of heterogeneous elements that coalesce into a functional whole while remaining in constant flux and being dynamically reconfigured through their multiple interrelations and distributive agency (Due, 2023). These (re)configurations trigger shifts from implicit spatial awareness to an explicit focus on a unified field, leading to complex interrelations between linguistic means and other multimodal elements shaped by the dynamics of mental simulations, including *dual processing* (Hartmann and Hofer, 2022). The architecture for interaction can be differently pre-structured and contain navigation and usability cues (Hausendorf and Schmitt, 2022, p. 442) as resources that evoke specific interactionally relevant knowledge and social meanings related to certain social practices (e.g., institutional or culturally specific architectural semiotics, such as columns in courthouses that symbolize stability, justice, and authority) and invite specific participation as architecture-for-interaction. Despite their projective potential, interactants sometimes selectively focus on features that are salient and aligned with their spatial pre-knowledge; however, they predominantly orient to spatial flags that establish the communicative framework for interactions occurring in that setting or activate cultural scripts that indicate what kind of interaction or activity type is likely to take place (Jucker et al., 2018; Berger et al., 2016). This is especially relevant in the virtual world, where usability cues undergo certain technical and functional transformations, causing potential cue collisions. The analysis of virtual games (Olbertz-Siitonen et al., 2021) demonstrates not only that the material ecology structures forms of participation in specific ways but also how different forms of co-participation, with a constant oscillation between structured and fluid activities, provoke players to reposition interactively in multiple interactional spaces, involving shifting orientations to various spaces. They constantly reconfigure spaces, oscillating between being here and there, and co-create the event, whereby the co-presence is multimodally accomplished and carefully balanced (Olbertz-Siitonen et al., 2021, p. 115).

Co-constructing space in virtual environments is particularly complex due to the parallel nature of presence and media awareness (Hartmann and Hofer, 2022). This feeling of existing in two worlds or partially inhabiting one while forming spatial perception from experiences in both (Neuberger et al., 2024, p. 122) also contributes to cognitive distancing (Hartmann and Fox, 2021, p. 720). This awareness maintains a constant, sensory-related baseline or fluctuates in response to new spatial experiences (Neuberger et al., 2024, p. 113). This is captured in the concept of *hybrid presence* (Senkbeil, 2024), which refers to a sense of perception, interaction, and co-presence in multiple communicative environments simultaneously, in both virtual and physical environments, emphasizing the blending of experiences and communicative patterns from two (or more) different spaces (Senkbeil, 2024, p. 217). The degree of hybrid presence varies based on usage experience and immersive reflexivity (Senkbeil, 2024), fostering creative solutions and the development of spatial situational models for virtual environments (Wiepke et al., 2024, p. 236). These models enable participants to make predictions and adapt their actions accordingly. According to Meyer and Jucker (2022), immersive virtual environments are spatially multilayered, involving several overlapping spatial domains. These include the immediate physical surroundings, extended by interactions with technological devices, the two-dimensional display of written text or animated images, and the quasi-three-dimensional virtual space, which consists of quasi-physical worlds with varying degrees of similarity to reality and potential for manipulating the environment, where users act as avatars in a virtual embodiment. Additional spatial layers may expand these environments, offering further content or interactions in separate spaces. This leads to spatial expansion and overlaps, requiring reflected navigation, coordination, and alignment with co-participants. This is intensified by limitations in avatar visibility, using it as a resource, and employing compensatory tools to offset a reduced sense of presence. Participants in virtual settings share a 'world in common' but encounter a fragmented visual environment (fractured ecology, Luff et al., 2003), which limits the interpretation of each other's actions. This fragmentation disrupts shared focus on objects in the virtual world and diminishes the sense of being together in a shared space (Haddington and Oittinen, 2022, p. 346). This can lead to fractures in the shared spatial ecology, manifesting as a lack of perceptual and actional synchronization, varying in intensity and tolerance among interactants across different contexts. Explicit synchronizations and signals of joint attention are essential for specific activity types (Berger et al., 2016). However, in some situations, a more spatially independent or temporarily blurred reference is preferred to broaden the scope of action and stimulate interaction by overcoming challenges of spatial alignment.

Situational anchoring involves three fundamental aspects: co-orientation, co-ordination, and co-operation (Hausendorf and Schmitt, 2022, p. 436), established in a triadic relationship among a referring participant, the addressee(s) of this referring action, and an entity to which the referring participant directs the attention of the other(s) (Auer and Stukenbrock, 2022, p. 23). The spatial-orientational relationship between participants, mutual perception, and the sense of being perceived (Hausendorf and Schmitt, 2022), alongside the acknowledgement of the reciprocity and interchangeability of perspectives (Auer and Stukenbrock, 2022), serve as a starting point within the preliminary shared interactional space. The central frames for spatial reference include the deictic frame, the intrinsic frame (based

on objects relative to which the spatial reference is established), and the absolute frame (defined by constant spatial coordinates across acts of spatial reference) (Auer and Stukenbrock, 2022). According to Fricke (2003), the deictic frame differentiates not only a single origo linked to the speaker's role but also multiple origos that can shift to other entities as communicative roles change. The instantiation of local deixis may differ between verbal and gestural levels (ibid.). Origo is always an interactional accomplishment (Keevallik, 2013, p. 368), particularly for a moving body, where the referent of "*here*" is always transformed into something else immediately, building a shared origo. In virtual environments, participants navigate multiple origos and a blended origo (Senkbeil, 2024), referring to the cognitive phenomenon of integrating the experience of hybrid presence (with both spaces *really here*) in speech acts involving deictics directed at interlocutors who may or may not share the same situation. "*Here* "may ambiguously refer to different locations in real or virtual environments or to the shared perceptual space of one or more participants (Senkbeil, 2024, p. 221 f.), which participants must disambiguate with semantic attributions, explanations, and specifications (Senkbeil, 2024, p. 223), building their shared blended origo.

Spatial reference involves a complex interplay of verbal and multimodal means, developed in a sequential and temporally coordinated, co-constructive manner. Stukenbrock (2010), for example, illustrates how *multimodal gestalts* involved in the constitution of interactional space and the regulation of aligned orientation provide more than mere perceptual instruction or directional reference; they provide procedural cues for perceptual processing (holistic, focused) in relation to knowledge processing (e.g., integration, procedural knowledge) and as a projection of a certain type of activity (e.g., evaluation, correction). Similarly, Stukenbrock (2018) demonstrates how deixis accompanied by pointing gestures relates to gaze, directing attention to relevant phenomena within the shared perceptual space. Preparatory intrapersonal self-calibration practices (Stukenbrock, 2018, p. 149), such as spatial repositioning and shifts of gaze between the reference space and addressees, play an important role. Gaze emerges as a central resource for monitoring visual attention and inviting the addressee to engage their gaze in specific ways, thereby reorienting toward the shared object of attention and regulating turn-taking. Meta-perceptive gaze practices (Stukenbrock, 2020) support demonstrative references and the development of joint attention. The embodied resources typically relied upon in co-presence, such as gestures, gaze, and bodily orientation, become unreliable directional cues in virtual realities. Interactions are thus extended (Hindmarsh et al., 2006; Luff et al., 2003) and require adaptive negotiation of compensatory resources, such as non-deictic means of directional reference that function independently of the avatar-mediated origo (Auer and Stukenbrock, 2022, p. 53).

According to a study by Gillian et al. (2024, p. 90). virtual interactions are strongly characterized by comments on spatial experiences shaped by emotions, self-talk addressing spatial relations, joint actions aimed at understanding affordances and regulating environmental constraints, and explicit attention-gathering to deliberately coordinate and reinforce the situatedness of language while co-producing actions (Gillian et al., 2024, p. 100). Participants in virtual environments develop a strong awareness of constraints that reflect their partial perceptual access (Haddington and Oittinen, 2022, p. 351). Neuberger et al. (2024, p. 117) further illustrate that participants focus on their virtual positioning and spatial self-discipline, avoiding blocking others' views or engaging in excessive explorations. They become attuned to how their avatars approach one another by inferring orientation from each other's avatars and using relative references of objects for coordination (Meyer and Jucker, 2022). They include orientation to deictic pointing (Meyer and Jucker, 2022) or engaging in self-identification to confirm their mutual participation (Haddington and Oittinen, 2022, p. 352). Co-orientation relies on a reflexive relationship with the local setup, shaped and synchronized according to ongoing activities (Olbertz-Siitonen and Piirainen-Marsh, 2021, p. 2). Pointing serves as a central, reliable resource for locating objects and coordinating interactional spaces. Olbertz-Siitonen and Piirainen-Marsh (2021) demonstrate various functions of virtual pointing as a situating practice, showing how it establishes shared referents, highlights objects, relates, illustrates, and characterizes them, organizes interactional space, coordinates action, manages collaborative orientation, and serves as a turn-organizational device to mobilize responses and advance joint activities. Pointing duration is influenced by its coordination with speech, reflecting how the pointing gesture responds to structural or semantic constraints, adds salience, or serves other functional distinctions (Cooperrider et al., 2021, p. 11). The present study builds on these previous findings and aims to analyze how co-participants in the virtual world *Wander* co-construct shared focus, navigate, and use space in instructional sequences. Before doing so, a brief overview of the current state of research regarding instructional actions will be provided.

3 Space and its dynamics in instructional practices

Instructive actions aim to encourage others to perform an action, to guide them in their performance, or to achieve a goal through more complex sequences of multimodal actions (Ehmer et al., 2021). They are co-constructive, collaborative achievements with varying degrees of complexity and variability in multimodal configurations, their density, complexity, and intensity (Stoeckl and Messner, 2021), according to recipient design, situational contingencies, and interactional framework. Instructive actions can be distinguished according to several aspects, including the immediacy of the action to be performed, the synchronization between the instruction and the instructed action (whether process-bound or out-of-process), and the coordination load between participants or co-instructors, which depends on spatial arrangements and the balance between static and dynamic perspectives (Haddington et al., 2013). In addition, differences arise from the level of shared knowledge, its asymmetry, and the epistemic and deontic dynamics (knowledge as mediated, activated, or co-constructively developed) (Schmidt and Deppermann, 2021). The specific goal of the action—whether it is practical implementation, creative tasks, or joint action—also plays a role, as does the complexity and duration of the action (e.g., sustained activities). To foster engagement and collaboratively construct instructional actions, in some contexts, there are prefatory actions (Arnold, 2012) that emerge as dialogic paired embodied actions that act as pre-enactments, illustrating the potential of collaborative co-construction in space through selective matching and following a pattern, which shapes the specific function of instruction and pre-structures the ongoing activity. The term 'in-structing' itself implies creating, constructing, or assembling knowledge and points to the process of perceptual, action-related, and interactive structure-building as well as to the process of relating internally (within), from

an inner perspective as well as at the interface and in a shared space (in between). Related to this is the constructivist metaphor of scaffolding, which encompasses the dynamics of constructing and deconstructing supportive structures. Deppermann (2018), for example, demonstrates the influence of interactional histories on the turn design of instructions, indicating that the complexity and length of instructions decrease according to the perception of learners´ increasing competence and knowledge.

A basic distinction can be made between instruction as a pedagogical activity and navigational directing instruction (De Stefani, 2018), which focuses on managing (visual and cognitive) attention and involves acts of locating, directing, guiding, and controlling moving objects. Even in closed spaces, for example, in the context of a museum (Kesselheim, 2012; Pitsch, 2012), spatial navigation, focusing, and concentrating joint attention serve as a preparatory step in the co-construction of a shared space for movement, perception, and action (Pitsch, 2012, p. 226). A further distinction can be made between *quasi-instructions* (Tekin, 2021), which demonstrate monitoring and shared action awareness without directly guiding but fostering co-experience and togetherness; *assisting instructions* (Simone and Galatolo, 2020), which play a supportive role by updating others with important information in the process that may be inaccessible due to changing circumstances; and *framing instructions* (Schmidt and Deppermann, 2021), which provide broader orientation alongside affirmative selection and co-construction, enabling co-design of creative actions rather than direct instructions. The empirical studies of pedagogical instructions show a wide range of different analysis contexts in relation to gestures, body and movement in space, including studies of instructions in school context (Kupetz, 2021; Putzier, 2012), in driving lessons (Deppermann, 2017; De Stefani, 2018; Helmer and Reineke, 2021), in orchestra and theater rehearsals (Stoeckl and Messner, 2021; Schmidt and Deppermann, 2021; Krug et al., 2020), in cooperative activities like climbing, dancing (Simone and Galatolo, 2020; Keevallik, 2013), while visiting museums (Pitsch, 2012; Kesselheim, 2012), in riding lessons (Szczepek Reed, 2023), in bicycle-repair shop (Arnold, 2012), in Pilates (Keevallik, 2020; Ortner, 2023).

Spatial resources, embodied experience, and negotiated spaces play a central role in providing instructions. Spatial knowledge is activated in a supportive manner but is also conveyed as interactional knowledge that guides actions and is relevant to understanding an action, providing processing and structuring cues for new knowledge structures and their re-contextualisations. For example, in dancing classes, the proximal deictic *here* (Keevallik, 2013, p. 368) does not only anchor the speech, organize referents interactionally, and constitute the essence of the ongoing activity of teaching but also acts as a cross-linguistic means of linking multiple modalities in a way that enables students to identify the target activity or movement, to establish the figure/ground distinction of the referential action, and to provide cues for relevant parsing of the ongoing activity.

The co-constructing space in instruction facilitates the adoption of the (joint) actional perspective, cognitive alignment, and shared mental simulation. It also supports overcoming epistemic asymmetry and reducing cognitive overload. Joint attention, understood as the reciprocal regulation of attention within a subject-subject-object relatedness (Eilertsen, 2014), emerges as central to maintaining intersubjectivity. This appears particularly relevant in contexts of abstract, practical, and not easily verbalized knowledge, as well as in intercultural contexts or in the case of different linguistic backgrounds, where spatial arrangements, dynamics of co-constructing space, and spatially bound metaphors fulfill an important bridging function. Intercultural differences can influence, among others, the perception and activation of potential spatial cues, as their different conceptual preformations can influence the activation of specific scripts, interpretative and inferential processes, and the co-construction of interactionally relevant spaces. Kupetz (2021), for example, demonstrates in content and language-integrated learning how gestural pointing to or presenting an object and enacting a verb serve as spatial co-constructions for illustration and explication, for negotiating meaning and disambiguation, and for the co-construction of the learnable or as offers for better interpretability of the instruction. Changes in spatial arrangements serve to structure processes and restructure interactive contexts and roles, helping to overcome challenges in addressing multiple participants. Similarly, the emphasis is placed on portioning and installments, which ensure better accessibility and intersubjectivity (Kupetz, 2021, p. 364). Similarly, Putzier (2012) illustrates how the demonstration space in chemistry lessons is co-constructed gradually, highlighting the role of spatial scaffolding during instruction. This includes the formation of a shared perceptual focus through the alternation of static and dynamic positions, as well as how spatial organization guides perception and knowledge structuring. Modality synchronization, as a specific relationship of different resources in terms of redundancy marking and their variation, proves to be functional: their changing relation supports transitions in action phases during perspective shifts and maintains the continuity of focus and dynamic change, depending on the epistemic and interactive repositioning and the process of knowledge construction. Similarly, Helmer and Reineke (2021) used the example of driving lessons to illustrate how spatial resources are used differently, depending on perspective and practical knowledge-building dynamics (whether static, holistic, macroscopic or moving, actional, and involving multiplied perspectivisation). The structural complexity and the need to verbalize certain spatial relations vary, as does the function of gestures, which range from identifying and illustrating new aspects to supporting a macro-perspective, prompting actions, supporting procedural and adaptive processes, and integrating different perspectives within a practical action sequence. Instructions in virtual reality contexts are also being increasingly explored but primarily focused on explaining words (Spets, 2023) or providing instructions in the initial phases of immersion. In these early stages, expert users guide novices on how to use VR equipment, establish controller functionality, and navigate game mechanics to develop agency in the virtual space (Olbertz-Siitonen and Piirainen-Marsh, 2023). These instructions are designed to bridge different spatial environments and fractured ecologies, with variations in how verbal elements, bodily adjustments, pointing gestures, and movements are coordinated, as well as in the use of the avatar body as a resource.

4 Materials and methods

4.1 Participants, setting, and data collection

The following study is conducted as part of the MA course 'Digital Learning Environments for German as a Foreign Language (GFL).' The seminar concept involved prospective GFL teachers (S1)

(Figure 1A) in the second semester of a master's program and GFL learners aged 24 to 30 years. In virtual tandems, they explored the space *Wander* (Meta Quest 3), focusing on (inter)cultural learning in episodes that included visits to culturally relevant sites, joint cultural reflections, and free interactions based on shared virtual experience (Lazovic, 2025). The learners (S2) are GFL learners at the A2 language proficiency level, with Spanish and Serbian as their first languages. They have spent more than 2 years in Germany, mainly in international academic environments with English as a dominant language. As they learn German autonomously in real-life contexts and adopt a strong multilingual approach, they primarily overcome expression difficulties through code-switching and by learning while initiating explanations related to new words. Both groups were new to the virtual world and underwent a brief 20-min training session immediately beforehand, supported by expert facilitators. For S1s, this participation is an integral part of the learning process to understand the didactic potential of this learning environment, while for S2s, it is a one-time, voluntary participation for research purposes aimed at exploring new learning methods. As novices, their interactions are characterized by highly expressive actions related to embodiment and simplification, shaped by the demands of immersion and related cognitive load.[1] The app *Wander* makes it possible to visit selected places (indoor and outdoor) in high immersion together, to move synchronously, and to co-construct the spatial experience. Mutual following allows one person to take the lead, managing the app's functionality (such as selecting locations, navigating and moving through the space, and determining the spatial arrangement of the co-constructive activities); the other person coordinates their position, aligns their perspectives through rotations and minimal movements, integrates pointing cues, and assumes the leading role in the next turn. Google Street View, featuring 360° photos (Figure 2), serves as the basis, allowing views in any direction, sometimes supplemented with additional information windows (Figure 1B).

Although the images are 2D, the ability to navigate changing perspectives—by rotating, repositioning, moving linearly along paths, rotating around objects, and using additional windows to manipulate them—creates the illusion of a 3D space, fostering a strong sense of immersion and spatial depth. The 3D space is created by dynamically changing perspectives and is brought to life by the co-construction of a shared space. However, participants dynamically transition between 3D and 2D, and vice versa, throughout the interaction. Their interaction space emerges as an assemblage of multiple elements in 2D and/or 3D interactions with the environment, interacting and constantly evolving. However, the 2D Street View esthetic makes strong assumptions about the co-orientation of anyone viewing it, which is why users tend to rely heavily on it when interacting, partly due to the balancing of cognitive loads by orienting to the environment in a 2D-based way. The combination of a macro-perspective (map with navigation area) and a situated, experiential micro-perspective proves to be interactively stimulating, as does the fact that, despite their authenticity, many of the photos were taken more than 10 years ago, so that the perceived differences serve as an impulse for an exchange about past and present, reasons for change and social developments (Figure 1C).

The data presented here are drawn from analyses of three case studies involving three tandem pairs. Although the cases differ in the contextualization of culture-related learning processes and the number of episodes, with each centered around a visit to a new place, they share a consistent structure. After an introductory phase, the tandem pairs first explore German-speaking locations, with the L1 speaker taking the navigational lead. This is followed by an exploration of the learners' home countries' cultural landscapes. While the prospective teachers (S1) approach the process from a didactic perspective, primarily grounded in cognitive cultural studies, the learners (S2) adopt a more subjective and experience-oriented approach. Although it would be equally important to consider the embodied actions in the real world during virtual interaction, the following study focuses only on the perspectives and actions in the virtual world to capture the virtually visible embodiments, shared perspectivisation and mutual adaptations. The interaction was recorded as a screen capture from both perspectives, which is synchronized frame by frame in Adobe Premiere Pro and exported as a single split-screen video. Each split-screen video lasts between 30 and 35 min, for a total of 100 min. The conversation was transcribed using Exmaralda according to GAT2 (Selting et al., 2009) and expanded with screenshots of videos, with a split-screen view and information about the participant's perspective in relevant instructional positions. The data is analyzed using methods from conversation analysis and interactional linguistics (Auer et al., 2020; Couper-Kuhlen and Selting, 2018; Imo and Lanwer, 2019).

The focus is on pre-service teachers (S1) and their practices in instructing and guiding learners with low language proficiency through the virtual space in a scaffolding way. According to Hammond and Gibbons (2005) model, the term scaffolding is used here as micro or interactional scaffolding in learning contexts, with diagnostic competence, adaptivity, and responsiveness (Koole and Elbers, 2014) as its central characteristics. It refers to a set of various supportive actions (like repetition, adaptation, and eliciting) or interventions aimed at stimulating, guiding, and supporting specific learning processes, problem-solving, or goal attainment, particularly in the anticipated zone of proximal development (Vygotsky, 1978). As different multimodal resources are used, scaffolding supports are adaptively provided, gradually built up, and then removed, illustrating the dynamic and adaptive nature of the process. The relevance of adaptive scaffolding in this data context arises not only from linguistic asymmetries and the support of specific learning processes but also from the varying perceptual and action differences within the action space, the distribution of roles in navigation, and the epistemic differences associated with a specific space and the associated cultural learning processes, as well as transitioning from 2D to 3D interaction with the environment and vice versa. Due to the various challenges

1 This study does not directly examine participants' orientation to cognitive load; however, the term is used in alignment with Cognitive Load Theory (for a recent overview, see Sweller et al., 2019), as the amount of cognitive processing demands is dedicated to the execution of a task, which shows to be higher in immersive technologies due to a higher amount of sensory information, activity and distraction effects or cybersickness. Empirical evidence highlights its importance, particularly for novices in virtual environments (Elkin et al., 2024; Juliano et al., 2022; Breves and Stein, 2022; Armougum et al., 2019), which decreases with growing routine and proficiency in VR use, ultimately shaping reflective thinking and perceived learning effectiveness (Sari et al., 2023).

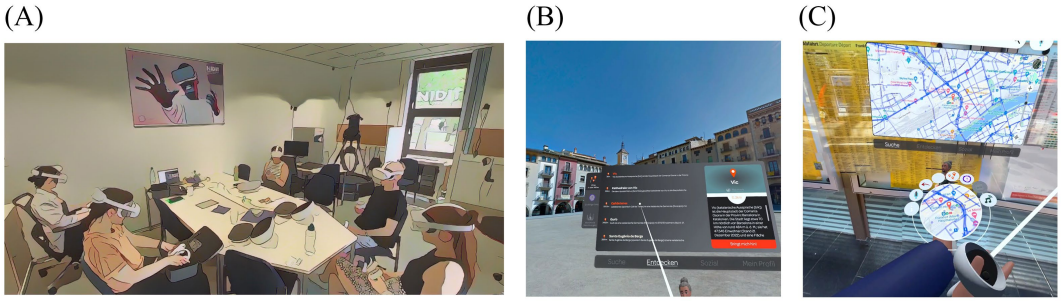


FIGURE 1
Representation of the research context, combining images from the real world and screenshots from the virtual environment.

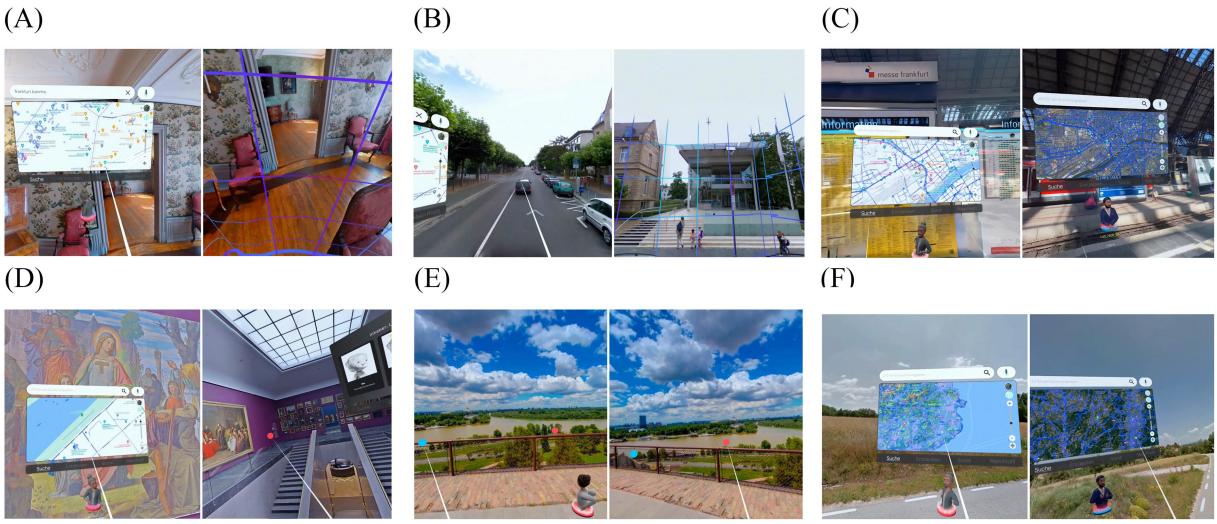


FIGURE 2
Screenshots illustrating the perspectives and differences between participants in the virtual world.

and asymmetries encountered when collaborating and solving problems related to fractures in virtual environments, I use this term to refer to all adaptive, supportive actions aimed at guiding or supporting collaboration and learning processes or mutual understanding as grounding for subsequent actions, specifically within instructional contexts.

The focus of the analysis is on the following instructive actions: First, elaborate instructions for the use of certain functions of the application (such as zooming) and the use of some objects noticed in the virtual world, with only four cases identified. This is the focus of the third analysis section, which illustrates two cases of S1's instructions as embodied spatial scaffolding to support S2's learning during instruction on how to use a specific app function and understand the function of a virtual object, thereby internalizing a new related concept and action pattern. Second, I focus on practices of instructive grounding for upcoming actions, which involve actions of establishing a joint or shared focus of attention and achieving alignment during spatial exploration and navigation. Due to the different positions, perspectives, and roles of participants in spatial management, these foundational activities play a crucial role in laying the groundwork for future actions, serving as directive instructions fundamental to other activities. Thirty cases of S1's instructive grounding actions provide a basis for analysis. These activities include, on the one hand, actions of co-construction of a shared focus related to the registering (Pillet-Shore, 2021) and, the establishment of a specific object or location as the shared focus of attention, referred to as focus alignment. This is addressed in the first part of the analysis. On the other hand situations of dynamic spatial perception and 3D interaction with the environment, characterized by the interplay between changing position, orientation, and floating attention (with some kind of movement of the avatar), without specific objects in focus, mostly in the context of exploratory spatial navigation, these directive instructions are different, which will be addressed in the second part of the analysis. I have adopted the term perspective alignment to describe alignment in dynamic spatial perception, exploration, and certain movement conditions. This refers to actions aimed at synchronizing, adjusting, and aligning with another person's perspective in situations involving dynamic spatial perception, shifting positions, orientations, and fluctuating attention, as well as in motion or due to relative spatial perception, where the relationship between objects in space is perceived from different vantage points (Steels and Loetsch, 2009). I also use this

term for situations in which S1s repair misalignments by dynamically changing orientation, pointing direction, and position or by adjusting their perspective, which involves dynamic movements and adjustments of the avatar's position to align with the other's perspective before they can establish joint attention and shared focus. Of particular importance is the distinction between the co-construction of a shared focus in static positions, related to salient and explicit objects or after object-focused interactions, and contexts in which orientations and positions change dynamically during spatial exploration, in cases of distractors or in transitional phases, with perspective alignment taking precedence over focus co-construction. Since these actions of instructional grounding serve as preparatory actions for upcoming activities, they are presented first in this study, followed by more elaborate instructions as more complex instructional practices.

4.2 Analytical preface

Before presenting and discussing the results of the analysis, it is important to consider some general observations about the spatial behavior of the participants. Virtual immersion creates a sense of spatial delimitation while simultaneously revealing fractures due to the division between physically tangible space, characterized by a stable ground, and the virtual world, which lacks a palpable ground. Technical limitations result in a restricted peripheral view and a distorted perception of proximity and distance. Spatial fractures become apparent in specific positions (Figure 2), particularly when participants move through defined areas in the real world. This movement can cause grids, visible as spatial boundaries in the virtual realm, necessitating continuous coordination between the statics of the real space and the dynamics of the other. This is also evident in the interaction with the technical device, when synchronizing real physical movements and activities of pointing or acting with specific outcomes in the virtual world. Users experience an internal spatial conflict caused by the illusion of movement through virtual space while being still, creating a challenging interplay between static and dynamic spatial relations. Internal spatial regulation refers to the alignment of the internal spatial interface, sometimes leading to spatial self-monitoring (Figure 2).

The sensation of floating, often accompanied by participants' humorous remarks about being unable to feel and see their own feet, results in a preference for stable objects that serve as spatial anchors, providing a solid point of orientation as a spatial anchor. Due to the novice's unfamiliarity with the environment, there is a preference for objects with absolute reference instead of the body and not to challenge the static nature of one's own physical anchors through action, leading to avoiding body-related action, such as references to turns or movements. They use salient and stable objects as spatial anchors (*in front/back of X*), navigate by using horizontal relations at a distance—usually to the right (*on the right side from*)—as well as upward perspectives based on objects with clear foundations (*upwards, on X*), while avoiding downward perspectives. Furthermore, up to 15% of all spoken words in recorded interaction are deictic expressions such as *that*, *there*, and *here*, combined with pointer references, utilized to establish object references and mitigate issues with local references through illustrative cues. Sometimes, overcoming the challenges of alignment leads to an excessive focus on negotiating objects and their relationships rather than facilitating further interaction, resulting in the overdoing of space as an overemphasis on spatial practices as central to interaction. The greatest challenge arises from uncertainty and managing partial (Figures 2A,C) or complete non-equivalence (Figures 2B,D) in perspective-taking (due to differences in angle, height, direction, rotation dynamics, and activity load when navigating and interacting), despite the same location and the resources for negotiating the focus (see the pointer in Figure 2D). Co-participants must actively, co-constructively solve this to ensure a shared or approximate perspective, which is, in the best case, not only shared in the mental domain but also mutually visible as such (Figure 2E). Activities for negotiating shared means of alignment (e.g., practices of pointing to objects, spatial flagging with different directions and forms of pointing, and circling) emerge, as well as supporting each other in spatial navigation (e.g., in the simultaneous use of navigation maps) (Figure 2F). The position of hovering over the world and the illusion of manipulating objects (e.g., by enlarging or marking them) create a sense of operating on a meta-level or a spatial-distance attitude as well, which is why real objects are also questioned or critically reflected upon (Lazovic, 2025). This is also influenced by the navigation maps and additional information windows that extend the virtual space into a meta-space level while simultaneously constraining spatial perception and establishing a certain degree of detachment (see Figures 2A,C,F). A significant challenge arises from the limited visibility of the other interactant, which affects perceptual awareness. This leads to an orientation toward the avatar (see Figures 2C,F) or trying to monitor his activity, typically within repair sequences, narrative, or phatic sequences (e.g., in small talk or expressive comments) as a means of communicating social presence and commitment and indicating expanded interactional space. In addition to turning toward the interaction partner or monitoring the relative spatial relations of the avatars, practices of spatial disengagement are also evident, such as lowering the head and looking slightly unfocused downward. This spatial disconnection occurs in relation to engagement load, for example, when S2 experiences language or formulation difficulties or when S1 intensely focuses on S2's utterances. In addition, increased proprioception and self-awareness can be observed, not only for better coordination and alignment of movements in the virtual world but also in the preparation of instructions in terms of co-simulation as a scaffolding for a specific action instruction. This is where the following study begins, looking at how S1 instructs S2, co-constructs a shared focus, and navigates toward aligned perspectives.

5 Results

5.1 Instructional grounding: co-constructing a shared focus

The *Wander* app allows simple joint movement through the virtual space, with one active person providing the technical scaffolding for further joint activities. However, different positions and orientations result in different foci, which have to be explicitly

co-constructed as shared focus if they are recognized as interactively relevant. The co-construction of shared focus is referred to here as instructional grounding and illustrated below with two typical practices (Examples 1, 2) and one challenging condition (Example 3). To begin with, awareness of fractured spatial perception promotes meta-regulatory strategies for co-constructing shared focus. This is evident at the beginning of interactions when participants jointly explore possibilities for interactive coordination using pointing gestures. The use of pointing is negotiated in simple contexts when focusing on objects that are already visually and semantically salient. This illustrates the following Example (1). After free exploration, focus is directed toward the prominent object. First (line 1), as a multimodal chunk consisting of 'pointer + prosodically emphasized local adverb (*here*)', followed by 'pointer + prosodically emphasized determiner (*this*) + noun' (line 2). While the first has more of a search command function, the second part is used for narrow focusing and is semantically specified. The pointer function precedes the linguistic realization slightly and runs parallel to it. However, due to technical constraints, some delays are noticeable for the other person, resulting in an asynchronous perception. Subsequently (lines 3–5), the functionality of the pointer is explicitly confirmed through an intersubjective meta-remark about shared perception, establishing it as a crucial interactional resource for mutual coordination. From this point on, the pointer is used as a situational anchor and navigation cue and to bring 3D space to life by co-constructing the shared space. Circular movements of the pointer serve as attention gatherers, indicating not only the focus but also its further thematization or use as a directional cue, while vertical and horizontal movements are used for visualization, emphasis, and disambiguation. However, due to the adaptation of the movement to the shape of the object, it was not possible to identify a shared systematicity in the forms of movements of the pointers, but some negotiations regarding the progressiveness of the focused objects and the nature of the co-orientation.

Moreover, pointing progressively develops into joint, synchronized pointing as aligned co-referencing, as illustrated in the following Example (2). Here, the tandem visits a place in S2's hometown, where S1, visiting this place for the first time, uses a distant, recognizable monument as a cue to elicit culturally relevant information and open a new interactional space. In contrast, S2 focuses more on the park as a usability cue. This difference in the orientation and activation of usability cues suggests a significant influence on individual or intercultural preferences. The directing of attention to the shared object involves several steps, starting from the W-question fragment (*what is*) as an attention-getter (line 2) to the unspecific W-question (*what is this*, line 3) as a re-start, with a pause after a prolonged *eine* as a search procedure for S2, orienting S2 to the use of a pointer and "waiting" until S2 finds the referent (cf. Goodwin, 1980), after which the semantically specific NP (*statue*) is realized in a syntactically and prosodically exposed way, as a narrow focus. The same pattern, 'pointer (with several vertical movements, depicting the object form) + unspecific question', is realized again (lines 4–5) as an established pattern to flag and locate the object, with a slight adjustment, shifting from a monument to a person-related reference (*who that is*). Laughter serves to bridge and smooth over the delicate moment of referential mismatch, facilitating the interaction and prolonging S2's search for the intended object. This sequential organization gives S2 the time and cues for alignment and opens the space for negotiation and transition from 2D to 3D experience. The circular movements indicate its topic function for the following sequence. Interestingly, S2 uses the pointer as a co-constructive

1 S1 genau. und HIER, **(a)**
Alright. And here

2 ist noch mal so: !DIE!ses **(b)** EUROzeichn;
is again this euro sign.

3 i:ch GLAUbe, du kannst **(c)** SEHen, (.)
I think you can see

4 wenn ich dir was ZEIge;
when I show you something.

5 (kreisende Bewegungen mit dem Pointer)
(circular movements with the pointer)

6 jA, (.) [ja:;]
Yes, yes.

7 S2 [jA_jA:;]
Yes, yes.

(a)

(b)

(c)

EXAMPLE 1
This sign (case 1, min. 2).

1 S1 jA, (rotiert sich nach rechts) (-) geNAU.
Yes. (rotates to the right) Exactly.

2 (-) und **(a)** WAS ist (.)
And what is,

3 was_ist DAS für eine::, (-) **(b)** !STATUE!,
what is this statue?

4 weißt DU: das, **(c)** wer DAS ist, (lacht)
Do you know who that is?(laughs)

5 (macht vertikale Bewegungen mit dem Pointer)
(vertical moves with the pointer)

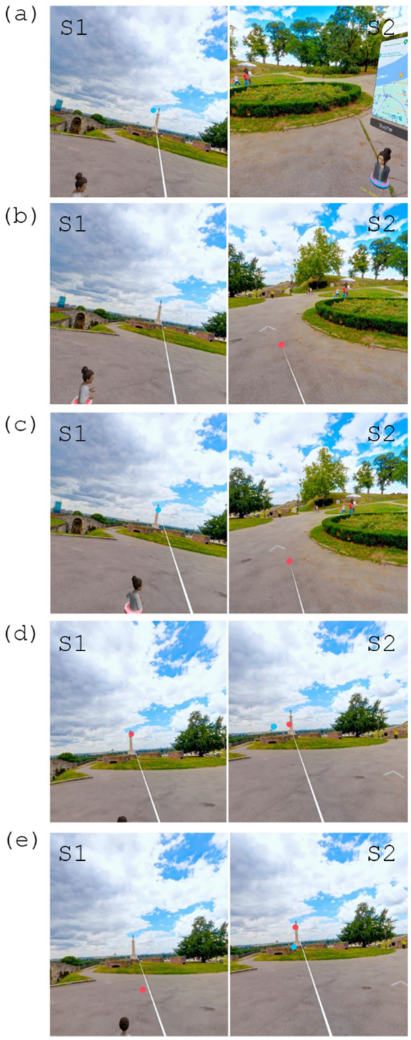
6 S2 (rotiert nach links) **(d)** !DIE!se?
(rotates to the left) This one?

7 (kreisende Bewegungen mit dem Pointer)
(circular movements with the pointer)

8 S1 jA, geNAU, **(e)**
Yes, exactly.

9 S2 jA, äh (-) it´s the (f) ähm (.) mhm, (-)
Yes uhm this is uhm mhm.

10 it´s the WINNER; (lacht)
This is the winner. (laughs)



EXAMPLE 2
The winner (case 2, min. 23).

resource despite having spatial and experiential knowledge of this location and lexical clarity. After the two pointers meet in the shared search field (lines 6–8), S2 confirms the recognized pointer and ratifies it as a shared focus. The focus is not fully established until they spatially converge in the virtual world with shared anchoring cues in the practice of joint, matched pointing, meeting through reference points, and generating a shared, blended origo with aligned, visible co-referencing. Vivid pointing gestures contour and animate the virtual environment, enlivening the frozen image, compensating for the lack of physical space, and creating a shared sense of material co-embodiment. This helps participants overcome the discrepancy between static and moving perspectives in real and virtual environments. Joint, synchronized pointing gestures serve as a central mechanism for co-positioning within the virtual space, enhancing immersion.

Finally, the following Example (3) illustrates some co-orientation problems when using verbs based on bodily movement without precise direction and spatial reference, but also due to the nature of the focus (marketplace) or when transitioning from 2D to 3D interaction with the environment. At the beginning of the sequence (line 1), both participants share the perceptual orientation by looking in the same direction; however, when the directing instruction includes the verb *turn around* (line 2) without any additional spatial cues, it leads to a rupture in the shared perspective and necessitates a reparative alignment. Instead of turning right toward the traditional marketplace as the target object, S2 turns left (b, d, e), preventing S1's situating pointer from serving as an orientation reference. Furthermore, an ambiguous search cue (*do you see here,* line 3) is accompanied by a prominently emphasized spatial descriptive adjective (*large*) to specify the focus. S2 orients to this spatial cue (line 4) instead of the nominal cue with a semantically unspecified and ambiguous reference (*place*). S2 turns left to identify a potential object, orienting to the spatial reference with the adjective, but expresses (line 6) his uncertainty about the actual focus (asking, *A big what?*). S2 initiates a repair by repeating part of the other's initial turn and using the question word !*WAS!*, which refers to the trouble source (Jefferson, 1972; Schegloff et al., 1977: 368; see also Kendrick, 2015). S2's perceptual focus (d, e) suggests that he anticipates a large house in front of him as a potential focus. However, even after S1 repeats the nominal anchor (*place*) in line 7 and provides pointer cues, S2's search

1 S1 u::nd, geNAU. **(a)** (-) ich wollte dich FRAGen,
And exactly. I wanted to ask you,

2 =wenn du_dich EINmal, (.) !UM!drehst- **(b)**,
When you once turn around,

3 siehst du HIER einen **(c)** !GRO:SS!en platz,
do you see a large place (square) here?

4 S2 (1s) einen GRO:SSEN?(dreht sich nach links, sucht) ja,
A large, (turns left, srching) yes.

5 S1 einen großen platz; **(d)**
A large place?

6 S2 ne_ne_ne:, einen GROSSen, (.) !WAS!? **(e)**
No, no, no a large, what?

7 S1 einen PLATZ, also ein
A place (square), so a

8 (kreisende Bewegungen mit dem Pointer)
(circular movements with the pointer)

9 S2 (2s) ähm (sucht weiter, rotiert sich nach links) **(f)**
Uhm (continues searching, rotates to the left)

10 S1 ein großer HO:F, **(g)**
A large square.

11 (macht kreisende Bewegungen mit dem Pointer)
(makes circular movements with the pointer)

12 S2 jA:, jA, jA, mit einer KIRche,
Yes, yes, yes. With a church.

13 S1 geNAU; weißt_du was !DAS! ist, diese:r PLATZ,
Exactly. Do you know what this is, this place?

EXAMPLE 3
Turn around (case 3, min. 5).

direction changes very little. Only after S2 decides to completely change his direction and turn toward a previously less utilized search direction (lines 9–10) does a significant shift occur.

In response to S2's search actions filled with processing indicators, S1 attempts to clarify with a synonymous expression (line 10), specifying, disambiguating, and narrowing the focus (*square*). After fully recognizing S1's circular pointer movements as situational anchors (line 11), S2 continues to negotiate the focus by offering a specification of the shared reference (line 12) with another prominent object (*a church*), relating it to a spatially clearly defined, salient object. Following this visually and semantically specified co-construction, the focus is established as a topic. This example highlights not only the challenges of co-construction due to misalignment as a result of the embodied activity of turning but also the use of adjectival cues for disambiguation (indicating spatial categories) when locating broad areas (places, markets) without reference to objects with clear spatial boundaries. In addition to semantic variation, spatial specification through relational references to other objects, along with alignment in dynamic movement (navigating through space during transitioning from 2D to 3D interaction with the environment), plays an important

scaffolding role, which will be explored in the next chapter on perspective alignment.

In summary, it can be said that immersive awareness, as demonstrated in Example 1, leads to the explicit negotiation and establishment of meta-regulatory practices for co-constructing shared focus, which are then reused as reliable patterns for co-construction. There is a preference for pointer gestures combined with prosodically emphasized local adverbs (which are less demanding in processing), which are subsequently semantically expanded, specified, or varied due to balancing cognitive load for semantic processing, their disambiguating function, and the transition from 2D to 3D interaction with the environment. As shown in Example 2, a distinct pattern can be identified, beginning with a pointing pre-invocation that is successively expanded verbally in a scaffolding manner. This starts with a simple W-question fragment as an attention-getter, followed by an unspecific W-question as a search procedure, after which the semantically complete noun phrase is realized in an emphasized manner, functioning as a narrow focusing procedure and further specified when relevant. In terms of descending scaffolds and reducing interactional asymmetry, the same pattern - 'pointer (indicating the object and its shape) + unspecific question' - is repeated, and sequential organization is important to provide the time and cues for alignment. The other person responds with a negotiation offer, which is mostly realized gesturally as a pointer (+ local adverb) or extended with further semantic specifications and explicitly co-constructed. Participants create a joint, matched pointing referencing a blended origo, using vivid gestures to contour and animate virtual space, enliven static images, and transition to 3D interaction with the environment, fostering a sense of co-embodiment and managing the shared origo-blending. This interactional achievement of locating each other through joint pointing and overcoming the feeling of being lost in space can sometimes lead to "overdoing" spatial engagement. Participants may become overly focused on frequent negotiations of shared focus, hopping from one potential object of reference to another in an impressionistic manner without delving deeply into conversation or expanding these references as usability cues. This is due to the reduced cognitive load associated with maintaining 2D interaction with the environment. Although experience-based, spatial prior knowledge provides a prestructured ground; the priority given to the interactively negotiated space becomes evident. Different orientations can be observed when orienting to possible usability cues and activating them for interactional purposes, likely influenced by both individual and intercultural factors. Objects with perceptual salience, without distractors, with clear spatial contours, or conceptually related are more easily negotiated and used as focal points than open spaces. References based on physical activities, movements of the avatar, rotations, initiating transitioning to 3D interaction with the environment, or adjectives that activate spatial relationality or depth tend to be problematic if not related to other objects or disambiguated.

5.2 Navigating through space: scaffolding and aligning perspectives

The following focuses on instructional grounding for upcoming actions in situations of dynamic spatial perception and 3D interaction with the environment, characterized by the interplay between changing position, orientation, and floating attention, occurring without or prior to focusing on specific objects in focus. The analysis illustrates this in the context of exploratory spatial navigation (Example 1), where S1 scaffolds S2's explorative co-navigation and aligns accordingly to establish a narrow focus, and in sequences aimed at repairing perspective misalignments (Example 2), or when S1 adaptively aligns orientation and direction to address distracting cues for S2 (Example 3). Perspective alignment here refers to actions aimed at synchronizing with another person's perspective in situations of dynamic spatial perception or increased interaction with the environment, due to mismatches in position and orientation as well as due to differences in relative spatial perception, where the relationship between objects in space as perceived from different perspectives causes misalignments or problems with focus formation or due to potential distractors. The first example (Example 4) illustrates how S1, in a location previously known to both, initiates broad spatial exploration and uses S2's perspectivization to scaffold toward a specific narrow focus. This fosters engagement in an exploratory stance and guides S2 to a more fine-grained, focused attention on relevant objects and relations that are relevant to upcoming actions, but it also shows to be very challenging. The initial elicitation has an unfocused, exploratory character (line 1), which is then expanded with prompts that activate existing spatial knowledge by using a familiar object as a situational anchor (line 2). It includes further a procedural prompt to encourage active rotation to overcome the static, frozen view (line 3, *turn around*) and intensify the 3D interaction with the environment. This is accompanied by a synchronized rotation activity that points to possible objects, is used as a perception trigger, and is used as a pre-invoice, which are still not recognized by S2 (b). There is a perceptual fracture, and S2 rotates to his own positively evaluated situational anchor (*park*), which serves as a starting point for alignment (line 4). This is followed by S1's perspective alignment and prompt for further perceptual exploration (line 5) and S2's perceptual descriptions (line 6), in which he identifies the new direction (*right side*), general location, and functionality (*for the tram*). S1 aligns accordingly (e) and uses this anchor to visually scaffold there and negotiate a possible topic (line 8). By initially focusing on perception-related aspects (*see*), with the referential focus realized as a pronominal adverb, S1 implies shared focus. This is subsequently (line 9) semantically specified to align with S2's previous statement and accompanied by pointing cues.

However, S2 does not recognize this as maintaining focus but reorients to the right (f, h). The use of an ambiguous pronominal adverb, along with addressing the spatial depth, relational spatiality (*behind it*) and non-object reference, leads to a short-term loss of focus as S2 turns to the right, changes perspective and searches for more objects (f), anticipating a topical shift. This is partially restored (line 9, g) during S1's semantic clarification and establishment of the reference to his last mentioned anchor, but the dimension of spatial depth (*behind*) and reference to a non-object appear to be problematic. Since S2 does not react, S1 disambiguates (lines 10–11) and reformulates the initial perception-related question in a dislocated topic with reference-explication (*this big building*) followed by knowledge-enquiry and not perception-related questions. The rightward rotation of S2 here reveals a renewed uncertainty about the actual focus (h) and potential topical shift. The relation *behind it* is – due to different positions, rotations, and perspectives – misunderstood. Instead of near focus, S2

1 S1 so:, was kannst du denn noch HIER **(a)** (-) entDECKen,
So, what else can you discover here?

2 fällt dir IRGENDwas auf,=was du noch KENNST;
Can you recognize anything that you already know?

3 und dich einmal so: **(b)** UMdrehst,
and turn around like this?

4 S2 **(c)** ahm jA, diese ähm dieses PARK ist schön,
(beide lachen)
But yes this park is nice (both laugh)

5 S1 und WAS noch? (lacht) **(d)**
And what else (laughs) yes ähm

6 S2 ja:: ähm in der rechte **(e)** ähm seite:,
On the right um side

7 es gibt die: ähm platz für_die STRASSENbahn (lacht)
there is the um space for the tram (laughs)

8 S1 mhm, und WAS siehst_du denn daHINTER? **(f)**
And what do you see behind it?

9 HINTER den straßenbahn (.) GLEIsen, **(g)**
Behind the tram tracks?

10 dieses GROSSE **(h)** gebäude;
There is this large building.

11 weißt du,=was DA:S ist? **(i)**
Do you know what it is?

12 S2 (--) !A! jA, das_ist die (.) oper,=vielleicht?
Ah, yes, that's opera. Maybe?

13 S1 JA, genau. das ist die OPER.
Yes, exactly. That's the opera.

14 und ähm genau hier !RECHTS!,
And right here on the right

15 da steht es ein bisschen DRÜBER-
It is written a little above it.

16 S2 das_ist die OPER,
That is the opera,

EXAMPLE 4
Can you recognize anything (case 1, min. 3).

anticipates a distant focus on the preferred right side, which also shows to be a *big building*, even higher than the intended one. The problem arises due to the virtual ambiguity of spatial cues (regarding spatial depth), but it is probably also interculturally relevant, with conceptual differences in the interpretation of *big*. The recognition of S1's pointer (line 11, i) serves as an alignment cue and restores the shared focus. After the focus, related to a specific object, has been established (lines 12–13), co-orientation based on horizontal orientations (*left/right*) and spatial "abowness" is unproblematic.

Even if S1 orients to S2's perspective, uses it as a spatial anchor, and aligns accordingly, the transition from aligning perspectives in flow to maintaining sustained shared focus and building the topic from it appears problematic. The scaffolding from a broad spatial navigation mode in flow (horizontally shifting, exploring with dynamic perceptual movement) to focusing on a spatially narrow aspect, building on a location, maintaining the focus and using it as a usability cue for expanding interaction is challenging, as is the implicitness, the ambiguous marking at the transition point, and the semantic reduction after briefly establishing the shared focus.

Further examples document S1's perspective alignments by monitoring and orienting to the position of S2's avatar and adjusting their own perspective accordingly to resolve focus mismatches. This practice develops gradually due to adaptation to the new environment, becoming a recognized resource for overcoming spatial fractures. This is illustrated firstly in Example 5: S2 refers in a semantically underspecified way, pointing to the new districts of her hometown in response to S1's question about the changes in the city and new districts (lines 1–4). However, S2's supportive, illustrative pointers are not visible to S1 due to their different positions, resulting in a rupture in the search space. S1 shows not only intensive search activity but also frequent rotations toward the avatar of S2 (cf. a, c, f, i) with subsequent attempts to negotiate a shared focus with pointing references (b, d, e, g) in relation to alignment with the avatar of S2. Relational spatiality appears to be used during these avatar observations and subsequent aligning actions, involving the continuous adaptation and approximation of perspectives with the orientation and position of the avatar. Turns to S2's avatar always occur after ambiguous determinatives and demonstratives, while the turns to the potentially shared focus, accompanied by the pointer reference, occur in positions parallel to the realization of nominal anchors, where pointer cues appear to be expected. S1 attempts to align with the orientation, direction or perspective of S2's avatars in ambiguous local references and to support the shared focus through the proper timing of pointer references in relation to nominal anchors. This aligning occurs as a kind of preparation for joint pointer actions. After several attempts (b, d, e) to implicitly negotiate the shared focus, S1 continues after the brief confirmation (line 5) and W-question (*which one*), with the offer of an anticipated focus (*here at the back*) (line 6), accompanied by the pointer reference and several circling movements, which S2 confirms (line 8). Here, the focus appears to be shared (h), but it is not yet certain since the pointer references of both are not explicitly visible to both. S1 then turns back to S2's avatar and uses it for a final adjustment of the perspective (i), after which the focus with both pointer references becomes clearly visible (j, k), and this is explicitly confirmed as shared for both (lines 9–12), as a joint activity of synchronized matching of pointing gestures. This negotiation is afterwards commented on as a recognized point or sharing of focus (lines 13–14), after which further questions for this topic follow. Meta-level comments point to the awareness of virtual co-presence, the need for explicit negotiation, visible, embodied alignment, and doing space as interactional achievement. Matching pointing in a joint action is a manifestation of blended origo and shared embodiment, resulting from monitoring the avatar's position, orientation, and perspective alignments based on it. These moments are accompanied by synchronized laughter and confirm the interactive value of the co-construction of the shared space. Monitoring the position and orientation of the other's avatar and aligning accordingly proves to be particularly relevant in situations without tight object focus, in transitional phases with dynamic movements, and before the formation of a new focus.

Example (6) illustrates another challenge, requiring reparative perspective alignment. In a transitional sequence, S1 guides the interaction toward a narrative based on personal experience but without a clear reference to the object. Meanwhile, S2 engages in exploratory perception, where a perceptual trigger causes disruption and highlights perceptual differences due to different positions and orientations (a, b, c). S2 interrupts the flow by addressing a break in his perceptual expectation and initiates a side sequence by focusing on an unknown object (pneumatic tube). S2 faces lexical difficulties in naming this object and uses semantically unspecific words (*these, things*), which makes the reference unclear (line 3). In this position, S1 needs to recognize the disconnection, adopt S2's perspective, adapt to the new situational anchor, and provide a context-specific, instructive explanation of the unknown object and its functionality. S1 automatically rotates toward S2's avatar (b) and aligns by reorienting (c) and pointing to a potentially disruptive trigger. Vertical pointing movements are used as cues to signal and invite alignment (line 4), isolating the object and creating an interactive space for negotiation. However, after initiating object clarification, S2 turns to S1 (d), indicating uncertainty and a possible turn-taking moment, which leads S2 to miss S1's gestural negotiation cues. By repeating the question and returning to the target object (lines 5–6), S2's vertical pointing movements and S1's horizontal pointing gestures become mutually recognizable. The co-referencing follows a standard scaffolding pattern: initially as a simple 'pointer-reference' (c, d), then as 'pointer and naming the object' (line 7, f) and finally syntactically integrated into an explicit question (line 8). A new scaffolding moment becomes evident as the pointer gestures are used in increasingly complex and dense forms to depict the object. Initially, the pointer reference, accompanied by vertical movements, is used to negotiate the shared focus by directing attention to the object (c, d), functioning as flagging in a one-dimensional manner. Subsequently, the pointer's direction shifts horizontally (e), serving as an exploratory visualization and a means of indicating spatial relations in a two-dimensional manner. This transition introduces a new dimensional understanding of the functional object. The object is then verbalized, combined with the demonstrative pointer reference, which depicts the object holistically and supports the semantical-conceptual composition in the sense of space enclosure, space deepening and enhancing spatial tangibility. When integrated into the question, this holistic pointer reference is made again, but also later (lines 10–11), the same holistic-depicting pointer composition is used as a ground for labeling the object to support the integration of the new lexical element into the spatial–visual concept. Gestural pointing supports new concept formation by facilitating the transition from registerings (Pillet-Shore, 2021) to the spatial understanding of situated objects. This process occurs through the successive addition of spatial dimensions in an instructive manner, moving from a vertical, here salient dimension (selection and isolation) to a horizontal one (spatial expansion and further specification). This progression enables participants to grasp the object more comprehensively, ultimately leading to a holistic understanding and three-dimensional conceptualization of the object—its complex spatial composition, spatial depth, and spatial relations. The

1 S2 also::, DIE:se hier, **(a)**
(macht kreisende Bewegungen mit dem Pointer)
Well these here (makes circular movements with the pointer)

2 (1s) DIE sind- **(b)** (1s) **(c)** NICHT zu alt;
(1s) they are (1s) not too old,

3 weil (-) da: **(d)** diese neue gebäude sind- **(e)**
because there these new buildings are

4 (-) FÜNF jahre alt- (.) oder so:,
five years old, or something.

5 S1 OKAY, **(f)** äh w_w WELCHE,
Okay, uhm w w which one

6 also **(g)** HIER hinten,
Here behind,

7 die (horizontale Bewegung nach rechts),
this one (moves the pointer on the right)?

8 S2 jA:: (--) **(h)** diese, (--) äh **(i)** (lacht)
Yes. These uhm (laughs)

9 S1 (lacht) !AH!, **(j)** DA:, [okay;]
(laughs) Ah, there, okay.

10 S2 [ja,]
Yes.

11 S1 [hier, **(k)**]
Here.

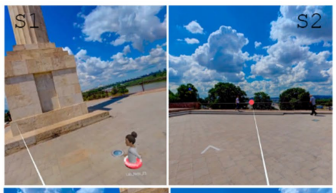
12 S2 [ja, ja,] (lacht)
Yes, yes (laughs)

13 S1 ich sehe [den punkt , ja]
I see the point, yes.

14 S2 [(lacht)]
(laughs)

15 S1 (lacht) okay; was äh (.) was wurde DA gebaut,
(laughs) Okay. What was built there

16 weißt du DAS,
Do you know that?

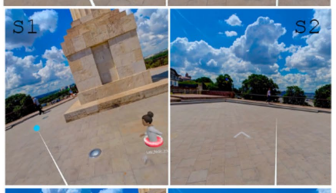
(a) 

(b) 

(c) 

(d) 

(e) 

(f) 

(g) 

(h) 

(k) 

(j) 

(i) 

EXAMPLE 5
These new buildings (case 2, min 25).

1 S1 genau, weil ich finde es **(a)**
Exactly, because I think it is

2 [ich_war EINmal]
I was once

3 S2 [was sind] **(b)** DIE:se::, (--) **(c)** (-)
What are these?

4 S1 jA:, **(d)** (macht vertikale Bewegung mit dem Pointer)
Yes? (makes vertical movement with the pointer)

5 S2 was sind diese **(e)** (-) dinge (--) ja,
What are these things, yes?

6 (macht mehrere vertikale Bewegungen mit dem Pointer, währenddessen S1 horizontale Bewegungen macht)
(makes several vertical movements with the pointer, while S1 makes horizontal movements)

7 S1 diese rohre, **(f)**
(führt mit dem Pointer drüber, um das Ganze zu zeigen)
These pipes? (S1 moves the pointer over it to show the whole object)

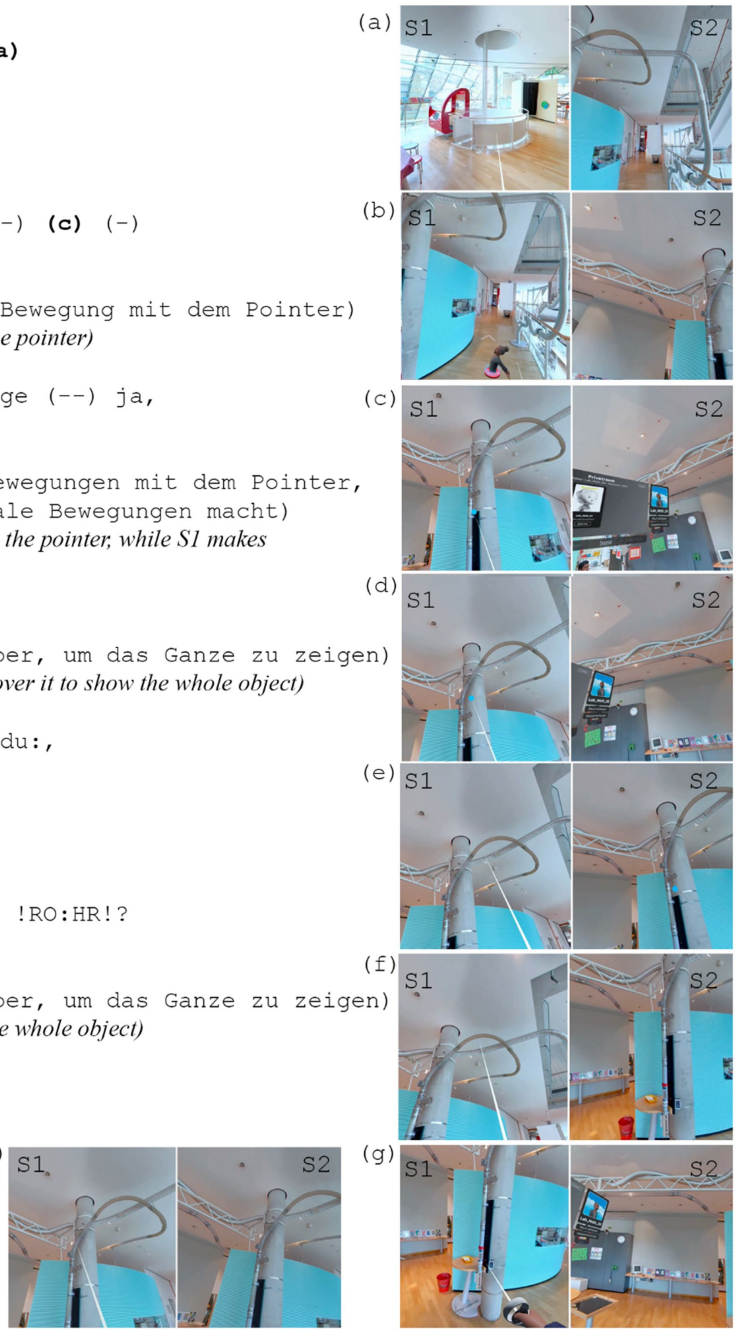
8 diese:: ROHre, **(g)** meinst du:,
Do you mean these pipes?

9 S2 wie heißt DAS, **(h)**
What is it called

10 S1 das ist äh (-) ähm (j) ein !RO:HR!?
This is uhm uhm a pipe?

11 (führt mit dem Pointer drüber, um das Ganze zu zeigen)
(S1 moves the pointer over it to show the whole object)

12 S2 ah, OKAY,
Ah, okay.



EXAMPLE 6
This pipe (case 1, min. 12).

continuation of this sequence, in which S1 further explains and demonstrates the object's functionality, is analyzed in the next section.

In conclusion, this section documented, on the one hand, practices of perspective alignment: first, while scaffolding the exploratory spatial navigation of the co-participant (Example 1), and second, by monitoring and aligning to the other's avatar perspective to repair mismatches and distractors (Example 2, 3). On the other hand, two practices of scaffolding through space were illustrated: one while broadly exploring through the space (Example 1), and the other while developing a new conceptual and spatial understanding of a new, here distracting, object (Example 3). The joint actions of synchronized, coordinated pointing (gesture matching), as an instantiation of blended origo and shared co-embodiment, demonstrated their significance in fostering a sense of co-presence and co-action, as well as spatial vivacity, to overcome static views and balance static-dynamic disparities. These actions serve as important disambiguation cues in dynamic constellations and represent key interactional achievements. The analysis of perspective alignments, prior to focus formation, showed first the elicitation of the other person's perspective during navigational landscaping and the use of adaptations to align with the other's perspective, to control distractors, but first to scaffold goal-oriented and navigate in a specific direction, and to develop mutual practices of adaptive co-orientation.

To minimize the costs and challenges of co-constructing a new focus—arising from differences in positions, orientations, pointing directions, or different orientations to usability cues—S1 elicits the other's perspective, aligns with it, and smoothly transitions to a new topic through scaffolding toward a goal. This reduces asymmetry, which is critical in linguistically diverse settings while increasing the experiential dimension, involvement, and agency and upgrading the epistemic and interactive position of the co-participant by giving perceptual priority to the other, expanding spatial triggers and controlling distractors. Example 1 shows, however, that even in situations of adaptive orientation, the transition from dynamic perspectives in flow to sustained shared focus and topic building appears to be problematic. This is due to the transition from exploratory perception with dynamic registerings (Pillet-Shore, 2021) in motion (focusing on horizontal perspective with intensive rotations) to a more focused perception of specific locations and object relations, loaded with spatial depth, relational spatiality and perception in the 3D sense, and ambiguously indicated (non-object) references through adjectives, pronominal adverbs, or deictics. Second, perspective alignment, based on monitoring the other person's avatar position, orientation, and relationship to other objects, is used as a resource for disambiguation, mismatch repair, and alignment, which is important for upcoming actions.

While scaffolding during broad spatial exploration without a specific object in focus proved more challenging due to the complexity of spatial relations and dimensions, Example 3 illustrates the use of verbal and pointer resources as scaffolding tools when introducing new concepts related to a specific object. It shows how understanding of spatial relationships evolves by gradually adding dimensions, moving from one salient dimension to another for spatial expansion and further specification. This process gradually supports a holistic understanding of the object as a three-dimensional concept, along with its complex composition and situational embedding while facilitating the mental integration of the new lexical element into the spatial–visual concept. As pointer activities become increasingly dynamic, complex, and dense, linguistic resources also expand, involving the scaffolding dynamics of constructing and deconstructing verbal cues with increasing semantic specification, syntactic complexity, and illocutionary explicitness.

5.3 Spatial resources in elaborate instructions

The following chapter shows two cases of elaborated instructions: first, on how to use the zoom function, and second, an instructive explanation about the function of an object. The first example (Example 7) begins with S1's uncertainty about the visibility of the Zoom option for the co-participant and the use of this option, which is relevant for mutual alignment. As this issue arises, S1 attempts to explain this function and instructs S2 to use it on her own. S1 starts by eliciting S2's perspective and gaining insight into her perceptual field, which is shown to be very general and static (line 1). Before S1 adopts an actional, simulating perspective in instruction, there is an observational reference to his own body (lines 2, b), functioning as a form of monitoring bodily interaction with the technical device while engaging in spatial self-alignment within the spatial interface. This occurs before assuming an action perspective, which involves anticipating S2's actions and inviting alignment. The physical experience of touching and feeling (lines 2–6, b-f) and, simultaneously, virtually looking at the joystick and its conduct is used in a leading position and as a simulation entry.

S1 adopts a generic you-perspective (lines 2–3) to define the starting point by naming the object with a possessive determiner as part of their own body, which serves as an embodiment call. However, this perspective shifts to a shared, actional perspective to trigger alignment (lines 4, 5). S1 initially focuses on the entire device (*right controller*) and then, from the actional, situated perspective (*there*), indicates the position of the button (*upwards*) as part of a device (d), which enables S2's synchronization. S1 first provides instructions for horizontal orientation (*right*) and then for the vertical dimension (*above/up*), which facilitates the interface with technical devices and regulates internal origo-blending. The enacted embodied experience is dynamic and visible through bodily self-monitoring with clearly indicative iconic movements of pressing upwards with the right hand. After adjusting the internal spatial interface, there is a procedural call for rotation (line 5), accompanied by exemplifying activities (line 6). This serves as a bridge from the static position of the internal interface to interaction within the virtual environment for immersive navigation and control of spatial interaction. As a novice in virtual space, potentially facing issues with spatial orientation and locomotion anxiety, S2 acknowledges but does not accompany this. However, it continues to be engaged within the interface with the technical device, maintains its spatial anchor, and does not rotate. In the next instructional step (lines 8–9), S1 encourages exploratory interaction with the virtual environment to support the proceduralisation of embodied practice between two worlds and sensing this spatial experience, firstly without any specific goal. This function is demonstrated through pointer cueing but without concrete spatial reference (*anywhere*).

To facilitate and achieve synchronization, the consequence of this action is explicitly elicited (lines 10–11; *do you see something in large*). The interactional space is expanded here through an invitation for S2 to transition from the output position of the action, presupposing that S2 is aligning and capable of adopting this embodied practice as a learnable and manageable action. S2 engages in an intensive rotational activity, which is accompanied by expressive comments while attempting (but still unsuccessfully) to test the new function (line 12). Meanwhile, S1 demonstrates and simulates the function several times, although it remains invisible to S2. In the next part of the instruction, S1 repeats the invitation for exploratory attempts at pointing without concrete object references (lines 13–14) but then specifies this with a reference to a textual object to establish a shared focus (line 15). The practice is first conveyed without a shared or narrow-focused object reference, as felt embodied action, forming a causal relation of two actions in hybrid space (tactile-embodied (pressing) + perceiving/focusing (pointing) = > enlarging) as a procedural embodied pattern, illustrated several times and then with a more concrete reference in its specific functionality. This highlights the balancing load when performing an action regarding the concreteness of the spatial reference, the dynamics of spatial interaction with the environment, and the immersion load. Although S2 initially fails to follow the action and struggles with the routinisation of this hybrid act, she appears to have recognized its functionality, allowing her to orient herself more freely and adaptively while developing a more spatially dynamic stance within the virtual environment. After several attempts, S2 successfully executes the

1 S2 äh:: (-) **(a)** stehen wi:r (.) i::m RAUM, ja,
Uhm we just stand in the room, yes.

2 S1 okay; **(b)** ähm (-)bei DEInem ja,
Okay. Uhm on your

3 bei_deinem **(c)** RECHTen controller,
On your right controller

4 DA (.) ist so ein joystick **(d)** OBEN,
(drückt und bewegt den Ponter nach oben),
There is a joystick above it (presses and moves the pointer upwards)

5 mit dem man_sich so DREHEN kann;**(e)**
Which is used to turn around like this

6 (drück mehrmals,rotiert sich nach rechts), ne? **(f)**
(presses several times, rotates to the right), right?

7 S2 mhm,
Mhm.

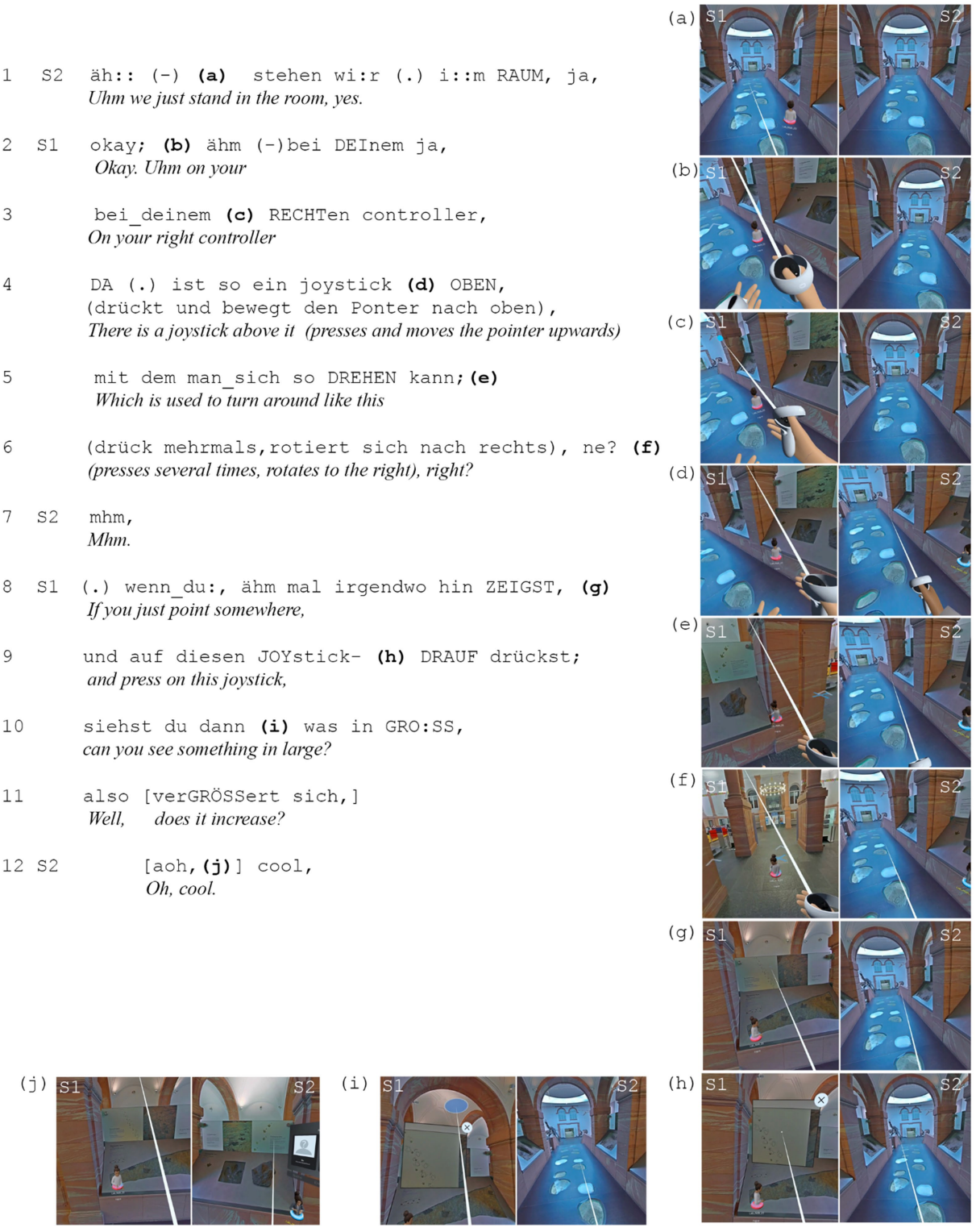
8 S1 (.) wenn_du:, ähm mal irgendwo hin ZEIGST, **(g)**
If you just point somewhere,

9 und auf diesen JOYstick- **(h)** DRAUF drückst;
and press on this joystick,

10 siehst du dann **(i)** was in GRO:SS,
can you see something in large?

11 also [verGRÖSSert sich,]
Well, does it increase?

12 S2 [aoh,**(j)**] cool,
Oh, cool.



EXAMPLE 7 (CONTINUED)

practice a few seconds later (line 21, p, q). However, due to the rapid thematic progression, this achievement holds no further interactive relevance beyond S2's expressive expressions and newly gained insights.

The next Example (8) presents an instructive explanation of the functionality of a previously discussed unfamiliar object for S2 (a pneumatic tube in the Museum of Communication). Instead of providing a simple explanation, S1 simulates an agentive, procedural perspective. The starting point for S1 is a shared focus and its localization (lines 1–5): First, the object and its distinctive part are identified using pointing gestures and adverbials (*here in the front*), suggesting a shared perspective and marked with circular movements. This is accompanied by an expression of epistemic uncertainty and a call for alignment (lines 2–3), inviting co-construction within the

13 S1 ja? **(k)** das heißt, (rotiert sich im Suchraum)
Yes, it means (makes rotations)

14 du kannst äh IRGENDwo (.) HINschauen, **(l)**
you can look anywhere,

15 manchmal auch auf den TEXT, **(m)**
Sometimes on the text

16 und dann kannst du dir das **(n)**
And, than you can have it shown to you

17 SO groß ZEIGen lassen; (lächelt)
This like in a large format. (laughs)

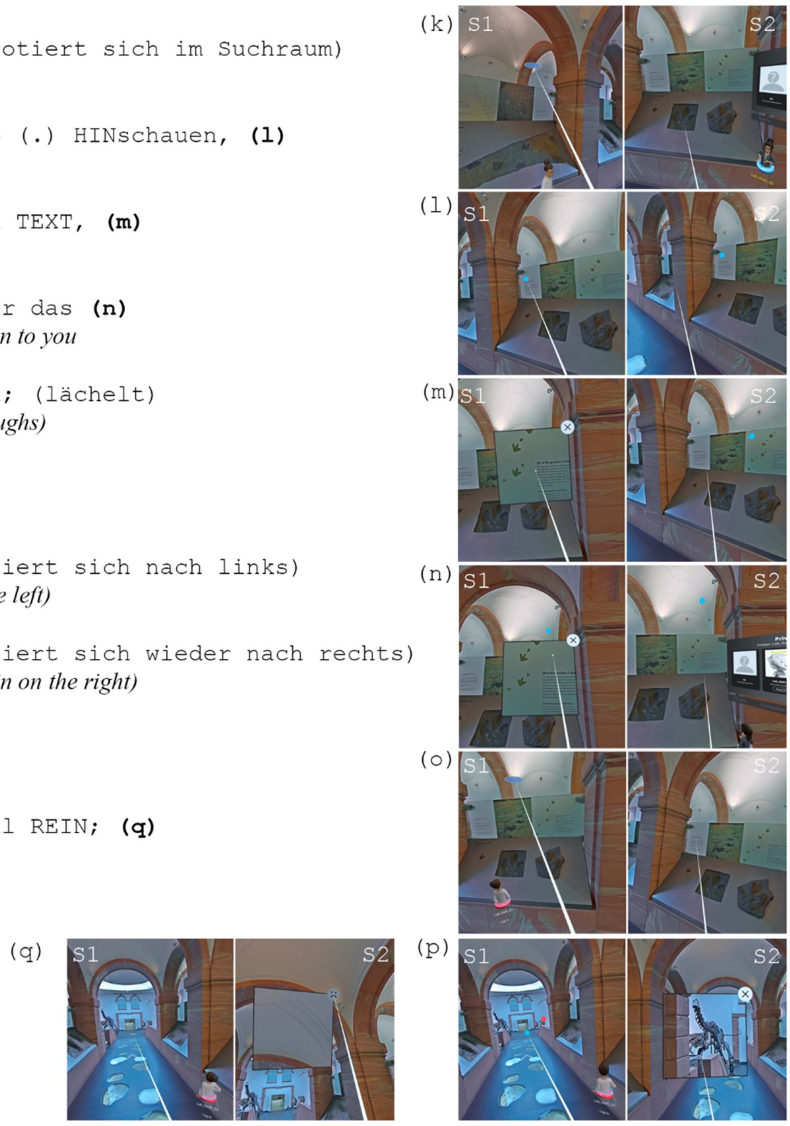
18 S2 cool, **(o)**
Cool.

19 S1 (lächelt) geNAU. (rotiert sich nach links)
(laughs) Exactly. (rotates on the left)

20 (lacht) jA, jA; (rotiert sich wieder nach rechts)
(laughs) yes, yes; (rotates again on the right)

21 S2 oh, aha! **(p)**
Oh, aha!

22 S1 SO:, wir können ja mal REIN; **(q)**
So we can go inside.



EXAMPLE 7
Use the controller (case 2, min. 8).

shared (actional) space. S2 aligns accordingly, recognizing the vague reference (*there*) specified by the pointing gesture. S1 uses this as grounding for disconnecting it from the goal/object functionality of the previous sequence and to establish a new spatial frame, where the focused object is becoming an instrument for action. This decomposition and spatial redefinition facilitate a focus on actions from an agentive perspective, where the spatial condition is changing, and the focused object transitions into an instrument for action. Visual indexing depicting its dynamic form in spatial flow (c, d, e) is used as a bridge to support the action in flow, particularly the verb *sending*, by using the object as an instrument (line 7). This action is systematically illustrated in its spatial flow in an emphasized manner, consistently accompanied by dynamically instructive gestures (c-f) that depict and illustrate multiple spatial relations at all stages of the action, upon entry, as flow around it, and at the exit. This dynamic visualization, featuring multiple rotations, directional changes, and turns, enhances the feeling of embodiment and participation as an acting subject, promotes an understanding of spatial relations in the action flow, and structures the explanatory sequence. The final repetition of the action with the object's functionality (line 12) returns to the starting point, marking the closing. S1 animates the space to enhance the understanding of actions in their progression from a processual perspective, highlighting both the object's functionality as an instrument and the co-construction of a relational understanding *in situ*. This includes transitioning from focus specification/spatial narrowing to transformative interventions to redefine the spatial frame, spatial fragmentation, and spatial synthesis in the action flow.

In the pre-closing part, when S2 attempts to connect the instruction to existing knowledge while coherently relating it to the other spatial structures surrounding them, this example illustrates another interesting phenomenon. S2 follows up with a humorous remark (lines 13, 16), suggesting that this post might have originated from the trash, associating it with another object visibly

1 S1 und DA: kann man- SO wie es aussieht; (a)
And there you can, it seems so

2 HIER vorne, **(a)** ich_weiß nicht,(kreisende Bewegungen)
here in the front, I don´t know (ciscular moves)

3 ob du DAS siehst,
if you can see that

4 ähm (-) !DA! **(b)** kann_man (.) ne POST reinstecken, (b)
uhm there you can put the post, a pipe

5 eine !RÖ:HRE!,
A pipe.

6 S2 ah, (c)
ah

7 S1 und die **(c)** SCHIESST man dann hier einmal,
and then you shoot it once

8 (macht Bewegung mit dem Zeiger der Röhre entlang)
(makes movements with the pointer along the pipe)

9 DRUM **(d)** herum, (d)
around it

10 ich weiß nicht, **(e)** wo sie RAUSkommt-
I don´t know, where it comes at the end

11 **(f)** (kreisende Bewegung) (e)
(circular moves)

12 ABE:R, **(g)** man kann damit POST verschicken; (lacht)
But, you can send post with it. (laughs)

13 S2 (lacht) aus dem **(h)** !MÜLL! (lacht) (f)
(laughs) from the waste (laughs)

[oder SO:, (lacht)] **(i)**
or something like that (laughs)

14 S1 [(lacht)]
(laughs)

15 ich weiß nicht- ob sie in den müll gehen; **(j)** (lacht) (g)
I don´t know, if it goes to the waste (laughs)

16 S2 (--) (lacht) aus dem ALTpapier mÜll; **(k)** (lacht)
(laughs) from the waste for used paper;

17 S1 genau, (lacht) (h)
Exactly. (laughs)

18 vielLEICHT gibts einen im anderen STOCKwerk nach unten;
Maybe there is one in another floor downstairs.

EXAMPLE 8
You can put the post (case, min. 13).

positioned at the periphery of the focused space (the *little red bin*). This is used as a potential usability cue to connect it with the interpretation framework of waste sorting, a popular theme in the GFL-didactic context. S2 turns to the S1 avatar (h, k), enlivening the interactive space in a phatic manner while commenting, as well as to the nearby object (i, j), which he attempts to conceptually link to this context and coherently interpret spatial relations. Since S1 has not recognized this object at the periphery and S2 is not explicitly addressing its position, S1 tries to conceptually adjust this (comment as a creative but unrealistic) spatial understanding (with the possibility of sending it to the trash) or to propose an alternative interpretation involving a lower floor (line 18). S1 is co-constructing the adjusted spatial understanding and scaffolding the language-specific use of the preposition in relation to actions presented previously. However, this example highlights the importance of spatial periphery and its distracting cues, as well as spatial coherence and intercultural differences in the use and integration of different spatial cues, and co-construction of spatial interrelations based on functionality.

In summary, virtual instructions involve pre-instructional spatial self-and other-monitoring activities that support internal spatial adjustment, such as self-alignment within spatial interfaces and blended origo management. Embodied spatial self-monitoring serves as an entry point for adopting and simulating the perspective of the situated other, facilitating the generation of a shared blended origo. This foundation provides a basis for mutual synchronization, which is expanded through *alignment calls* as practices embedded in key instructional positions (preparing and accompanying actions) that deliberately focus on specific action steps within particular spatial relations, which also vary in their degree of explicitness, prompting, and directiveness. When instructing on device use, there is a successive increase in the concreteness of spatial references and the dynamics of spatial interaction with the environment. This progression moves from adjusting the internal spatial interface, bridging from a static position to more dynamic interaction within the virtual environment. Initially explorative, it broadens and senses the space in different directions without specific object references to experience the spatial dynamics and proceduralize embodied patterns. It then transitions to more concrete references, focusing on specific functionalities to gradually develop spatial awareness, control spatial interaction, and immerse in the spatial flow. Instructions on the functionality of virtual objects tend to animate space and adopt an agentic, procedural, participatory perspective. After constructing a focus, the spatial frame is first deconstructed and redefined, suggesting changes in the frame of action and spatial relationships, which are then reformulated or recomposed into a new understanding of spatial relations and action flows. In addition, spatial referencing through the use of a pointer (flagging, depicting, indicating relations) not only illustrates and represents separate phases of an action cycle and their dynamics but also serves as a mental bridge for complex meanings and actions. This instructive explanation includes spatial narrowing, transformative interventions to redefine the spatial frame, spatial fragmentation, and spatial synthesis and recomposition to create dynamic spatial relations in action flow. Spatial periphery and coherence are particularly relevant due to intercultural differences in the use and integration of different spatial cues and their connection to various interpretative frameworks.

6 Discussion

This study, based on video-recorded tandem interactions between prospective teachers and GFL learners in the virtual world of the app *Wander*, examines the instructional practices of pre-service teachers and their use of spatial resources in elaborated instructions and instructional grounding. Embedded in partially familiar environments, the analysis shows that users prioritize interactively negotiated space despite their experience-based spatial knowledge. As novices in a virtual environment, participants tend to prefer salient objects and orient themselves to the immersive environment in a 2D manner. They also exhibit a tendency to avoid body-related actions, such as references to turns, movements, or dynamic interactions with the environment. Navigation typically favors horizontal relationships positioned to the right and upward, based on objects with clear foundations and contours while avoiding downward perspectives. The 2D esthetics provide a stable foundation for co-orientation, which is why participants tend to rely on it extensively during interactions instead of engaging in highly immersive interaction in a 3D sense. This preference also stems from the need to reduce cognitive load, which is lower in 2D mode than in 3D. The 3D space is still created through dynamic perspective shifts, navigation, or app features with spatial extension. It comes to life through the co-construction of the shared space, with participants constantly shifting between 3D and 2D interactions with the environment. Their interaction space evolves into an assemblage of multiple elements shaped by changing 2D and/or 3D engagement with the environment. However, L2 speakers experience difficulties transitioning from their preferred static 2D perspectivization to dynamic 3D perspectives using spatial relationality, depth, and navigation transitions. This is why L1 speakers adaptively scaffold and support transitions from 2D to 3D, as shown in analytical parts 2 and 3. The data revealed increased proprioception and self-awareness (spatial self-monitoring) related to alignment within the internal spatial interface and co-simulation, which serves as preparation for instructions or scaffolding actions. In addition, there are practices of spatial disengagement related to language-related difficulties arising from intense focus on the activities of the co-participant. The spatial extension in the virtual realm also facilitates the development of a sense of operating on a meta-level or adopting a spatial distance attitude, which is shown to be important for deepening perceptual and reflective activities (Lazovic, 2025). Due to the specific interplay of language competence, individual and intercultural factors, different spatial orientations and the activation of usability cues can be observed. These include how participants manage distracting cues, use the spatial periphery, connect cues, and establish spatial coherence.

The first part of the analysis confirmed findings from other studies in the field of virtual reality (Neuberger et al., 2024) and showed that immersive awareness leads to meta-regulatory practices for mutual co-orientation early at the beginning of the interaction, which are subsequently employed as reliable practices for focus-negotiation. Due to competence asymmetries and general cognitive load in the virtual environment for novices, there is a preference for

pointer gestures in combination with prosodically emphasized local adverbs such as *"here"* and *"there"* (which are less demanding to process). These are then successively semantically expanded, specified, or varied due to the balancing of cognitive load and its disambiguating function. A typically emerging pattern for co-orientation begins with a pointing pre-invocation, which is successively expanded in a scaffolding sense (as installments), first as a simple W-question fragment serving as an attention-getter (a *call for alignment*), followed by a non-specific W-question (a *call for search*) to focus on the object. After that, the semantically complete noun phrase is realized in an emphasized manner as a spatial anchor (*narrow focusing*), being further specified if relevant. In terms of descending scaffolds, reducing interactional asymmetry and potential tension by providing search time for the co-participant and alignment, the same pattern - 'pointer (indicating the object's and its shape) + unspecific question' - is repeated. The other person ratifies but follows in a negotiating manner, gesturing with a pointer (+ local adverb) or extending with further semantic specifications. Throughout the interaction, synchronized and coordinated co-referencing practices, such as joint and matched pointing, are increasingly employed as negotiating practices. Vivid gestures animate the space and help overcome static perspectives in a 2D sense, create a shared blended origo, and foster a sense of co-embodiment and co-action in 3D.

In some cases, these activities, as joint transitions to higher immersion, become the primary focus of the interaction. By 'overdoing' spatial engagement, participants shift impressionistically from one potential reference object to another without developing these references further. In this sense of 'overdoing space,' the co-location of objects and negotiation of their spatial relations in an immersive sense become a goal without leveraging these usability cues for more advanced interactions.

The second part of the analysis, focusing on dynamic spatial exploration with higher immersion, revealed two practices that are increasingly used during interactions. These practices indicate a specific interplay among environmental accommodation, interactional co-adaptation, and interactive learning processes: the elicitation of the other person's perspective and corresponding alignment for goal-directed navigation, as well as scaffolding spatial interaction while managing perceptual distractors and unfocused exploration. The second practice includes monitoring the other person's avatar position, orientation, and relation to surrounding objects as resources for disambiguation and alignment. The orientation toward the avatar is typically conveyed within reparative, narrative, or phatic sequences to communicate social presence and commitment, indicate the expansion of interactional space, and extend spatial engagement in a 3D sense. In cases of conceptual or cognitive distractions and misalignment, scaffolding is required to support new conceptual construction through spatially dynamic instructions. This approach involves gradually developing an understanding of spatial relations by sequentially adding dimensions and transitioning from one salient aspect to another for spatial expansion. This process fosters a holistic, three-dimensional understanding of an object and its complex composition for situational embedding and action while facilitating the mental integration of the new lexical element into the spatial–visual concept.

The third part of the analysis on virtual instructions highlights pre-instructional spatial self-and other-monitoring activities designed to support internal self-alignment and blended origo in the spatial interface. Embodied spatial self-monitoring provides the basis for adopting and simulating the other's perspective. This is further reinforced by *calls for alignment* embedded within key instructional phases, which deliberately focus on specific action steps in relation to spatial configurations. During instruction, there is a gradual increase in the concreteness of spatial references and the dynamics of spatial interaction with the environment, representing a smooth transition from 2D to greater immersion in a 3D sense. This progression moves from aligning the internal spatial interface, bridging its static position to a more dynamic interaction within the environment, initially exploring, expanding, and sensing spatial immersion during rotation activities without specific object references to experience spatial immersion and proceduralize embodied patterns. It then moves to more concrete references, focusing on specific functionalities, to gradually develop spatial awareness and control of one's own spatial interaction in the action flow. In instructions on the use of virtual objects, users tend to animate space and adopt an agentive, procedural, and participatory perspective. The spatial frame of an object is first focused in 2D, then spatially redefined to create an instructional basis that defines the object's specific functionality in 3D. This process involves transforming spatial relations from an action-oriented perspective, adding spatial depth and relationality, and recomposing the space within the new action framework. This instructive explanation includes spatial fragmentation and focusing in 2D, transformative interventions to redefine the spatial frame in 3D, and spatial synthesis and recomposition, all aimed at creating dynamic spatial relationships embodied in the flow of action. As spatial markers, the pointers not only illustrate, flag, and represent spatial relationships and their dynamics in action flow, but they also have a priming and bridging function for complex meanings or actions, supporting the transition to higher immersion.

First, when comparing the activities of L1 and L2 speakers in interactions involving unfamiliar objects (Examples 2, 6), the use of the same resources (pauses, emphasis, structures, and the use of pointers and transitional indicators) can be observed. However, there are some differences: the orientation to objects in 2D, the timing of pointer usage (earlier and as a primary resource for the L2 speaker), the practices of disambiguation and semantic specification, and the design of the sequence, which features greater complexity, scaffolding character, smoother transitions, and openness to co-construction in L1 usage. The L2 user focuses more on the new word (engaging in learning) rather than on a shared focus and tends not to react or follow up in a co-constructing manner. Conversely, the L1 participant uses it as a thematic entry point to activate the co-participant and transitions by inviting co-orientation to a more dynamic 3D interaction with the environment. Similarly, the cases in which new references are introduced, followed by more detailed explanations (Examples 3, 5), reveal differences: While the L2 speaker uses discourse markers to indicate transitions and presupposes a shared blended origo, relying mainly on intensive pointing activity accompanied by minimal and ambiguous deictics (e.g., this one here), the L1 speaker uses an extended transition with pre-announcement and preparatory addressee orientation in a multi-step sequence to cooperatively develop a shared focus. The L2 speaker continues without considering whether the focus is shared, resulting in additional clarifications and disruptions of discursive flow, while the L1 participant adopts the addressee's perspective, simulates

anticipated physical activity, and aligns further elaborations with the steps taken by the L2 speaker. Another important difference is reactivity, defined as the ability to adapt to new environmental demands, the transition from 2D to 3D interaction within the environment, and employing a broader range of practices, which is more pronounced in L1 speakers but limited in L2 users. However, this should be examined more systematically with different L2 speakers at different proficiency levels, considering actions in the L1 within similar (virtual) contexts. This approach would facilitate a better understanding of the dynamics of pragmatic transfer, environmental accommodation, interactive co-adaptation, and microscopic interactive learning processes.

This study presents qualitative analyses across three settings with pronounced asymmetries among participants, highlighting the development of adaptive forms of action to ensure insubjectivity. To comprehensively understand the multifaceted nature and dynamics of co-constructed space in virtual interactions, further studies should investigate these practices in various settings and participation frameworks while also considering intercultural differences in the use of usability cues. In addition, they should consider longitudinal practices that provide insights into participants' local preferences, the emergence of new creative practices, and developmental dynamics. Case study comparisons of practices in other contexts involving the same interactants could yield valuable insights into individual preferences and adaptive dynamics, as well as changes brought about by virtual reality usage. Reliable eye-tracking data on activities are also necessary to establish a fine-grained basis of different cases that can reveal the systematicity of context-dependent variations in practices and their interplays. The analysis should expand to include a systematic examination of the interaction between shared focus practices in static positions and those in dynamic movement and navigation, as well as across different instructional contexts (distinguishing between various practical or joint activities) to draw generalizable conclusions on how these practices change in relation to one another over time. Future studies should also include comparisons of focused training across different professional groups to examine the benefits and impact of the medium in promoting specific areas of practical knowledge and professional development.

Data availability statement

The recordings for this study are not publicly available because of restricted informed consent of the participants. Requests to access the recordings should be directed to ML, milica.d.lazovic@gmail.com.

Ethics statement

The requirement for ethical approval for studies involving human participants was waived by the Lab for Innovative Teaching within the framework of NiDIT (Network for Impactful Digital International Teaching Skills) at Philipps University Marburg and Justus Liebig University Giessen. All studies were conducted in accordance with local legislation and institutional guidelines. Participants provided written informed consent to take part in the study. Additionally, written informed consent was obtained for the publication of any potentially identifiable images or data included in this article.

Author contributions

ML: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. Open Access funding provided by the Open Access Publishing Fund of Philipps-University Marburg. The author declares no additional financial support for this research.

Acknowledgments

I gratefully acknowledge the technical support for this research from the lab for innovative teaching of the NIDIT Center (Network for Impactful Digital International Teaching Skills), Justus Liebig University Gießen, and Philipps-University Marburg. I would like to thank the participants of the recordings and reviewers for their valuable feedback.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

References

Ahlers, T., Lazović, M., Schweiger, K., and Senkbeil, K. (2020). Tandemlernen in social-virtual-reality: immersiv-spielebasierter DaF-Erwerb von mündlichen Sprachkompetenzen. *Z. Interkulturellen Fremdsprachenunterricht* 25, 237–269.

Ahlers, T., Bumann, C., Kölle, R., and Lazović, M. (2021). *Hololingo! - a game-based social virtual reality application for foreign language tandem learning*. DELFI 2021. Bonn: Gesellschaft für Informatik e.V. pp. 37–48.

Armougum, A., Orriols, E., Gaston-Bellegarde, A., Joie-La, M. C., and Piolino, P. (2019). Virtual reality: a new method to investigate cognitive load during navigation. *J. Environ. Psychol.* 65:101338. doi: 10.1016/j.jenvp.2019.101338

Arnold, L. (2012). Dialogic embodied action: using gesture to organize sequence and participation in instructional interaction. *Res. Lang. Soc. Interact.* 45, 269–296. doi: 10.1080/08351813.2012.699256

Auer, P., Bauer, A., Birkner, K., and Kotthoff, H. (2020). Einführung in die Konversationsanalyse. Berlin: De Gruyter.

Auer, P., and Stukenbrock, A. (2022). "Deictic reference in space" in Pragmatics of space. eds. A. H. Jucker and H. Hausendorf (Berlin: de Gruyter), 23–62.

Berger, M., Jucker, A. H., and Locher, M. A. (2016). Interaction and space in the virtual world of second life. *J. Pragmat.* 101, 83–100. doi: 10.1016/j.pragma.2016.05.009

Breves, P., and Stein, J. P. (2022). Cognitive load in immersive media settings: the role of spatial presence and cybersickness. *Virtual Reality* 27, 1077–1089. doi: 10.1007/s10055-022-00697-5

Cooperrider, K., Fenlon, J., Keane, J., Brentari, D., and Goldin-Meadow, S. (2021). How pointing is integrated into language: evidence from speakers and signers. *Front. Commun.* 6:774. doi: 10.3389/fcomm.2021.567774

Couper-Kuhlen, E., and Selting, M. (2018). Interactional linguistics: Studying language in social interaction. Cambridge, UK: Cambridge University Press.

Deppermann, A. (2017). Instruction practices in German driving lessons: differential uses of declaratives and imperatives. *Int. J. Appl. Linguist.* 28, 265–282. doi: 10.1111/ijal.12198

Deppermann, A. (2018). "Changes in turn-design over interactional histories: the case of instructions in driving school lessons" in Time in embodied interaction: Synchronicity and Sequentiality of multimodal resources. eds. A. Deppermann and J. Streeck (Amsterdam: John Benjamins), 293–324.

De Stefani, E. (2018). Formulating direction: navigational instructions in driving lessons. *Int. J. Appl. Linguist.* 28, 283–303. doi: 10.1111/ijal.12197

Döring, J., and Thielmann, T. (2008). Spatial Turn. Das Raumparadigma in den Kultur-und Sozialwissenschaften. Bielefeld: Transcript Verlag.

Due, B. L. (2023). Assemmethodology? A commentary. *Soc. Interact. Video Based Stud. Hum. Soc.* 6:1. doi: 10.7146/si.v6i1.137001

Ehmer, O., Helmer, H., Oloff, F., and Reineke, S. (2021). How to get things done. Aufforderungen und Instruktionen in der multimodalen Interaktion. *Z. Gesprächsforschung* 22, 670–690.

Eilertsen, L. J. (2014). Maintaining intersubjectivity when communication is challenging: hearing impairment and complex needs. *Res. Lang. Soc. Interact.* 47, 353–379. doi: 10.1080/08351813.2014.958278

Elkin, R. L., Beaubien, J. M., Damaghi, N., Chang, T. P., and Kessler, D. O. (2024). Dynamic cognitive load assessment in virtual reality. *Simul. Gaming* 55, 755–775. doi: 10.1177/10468781241248821

Fricke, E. (2003). "Origo, pointing, and conceptualization - what gestures reveal about the nature of the origo in face-to-face interaction" in Deictic conceptualisation of space, time and person. ed. F. Lenz (Amsterdam: John Benjamins), 69–94.

Gibson, J. J. (1979). The ecological approach to visual perception. New York: Taylor and Francis.

Gillian, M., O'Rourke, B., and Werner, S. (2024). "The language of situated joint activity: social virtual reality and language learning in virtual exchange" in Virtual Reality in den Geisteswissenschaften. Konzepte, Methoden und interkulturelle Anwendungen. eds. K. Senkbeil and T. Ahlers (Berlin: Peter Lang), 81–106.

Goodwin, C. (1980). Restarts, pauses, and the achievement of a state of mutual gaze at turn-beginning. *Sociol. Inq.* 50, 272–302. doi: 10.1111/j.1475-682X.1980.tb00023.x

Haddington, P., and Oittinen, T. (2022). "Interactional spaces in stationary, mobile, video-mediated and virtual encounters" in Pragmatics of space. eds. A. H. Jucker and H. Hausendorf (Berlin: de Gruyter), 317–362.

Haddington, P., Mondada, L., and Neville, M. (2013). Interaction and mobility. Language and the body in motion. Berlin: de Gruyter.

Hammond, J., and Gibbons, P. (2005). Putting scaffolding to work: the contribution of scaffolding in articulating ESL education. *Prospect* 20, 6–30.

Hartmann, T., and Fox, J. (2021). "Entertainment in virtual reality and beyond: the influence of embodiment, co-location, and cognitive distancing on users' entertainment experience" in The Oxford handbook of entertainment theory. eds. P. Vorderer and C. Klimmt (Oxford: Oxford University Press), 717–732.

Hartmann, T., and Hofer, M. (2022). I know it is not real (and that matters) media awareness vs. presence in a parallel processing account of the VR experience. *Front. Virtual Real.* 3, 1–16. doi: 10.3389/frvir.2022.694048

Hausendorf, H., and Schmitt, R. (2022). "Architecture-for-interaction: built, designed and furnished space for communicative purposes" in Pragmatics of space. eds. A. H. Jucker and H. Hausendorf (Berlin: de Gruyter), 431–472.

Heller, V. (2022). "Imaginary spaces in storytelling" in Pragmatics of space. eds. A. H. Jucker and H. Hausendorf (Berlin: de Gruyter), 209–250.

Helmer, H., and Reineke, S. (2021). Instruktionen und Aufforderungen in Theorie und Praxis–Einparken im Fahrunterricht. *Z. Gesprächs* 22, 114–150.

Hindmarsh, J., Heath, C., and Fraser, M. (2006). (Im) materiality, virtual reality and interaction: grounding the 'virtual' in studies of technology in action. *Sociological* 54, 795–817. doi: 10.1111/j.1467-954X.2006.00672.x

Imo, W., and Lanwer, J. P. (2019). Interaktionale Linguistik. Eine Einführung. Stuttgart: Metzler.

Jefferson, G. (1972). "Side sequences" in Studies in social interaction. ed. D. N. Sudnow (New York: Free Press), 294–333.

Juliano, J. M., Schweighofer, N., and Liew, S. L. (2022). Increased cognitive load in immersive virtual reality during visuomotor adaptation is associated with decreased long-term retention and context transfer. *J. Neuroeng. Rehabil.* 19:106. doi: 10.1186/s12984-022-01084-6

Kendrick, K. H. (2015). Other-initiated repair in English. Open. *Linguistics* 1, 164–190. doi: 10.2478/opli-2014-0009

Jucker, A. H., and Hausendorf, H. (2022). Pragmatics of space. Berlin: de Gruyter.

Jucker, A. H., Hausendorf, H., Dürscheid, C., Frick, K., Hottiger, C., Kesselheim, W., et al. (2018). Doing space in face-to-face interaction and on interactive multimodal platforms. *J. Pragmat.* 134, 85–101. doi: 10.1016/j.pragma.2018.07.001

Kesselheim, W. (2012). "Gemeinsam im Museum" in Materielle Umwelt und interaktive Ordnung, in Raum als interaktive Ressource. eds. H. Hausendorf, L. Mondada and R. Schmitt (Tübingen: Narr), 187–232.

Keevallik, L. (2013). "Here in time and space: decomposing movement in dance instruction, interaction and mobility" in Language and the body in motion. eds. P. Haddington, L. Mondada and M. Neville (Berlin: de Gruyter), 345–370.

Keevallik, L. (2020). "Linguistic structures emerging in the synchronization of a Pilates class" in Mobilizing others: Grammar and Lexis within larger activities. eds. C. Taleghani-Nikazm, E. Betz and P. Golato (Amsterdam: John Benjamins), 147–173.

Koole, T., and Elbers, E. (2014). Responsiveness in teacher explanations: a conversation analytical perspective on scaffolding. *Linguist. Educ.* 26, 57–69. doi: 10.1016/j.linged.2014.02.001

Krug, M., Messner, M., Schmidt, A., and Wessel, A. (2020). Instruktionen in Theater- und Orchesterproben. *Z. Gesprächsforschung* 21:11.

Kupetz, M. (2021). Multimodalität und Adressatenorientierung in Instruktionen im DaZ-und fach-integrierten Unterricht. *Z. Gesprächs* 22, 348–389.

Lazovic, M. (2025). "Erlebte Landeskunde in SVR-Tandems: (Inter-)Kulturelle Räume immersiv erleben und kulturreflexiv erkunden," in *Ege Germanistik. Forschungen zur deutschen Sprache, Literatur und Kultur*. Band 2. ed. Ö. Gencer Citak (Izmir: Ege Üniversitesi Yayınlari), 39–63.

Luff, P., Heath, P., Kuzuoka, H., Hindmarsh, J., Yamazaki, K., and Shinya, O. (2003). Fractured ecologies: creating environments for collaboration. *Human Comput. Interact.* 18, 51–84. doi: 10.1207/S15327051HCI1812_3

Meyer, N., and Jucker, A. H. (2022). "Co-presence and beyond: spatial configurations of communication in virtual environments" in Pragmatics of space. eds. A. H. Jucker and H. Hausendorf (Berlin: de Gruyter), 579–610.

Mondada, L. (2013). "Interactional space and the study of embodied talk-in-interaction" in Space in language and linguistics. Geographical, interactional, and cognitive perspectives. eds. P. Auer, M. Hilpert, A. Stukenbrock and B. Szmrecsanyi (Berlin: de Gruyter), 247–275.

Neuberger, O., Limpinsel, I. L., and Aßmann, S. (2024). "Zwischen zwei Welten" in Zum Verhältnis von Präsenz und Media Awareness während einer virtuellen Geländeführung, in Virtual Reality in den Geisteswissenschaften. Konzepte, Methoden und interkulturelle Anwendungen. eds. K. Senkbeil and T. Ahlers (Berlin: Peter Lang), 107–130.

Olbertz-Siitonen, M., and Piirainen-Marsh, A. (2021). Coordinating action in technology-supported shared tasks: virtual pointing as a situated practice for mobilizing a response. *Lang. Commun.* 79, 1–21. doi: 10.1016/j.langcom.2021.03.005

Olbertz-Siitonen, M., Piirainen-Marsh, A., and Siitonen, M. (2021). Constructing co-presence through shared VR gameplay. *J. Med.* 4, 85–122. doi: 10.21248/jfml.2021.31

Olbertz-Siitonen, M., and Piirainen-Marsh, A. (2023). Bridging physical and virtual ecologies of action: giving and following instructions in co-located VR-gaming sessions. *PRO* 20, 137–166. doi: 10.33352/prlg.121525

Ono, T., and Thompson, S. A. (2024). "The indeterminacy and fluidity of reference in everyday conversation" in In (non) referentiality in conversation. eds. M. C. Ewing and R. Laury (Amsterdam: John Benjamins), 123–140.

Ortner, H. (2023). "Sprache–Bewegung–Instruktion" in Multimodales Anleiten in Texten, audiovisuellen Medien und direkter Interaktion. ed. H. Ortner (Berlin: de Gruyter).

Pillet-Shore, D. (2021). When to make the sensory social: registering in face-to-face openings. *Symb. Interact.* 44, 10–39. doi: 10.1002/symb.481

Pitsch, K. (2012). "Exponat–Alltagsgegenstand–Turngerät: Zur interaktiven Konstitution von Objekten in einer Museumsausstellung" in Raum als interaktive Ressource. eds. H. Hausendorf, L. Mondada and R. Schmitt (Tübingen: Narr), 233–274.

Plötner, K., and Nowotny, F. (2023). Fremdsprachendidaktik meets 360° and virtual reality. *Studierenden* 51, 140–159. doi: 10.21240/mpaed/51/2023.01.15.X

Putzier, E. M. (2012). "Der 'Demonstrationsraum' als Form der Wahrnehmungsstrukturierung" in Raum als interaktive Ressource. eds. H. Hausendorf, L. Mondada and R. Schmitt (Tübingen: Narr), 274–316.

Sari, R. C., Pranesti, A., Solikhatun, I., Nurbaiti, N., and Yuniarti, N. (2023). Cognitive overload in immersive virtual reality in education: more presence but less learnt? *Educ. Inf. Technol.* 29, 12887–12909. doi: 10.1007/s10639-023-12379-z

Schegloff, E. A., Jefferson, G., and Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language* 53, 361–382. doi: 10.1353/lan.1977.0041

Schmidt, A., and Deppermann, A. (2021). Instruieren in kreativen Settings – wie Vorgaben der Regie durch Schauspielende ausgestaltet werden. *Z. Gesprächsforschung* 22, 237–271.

Selting, M., Auer, P., Barth-Weingarten, D., Bergmann, J. R., Bergmann, P., Birkner, K., et al. (2009). Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung* 10, 353–402.

Senkbeil, K. (2024). "Communication in hybrid presence–concepts for the analysis of social virtual reality in the humanities" in Virtual Reality in den Geisteswissenschaften. Konzepte, Methoden und interkulturelle Anwendungen. eds. K. Senkbeil and T. Ahlers (Berlin: Peter Lang), 205–234.

Simone, M., and Galatolo, R. (2020). Climbing as a pair: instructions and instructed body movements in indoor climbing with visually impaired athletes. *J. Pragmat.* 155, 286–302. doi: 10.1016/j.pragma.2019.09.008

Spets, H. (2023). Environmentally coupled gestures as a communicative resource in the word explanation activity: a multimodal analysis of interaction in social VR. *PRO* 20, 167–192. doi: 10.33352/prlg.120936

Steels, L., and Loethsch, M. (2009). "Perspective alignment in spatial language" in Spatial language and dialogue, explorations in language and space. eds. K. R. Coventry, T. Tenbrink and J. Bateman (Oxford: Oxford Academic), 70–88.

Steinbock, J., Hein, R., Eisenmann, M., Latoschik, M. E., and Wienrich, C. (2022). Virtual Reality im modernen Englischunterricht und das Potenzial für Inter- und Transkulturelles Lernen 47, 246–266. doi: 10.21240/mpaed/47/2022.04.12.X

Stoeckl, H., and Messner, M. (2021). Tam pam pam pam and mi – fa – sol: constituting musical instructions through multimodal interaction in orchestra rehearsals. *Multimodal Commun.* 10, 193–209. doi: 10.1515/mc-2021-0003

Stukenbrock, A. (2010). Überlegungen zu einem multimodalen Verständnis der gesprochenen Sprache am Beispiel deiktischer Verwendungsweisen des Ausdrucks so. *InLiSt Interact. Linguist. Struct.* 47:11.

Stukenbrock, A. (2018). Blickpraktiken von Sprecher Innen und Adressat Innen bei der Lokaldeixis: Mobile Eye Tracking-Analysen zur Herstellung von joint attention. *Gesprächsforschung* 19, 132–168.

Stukenbrock, A. (2020). Deixis, Meta-perceptive gaze practices, and the interactional achievement of joint attention. *Front. Psychol.* 11:1779. doi: 10.3389/fpsyg.2020.01779

Sweller, J., van Merriënboer, J. J. G., and Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educ. Psychol. Rev.* 31, 261–292. doi: 10.1007/s10648-019-09465-5

Szczepek Reed, B. (2023). 'Go on keep going': the instruction of sustained embodied activities. *Discourse Stud.* 25, 692–717. doi: 10.1177/14614456231153578

Tekin, B. S. (2021). Quasi-instructions: orienting to the projectable trajectories of imminent bodily movements with instruction-like utterances. *J. Pragmat.* 186, 341–357. doi: 10.1016/j.pragma.2021.10.018

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Massachusetts: Harvard University Press.

Wiepke, A., Belli, F., Fischer, M. H., and Miklashevsky, A. (2024). "Embodied learning in virtual reality" in Virtual Reality in den Geisteswissenschaften. Konzepte, Methoden und interkulturelle Anwendungen. eds. K. Senkbeil and T. Ahlers (Berlin: Peter Lang), 235–254.