A Future Direction for Virtual Reality Language Learning Using Dialogue Generation Systems

Haimingyue Xu

Faculty of Engineering, The University of Sydney, NSW, Australia

kaxu2041@uni.sydney.edu.au

Abstract. The ability to speak multiple languages is crucial in the modern world. This paper sets a future direction for second language learning and introduces a new design incorporating recent technological advances such as virtual reality and dialogue generation systems. The design will be analysed from different perspectives, including the affordances of virtual reality systems, the learning objectives it can achieve, the learning strategies it utilises, and the language learning competencies learners can obtain using the design.

Keywords: Virtual Reality, Second Language (L2) Acquisition, Dialogue Generation System.

1. Introduction

As recent technology advances, globalisation and multiculturalism become essential in people's work and personal lives. Thus, the importance of learning a second language (L2) and fostering interpersonal communication comes to light not only in academia but in the whole world. As a result of the new technologies, the focus of L2 learning is slowly transiting from traditional pedagogy to pedagogies of computer-assisted language learning such as Virtual Reality (VR), Augmented Reality (AR) and Mixed Reality (MR). Specifically, the affordance of VR environments makes it a helpful tool for both language learning and cultural immersion, a prominent aspect of L2 acquisition [1]. Furthermore, the recent advancements in dialogue generation systems that mimic human conversations brought the possibility of implementing such procedures in L2 learning environments. A real-life simulation could be created to make language exchange more accessible to learners by using the dialogue generation systems.

It is essential to understand the affordances of VR environments, i.e., the possible actions these environments provide for the interaction between learners and the virtual world to fully explore its possibilities for language learning. Dalgarno et al. listed five affordances of 3D Virtual Learning Environments, including facilitating spatial tasks, experiential learning tasks that are impractical to implement, leading to intrinsic motivation and engagement, contextualisation of learning, and richer collaborative learning [2]. In the case of L2 education, the primary affordance is its immersive nature. Immersion is achieved by revealing spatial relationships and experiencing an environment where all conversations are in the target language. Immersing in a completely different world not only provides the setting for practicing but also leads to "higher cognitive engagement" [3]. However, we need to consider how these affordances benefit L2 acquisition through the learning strategies to investigate how VR helps with the process.

Bialystok highlighted four strategies to increase the effectiveness and proficiency of L2 acquisition: formal and functional practicing, monitoring, and inferencing [4]. Practicing refers to the language learner's exposure to the language, whereas formal practicing focuses on the linguistic and accuracy of the use of language, while functional practicing emphasises the communicative, immersive aspects of the language. Monitoring includes examining the linguistic output, which is a formal strategy, while inferential comprehension means that the learner understands the passage while the meaning of language is not explicitly explained. Apart from the learning strategies, language learning is shaping and being shaped by society. Therefore, it is essential to understand the culture in relation to a language. Culture includes ideas, customs, skills, arts, tools, beliefs, and values; it provides the context for people's cognitive and affective behaviour [5]. By using VR environments, learners can practice functionally, inferentially interpret conversations, and immerse in the culture simultaneously.

There are many related works that talk about how learning L2 in a VR benefits intercultural competence and how they benefit language learning in turn [6-8]. While these studies only made students watch videos using VR equipment, they were able to empirically suggest that using VR can facilitate cross-cultural learning and showcase authentic cultural experiences. However, their focus was solely on one-way communication, such as video media. In this paper, it is proposed that by using the dialogue generation systems that mimic a real conversation with a native speaker, learners can immerse themselves in the foreign culture in two-way communication, practicing conversations in real-time.

Dialogue generation systems are personalised systems using deep learning technology to provide applications with the ability to interact with users naturally. Chen et al. proposed a hierarchical variational memory network that enables the system to capture the variation of hypothetical concepts and randomly retrieve long-term memories to create a dialogue system that resembles real-life conversations [9]. Similarly, the work of Yang et al. discussed three types of rewards and measured their conversation, such that the system they produced was able to "generate topic-coherent, informative and grammatical responses" [10]. The two dialogue generation systems mentioned above are primarily helpful for artificial intelligence agents such as Siri and Alexa; they focus mainly on the functional aspect of communication. On the other hand, Huber et al. investigated how dialogues can be generated such that the system can express emotions, rather than purely language, to provide informative responses [11]. More recent work by Liu et al. created personas that aim to understand the interlocutors to mimic human-like responses based on the conversation history [12]. Both Huber et al. and Liu et al. demonstrated the possibility of generating personalised messages using dialogue generation systems.

Combining the recent discoveries in dialogue generation systems and the affordances of VR environments, a hypothesised VR project could be established to discuss the future directions of L2 learning: an open-world virtual environment mimicking real-life scenarios that embody the dialogue generation technology. This paper demonstrates such design and will use English language learning as an example. Still, it is essential to note that the model can be applied to any language, with modifications in place. The paper will highlight the connection of the design to different affordances and types of knowledge and introduce the language learning competencies to demonstrate potential implementations of such systems in the future.

2. The virtual reality learning design

2.1. The Design

This design consists of an open-world environment where users can freely interact with the people around them, just like in the real world. The main principle is that through the mimicking of real-world conversations in real-world scenarios with dialogue generation, learners gain language abilities while improving their intercultural competence and can produce conversations suitable to the context.

One example of when the occasion sets the context of conversation would be formal and informal situations. When talking in everyday situations, such as talking to friends, people tend to use colloquial language, abbreviations, and slang. On the other hand, in formal cases, such as in the workplace or at school, they need to pay close attention to the choice of words to be respectful to other people. In this virtual environment, the design of learning modules for different places should use the appropriate formality, introducing new slang and words to mimic a real-life experience.

In addition to scenario-based learning, the design should also allow a "building" or a "room" for language exchange. This is where users can create custom chat rooms to communicate freely. It can be divided into organisation-hosted rooms and private rooms, where private rooms can only be accessed by users only if they achieve a basic level of fluency in the language to protect the user experience.

2.2. Learning Objectives

According to Krathwohl's revised two-dimensional taxonomy based on Bloom's taxonomy, learning outcomes can be divided into the knowledge and cognitive process dimensions [13]. The knowledge dimension can be further sub-divided into four sections, as listed in Table.1 below.

Table 1. Types of Knowledge

Types of Knowledge	Examples
Factual	Terminology, details, and elements
Conceptual	Classifications, principles and generalizations, theories, models, and structures
Procedural	Subject-specific skills, algorithms, techniques, methods, and determining appropriate procedures to use under different circumstances
Metacognitive	Strategic, contextual, conditional, and self-knowledge

On the other hand, the cognitive process dimension was divided into six levels, including remembering, understanding, applying, analysing, evaluating the materials, and creating based on materials.

This design mainly concerns the factual, procedural, and metacognitive types of knowledge and focuses on remembering, understanding, applying, and creating meaningful sentences. Using programs based on this design, learners can memorise new vocabulary and grammar items, understand conversations, apply their learning to exchanges in suitable environments, and create meaningful sentences when talking to others. Table 2 summarises the learning objectives that can be achieved by using this design.

Table 2. Learning Objectives

Learning Objectives	Usage in this design
Remember	• Vocabulary items • Approximate grammar usage (can have minor mistakes)
Understand	• Sentences • Conversations with context
Apply	• The information they remembered and understood in conversations in the virtual world and extends it to the real-world
Create	• Appropriate, meaningful sentences under a particular context

2.3. Affordances

Many previous papers studied the affordances of VR, which can be summarised into six main affordances: spatial cognition, depicting the non-existent, engagement, immersion, contextual learning, and collaborative learning [14-16]. Each affordance can be employed in different language learning scenarios.

When learning a language, spatial cognition is necessary to express spatial experiences [17]. When learners learn locational words, such as the differences between "here" and "there", although direct translations to their native language can explain the differences, sometimes it is not the case. For example, in Japanese, there are words such as "koko" (here), "soko" (there), and "asoko" (that over there). These words can be literally translated. However, the area of "soko (there)" for Japanese people grew larger as the distances between the speaker and the listener increased, whereas the "here" area for English speakers grew larger [18]. The knowledge of these subtle differences cannot be acquired via traditional learning methods.

Depicting the non-existent means that the learners can practice naming different objects even though these objects do not exist in the physical world where they are present. This can be employed when learning about specific cultural objects on a particular occasion. For example, people will give

out "red packets", red envelopes containing money to children and minors, on the Chinese New Year [19]. This word is pivotal in Chinese culture, and people usually mention it in daily conversations. Still, it is difficult for an L2 learner in a different environment to encounter these words and know in which context they are used.

VR supports engagement and immersion. With dialogue generation technology, L2 learners can engage in personalised conversations where their dialogue can also depend on their chatting history. By implementing such design, learners can immerse themselves in the virtual world not only because of the cognitive idea of "being in the world" but also because of the high similarity between the virtual world and the real world.

The meaning of words can change over time, called semantic change. It happens because people extend words' meanings creatively, which can cause different interpretations of the exact words. Semantic change is mainly influenced by social factors, in which if the new definition is used by a speaker whom other people are willing to be associated with, it is more likely to spread [20]. The online feature allows L2 learners to talk to native speakers to shape their understanding of words and phrases, which might have gone through semantic changes over time. The dialogue generation systems could also connect to the Internet and improvise dialogues using the common usage in social media, filtering out sensitive words.

2.4. Learning Strategies

Learning strategies, as mentioned above, includes formal and functional practicing, monitoring, and inferencing [4].

The design should not reward learners based on formal applications of grammar. It focuses on the cognitivism and social constructivism aspects rather than behaviourism aspects of learning because language is all about communicating successfully and delivering thoughts and ideas to other people. This means that the grading system for monitoring output, if there is one, should be based on whether the language produces the meaning that the speaker intended to deliver rather than the correctness of grammar structures. For example, "I will go to the library on Sunday" is semantically equivalent to "Me go library, Sunday" or "I go library at Sunday", even though the latter two are not formal English and have obvious grammatical errors. A native speaker can, however, still understand the meaning of the sentence because the keywords are included in the sentence.

Therefore, a formal practice should not be the primary evaluation method in this design. Instead, the design should focus on practical language learning, focusing on the inferential comprehension of language communication.

2.5. Language Learning Competencies

Four essential skills for successful language learning include listening, speaking, reading, and writing [21]. In addition to these skills, cultural competence is also a crucial aspect of learning for an L2 learner.

The design focuses primarily on listening and speaking a language. Our communication with others in real life is usually verbally based. We listen to others, comprehend the meaning of what we heard, and generate a response in which we speak it out loud for others to comprehend. This is precisely what a traditional approach to language education cannot achieve: studying by textbooks and doing exercises usually consolidates reading and writing during the individual study, rather than verbal communication with others. It is designed to compensate for what was not achieved using traditional methods and focuses on interpersonal communication in real-life.

Culture is deeply related to language, where language communicates the beliefs, ideas, and knowledge of customs and values formed by culture [22]. Suppose the cultural context is not introduced to the learner. In that case, it will lead to misinterpretation of language itself "because language is not simply sending or receiving information, but it functions as a social behaviour in a certain cultural context" [22]. Therefore, learning about the target culture is as important as knowing the language itself. The immersive nature of the VR system and dialogue generation, which mimics

real humans, are ideal for cultural immersion. The learners can be in the virtual world where every avatar, or computer-generated personas, speaks in the language they want to learn and uses appropriate words under correct contexts. As mentioned earlier, the system can also demonstrate the underlying semantic nuances of words, which improves the learner's ability to identify and apply the right terms under correct situations.

3. Discussion

As discussed earlier, the design has many advantages in L2 learning, but it is necessary to examine its generalisability to different languages. The general strategy should be able to apply to most languages, as every language has its own culture, and learning modules can be developed for every language. However, since a primary focus of this design is cultural immersion, which differs from place to place, cultural aspects need to be tailored to every language where the system is being developed. Even if it is the same language, there might be differences in speaking and writing. For example, Simplified Chinese is used in mainland China, and Traditional Chinese is used in Taiwan. This increases the cost of developing such systems. Developers could consider either implementing the system in a popular language, such as English or implementing smaller learning modules that can be developed first, with new content added in gradually.

The design also does not have to be restricted to a realistic model. By employing the ability to create a virtual world, developers can add fantasy or fiction design elements to make the system more appealing and attractive. They can also add story plots to keep the learners engaged. As long as the principle of appropriate language in appropriate contexts, cultural immersion, and realistic dialogue generation can be achieved, the design should achieve its intended goal.

It is also possible to implement the system in two-dimensional (2D) environments to reduce the cost of development. However, if this is implemented in 2D, experience in spatial cognition and immersion could be compromised. It is up to the developers to choose the implementation of this design.

4. Conclusion

The design specified in this paper could benefit second language learners in the learning process by using a natural way similar to first language acquisition, where culture and sensemaking – developing situational awareness to understand materials, play a significant role. The primary advantage of using this system is that it provides experiences similar to real-life scenarios, which allows more accessible application of knowledge. The design concerns remembering, understanding, applying, and creating factual, procedural, and metacognitive knowledge. The affordance of the virtual environment allows designers to design engaging, immersive virtual worlds that consolidates the verbal communication skills of L2 learners. Implementing dialogue generation systems with internet access can also make artificial intelligence characters have conversations nearly as fluent and natural as real-world conversations.

Although this design is helpful for L2 learning, it can only be implemented once the dialogue generation system becomes thoroughly tested and widely used. Aside from the immersion in the target language and culture, there is also a need to consider gestures and body language when communicating. Different cultures might have different body languages to express ideas, which is challenging to demonstrate and learn using VR headsets. Further research is also needed to precisely determine whether a native speaker can comprehend a sentence with grammatical mistakes – to which extent can the sentence be altered for native speakers to understand the sentence? The parts and words which affect the meaning of sentences also need to be defined to detect accuracy in the system. This design sets a future direction for L2 learning, and once these technical issues are solved, an effective, immersive, and engaging solution can be developed.

References

[1] A. Bahari. Affordances and challenges of teaching language skills by virtual reality: A systematic review (2010–2020). E-Learning and Digital Media. SAGE Publications, 2021, 19 (2): 163–88.

[2] B. Dalgarno, M. J. W. Lee. What are the learning affordances of 3-D virtual environments? British Journal of Educational Technology. Wiley, 2009, 41 (1): 10–32.

[3] K. Schwienhorst. Why virtual, why environments? Implementing virtual reality concepts in computer-assisted language learning. Simulation & Gaming. SAGE Publications, 2002, 33 (2): 196–209.

[4] E. Bialystok. A theoretical model of second language learning 1. Language Learning. Wiley, 1978, 28 (1): 69–83.

[5] M. M. Kuo, C. C. Lai. Linguistics across Cultures: The Impact of Culture on Second Language Learning. Online Submission, 2006, 1 (1).

[6] R. Shadiev, Y. M. Huang. Facilitating cross-cultural understanding with learning activities supported by speech-to-text recognition and computer-aided translation. Computers & Education. Elsevier BV, 2016, 98: 130–41.

[7] R. Shadiev, X. Wang, Y. M. Huang. Promoting Intercultural Competence in a Learning Activity Supported by Virtual Reality Technology. The International Review of Research in Open and Distributed Learning. Athabasca University Press, 2020, 21 (3).

[8] R. Shadiev, X. Wang, Y. M. Huang. Cross-cultural learning in virtual reality environment: facilitating cross-cultural understanding, trait emotional intelligence, and sense of presence. Educational Technology Research and Development. Springer Science and Business Media LLC, 2021, 69 (5): 2917–36.

[9] H. Chen, Z. Ren, J. Tang, Y. E. Zhao, D. Yin. Hierarchical Variational Memory Network for Dialogue Generation. Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW'18. ACM Press, 2018, 1653-1662.

[10] M. Yang, Q. Qu, K. Lei, J. Zhu, Z. Zhao, X. Chen, J. Z. Huang. Investigating Deep Reinforcement Learning Techniques in Personalized Dialogue Generation. Proceedings of the 2018 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, 2018, 630–8.

[11] B. Huber, D. McDuff, C. Brockett, M. Galley, B. Dolan. Emotional dialogue generation using image-grounded language models. Emotional Dialogue Generation using Image-Grounded Language Models. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, 2018, 1-12.

[12] Q. Liu, Y. Chen, B. Chen, J.G. Lou, Z. Chen, B. Zhou, D. Zhang. You Impress Me: Dialogue Generation via Mutual Persona Perception. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2020.

[13] D. R. Krathwohl. A Revision of Bloom's Taxonomy: An Overview, Theory Into Practice, 2002, 41 (4): 212-218.

[14] H. H. Liou, S. J. Yang, S. Y. Chen, W. Tarng. The influences of the 2D image-based augmented reality and virtual reality on student learning. Journal of Educational Technology & Society, 2017, 20 (3): 110-121.

[15] A. Elliott, B. Peiris, C. Parnin. Virtual Reality in Software Engineering: Affordances, Applications, and Challenges. 2015 IEEE/ACM 37th IEEE International Conference on Software Engineering. IEEE, 2015.

[16] Dalgarno, Barney, Mark J. W. Lee. What are the learning affordances of 3-D virtual environments? British Journal of Educational Technology. Wiley, 2009, 41 (1): 10–32.

[17] B. Landau, R. Jackendoff. "What" and "where" in spatial language and spatial cognition. Behavioral and Brain Sciences. Cambridge University Press (CUP), 1993, 16 (2): 217–38.

[18] A. Higashiyama, H. Ono. "Koko," "soko," and "asoko" ("here" and "there") as verbal dividers of space. Japanese Psychological Research, 1988, 30 (1): 18-24.

[19] K. W. M. Siu. A Traditional Object in the Modern World. The Journal of Popular Culture. Wiley, 2001, 35 (3): 103–25.

[20] W. B. Hollmann. Semantic change. 2009, Chapter 35, 525-537.

[21] L. M. Sadiku. The Importance of Four Skills Reading, Speaking, Writing, Listening in a Lesson Hour. European Journal of Language and Literature. EUSER, 2015, 1 (1): 29.

[22] S. A. Soomro, B. Kazemian, I. H. Mahar. The Importance of Culture in Second and Foreign Language Learning. SSRN Electronic Journal. Elsevier BV, 2015.