KATHY A. MILLS

Potentials and Challenges of Extended Reality Technologies for Language Learning

1. Introduction

Extended reality technologies have a range of distinctive potentials to support and augment language and literacy practices in a digital age, while bringing new challenges. Extended reality (XR) is a classification of technologies that includes virtual reality (VR), augmented reality (AR), mixed reality (MR) technologies, and future applications that combine virtual and real elements. VR is a completely 3D immersive digital world, AR overlays virtual content on real world, and MR interactively merges virtual with the real or possible. Extended reality also includes future (X) technologies that combine various elements of real, virtual, and potential environments, often involving wearable technology and human-machine interaction (Palmas and Klinker 2020).

Extended reality is one of the fastest growing educational technologies for classroom use (Eldequaddem 2019), yet research of VR, AR, and MR technology for language and literacy learning, including foreign language learning, is scarce (Ok et al. 2021). This is because the recent wave of extended reality technologies, such as virtual reality immersion with head-mounted displays, was developed as recently as 2013 (Pottle 2019). These technology platforms are often designed for commercial use, business productivity, and recreational gaming, which, as educators have seen, has created a significant gap in terms of learning theories to effectively guide classroom applications (Radianti et al. 2020).

Currently, only a small proportion of research of virtual, augmented, and mixed reality technologies in education has been published specifically with a view to language and literacy learning, with a predominance of VR, AR, and MR technology research applied in technology and STEM education disciplines. For example, in a systematic review of original virtual reality studies, only 27 of 167 had any relevance to language learning (Reisoglu et al. 2017).

Even when research of extended reality technologies is driven by a language and literacy focus, the emphasis has been on rote learning of decontextualised basic language skills, such as phonics and letter recognition (Ok et al. 2021), words in different languages (Barreira et al. 2012), and sight words and spelling, typically presented in a quiz or puzzle format (Fan et al. 2020). The limitations of such applications are that they are driven by the technology development industry, based on common-sense understandings of language-learning, synthetic phonics, and behaviouristic stimulus-response models of education, rather than designed to support the kinds of authentic communication practices that are used in foreign language education and society.

Augmented reality (AR) technologies overlay virtual content on the physical world, including animations, graphics, advertisements, and other digital information. A key
novelty is that elements of the virtual and real world coexist (Chang et al. 2013), often in useful ways for language learning, social practices, digital making, and text creation (Mills, Scholes, and Brown, 2022). The rapid development of AR platforms has been fortified by the advent of mobile devices with cameras, internet access, accelerometers, gyroscopes, and GPSs to trace the rotation and global positioning of mobile devices (Dunleavy et al. 2008).

These AR applications typically require the use of a handheld device, or less commonly, a head-mounted display, including the functionality of a webcam or projector-camera system (Fan et al. 2020). Augmented reality reading can include activities such as viewing augmented virtual web content or advertising through an internet browser. Augmented reality interactive books often overlay digital content, which involves holding a smartphone over a physical book or other tangible learning objects to connect to virtual content, such as 3D models.

Virtual reality (VR) also simulates artificial or computer-generated imagery, but typically without anchoring the virtual to the physical world – as occurs for augmented and mixed reality environments. Virtual reality technologies provide a synthetic or computer-simulated environment for user-immersion generally requiring the use of a head-mounted display (HMD) for stereoscopic vision. Virtual reality technologies are interactive and use motion tracking controls to supply haptic feedback in the virtual environment that differs, for example, from viewing a three-dimensional film. Virtual reality texts are fully immersive, three-dimensional spaces, enhancing the user’s sense of presence – of being there – with a substantial degree of immediacy, while the physical world is blocked from view (Heeter 1992; Jensen and Konradsen 2018; Velev and Zlateva 2017). This renders the physical or real-world context as remotely backgrounded in the virtual reality experience, perceived only by incidentally touching objects in the real world that are concealed from view (Mills and Brown 2021).

Virtual technologies can be used for communicating, storytelling, learning through play, and socialising, depending on the application (Marsh and Yamada-Rice 2018; Moran and Woodall 2019). VR applications have been shown to aid immediate and direct learning, support memory, increase learner attention and motivation, and enable better learner decision-making in simulated environments (Elmqvaddem 2019; Pottle 2019; Radianti et al. 2020). Because VR systems allow the user to touch and manipulate objects in the virtual environment, and to use locomotion, virtual reality technologies have new potentials for bodily engagement compared to the haptic affordances of other technologies for text creation, such as writing and drawing on paper (Mills and Brown 2021), or screen-based computing and tablet use (Crescenzi et al. 2014).

Mixed reality (MR) refers to environments that merge virtual and real elements, and interactive virtual objects that respond to the user’s movement, including virtual content that facilitates interaction between realities as a consequence of blending realities or fusing the virtual and the real (Maas and Hughes 2020; Palmas and Klinker 2020).
Recent examples are smart glasses, including the Microsoft HoloLens 2, utilising mixed reality features that are responsive to hand and eye movements of the user. These simulation technologies generate an enhanced sense of immediacy, presence (the sense of being there), and in some applications, immersion (Bronack 2011), with emergent potentials for literacy and media practices.

2. Extended Reality Video Games for Language Learning

Theorists of the new literacy studies (Gee 2004; Street 2003), multiliteracies (New London Group 1996), and multimodality (Kress 2000), have acknowledged for several decades that the ways in which people learn, play, and work are changing, so much so that the very nature of language and literacy requires reconceptualization (Mills 2010). Digital platforms facilitate unstructured experimentation with digital media, gaming, coding or making games, and developing and sharing content on the social web. Many of these interest-driven practices of youth involve games-based media (Ito et al. 2009), while augmented, virtual, and mixed reality technologies intersect with these social practices. While some extended reality technologies are designed for productivity and commercial use rather than gameplay, many virtual, augmented, and mixed reality platforms are typically supported by gaming wearables and hardware that is marketed for the social purpose of entertainment (Mills, Scholes and Brown 2022). Games are defined as practices involving players who typically engage in artificial conflict within a system of rules with "quantifiable outcomes" (Salen Tekinbaş and Zimmerman 2004, 80). This definition can encompass both digital and non-digital games.

From this perspective, virtual, augmented, and mixed reality practices can potentially be built on long-established theoretical foundations of digital gaming as literacy practices and communities (e.g., Beavis 2014a; Gee 2004; 2007). When we consider the variety of gaming contexts in which virtual, augmented, and mixed reality technologies occur (e.g., Microsoft HoloLens), many of these practices are playful spaces that may contrast with the mundane nature of ordinary life (Mills, Unsworth and Scholes 2022). Video games create an environment in which identities, subcultures, and events are governed by their own progression of time, transporting the player into another world (Garcia 2018). Researchers have argued that a key advantage of games for learning is the provision of space for experimentation, role play, and simulation that is less constrained by conventional social norms (Salen Tekinbaş and Zimmerman 2004). Virtual, augmented, and mixed reality applications for language learning can be seen as another kind of hybrid, games-based literacy that is driven by a range of social interests.
3. New Potentials of Extended Reality Research for Language Learning

Given that researchers are now turning their attention to the benefits or potentials of extended reality technologies for language learning, this section explores some of the key affordances that have been found in recent years. This is by no means an exclusive list, particularly given that extended reality technologies and learning theories to support learning designs are rapidly developing. However, there are some unique features of these technologies that have been shown to support learning in particular ways.

3.1 Focusing Learner Attention and Hands-on Interaction

Researchers have observed the benefits of augmented reality technologies in terms of the strategic provision of virtual content that may be presented in conjunction with tactile learning materials, which has been found to draw early learners’ attention to relevant visual-audio information (Fan et al. 2020). A related benefit is the hands-on interaction that is provided by interacting with physical learning materials, such as interactive augmented reality books, blocks, wall charts, and other materials, while offering related virtual content to anchor key concepts in concrete, interactive, and embodied ways (Barreira et al. 2012). The facility for hands-on interactive language learning is important, particularly given recent research on haptics. Haptics refers to the use of touch to perceive objects or to communicate non-verbally, and is an essential part of language learning involving gestures. Research of haptics in current research concerns studies of human touch to interact with the external environment (Minogue and Jones 2006).

When augmented reality technologies have been used to overlay abstract language symbols on physical or real learning objects, such as blocks, books, and flashcards, these virtual-physical correspondences are highly effective for language learning. Theorists suggest that learners are limited by the amount of information that can be processed at one time, so augmented reality applications that draw the learners’ attention to specific language information, such as letter or word knowledge overlaid virtually, supplies a valuable attentional cue to isolate the most relevant and focal language knowledge. Likewise, augmented reality applications for language learning have an added advantage when they require the learner to haptically interact with both physical learning materials and 2D or 3D augmented virtual overlays (e.g., rotate, grasp, or line up), which supports learners to connect abstract language knowledge with everyday objects and environments (Fan et al. 2018).

A related advantage of augmented reality virtual overlays of content upon the real world is that language use becomes situated in an authentic environment. Exemplifying the pedagogy of situated practice (Brown et al. 1989), when designed in authentic ways, the anchoring of virtual, abstract language content in situated real-life contexts has been
found to strengthen learning and memory. For example, the teaching of spelling and vocabulary can be taught by connecting the digital and physical space, for example by virtual words being positioned near the physical objects in the learner's real world, while hearing the associated sound or other digital feedback (Fan et al. 2020). An example of this is situational, interactive games that have been found to improve preschool children's performance in learning English vocabulary (Pu and Zhong 2018).

3.2 Sensory and Embodied Language Practices for AR, VR and Mixed Reality

Augmented, mixed, and virtual reality platforms offer new multisensory ways of interacting with sophisticated, three-dimensional virtual environments and objects, involving the user's orchestration of vision, haptics, sound, and large body movements, such as locomotion, in language learning (Mills and Friend 2022; Yu and Smith 2012). In the case of virtual reality, the use of motion sensing technologies permits tracking of movement within the virtual environment, while augmented reality has a particular advantage for movement and location-based learning that involves locating virtual content in geolocations that are physically traversed by users (Hsu 2017). Image-based AR utilises marker-based or markerless technologies that use image recognition techniques to track the position of an object, while location-based AR draws on geolocational or positional data to identify the location of objects (Cheng and Tsai 2013). Augmented reality language learning has benefits for directing the embodied attention of toddlers to support the learning of new vocabulary (Yu and Smith 2012). Similarly, mobile-based AR applications have been found to support the vocabulary learning of older ESL students (He et al. 2014). Given that much of human cognition is dependent upon sensorimotor processes, such as movement, and that even abstract mental processes are fundamentally body-based (Wilson 2002), the location-based and sensory affordances of virtual and augmented reality technologies are charged with possibilities for language learning.

Virtual and mixed reality technologies are responsive to haptics, supported by movements of the head and body, with high simulator fidelity (Jensen and Konradsen 2018). Theorists of embodied cognition point to an integrated view of cognition and sensorimotor visual systems, and between perception and action, with these systems functioning as a single neurological process. Perception and learning do not operate exclusively in the brain, rather, the learner uses perceptually guided explorations of a learning environment (Gibbs 2005). Multisensory research from the cognitive sciences has demonstrated that the restricted use of the senses to involve only vision results in perceptual processing deficits for learners. Not only this but restricting learning to vision alone results in reduced efficiency for both memorising and communicating information, compared to multisensory input and outputs (Shams and Kim 2012). Thus, virtual, augmented, and mixed reality technology platforms (e.g., MS HoloLens 2) that
harness combinations of vision and movement (and often auditory learning), have a distinct advantage for language learning.

New sensory affordances of virtual reality technology for digital text creation were shown in recent research by Mills and Brown (2021), who observed the multimodal literacy learning of three classrooms of upper elementary students. The students used a three-dimensional virtual painting program (Google Tilt rush) to produce creative digital designs using the HTC Vive headset and sensors. The students designed immersive, three-dimensional texts or virtual paintings that were created by proprioceptive bodily action. Multimodal interaction analysis of the video and screen capture data was used to map the sensory and embodied patterns of action during the students' composing of virtual designs. Compared to conventional, screen-based digital text making practices, virtual painting required much more dramatic body movements of extensive variety, including haptics, head movement, and locomotion (Mills and Brown 2021). It points to new affordances for students to use the full sensorium in digital designing, extending current understanding of how perceptual and motoric experiences can optimise word learning in non-digital contexts, such as using gestures and actions to anchor new vocabulary in the mind (Hald et al. 2016).

Other studies of immersive virtual reality technologies have demonstrated learning potentials that have implications for language education, particularly in terms of skill acquisition and factual learning (Rasheed et al. 2015). Similarly, immersive technologies have clear advantages for assisting learners to remember and understand visual and spatial information (Jensen and Konradsen 2018). Virtual reality technologies have been found particularly useful for supporting learners' use of visual scanning and observational skills (Ragan et al. 2015), as well as affective learning development for situations involving emotion management (Pallavicini et al. 2016), each having potentials for optimising the design of language learning activities and environments.

3.3 Transmediation across Modes in VR, AR, and Mixed Reality Technologies

Virtual, mixed, and augmented reality applications provide particular opportunities for learners to transmediate meanings across sign systems or modes. Extended reality environments are multimodal, combining two or more modes, such as two and three-dimensional imagery, written words, sonic elements (e.g., music, silence, dialogue, or sound effects), mathematics, gesture, kinesics, locomotion, and proxemics (Mills 2016). Transmediation is the process of mapping meanings from one expression plane, symbolic system, or communicative mode to another (McCormick 2011), such as using a drawing to depict a scene from a novel.

The process of transmediation was conceptualised and named by Suhor (1984), who saw that each sign-making system has unique organisational principles so that while
concepts can be shifted across modes to show roughly equivalent meanings, there is never an exact or precise interpretant for mapping one meaning onto the expression plan of another system (e.g., drawing to writing, music to sculpture, image to sound). Transmediation is always a complex process because each sign-system or expression plane has different elements that enable or constrain the way ideas are communicated, symbolised, and remembered (Siegel 2006; Forman 1994).

Extended reality applications offer new opportunities for language learners to transmediate semiotic content across modes, as proven in a recent study by Mills and Brown (2021). Upper elementary students were introduced to Greek mythology over several lessons that used a range of media (e.g., books, presentations, internet), becoming familiar with several Greek myths, including Icarus and Daedalus, Daedalus' Maze (Labyrinth) and the Trojan war. The students authored story retellings of the Greek myths and illustrated these using pencil and paper. Learners then translated their written and illustrated stories to the mode of virtual reality painting using Google Tilt Brush and the HTC Vive headset and sensors. Multimodal interaction analysis of video, screencast, and think-aloud data was used to identify the features of the virtual environment that enabled and constrained students' transmediation of content. Because the process of transmediation involves mapping meaning across non-analogous expression planes or symbol systems, there were generative possibilities for learners to adapt their knowledge to communicate these ideas creatively and effectively in an immersive virtual painting mode.

One of the unique features of the virtual environment painting mode was the simulated, three-dimensional environment, which differed visually from the two-dimensional elevation drawings. Students were able to create immersive scenes where they could view both inside and outside the Trojan horse, and walk inside their artwork, as students noted: "You have all this space to create around you, instead of working on one piece of paper" (Mills and Brown 2021). Similarly, the virtual mode created a unique sense of presence in their virtual paintings – the experience of being there in the painting. Haptic and bodily interactivity was dramatic in the virtual painting mode, which differed significantly to sitting at a desk drawing and writing. The immersive, virtual painting mode was not equivalent to sign making using conventional writing, drawing, or painting, but required users to generatively invent and adapt ways to simulate literary content through meaningful and readily interpreted virtual symbols in a three-dimensional, life-sized expression plane (Mills and Brown 2021).

4. New Challenges for Extended Reality in Language Learning

The following section is divided into two sections that discuss key challenges for the future of extended reality research in language learning. The first concerns the pedagogical pitfalls for extended reality language learning. The second is the challenge
of digital divides that are anticipated to stratify access and quality of use of extended reality platforms for language learning and literacy practice.

4.1 Pedagogical Pitfalls of Extended Reality Language Learning

Extended reality technologies vary widely in their social purposes, while gamified applications designed for educational use need to be guided by research-based learning theories. For example, as noted earlier in this paper, currently the range of augmented reality language applications in a game-based format rely heavily on quizzes, rather than involving role plays, narratives, virtual tutors, and creative designing, resulting in a narrow range compared to the range of game genres for other social purposes (Fan et al. 2020). A further criticism of current extended reality programs for language learning is an emphasis on teaching rudimentary elements of language, such as sight words and letter sounds, over supporting authentic language use. In addition, there is currently little research on the long-term language learning gains of augmented and virtual reality applications, with most studies involving short-term pre- and post-study designs (Li et al. 2017). Currently, many virtual and augmented reality applications are too inflexible for teachers to be able to adapt to a wide range of curriculum objectives, which is a failing of software focused on delivering finite and unmodifiable content, vocabulary words, historical simulations, and so on, rather than supporting open-ended, multi-purpose language uses that reflect real-world language uses and social interaction (Fan et al. 2020).

Extended reality technologies need to support users’ creative production of multimodal texts, while providing opportunities to develop critical literacy, playful creativity, inquiry-based learning, and collaboration. For example, Aurasma was an augmented application that supported creators to overlay a virtual text or object of any variety (e.g., image, video, animation) over a physical text of any kind or genre. One study used this software to link videos the children had created to accompany their physical texts. Parents could use smartphones to view the children’s videos from the written stories (Marsh and Yamada-Rice 2018).

Virtual reality applications, such as Google Tilt Brush and Let’s Make Pottery VR, provide opportunities for learners to create three-dimensional multimodal designs, attending to the grammar of visual imagery. In the case of Google Tilt Brush, a virtual painting program, users create two- or three-dimensional representations through in-air haptics, which can involve making semiotic choices about narrative and conceptual depictions, viewer positions, visual reading pathways in immersive 3D texts, grammars of presentation and interaction, modality and validity, compositional meanings, and colour (for more on visual grammars, see Kress and van Leeuwen 2021). Using the software Let’s Make Pottery VR, users can shape virtual clay on a pottery wheel using in-air haptics, and then bake, paint, and embellish the pottery with patterns from various...
cultural origins and historical periods to support the integration of knowledge from visual arts, media arts, multimodal design, culture, and history.

These are examples of extended reality applications that support learner's multimodal designing in more open-ended ways than closed-answer quiz formats, which can be readily adapted to the curriculum, and which reflect the way meaning making is used in society beyond schools. Any technology can be used in ways that simply reinforce retrograde pedagogies of behaviouristic, lockstep, stimulus-response learning, or technologies that position the teacher or technology as a dispenser of information and learners as passive recipients. The development of new extended reality technology uses for language learning is not untouched by the perpetuation of ideological, autonomous skillsets that are decontextualised from real-world language uses, based on a 'one-size-fits-all' model of language and literacy curriculum.

4.2 Digital Divide for Extended Reality Language Learning

One of the potential challenges for the use of extended reality technologies in the language classroom is the fault lines of social inclusion upon which digital divides will rupture, the enormity of the gap between those who have ready access to extended reality technologies and those who do not, at what age, for which groups, for how long and how often, and for what social purposes. Digital inclusion is complex and multifaceted, and the accessibility of particular kinds of extended reality technologies already varies considerably. For example, many people in contemporary society have access to the mobile phones through which to view augmented reality products. However, access to immersive virtual reality gaming at home, involving high quality system requirements, is currently less common.

Historically, some of the key differences for digital users pertain to home and school access, and the quality of home and school use of these platforms (Warschauer 2007). Access to digital technologies for information and digital text production has become part of everyday literacy practices in contemporary societies. Given the vast disparity between the "haves" and "have nots" in relation to sophisticated, digitally mediated literacy practices at home, education providers are essential in bridging the gap (Warschauer and Tate 2018). Virtual reality technologies are rapidly becoming more accessible for group applications in education, yet the research lags behind, particularly in terms of the learning theories to guide practice (Radianti et al. 2020), including technologies specifically designed to guide applications in foreign and English language classrooms. We know from research of digital inclusion in technology education, that key issues for access concern physical access to computer hardware and software, access to reliable sources on the Internet, human resources to facilitate training in the use of digital technologies, and finally, social resources, including
infrastructure institutional and national structures to support access (Warschauer and Tate 2018).

It is expected that these challenges will similarly influence take-up of extended reality technologies for language learning, with limited opportunities for some learners to experience presence – a sense of being there. Some will have deeper levels of immersion in more expensive and less accessible virtual reality environments, and opportunities to use sophisticated computer-simulated sensory reception for learning. For example, virtual reality has been used to simulate life inside a 1930s house to develop a deeper understanding of how time and place influenced events in the novel, To Kill a Mockingbird, and to empathise with the characters in their historical context. As Moran and Woodall (2019, 90) argue: "VR is not a space-age novelty for intense video gamers. It is a legitimate digital tool" to enhance English language learning.

Extended reality technologies offer interactive and experiential language learning experiences that support hybrid forms of abstract thinking and communication (Fernandez 2017). Likewise, meaning making using three-dimensional textual formats and immersive representations could enhance learning for all, rather than few. In the case of augmented reality, language learning is enhanced when language knowledge is presented in ways that anchor virtual content in relevant and authentic ways to everyday objects and real-world locations. Many augmented reality language applications provide opportunities for learners to use language in both virtual and physical spaces, including in location-based contexts, supported by three-dimensional multimedia content (Fan et al. 2020). Some researchers anticipate that augmented reality applications and access to the immersive web from mobile devices may go some way toward bridging the digital divide. This is due to low-costs and commercial availability of handheld devices that can be easily deployed at schools or at home with Wi-Fi (Fan et al. 2020).

The distribution of access to digital technologies has tended to disenfranchise many who are rural and remote, those with low household incomes, and in the USA and Australia, certain racial groups, such as African Americans and Indigenous Australians (Horrigan and Duggan 2015). A further issue is ableist assumptions that often drive technology development, including an emphasis on the visual mode that is prominent in most extended reality technologies. The design of extended reality technologies needs to optimise principles of universal design for learning – accessible for all learners (Storey 2007).

5. Conclusion: Extended Reality Futures for Language Learning

The digitalisation of language and literacy practices in education has been long recognised, with researchers as early as the 1990s observing these dramatic changes through publication titles such as: Writing Differently in the Post-Age (Green 1993),
"Technologising Literacy" (Bigum and Green 1993), Page to Screen: Taking Literacy into the Electronic Era (Snyder 1997). Computer games and word processors took the home market by storm in the 1980s, changing the way families used technology for school, work, and play.

By the late 1990s, theorists of computer-mediated communication were exploring the changing nature of texts, play, and performance on the internet (Danet 1998) and the identity reconstruction and anonymity of participants in virtual communities (Donath 1999). In the technological and media shifts of the past four decades, the participation of children and young people in these social practices is undeniable (Beavis 2014b), particularly since the rise of web 2.0 or the social web (Mills 2015). Now in the 2020s, language and literacy practices will continue to be transformed with the new potentials of the immersive web (Medley 2020) and its potentials for hybrid social interactions using augmented reality.

Language and literacy practices have always been multisensorial, intrinsically connected to perception, cognition, and the body, constituting emplaced human experience in the world (Mills 2016). However, since the advancement and increasing accessibility of immersive multimedia, technologies for making meaning can now respond to users' physical movement and presence, requiring the use of multiple senses and significant, whole-body movement in computer-simulated, three-dimensional environments (Velev and Zlateva 2017).

The rapid advancement of virtual reality technologies and their adoption for learning has opened up new possibilities for shifting concepts and semiotic content across modes, with underexplored features for translating narrative content in visual, haptic, and auditory ways. There are human-computer technological developments already underway to simulate food texture or virtual chewing (Nijjima and Ogawa 2016), and tasting sweetness (Ranasinghe and Do 2016). Likewise, thermoreception is now possible with the simulation of ambient temperature and wind conditions (Ranasinghe et al. 2016). Each of these developments suggests that we are rapidly moving toward more holistic uses of the full sensorium in language learning and literacy practices through virtual, augmented, mixed, and extended reality technologies of the not-too-distant future.

Works Cited


Fan, Min, Alissa N. Antle, and Shubhra Sarker. "From Tangible to Augmented: Designing a PhonoBlocks Reading System Using Everyday Technologies." *CHI EA '18: Extended Abstracts of the 2018 CHI Conference on Human Factors in
EXTENDED REALITY TECHNOLOGIES FOR LANGUAGE LEARNING


