Vocabulary Acquisition in New and Learned Contexts Using Immersive Virtual Reality

Thérèse Bergsma
M.Sc. Thesis
January 2022

UNIVERSITY OF TWENTE.

M.Sc. Interaction Technology

Human Media Interaction Group

Faculty of Electrical Engineering, Mathematics & Computer Science

University of Twente

Author

Thérèse S.L. Bergsma

Examination committee

dr. Mariët Theune

Human Media Interaction Group

University of Twente

dr. ir. Robby van Delden

Human Media Interaction Group

University of Twente

dr. ir. Wouter Eggink

Interaction Design Group

University of Twente

January 2022

Acknowledgements

Robby -Many years ago, I met you as a PhD-er giving their first course. There were many things that, as students, caught our attention during that time. The course was well-prepared and thought out, and most surprisingly, everything that was expected of us had been put on paper in detail. We never had to guess what an assignment was about, which was a welcome change. And whether it's a full course you've never given before or, as I've now also discovered, supervising a master thesis, your effort, enthusiasm and positivity know no bounds. It has helped me to see the potential of my chosen topic and to stay positive throughout the journey, because even at times that I wasn't all that happy with myself, you never seemed to share that feeling. And for that I couldn't be more grateful.

Mariët -Your advice and insights for this thesis were more than welcome, as it was always clear and to the point, while keeping me on track. But I'm even more appreciative of your openness and kindness. To begin with such a large task as a master thesis was daunting and difficult, and at first I wasn't able to figure out how to start. But you stayed supportive and positive, which provided me with the space to try out different approaches until I understood what to do to make it work. Later on there were some smaller struggles of staying motivated while working in a pandemic, but there was nothing but support while I also worked through that. And that really made a difference, as I went through my thesis with much more confidence and enjoyment than I otherwise would have had.

Wessel - For not only making my life so much better, but my thesis as well.

Mom - For always being there.

Abstract

There are many word aspects a language learner can learn about a word to deepen their word knowledge, but all word aspects cannot be learned simultaneously. Thus language learners need multiple encounters with a word so unknown word aspects can be added to their word knowledge or to strengthen the knowledge of previously learned word aspects. Combining vocabulary acquisition with immersive virtual reality (IVR) creates the possibility to present language learners with such multiple encounters of second language words, as the virtual environment (VE) in IVR can be quickly changed to recycle words. When recycling words in IVR it is possible to repeat target word objects in the same context (i.e. learned context), but it is also possible to change the context for each learning session (i.e. new context). To look at learning words in either a new or learned context in IVR we build a dynamic IVR system called Wics where participants learned 32 Japanese words in three learning sessions with small breaks in between. In the second and third learning environment all 32 target words presented in the first learning environment were recycled in either a new context, in which the visual representation of objects and the VE was changed, or a learned context, where the VE of the first learning environment was repeated and only the placement of target word objects was changed. Participants were tested in IVR on a posttest and one week later on a delayed posttest, which were both provided in a new context. Scores were calculated for both tests as performance by looking at correct and almost correct answers and providing them with points, and correcting for target words known prior before participating. The performance scores for the first posttest was compared for both new and learned participant groups, and the word retention for each participant was calculated by dividing the delayed posttest score with the posttest score. Participants were allowed to decide when they were satisfied with learning, so learning duration was also taken into account. There was no significant difference for performance on the posttest between conditions, nor on retention. There was also no 95% significant difference for total time duration spent in learning environments but there was a 90% significant difference seen where participants in the learned context condition learned longer. For as far as we are aware, Wics is the first IVR system to provide learners with multiple learning sessions where words are recycled, where learners have control over their own learning by being able to activate indicators for missed words, and where the posttests are also inside IVR.

Contents

B Overview visual representation of target words

C Overview VEs

Abbreviations

CALL - computer-assisted language learning

EFL - English as a foreign language

ESL - English as a second language

IVR - immersed virtual reality

IVRALL[1] - immersed virtual reality-assisted language learning

IVRALL+VA[1] - immersed virtual reality-assisted language learning with a focus on vocabulary acquisition

L1 - first language

L2 - second language

LL - language learning

MMO - massively multiplayer online

MMORPG - massively multiplayer online role-playing game

MOO - MUD, object oriented

MUD - multi-user domain/dimension

NPC - non-playable character

SIE - synthetic immersive environment

VA - vocabulary acquisition

VR - virtual reality

VRALL - virtual reality-assisted language learning

WoW - *World of Warcraft*

[1] Abbreviation introduced in this thesis.

1 Introduction

Between the time of the Renaissance and just forty year ago, a great importance was placed on specifically grammar in language learning, but slowly from that time on an awareness arose more and more that vocabulary might be important too [1], [2]. This awareness only increased further and the notion that vocabulary is important and worthy of study seems well-established around 2010 [3]. Now researchers are involved in many different ways in the complex and multi-dimensional nature of vocabulary acquisition (VA) research [4], [5].

What can be learned about a word is more than just the meaning that is connected to it. A word has its own pronunciation and a specific text representation, but to know which other words can also be used as a synonym or in what context you can expect to meet a word is also knowledge which belongs to a word. Such knowledge pieces about a word are called *word aspects*, and the word aspects that contribute to the *word knowledge* of a word were put into an eighteen word aspect framework by Nation [3]. Following Nation, if more word aspects are understood then the *depth of word knowledge* increases. Not all word aspects can be learned at once, so multiple encounters with a word are necessary to add to a learners' word knowledge. Therefore, learners should be repeatedly exposed to a word, which is called word *recycling*, to provide an opportunity to have multiple encounters. If words are recycled in the same context, called a *learned context*, then the previous knowledge of those word aspects is strengthened. A *context* is here everything that surrounds a word when it is encountered. To also add to existing word knowledge, words must also to be recycled in a *new context*, so previously unknown word aspects can also be learned.

An emerging focus within language learning is the combination of language learning with immersive virtual reality (IVR), so immersive virtual reality-assisted language learning (IVRALL). IVR is a technology which provides a person with a full 360-degrees experience by wearing either a head-mounted display (HMD) or by standing inside an encompassing area that can project images called a cave automatic virtual environment (CAVE). We make in this research a clear distinction between non-immersive virtual reality (VR) and IVR, where VR does not provide a virtual environment (VE) that encompasses the user completely, but where the VE is seen on a screen like a monitor, tablet or smartphone. IVRALL is a term introduced in this research to make a clear distinction between studies with a focus on VA and IVR (i.e. HMDs and CAVE) and studies with a focus on VA and VR (e.g. screen-based) for which the existing term virtual reality-assisted language learning (VRALL) is used.

Users in IVR describe a feeling of spatial immersion or *being there*, making it a technology that can trick the brain into thinking that it is inside a virtual environment instead of being in its actual real life location [6]. Vocabulary acquisition research with a focus on IVR (IVRALL+VA) focuses on how learning words is affected when participants are placed in such a VE in IVR. When an object is perceived as being close to a person, then the object is stored in memory as a *simulation* which holds multiple modalities of sensory information[7]. For example how an object feels, how it looks and how your body would feel when interacting with it. Learning words in IVR connects second language (L2) words directly to such a simulation and can create a stronger connection to the learned L2 word which can help with word retrieval.

The most common approach that is seen in IVRALL+VA research is an exploration-based approach where the VE is explored to learn words. In the learning phase many objects are placed around the VE. By coming close or interacting with an object a second language (L2) word aspect can be learned. This L2 word aspect can be either the textual form, pronunciation or both. Some IVRALL+VA studies also have, next to an IVR condition, a non-IVR and more traditional condition to compare results. For example, by providing the first language (L1) word on paper and expecting the written L2 word in return. Both non-IVR participants as IVR participants can provide throughout different studies either a better score for the posttest when compared to the other condition, but research that also looked at retention found that IVR conditions often have almost no or little retention loss, while non-IVR conditions scores drop rapidly over time [8], [9]. Combining IVR with vocabulary acquisition has therefore the possibility to be beneficial for retaining learned words.

IVR has different characteristics that create possible opportunities for an L2 learner, like creating a feeling of interacting with an object inside the VE or conversing with an NPC to counter learner anxiety [10]. IVR is also able to quickly bring variations into a VE or to change the whole VE. However, all found IVRALL+VA research work per condition with one VE for each participant, which means that participants are all placed in a new context, but not more than once. Building on the earlier mentioned necessity for word encounters and recycling for adding to the depth of word knowledge, we investigate in this research if a learner would benefit from either repeating the same context again as a learned context, or from continuing on to a new context. To understand the effect learned and new contexts could have on specifically IVRALL while both contexts provide the exact same word aspects, the following research question is formulated:

> What are the effects of recycling words in IVR in learned or new contexts for retrieving words when encountered in a new context?

The context is here defined as the virtual environment and the visual representation of words as objects.

To allow the recycling of words in IVR in new and learned contexts, and its evaluation in a new context, an experiment was set up for which a system was built called *Wics*, which is derived from *words in contexts system.* For the experiment, participants go through three learning environments in IVR where they recycle 32 words in each environment, followed by a test environment in IVR with a posttest, and a week later a test environment in IVR with a delayed post to measure retention. For the learned context condition, users encountered the same uninhabited island VE as learning environment, where all target word object visualisations were also the same. For the new context condition, users encountered for each learning environment a different VE, namely an uninhabited island VE, a bedroom with garden VE and an apartment VE, and different word object visualisations. The posttest and delayed posttest VEs provided always a new context in both conditions, with a theatre VE for the posttest and a barn with a bar VE for the delayed posttest. Scores were calculated for both tests as performance by looking at correct and almost correct answers and providing them with points, and correcting for target words known prior before participating. The performance scores for the first posttest was compared for both new and learned participant groups, and the word retention for each participant was calculated by dividing the delayed posttest score with the posttest score.

This research is structured as followed. In Chapter 2 we look specifically at VA in language learning and its history and theory, while taking a closer look at how IVR can provide a contribution to VA. In Chapter 3 we look at the origins of IVRALL, which lies in computer-assisted language learning (CALL) and where VEs where first used as environments in which language learning takes place. We also discuss commercial VRALL and IVRALL games for VA. In Chapter 4 we discuss contemporary IVRALL+VA studies and their approach for presenting words and their evaluations. We conclude the chapter with observing that words are presented per condition once in a VE in contemporary IVRALL+VA studies, but as learners need multiple encounters with a word, propose to look at the effects of recycling words in new and learned contexts. In Chapter 5 we present an experiment plan to study the effects recycling words in a new or learned context for retrieving words in a new context. In Chapter 6 we discuss the requirements for a system to perform the experiment and detail design choices and implementations. In Chapter 7 the method is discussed for doing the experiment and in Chapter 8 the results are stated. In Chapter 9 the results and the limitations are discussed, followed by suggestions for future work and a conclusion.

2 Background

The field of VA focusses on the processes that are involved in learning lexical items, so not only single words but also words that are often found together (i.e. formulaic language), where researchers look at how, when and what is learned and at how a mental lexicion is constructed and employed [4]. IVR places a user inside a virtual constructed world that can activate parts in the brain that are consistent with brain activity in real reality [11]. Understanding the concepts behind VA and behind IVR makes it possible to discern the possible benefits and options for combining the topics of VA and IVR. For example, one encouraged option in vocabulary acquisition and language learning in general is *total immersion*, where the learner places themselves in the L2 culture. IVR cannot attain total immersion, but can provide *spatial immersion* in a simulated environment [12], which opens up possibilities for language learners to utilize. The theory behind both VA and IVR is also relevant for the requirements and design choices that are discussed in Chapter 6.

2.1 Vocabulary acquisition

The field of VA was before 1980 of a small scale and largely neglected, as grammar was considered the important and difficult aspect of language learning, which made it not surprising that there were no clear theories regarding the workings of VA at that time [1]. However, a few researchers saw VA as a necessity for good language acquisition and advocated for studying VA as a separate and serious field to gain an understanding of the workings of VA [1], [2], [13], [14]. This call found a resonance in other researchers and the field of VA is now well-established with a central role in language learning [4] with its own well-established concepts as *word knowledge*, *mental lexicon* and *recycling*.

2.1.1 History

The earliest record of human curiosity about learning another language dates back to around 1600 BCE during the ancient Mediterranean world and a word list from that period which compares Sumerian and Akkadian morphological facts [15]. An interest in language was further seen in the philosophical scholarship of the ancient Greece with the introduction of a descriptive metalanguage of which the developed concepts were later exploited by the Romans. However, the Romans also looked at a more practical approach to language learning with a focus on educational goals, as the Romans had substantial direct experience with (foreign) language learning. It is speculated that the Romans also placed great emphasis on vocabulary [16], as rhetoric was held in high esteem. In the medieval period the focus of language learning, where most students studied

Latin, shifted more strongly to grammar, a trend that continued during the Renaissance. Although several persons and different movements tried to establish the importance of vocabulary acquisition during those times and later on, the emphasis in language learning remained strongly on grammar and vocabulary was not addressed in a principled way.

Levenston [2] suggested in 1979 that the neglect regarding vocabulary acquisition could almost be called discrimination, as research frequently used the term *language* while only *grammar* was meant. Meara [1] followed this observation of Levenston by stating in 1980 that there are no clear theories regarding vocabulary acquisition and that the level of research activity is fairly low, small-scale and largely neglected by, for that period, recent developments in research. Meara found this negligence especially remarkable because vocabulary was identified by learners as their greatest single source of problems in their later learning stage. It was not that there was no focus at all on vocabulary, in contrary, Meara identified two major areas of research into the field, namely (i) vocabulary control and selection, where it is attempted to justify the selection of vocabulary items on the basis of frequency counts, and (ii) mnemonic techniques, where L2 words are connected to keywords, for example phonetically similar L1 words, to help with learning. However, Meara's critique regarding the focus of (i) is that the work concentrates more on the management of learning, than on the actual learning process, because the work focuses on deciding what should be taught, and not on how such words are actually learned. He also concluded how the work is based on a whole set of assumptions of which the validity has never been called into question. The identified problem by Meara for (ii) is the treatment of vocabulary items as a simple problem that can be helped with only pairing words with their translation equivalents, which ignores the already existing knowledge that learning vocabulary is more than a simple word pairing, as semantic relationships are built up in complex patterns when learning a language, and languages rarely map their lexical items onto each other. Meara proposes to first look at the larger body of experimental work that focuses on the mental dictionaries of bilingual speakers, due to the deficiency on the topic for language learners, but also identifies topics as a starting point for future research into vocabulary acquisition.

Meara and Levenston were not the only researchers who started to identify the need to study the processes of vocabulary acquisition around that time (e.g. [13], [14]), and slowly the interest in vocabulary increased, with Laufer [17] noting in 1989 that vocabulary could no longer be seen as a *victim of discrimination*, as research questions addressing vocabulary acquisition were growing in numbers while also addressing different aspects of vocabulary learning. However, it was not until publishment of the book *Teaching and Learning Vocabulary* in 1990 by Nation [18], and the introduction of a principled and systematic approach to vocabulary instruction, that the interest in vocabulary started to gain real momentum [4].

The emphasis in language learning was for a long time only placed on grammar [19], but with the emergence of more research into how vocabulary acquisition works, a new line of thought emerged around 2000 that grammar and vocabulary are not two separate entities, but that they are fundamentally linked [16]. Awareness was raised that learning another language could not be successfully acquired without addressing both areas explicitly. Folse [19] also argues that vocabulary might be even more important than grammar when conversing in another language, as a lack of grammar knowledge might hinder a conversation, but not knowing the words can also stop a conversation.

The notion that vocabulary is important for language learning and is worthy of detailed examination seems well-established around 2010, with vocabulary learning being called an essential part of mastering a second language [20], being just as important as the acquisition of grammar [21], and being called the *building blocks* upon which language learning largely depends [22]. Nation also calculated in 2013 that 30% of all research from 1900 until 2012 on language learning has appeared between 2001 and 2012, and notices how there is now an international community of vocabulary researchers and how research on vocabulary is clearly alive and well.

Vocabulary acquisition has now in contemporary research a central role in language learning [4] and its complex and multi-dimensional nature are a topic with which many researchers are involved [5].

2.1.2 Theory

L2 learners need many words to communicate successfully in everyday informal situations, but also to read texts like newspapers or novels [4]. There is not enough research to determine exactly how many words are necessary for each activity, but most studies find that for successful conversations somewhere between 95% and 98% of the vocabulary in a conversation should be known, and that around 98% of the words in a text must be understood for full comprehension of the text. For English this would result for communication in needing to know somewhere between 2,000 and 3,000 word families, where one word family consists of all forms that share one core meaning, to cover 95% and between 6,000 and 7,000 to cover 98%. For reading it would be as high as 8,000 or 9,000 word families to provide 98% coverage [20].

L2 learners can learn words explicitly or implicitly, where explicit learning is a more conscious operation where the learner focuses on specific words to study, while implicit learning happens naturally and without conscious operations [23]. Study results regarding how many written encounters a learner should have with a word before it is learned implicitly vary widely, but most numbers are located somewhere between five and sixteen exposures [22], although repetition is not the only factor to take into account, as similarity to words in L1 and

TABLE I

Framework of the aspects involved in knowing a word [3], where R = receptive knowledge and P = productive knowledge

Form	spoken	R	What does the word sound like?
		P	How is the word pronounced?
	written	R	What does the word look like?
		P	How is the word written and spelled?
	word parts	R	What parts are recognisable in this word?
		P	What word parts are needed to express the meaning?
Meaning	form and meaning	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	concepts and referents	R	What is included in the concept?
		P	What items can the concept refer to?
	associations	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
Use	grammatical functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	collocations	R	What words or types of words occur with this one?
		P	What words or types of words must we use with this one?
	constraints on use	R	Where, when, and how often would we expect to meet this word?
		P	Where, when, and how often can we use this word?

the meaningfulness of the context also affects learning [24]. Extensive reading can lead with implicit learning to many high-frequency words being taught, but since many incidental encounters are necessary it is more difficult to also learn low-frequency words through implicit learning [4]. It is therefore necessary that the L2 learner also learns vocabulary explicitly.

As with implicit learning, it is also important for explicit learning that a learner has multiple encounters with a target word. Initially so the L2 learner can establish the form-meaning link, which enables the learner to see the written form or hear the spoken form of a word and to know the meaning that accompanies that form [20]. However, the form-meaning link is only the starting point of the knowledge a learner can have about a word. Other knowledge is for example how a word is pronounced, knowing the opposite word or equivalent words, knowing which words typically accompany a word, understanding how the meaning changes in different contexts, etc. This *word knowledge* that accompanies a word was put in a framework of eighteen aspects by Nation [3], as can be seen in Table I, where for each aspect it is also specified if it helps with the receptive knowledge of a word (i.e. reading and listening) or the productive knowledge (i.e. speaking and writing). Multiple encounters are also necessary to gain more word knowledge aspects and to deepen the understanding of a word. This deepening of word knowledge by understanding many aspects of a word is called *depth of word knowledge*, while a limited understanding of a word, for example by only knowing the form-meaning link, only pertains to the *breadth of word knowledge*.

The order in which each word knowledge aspect is acquired is not a fixed process and is different for each word and each learner, making vocabulary acquisition a dynamic process. VA is also incremental in nature, because it is necessary that learners have multiple encounters with a word, as it is not possible to learn all word knowledge aspects simultaneously [16]. Therefore it is also important that words are recycled during explicit learning and that it is possible to have multiple encounters with a word. This variable process of vocabulary acquisition makes it difficult to examine the links that exist between words in a mental lexicon (i.e. the mental dictionary where words are stored connected to each other in an intricate system), which results in a lack of generally accepted theory on how the mental lexicon is built and how it functions [4].

2.2 Learning conditions

Words can be encountered in a *new context* or a *learned context* [25], where the former happens when a learner has not seen a word in the same context before and the latter occurs when the learner has previously encountered and processed the word in the same context.

Text-based learning in VA is one type of learning that most often revolves around learning in a learned context since it is based on rote associative memorization learning. The most basic form of text-based vocabulary learning in explicit learning is reading the L1 word and the learner tries to produce the L2 word from memorisation, or vice versa. The first time the learner encounters such a combination the context can still be considered a new context, but most text-based learning exercises will repeat this process so the context will become a learned context for the learner and the learner will eventually be able to provide the L2 counterpart or the L1 translation. The context in which this takes place can be considered a learned context because the information around the L2 word does not change. The context of the L2 word here is, in case of word-to-word learning, a single written word or formulaic language (i.e. types of vocabulary which operate as multiword units, where the meaning cannot be derived without the words being together, e.g. *high five*) and how it looks written, and the learner is offered that same context each time the word is provided. It is, however, possible to make changes to the traditional text-based learning and to also create a variant that is new context oriented. For example, with picture-word association, it could be possible to change the picture symbolising the word each time it presents itself to the learner, which creates for each encounter a new context for the target word. One of the most common methods for implicit learning during text-based learning is reading a book and guessing the unknown words that are encountered.

A counterpart to text-based learning is *situated learning*, where words are provided in the context in which they can be applied or encountered, so during real-life situations or social interactions [26]. During explicit learning these contexts are created specifically for L2 learning, for example, by trying to present real-life-like situations. Examples are showing a video of someone assisting an-

TABLE II
Overview

	implicit learning	**explicit learning**
text-based learning	reading/listening to non-education specific material common context: new	rote association memory exercises common context: learned
situated learning	total immersion (e.g. visiting the L2 country) common context: new	fabricated contexts for L2 practice common context: new

other person and then encoding that scene with the corresponding L2 word *to assist* [25] or trying to create an authentic context environment for the learner to participate in, like showing drama scenes and asking the learner to participate by showing how they wished a character would act [27]. Situated learning goes often together with learning in a new context, as most authentic-like contexts offered are used once and then followed with a new context. This especially occurs with situated learning while being in the country where the L2 language is spoken (i.e. *total immersion*) and the learning is implicit, as most situations will be unpredictable and new. However, as with text-based learning, situated learning can happen in both contexts, for example, if the same video from [25] would be shown not once but multiple times to encode the word *to assist* then the word would also be learned in a learned context. An overview of one example implementation for each learning combination and the more common context it is used in, so in either a used context or a learned context, can be found in Table II.

Note that no combination is named *the best* and that no suggestions are made to choose for successful L2 learning a certain method, learning type or context over another or in a specific combination. Different encounters are necessary to learn different aspects of a word to acquire more word knowledge depth, so L2 learners are encouraged to have multiple encounters in a variety of ways, and each combination of learning can provide such a variety. Recommendations should also fit within the learning strategy of a teacher or within the personal goals of an L2 learner, providing a second reason that naming a strategy the best is undesirable. The focus in this thesis lies on explicit situated learning in both new and learned contexts, but it is desirable to emphasize that all learning combinations can be valuable for language learning. However, before fully focusing on explicit situated learning it is also worthwhile to look at implicit situated learning which explicit situated learning tries to emulate.

2.2.1 Research regarding new and learned contexts

Jeong *et al.* [25] also performed a study regarding new and learned contexts, where they looked at situation-based learning using new contexts, and text-based learning using learned contexts. Their situation-based learning condition consisted of 5 second videos that visualised the target word in action and their text-based learning condition consisted of a video of a person holding up a piece of paper with the target word for 5 seconds. Each target word had multiple situation-based videos where the actor and location of the videos changed, creating a new context for the target word each time it was encountered by a participant, while the context did not change for the text-based target words. All participants tested their knowledge in a post-test by doing both a text-based learning test and a situation-based learning test with a video that they had not yet encountered. The text-based participants performed exceedingly well when doing the text-based test, where they saw their learned context, but performed badly when they had to recall the word through a new context in the situation-based test. The situation-based learners did relatively well in both learned and new contexts. The abundance of contexts associated with the target words might help with comprehending the word when encountering it in a new situation.

2.2.2 Total immersion

The most complete implementation of implicit situated learning is *total immersion*, where the learner travels or moves abroad to be situated in the L2 culture, walk the local streets and to interact with the people living there [28]. Total immersion is often seen by teachers as the highlight of students' careers where acquired knowledge becomes immediately relevant and connected to lived experience [29]. Likewise, the hopes and expectations of L2 students often revolve around the envisioned necessity to speak L2 during total immersion in its natural context which should force the L2 learner to adapt and become fluent [28]. In these hopes and expectations the assumption lies that the L2 learner will take full advantage of the offerings of the L2 culture to gain an immediate and beneficial effect on their language proficiency, but also to gain a deep L2 cultural understanding. However, students returning from total immersion programs were less accepting of these assumptions, as their experience taught them that the actual total immersion experience can be very different from the beforehand fantasized experience, and researchers are now not only emphasizing the favourable learning outcomes of total immersion, but are also studying the nature of the immersion context itself and possible impediments to L2 learning.

Two primary impediments to utilizing the opportunities of total immersion are (i) motivation and (ii) social personal risk. Motivation is seen as extremely important for L2 learning and directly influences how often students interact with their L2 language [30]. Students with high motivation in total immersion make an effort to go out and explore the environment, interact with native speakers and are actively trying to incorporate new words or phrases that they encountered into their own L2 speech [31]. In contrast, unmotivated students are insufficiently involved which hinders them in developing the potential of their L2 skills [30]. Additionally, learning is also obstructed if L2 learners feel a social personal risk, which can occur when an L2 learner believes there is a contrast between their actual self and how the L2 learner thinks their self is seen by native speakers, resulting in not only feeling misunderstood linguistically but also personally [32]. This perceived contrast can arise when the L2 learner believes that they cannot express their true thoughts or sense of humour sufficiently, combined with the idea that as a result native speakers think the L2 learner is childish or stupid. Such social personal risks can result in a low self-esteem or anxiety which can hinder L2 use and subsequent L2 learning.

Mendelson [28] therefore recommends to inform L2 learners about the challenging nature of total immersion before leaving, to offer guidance and support throughout the actual experience, and to put the experience into perspective at the close of the venture. Savage and Hughes [31] conducted a study where L2 learners were able to make use of the possibilities created by total immersion. The results of the study showed a statistically significant improvement in listening and reading from tests before and after the total immersion experience of the learners. However, it was also found difficult to maintain their initial language proficiency gain once they returned home.

Total immersion can give an L2 learner a language proficiency gain due to (i) the implicit learning aspect and (ii) the situated environment in which it takes place. Total immersion is implicit because interactions and encounters happen naturally without a preconceived educational setup. This unpredictability makes it impossible for the L2 learner to prepare for what they will hear and wish to say, resulting in new word encounters and situations that differ strongly from the classroom environment. Wilkinson [33] did note that L2 learners staying with a host family tended to still employ classroom norms while being outside the classroom environment, so Wilkinson recommends a focus by teachers on behaviour that is often ignored during classroom sessions, like using non-interrogative topic initiators, to enable students to express themselves during total immersion more like they would have in their L1.

2.2.3 Situated environment

Embodied cognition theory revolves around the notion that there is a close relationship between the sensorimotor system and cognitive processes, and that words and concepts have multimodal representations in the brain [34], [35]. The multimodal representations are acquired with objects that are close to a person, and are processed in multiple modalities of sensory information [7]. For example, a *permanent marker* has a perceptual component (e.g. how a it can look, how a person using it looks), a somato-sensory component (i.e. how the different materials feel), and a motor component (i.e. what a body does with it), but also a smell and sound component and more [34]. All these different components, when encountered through an object close to a person, contribute to the multimodal representation of a word or concept in the brain. These multimodal representations are stored in memory and when later there is a need for that knowledge, for example when reading a text, then the brain reactivates these multimodal representations by simulating how it represented the perception, action, and introspection of that word when it was encountered in real life [35]. Such a reenactment in the brain of previous perceptual, motor, and introspective states is called a *simulation* or *semantic representation*, and simulations appear central to the representation of meaning. Simulations are already in place for L1 words, but there is an added value for language learners to also have simulations directly connected to L2 words. Text-based learning often first creates a link between L1 and L2, but situated learning can be helpful here to create a direct link between an L2 word and a simulation [36], to help with word retrieval. However, a situated learning variant such as total immersion, where real objects are close by for creating multimodal representations, is often not a feasible option for L2 learners. IVR can provide, as a digitalised environment, an alternative for learning in total immersion, but also has its own unique properties that can have an added benefit for language learning.

2.3 IVR

An IVR environment can create a feeling of spatial immersion [12], which occurs when the brain is tricked into believing that what it sees is actually present [37]. This activates the visual and motor channels of the brain which also allows for the creation of simulations, making it possible to store words as multimodal representations when learning them in IVR. Thus VA in IVR can act as a variation of total immersion where it is possible to train at any time [38] and by anyone, as people for whom travel is difficult, impossible or unrealistic can obtain spatial immersion through IVR [39]. As IVR is more suitable for explicit learning than the implicit learning that goes with total immersion, there are also possibilities to personalise language training [11].

Motivation is seen as an important necessity for L2 learning [30], where motivation is seen as a key feature that is supported by IVR [11]. Hastings and Brunotte [40] also found that their students immersion into IVR made them excited which increased motivation to engage with their task in IVR, similarly to Alfadil who noted that students were excited to go back into IVR for language learning. Another explanation for students staying on task and focused is that the only visual input in IVR is the learning environment [40], where the learner is isolated from the real world and its distractions [42]. IVR might also be an opportunity for people with language anxiety, where people feel anxiety when they try to practice the L2 language with real persons. Language anxiety has a negative influence on the motivation for learning and disrupts the communicative process that leads to L2 development as their is a lower willingness to communicate in L2 [43]. In IVR learners could be in a safe context without the perceived social personal risk to lower their anxiety and thus improving their performance [44]. IVR might also aid total immersion, where IVR could prepare L2 learners for the locations they are about to visit and to decrease anxiety [40]. For example by learning how to find important buildings or to see how a possible home-stay family's residence could look like before actually arriving in a similar environment.

Lan [44] states that there are many potential benefits for language learners from learning with IVR, but noted that more studies should be conducted to look at those possibilities and challenges for IVR when combined with language learning. Especially now IVR sees an increase in popularity with affordable IVR equipment as Google Cardboard, where a cell phone can be used to provide the IVR images, and becomes thus more accessible. They also noted that participants of IVR studies should be more diverse than they currently are, as most IVR studies are done with university students, so IVR studies should include more age ranges, and individual differences and motivation levels should be taken more into account.

A possible disadvantage of using IVR is the chance of users experiencing symptoms of motion sickness, where symptoms can include disorientation, nausea, eye fatigue and more [45]. Users with no prior experience in IVR report greater discomfort and have poorer task performance in IVR than users who have repeatedly experienced the same IVR content.

2.4 Conclusion

A word consists of more than the meaning that is expressed with it. How it sounds, what synonyms are of the word, or which words are often grouped together with it are more examples of what makes up a word. Nation made a framework with 18 of such identified word aspect. Encountering a word multiple times helps to strengthen the word aspects that have been previously seen and to see new word aspects of a word to deepen the word knowledge of that word further. Therefore it is important that a learner is repeatedly exposed to a word, which is also called word *recycling*, and thus have multiple encounters with words.

Words can be encountered during explicit learning and implicit learning in combination with either text-based learning or situated learning and where the words are encountered in either a new context or a learned context. If a learner is purposefully learning specific words or does an exercise that is designed to teach specific words, then learning is explicit. If a learner has no control over the specific words that they get in contact with or if there is no learning goal, then the learning is implicit. Text-based learning often includes rote association memory exercises, for example when learning with flash cards. Situated learning takes often place during social interactions or other real-life-like situations. The distinction between a new and learned context is dependent on if a word has been encountered in that specific context before. If this is the first time that a word is seen in a context, then the context is new, but if it is not, then the context is learned. Common combinations in VA are explicit text-based learning in learned contexts and implicit situated learning in new contexts.

Total immersion is often seen as one of the most worthwhile methods for language learning due to its richness of interaction possibilities with L2 and the L2 culture and how it allows learners to create multimodal representations. However, downsides to the method are also mentioned. Learners must be motivated to learn the language while being on location and need to battle possible hesitations caused by a feeling of social personal risk to experience clear benefits. Furthermore, total immersion can also cost relative more time and money than other VA strategies. It is also recommended to first mentally prepare before making the journey, and to try to keep the language proficiency gain up after returning to avoid losing the gain they achieved during total immersion.

A strongly related method to total immersion is the use of IVR for VA, where users are spatial immersed which enables them to also create multimodal representations for words when learning. Some of the advantages of learning in VA in IVR are the time needed to make use of IVR, the lack of social personal risk, fewer distractions to get off task with a headset on, creating motivation to learn by being considered fun to pertain in, and its increasing availability. IVR might also help with preparing learners for their total immersion trip or to help with keeping up the language proficiency gain achieved during such a trip after the learner has returned.

Researchers into IVR and VA are relatively new and upcoming and are further discussed in Chapter 4, but the concept of learning words in a virtual environment started much earlier within the research field of CALL. We discuss CALL in the next chapter together with existing VR and IVR commercial programs that focus on VA, to understand the workings of VA in IVR better and for possible design choices.

3 Related work

IVR research on language learning originates from much earlier research on different and non-immersive technologies, starting with text-based 2D VR and followed by 3D graphics VR. Building language learning projects is, however, not only done by academic researchers, and VA games are also commercially available.

This chapter starts in Section 3.1 with an explanation of why only research and applications with a VE are included in this thesis, followed in Section 3.2 by a short history into earlier relevant research fields with a focus on VR. The commercial side of language learning games in VR is discussed in Section 3.3, while language learning games in IVR are discussed in Section 3.4. The first research project that combined IVR with language learning is then discussed in Section 3.5, and a short conclusion of this chapter follows in Section 3.6.

3.1 Virtual environments

This thesis focuses on spatial immersion, which can be obtained in IVR, and potential benefits of IVR features for language learning. To be immersed implies that the immersed person is surrounded by something. During total immersion a person is surrounded by L2 speaking people and the L2 culture, while during spatial immersion in IVR a person is surrounded by a virtual environment. Such a VE is not exclusive to IVR, and this thesis follows the definition provided by Schroeder [46] for *virtual environments* and the technology that displays it:

> a computer- generated display that allows or compels the user (or users) to have a sense of being present in an environment other than the one they are actually in, and to interact with that environment.

So a VE provides a user with a strong sense of *being there* [47], and it is the VE in which users are immersed. Because spatial immersion, and therefore VEs, are a core aspect of this thesis, we exclude in this thesis research, projects and commercial applications that make no use of a VE or that have no possibility for interactions. Therefore well-known language learning applications like *Duolingo*[2] are not included, while research where only videos are played in IVR[3]

[2] *Duolingo* offers lessons via mobile application and their website for different languages. These lessons resemble a quiz where the user either selects a multiple choice answer, types in the answer, or uses voice recording to provide the answer. Lessons can focus on pronunciation, forming phrases by ordering words, matching words to images, or reading sentences. *Duolingo* also combines language learning with gamification by rewarding users when they learn a lesson each day, by working with an experience point system to level up, by using leaderboards, and by awarding badges for completing specific objectives.

[3] When playing a video in IVR a 360 degree movie is shown within IVR where the user can

[48] is also excluded.

3.2 From CALL to IVRALL

The research field of computer-assisted language learning (CALL) that focuses on virtual reality (VR) is called virtual reality-assisted language learning (VRALL). VRALL became slowly more prominent around the early 2000s [49] and has seen multiple focus shifts these past twenty years from text-based VR to 3D graphics VR to IVR. To make a clear distinction between VRALL with a focus on non-immersive VR and VRALL with a focus on immersive VR, the term immersive virtual reality-assisted language learning (IVRALL) is used in this thesis when discussing the latter.

3.2.1 Text-based 2D VR

The first virtual environments used for helping students with language learning were simple text-based 2D virtual environments called MUDs (multi-user domains/dimensions) or MOOs (MUD, object oriented), which are database programs that are run on a server. Everything that happens in the virtual environment is described in texts and obtained with the use of commands [50]. A MUD or MOO environment consists of multiple rooms which can contain objects, and players can travel to these rooms and interact with objects and change their appearance descriptions (e.g. edit a plant description from *a small green plan showing a budding sprout* to *a small green plant with a newly bloomed red flower*) and talk with each other. There is often also a map of the environment that is drawn with characters due to its text-based limitations, see Figure 1. One educational MOO that was created in 1994 and that is still running today, is *schMOOze University*,[4] which focuses on English as a second language (ESL) and English as a foreign language (EFL).[5]

Advantages for MOOs in combination with language learning are the feeling of safety they can provide, and that they can enhance motivation [51], that they can provide learner autonomy [49], [52], and can enable tandem learning[6] [54], which are also advantages that continue to be listed for 3D VR and IVR. Another MUD-specific advantage, which has now become outdated as an ad-

turn around to watch the video from different degrees but where other types of movements are not registered. The user also cannot interact with the video.

[4] http://schmooze.hunter.cuny.edu/

[5] The (British) term EFL is used when English is taught to a non-native English speaker in a non-English-speaking country, and the (American) term ESL is used when English is taught to a non-native English speaker in an English-speaking country.

[6] *Tandem learning* is autonomous learning between two persons, where one person has as native language the language the other person wants to learn, and vice versa and who practice talking in both languages with each other [53].

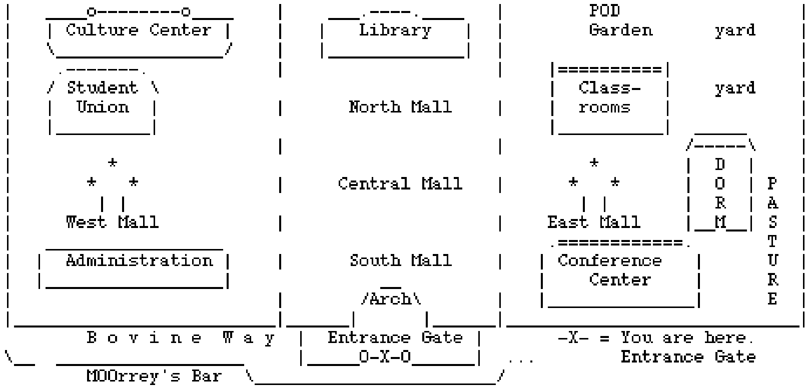


Figure 1: Map of *schMooze University*.
Source: https://pages.uoregon.edu/dhealey/ed596/schmooze.html

vantage, is the low bandwidth that it uses[7] in comparison to other non-text based or mixed media forms, which made it more reliable, and the low cost [51]. These advantages resulted in MOOs being studied for several more years, even though 3D graphics were becoming rapidly more common, before the focus shifted strongly to 3D VR environments.

3.2.2 3D graphics VR

A strong focus shift towards 3D graphics VR happened in the late 2000s when virtual environments became more interactive and the graphic quality became more visually appealing [56]. Especially online virtual worlds warranted significant attention. Sykes *et al.* [57] name three online virtual worlds as particularly interesting to language learners: (i) open social spaces, (ii) massively multiplayer online (MMO) gaming spaces, and (iii) synthetic immersive environments (SIEs). The most popular of (i) used in education at the end of the 2000s is *Second Life* (2003) [47], a virtual world that tries to resemble the real world and that has no set objective for its players, and is still used in contemporary research [58]–[60]. Reasons for its popularity are the possibility to build complex objects and environments, and its low cost of entry [47]. Research with open social spaces such as *Second Life* often focuses specifically on the social aspect that such a virtual world can provide, and can consist of quest-like projects where players have to solve a problem together which stimulates communication and enables learning through discovery learning [61], studying listening comprehension [62], and doing activities together to enable tandem learning [63].

[7] The adjective *low* is relative. In the early nineties it was not uncommon for universities to restrict access to ports that were accessed by MUDs or MOOs to prevent them from taking up too much Internet access [55].

The MMO that is most often used for studies into (ii) is a massively multiplayer onlne role-playing game (MMORPG) called *World of Warcraft* (WoW) (2004) [64], which is the most popular MMORPG today with an estimated 5 million active players each month[8] and is currently available in nine languages.[9] Sylvén and Sundqvist [65] argue that games like WoW can accomplish fundamental L2 components such as immersion, authenticity (i.e. not adapted to cater to wishes of L2 learners) and motivation. While playing, learners will be setting self-directed in-game tasks which creates the opportunity for self-learning. During such in-game tasks learners will interact with NPCs and the environment, resulting in a constant exposure to the L2 language, with many words being recycled within the game [66]. Next to interacting with NPCs the game also strongly encourages players to interact with each other by including strong enemies that need a party of players to be defeated [67] and by inviting players to communicate with each other to discuss tactics.

The language learning benefits of MMOs like *World of Warcraft* and social worlds as *Second Life* are sought after the creation of the platforms and games, which is contrary to synthetic immersive environments, (iii) of the online virtual worlds mentioned by Sykes *et al.*, which are specifically developed with their educational purpose in mind [56]. SIEs can therefore integrate the benefits of online gaming while developers target specific skills and educational objectives [57].

3.3 Commercial VRALL games

There are many research projects that look at the educational potential of existing virtual environments or that create SIEs of their own for such purposes, but there exist also a few commercial games that are developed for the purpose of language learning in a VR virtual environment. Two of these VR games are *Influent*[10] and *Lingotopia.*[11]

Influent focuses on vocabulary acquisition and pronunciation. It puts the player in an apartment with only a few rooms, see Figure 2, where the player can click on items to learn the accompanying word. There are mini-games, for example where you have to find a number of words in a specified time, and you fill a booklet with the words that you have learned. There are 420 words to learn in total.

In *Lingotopia* you are placed in a city, see Figure 3, where inhabitants only speak the language you wish to learn. Some residents teach you the meaning of a word, which lets you slowly built a vocabulary. Once a word is learned, you can hover

[8] https://activeplayer.io/world-of-warcraft/

[9] https://wow.gamepedia.com/Localization

[10] http://playinfluent.com/

[11] https://store.steampowered.com/app/860640/Lingotopia/

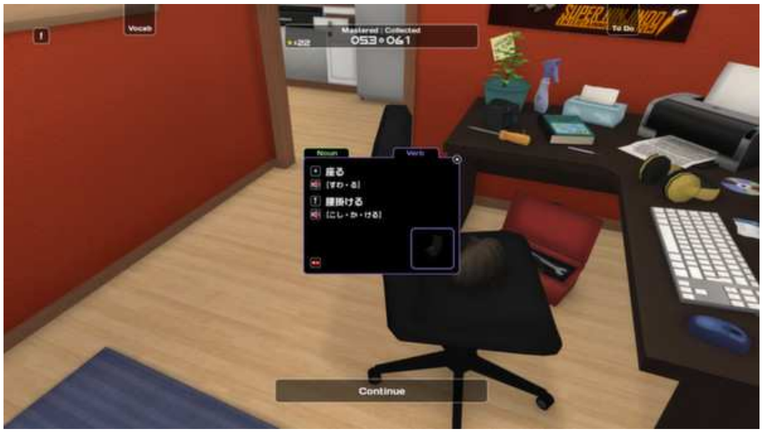


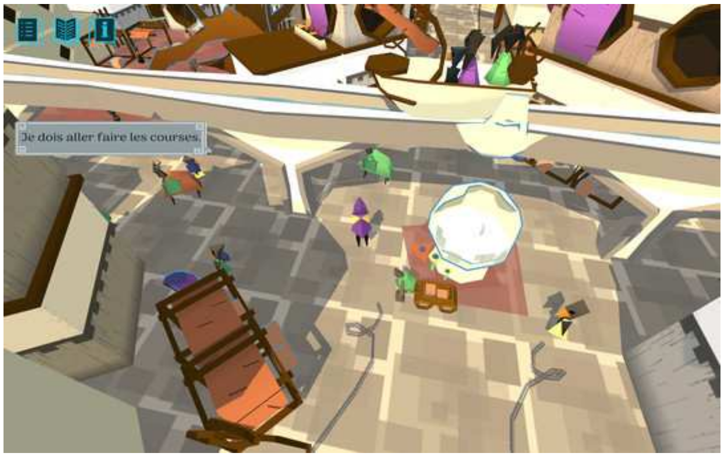
Figure 2: Screenshot from *Influent*.
Source: https://store.steampowered.com/app/274980/Influent/

over it to see its meaning when you read the text of a non-playable character (NPC). This should enable the player to, in the end, understand everything that is being said in the city.

3.4 Commercial IVRALL games

There are a few different language learning games also published for IVR, that focus on different language learning aspects. *House of Languages VR* is an IVR game for the *Samsung Gear VR* and *Oculus Go* with a focus on vocabulary acquisition. Players are placed in a VE (e.g. airport, cafe, cinema, zoo, museum) and a raccoon NPC named *Mr. Woo* asks players to find specific objects in the VE. Players can meanwhile look around and hear and read the L2 word for each target word in the VE. Applications such as *Mondly VR*, *VR Speech*,[12] and *busuu's Spanish Learning Game*[13] focus more on conversing in L2 by placing the player in a VE in scenarios (e.g. eating in a restaurant, taking a cab) where they can interact with NPCs. Speech recognition allows the applications to comment on the player's pronunciation and to determine the follow-up comments of NPCs. Open social spaces, like *Second Life* for VR, now also exist for IVR and are called virtual reality social networks (VRSNs) [68]. One such VRSN is *AltspaceVR*[14] which is not designed for educational learning, but where individuals do host events from time to time for people who want to practice their language skills with each other.

[12] https://www.vrspeech.app/
[13] https://www.oculus.com/experiences/go/1644221912279007/
[14] https://altvr.com/



Je dois aller faire les courses

Figure 3: Screenshot from *Lingotopia*.
Source: https://store.steampowered.com/app/860640/Lingotopia/

3.5 An early IVRALL project

Zengo Sayu[15] is a project from 1995[16] by Rose and Billinghurst [69], and likely the first to combine language learning with IVR. Its IVR environment is designed to teach Japanese prepositions to students with no prior knowledge of the language and consists of a Japanese house with furniture and several boxes and orbs. These shapes have different colours and by touching an object a student will hear the name and, if applicable, colour of the object. The relation between these two objects (e.g. *the red box is next to the blue box*)[17] is spoken in Japanese when the student touches two objects. Next, animation sequences are shown in combination with the accompanying voice command (e.g. a yellow box that is placed under a black box and the accompanying Japanese voice command *put the yellow box under the black box*)[18]. After seeing all sequences the student will hear a command and is expected to manipulate the environment to follow the command. Lastly, the student will see a stack of blocks which they must match with blocks of their own, but they can only build by providing voice commands. They can, however, still point at objects and ask in Japanese *what is that*[19] when they forget words or prepositions.

Rose and Billinghurst were far ahead of their time, as they noticed that there were no educational IVR applications for foreign language learning and asked

[15] The project name *Zengo Sayu* originates from the Japanese word 前後左右 (*zengosayuu*) which can be translated as *in all directions*.

[16] https://www.youtube.com/watch?v=oPu0Hn4Sjgs

[17] *Akai hako wa aoi hako no tonari ni arimasu.*

[18] *Kiiroi hako o kuroi hako no shita ni oite kudasai.*

[19] *Are wa nan desu ka.*

for greater research into the efficacy of IVR for language education, where *Zengo Sayu* could be the first to address that deficiency. They listed as possible advantages addressing learner's anxiety, a higher level of immersion and stimulation, and giving access to both the meaning and the experience to develop an intuitive understanding of abstract concepts. However, for years no one within CALL continued in that direction and instead, as described earlier, started with exploring 2D environments which was then followed by researching 3D environments. The exploration into CALL with IVR (IVRALL), both with and without a focus on VA, started only recently in the second half of the 2010s, so twenty years after *Zengo Sayu*.

3.6 Conclusion

Virtual environments in the early days of CALL emerged as simple 2D text-based VR VEs before evolving to more complex 3D graphic VR VEs due to hardware upgrades and being able to create a more immersive VE with more advanced input options. A similar trend can be seen when comparing 3D graphics to IVR, since IVR is able to create an even more immersive VE with even more advanced input options. However, this should not diminish the value 3D graphic VR VEs still have for language learning, as purchasing IVR hardware can still be considered expensive and has a high chance of causing nausea with new users. Furthermore, if a learner wants to learn for an extended period of time, then a 3D graphics VR VE might also be more suitable. Additionally, 2D text-based VEs might not offer any inspiration for design practices for language learning with how much everything has changed when comparing it to an IVR VE, but there is still much overlap between 3D graphics VR characteristics and IVR. For example, letting a player interact with an item by activating through a click interaction as occurs in *Influent*, can still be a viable method for target word object interaction in an IVR VE, as is also shown when looking at some of the contemporary IVRALL research projects discussed in the next chapter.

4 Contemporary IVRALL+VA research

Language learning has many aspects that are necessary for learners to pay attention to and that are interesting for researchers to study. IVRALL research also focuses on multiple language learning aspects, like listening comprehension [70], conversational practice [71] or vocabulary acquisition. Because of the focus of this thesis on specifically the vocabulary acquisition aspect of language learning, we will only discuss IVRALL research with also a specific focus on the VA aspect of language learning (IVRALL+VA).

This chapter starts in Section 4.1 with defining which research is included in our literature search and how research is selected, followed by the rational of research projects for selecting a specific L2 in Section 4.1.1. The sixteen selected papers regarding IVRALL+VA can be divided into four categories, where all research projects placed in the exploration-based category are discussed in Section 4.2, followed by all research projects in the conversation-based category in Section 4.3, one research project in the location-based category in Section 4.4 and all research projects in the movement-based category in Section 4.5. In Section 4.6 a summary is given on different aspects of the research projects, like reoccurring phases, retention, condition comparisons, choices regarding target word objects, different contexts and participants, which are followed by an outline for our research design based on the state of the art work discussed in this chapter.

4.1 Paper selection

We define here IVRALL+VA research as research that makes use of a VE in IVR and that fulfils at least one of the two following conditions: (i) the research project presents participants with predetermined target words inside a VE to study and learn, and (ii) participants are tested and evaluated on words that they have come across as visual representation, heard pronounced or read as text during their time in the VE. Target words in condition (i) are predetermined and emphasised in some manner within the VE to make the participant aware of them, therefore condition (i) filters for explicit vocabulary learning. For explicit vocabulary learning it is also possible to adhere to both (i) and (ii). By also including research that only adheres to (ii) it is possible to also find research with an implicit vocabulary learning approach. For example a research project that lets participants play an IVR game that is not a SIE (i.e. a game designed without any educational purposes in mind) and which evaluates participants on the N nouns they heard most often during their playthrough.

An unstructured literature review was conducted for finding research that adheres to the IVRALL+VA definition and at least one of its conditions. Searches with combinations and variations on keywords for *immersive virtual technology*, *language learning* and *vocabulary acquisition* were applied in the bibliographic databases *Scopus*, *ScienceDirect* and *Google Scholar*, which resulted in fifteen articles of which one is a master thesis (i.e. [72]) and another is the aforementioned project *Zengo Sayu* from 1995 by Rose and Billinghurst [69].

A literature review by Palmeira *et al.* [42] with also a focus on IVR and VA, although less strict with the particulars for VA and without the requirement of a VE for the IVR, identified nine papers. Seven of the nine papers had overlap with our search, of which five (i.e. [8], [73], [74], [9] and [6]) were also included in our fifteen selected articles and two (i.e. [71] and [75]) had been discarded. The research project of [71] lets participants use the commercial program *Mondly VR* for language learning. This program revolves around conversations between the user and NPCs and has here no focus on specific target words (i.e. not fulfilling (i)), and the evaluation of [71] is on engagement and attitude towards language learning in IVR (i.e. not fulfilling (ii)). In [75] participants prepare and give an oral presentation in IVR where they are free to choose what they want to convey (i.e. not fulfilling (i)), and while there is an evaluation on the vocabulary quality, there is no evaluation on specific words (i.e. not fulfilling (ii)).

Of the remaining two papers of Palmeira *et al.* (i.e. [38] and [76]) one paper was added to our selected papers and one was not. XU *et al.* [76] did not make use of a VE, but showed instead a video for each target word (i.e. not fulfilling (i)), and only evaluated on design needs (i.e. not fulfilling (ii)), and was therefore not added to our research. Repetto, Colombo, and Riva [38], however, both uses target words in their VE (i.e. fulfilling (i)) and evaluates participants on them (i.e. fulfilling (ii)), and was therefore included in our research, resulting in sixteen papers with an IVRALL+VA focus.

Of these sixteen papers none fulfils only condition (ii), meaning that no research with implicit vocabulary learning was found, and that all found papers work with explicit vocabulary learning (i.e. fulfilling both (i) and (ii) or only (i)). All sixteen papers do adhere to condition (i), with thirteen papers also fulfilling condition (ii). With the exception of *Zengo Sayu* [69] from 1995, all papers were published from 2015 onwards, with one paper published in 2015, 2016 and 2017 each, three papers published in 2018 and in 2019 and six papers published in 2020. See Figure 4 for an overview.

4.1.1 Language choice

A total of ten different languages are used among the sixteen projects, with English being chosen five times, and Japanese and Spanish two times. Chinese, Czech, Finnish, Irish, Korean, Swedish and an artificial language called *Vimmi* have all been selected once. See Figure 5 for an overview of the frequency of languages.

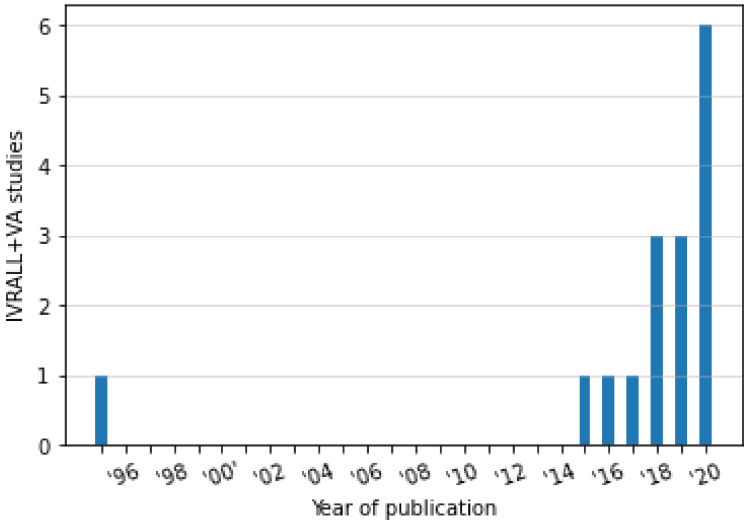


Figure 4: Sixteen selected IVRALL+VA studies ordered by their publication year.

A few researchers chose a specific language with their target audience or participants in mind. Jia and Liu [77] from *Words in Kitchen* explain why English is most worthwhile to learn for their Chinese participants, as English is a compulsary course in China while there is a lack of daily English conversation possibilities. Collins *et al.* [10] from *IrishSuper** state in their turn that the Irish language is only spoken by a small part of the population in Ireland, and see possibilities in IVR and situated learning to increase participants motivations to engage with the Irish language.

Other researchers link their language choice to their project design or the importance of participants being unfamiliar with a language. Repetto *et al.* [38] from *LimbVerbs** chose the Czech language because it is highly unknown in Italy, and therefore hopefully also for their Italian participants, while its phonology is quite comprehensible for Italian speakers, which is helpful since there are only audio target words in *LimbVerbs**. Fuhrman *et al.* [78] have a similar rationale as Repetto *et al.*, by stating that Finnish was chosen for *ObjectManipulation** because Israelis are rarely proficient in or exposed to the language. However, contrary to Repetto *et al.*, Finnish was also chosen because its phonetic structure does not resemble that of the languages that are widespread in Israel. Cheng *et al.* [73] name no rationale for choosing Japanese for their IVR *Crystallize* game. However, in earlier research where *Crystallize* was still a non-IVR game, Culbertson *et al.* [79] stated that Japanese grammar is very different from English grammar which made it useful for evaluating how the game performed in facilitating learning of unfamiliar grammatical structures.

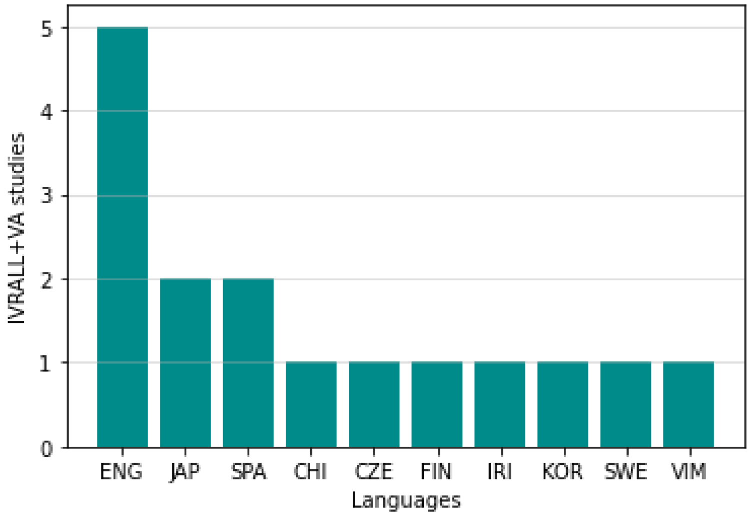


Figure 5: Chosen L2 languages of the sixteen selected research projects.

Legault *et al.* [6] from *ZooKitchen** did not link their language choice explicitly to how it relates to their participants knowledge, but did state as a requirement for their participants that they should not have any knowledge of the target language. Cho [72] from *ClassroomVR** follows Legault *et al.*, but elaborates on their participant requirement further by explaining how they want participants to start from an equal initial condition since their project aims to examine how a new, and therefore unknown, language can be learned with the use of their selected technology. Vázquez *et al.* [9] from *Words in Motion* did not require participants to have no knowledge of the target language, but did want to ensure that their project would be challenging for all participants, so selected Spanish words that had a low frequency counter according to a frequency list with over 450 million Spanish words. Macedonia *et al.* [80] from *CaveGrasp** removed both the possibility of participants being familiar with the target words and with them associating the target language with their first language German, by using *Vimmi*, an artificial corpus designed for research purposes. The artificial words in Vimmi were created according to Italian phonotactic rules, and were first randomly generated and then manually evaluated and selected [81].

Lastly some researchers chose a language based on location-based possibilities. Alfadil [41] with *House of Languages* asked students from a nearby school who were following an English course as participants, and Monteiro and Ribeiro [82] from *MuseumTour** used students of English from a nearby university and from a private English course that one of the authors taught.

4.1.2 Four approaches in IVR for VA

All projects can be roughly divided over four categories: (i) an exploration-based approach for VA, (ii) a conversation-based approach for VA, (iii) a location-based approach for VA, and (iv) a movement-based approach for VA, where all include either a *learning phase* where target words can be learned in IVR, a *quiz phase* where target words can be tested in IVR, or both. All papers are first discussed along their assigned category, where *VirtualCustoms** by Dobrova *et al.* [74] is categorized both under (i) and (ii) as their *learning phase* has an exploration-based approach and their *quiz phase* a conversation-based approach, and *ZooKitchen** by Legault *et al.* [6] is regarded as belonging to (i) and (iv), as participants can explore freely to learn target words (i.e. exploration-based approach), while Legault *et al.* also compare conditions with and without manipulation possibilities (i.e. movement-based approach). Of the remaining fourteen papers have six an exploration-based approach, three a conversation-based approach and four a movement-based approach. See Figure 6 for an overview of the number of papers in each category.

Projects without a predetermined project name have been provided here with made-up names to improve readability. These made-up names are always indicated with an asterisk symbol at the end (e.g. *ZooKitchen**). *Zengo Sayu* from Rose and Billinghurst [69] is left out of the discussion due to its detailed description in Section 3.5, and the design of the commercial program itself in *House of Languages* by Alfadil [41], which is discussed in depth in Section 3.4, is only briefly touched upon.

4.2 Exploration-based approach

The VEs in the exploration-based category are familiar environments like a living room [8], [83], kitchen [6], [77], zoo [6], customs point at an airport [74], classroom [72] and a supermarket [84], in which target words are placed as visual objects. Most projects in the exploration-based category let participants explore freely in the VE and employ different strategies to make participants aware of the target words within. For example, in *Ogma* [8] there is an exclamation mark hovering over each target words, in *ProtoQuiz** [83] target objects are highlighted in blue when facing in their direction, while in the zoo in *ZooKitchen** [6] there is a gem floating next to target words. Contrary, the kitchen VE in *ZooKitchen** [6] is purposefully not calling attention to target words to encourage exploration, and is only pointing arrows to missed target words after some time has passed. In *ClassroomVS** [72] participants are least free in their exploration as they are guided past all target words with arrows placed on the floor for walking directions, but they were free in studying the target words at their own pace.

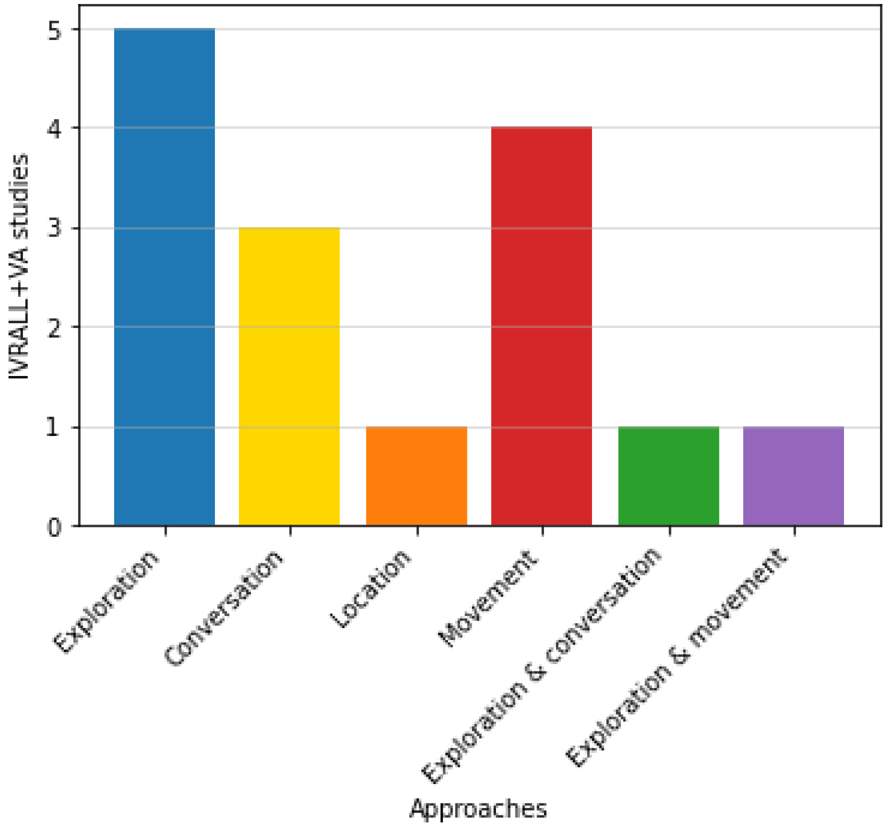


Figure 6: Division of the sixteen selected research projects for each self-identified approach of exploration, conversation, movement and location. *ZooKitchen** and *VirtualCustoms** are counted for each of their assigned approaches.

4.2.1 VirtualCustoms*: exploration

Dobrova *et al.* [74] name the learning environment and motivation as most important learning factors for effective language learning and notice how in Russia IVR is only used on a few schools and universities. They therefore introduce *VirtualCustoms**, an IVR environment of a customs point at an international airport that was designed by a group of teachers, to promote using IVR for learning by providing an example of how IVR technologies can create an ideal learning environment. Thus an evaluation on *VirtualCustoms** is left out and it is only the idea of *VirtualCustoms** that is presented by Dobrova *et al.* in their paper.

There are two main VEs in *VirtualCustoms**: (i) a VE called the *green channel* where the language learning is conversation-based, which is further discussed in Section 4.3.4, and (ii) a VE called the *red channel*, see Figure 7, where language learning is exploration-based. In the *red channel* target words of customs equipment are placed as visual objects in the VE and users can approach an object to read the text belonging to the target word to learn it.

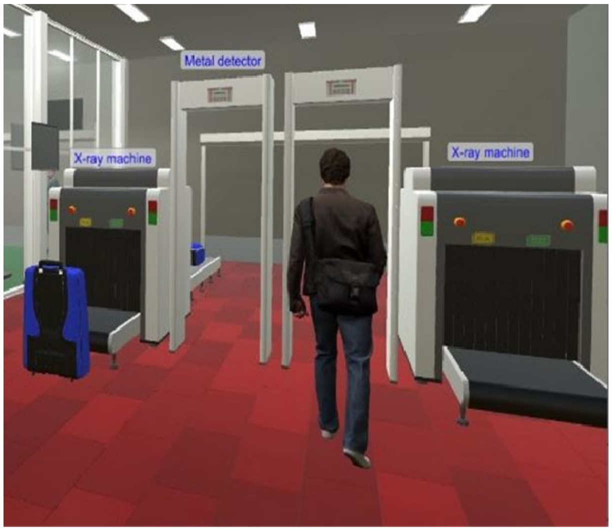


Figure 7: Image of the *red channel* in *VirtualCustoms** from Dobrova *et al.* [74].

4.2.2 ProtoQuiz*

Garcia *et al.* [83] see the possibilities that IVR can provide for language learning, but find the research into IVRALL lacking. Therefore they present an interactive IVR experience with *ProtoQuiz** of which they detail the design and development. Garcia *et al.* also test *ProtoQuiz** with a focus group of four persons aged between 18-24 to study user experience feedback.

The VE of *ProtoQuiz** is a living room, see Figure 8, and a voice asks participants in Spanish, the target language, to find a specific object (e.g. *where is the bed?*).[20] This question is also displayed in text at the bottom of the participants vision. The participant is then expected to select the correct item, where items that can be selected become blue when hovered over with a gaze pointer, and the item will become red if the wrong answer is selected. If the answer is correct then the object name is repeated audibly, ten points are added to the score that is displayed at the top of their view while no points are subtracted for a wrong answer, and a prompt appears for the next item until all items are found.

[20] *¿Dónde está la cama?*

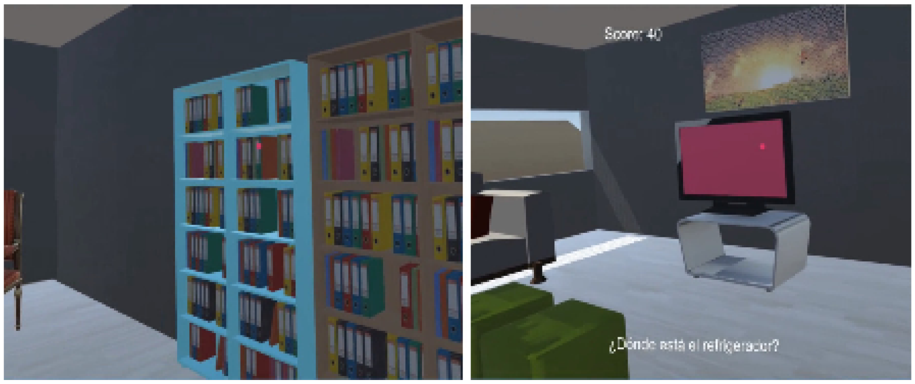


Figure 8: Living room environment from the project of Garcia *et al.* [83]. The red dot on the fridge is a gaze pointer that is controlled with the participant's line of vision. Items that can be selected and that are hovered over with the gaze pointer become blue.

According to the focus group the system was more enjoyable than traditional language learning methods, but needed a more spacious environment, more diverse objects, and an improved user interface. Participants also mentioned feeling nauseous. Garcia *et al.* plan for their next iteration to change item selection from using participant's line of vision to using touch controllers and adding maybe a friendly speaking avatar to avoid the sensation of a voice coming from out of nowhere. They also mention that studies are necessary regarding language retention and vocabulary gain to see how successful it can teach information to participants.

4.2.3 Words in Kitchen

Jia and Liu [77] also propose an IVR system for vocabulary learning called *Words in Kitchen* and conducted a pilot study to identify its usability and enjoyment factor. *Words in Kitchen* has a standard approach for the exploration-based category; it uses a kitchen as VE in which virtual objects of target words are placed and where participants can interact freely with the objects while exploring the kitchen. However, similar to *House of Languages* with *Mr. Woo*, they added a virtual character named *Tony* as a learning companion, which can be seen in Figure 9. In other exploration-based projects the user learns the audio or text L2 word through a sourceless recording that is played or through a text field hovering near the visual object. Here the L2 word accompanying the visual object is provided by the learning companion when approaching an object, where the learning companion says the word out loud while there is also a text box near them with the text of what they are saying. The virtual character could also ask participants to search for a specific object. The virtual charac-

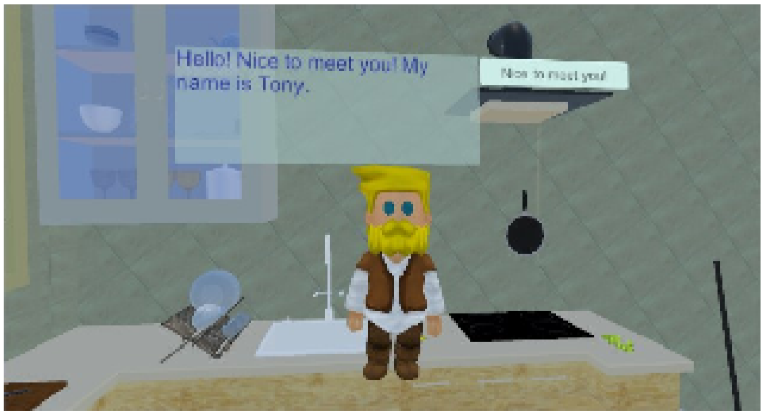


Figure 9: Image of *Tony* from *Words in Kitchen* from Jia and Liu [77].

ter has multiple functions; it guides participants to prevent aimless roaming, it improves interactivity, the *quiz phase* can be hidden by making it a request for help, and it can provide feedback by showing a happy expression when choosing correctly and a sad expression when choosing incorrectly.

A small pilot study with three children between 8 and 10 years old and two persons of 24 years old was conducted to test the design of *Words in Kitchen*. All participants expected that the system could increase learning interest and that it was easy to use. One participant mentioned how they also associated target words with their location in the VE. Participants also thought that the learning companion made the learning process more lively.

4.2.4 Ogma

Ebert *et al.* [8] built a system called *Ogma* which consists of an IVR living room environment in which participants can move around by using a North Myo armband[21] and pointing their arm in the desired direction. They tested their system by letting half of their nineteen participants work with *Ogma* and by giving the other half a traditional vocabulary acquisition method.

The experiment for the virtual reality group consisted of both a *learning phase* and a *quiz phase*. The living room, see Figure 10, has ten objects marked with an exclamation mark for the first phase, which indicates a word of interest. The Swedish word that belongs to such an object is hovering in text above the object and pronounced when the object is approached and looked at. Participants had

[21] A wireless gesture recognition device that is worn on the forearm. The Myo armband can sense electrical activity in the forearm muscles with a set of electromyographic sensors and has a gyroscope, an accelerometer and a magnetometer to determine its position. These signals enable the user to control technology by using hand motions.

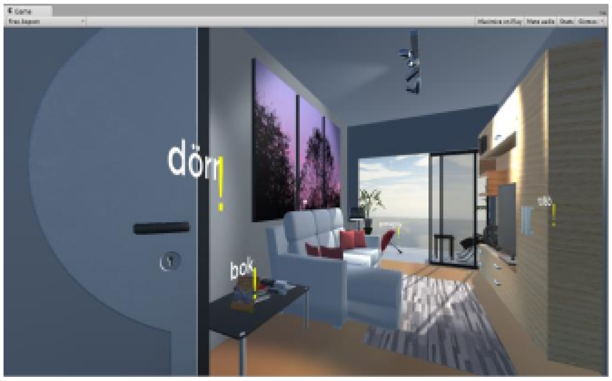


Figure 10: Image of the living room environment from *Ogma* [8]. Target words are indicated with an exclamation mark during the first phase.

five minutes to get familiar with the ten words of interest during the experiment. Next followed the second phase, where all exclamation points were removed and participants saw a word in text displayed in their field of view and heard it pronounced. Participants were then expected to walk towards the accompanying object in the living room and point towards it. The experiment was concluded after all ten words were identified.

The experiment for the traditional group consisted of receiving a written list with the same ten Swedish words as the virtual reality group encountered and their translation to English. All words were also read aloud and the participants received flashcards to help them. Participants could study the words until they felt ready for the follow-up test.

Both groups were given as posttest a written list with L1 words after their experiment and had to write and pronounce the corresponding L2 words. A week later they made as delayed posttest the same test again. The posttest was made significantly better by the traditional group, while the score for both groups was similar for the delayed posttest. However, the retention rate, which was calculated as a percentage of the correct words from the posttest which were also correct during the delayed posttest, was significantly higher for the virtual reality group. Some virtual reality participants also mentioned that they were able to visualise the living room and the objects in it to help them during the test.

Participants were also asked to rate their experience on enjoyability and effectiveness, where the virtual reality condition received much higher ratings for both, even though many participants experienced dizziness from using the Oculus Rift.

4.2.5 ZooKitchen*: exploration

Legault *et al.* [6] created two different IVR environments: a zoo and a kitchen. The main difference between the two environments is the navigation method, where teleportation is needed in the zoo while the kitchen can be traversed in real space, and the level of interaction, where the kitchen has more items to interact with.

64 participants with an average age of 19.05 first took a cognitive test before the actual experiment, where they were assessed on inhibitory control[22], language history background, phonological working memory, first language (L1) proficiency, and spatial abilities. All 64 participants were then evenly divided over four groups, so 16 participants per group, where the first two groups learned thirty zoo words in IVR and thirty kitchen words in a traditional manner, but the first group started in IVR and the other group with the traditional method. The last two groups were similarly divided but learned the kitchen words in IVR and the zoo words with a traditional method.

The traditional method consisted of an English word presented on a computer screen where the participant could hear the corresponding Chinese word after which they could go to the next word. After seeing all thirty words the participant continued going through the words until twenty minutes had passed. Participants in the IVR kitchen environment had to find the target objects themselves by pointing at an object and then hearing the corresponding Chinese word. After ten minutes, arrows started appearing to indicate any missed objects to ensure that all thirty words had been heard before twenty minutes were up.

Target objects in the IVR zoo environment, see Figure 11, were indicated with a gem next to the object and the accompanying Chinese word could be heard by pointing towards the object. Participants had again twenty minutes to hear all thirty words.

After the experiment a post-test took place where participants would hear the Chinese word and then had to choose the corresponding word from either four English words in text for the traditional condition or four screenshots of target words for the IVR condition. Participants were thus only tested on learned contexts and not on new contexts, since participants of the IVR condition were not presented with the written target words while participants of the traditional condition did not see a picture of their target words. Results showed that accuracy was significantly higher for words learned in the IVR environment. However,

[22] An inhibitory control test looks at the degree in which a person can suppress the dominant behaviour impulse they feel when seeing a stimuli so a more appropriate behaviour can be selected that helps with completing their goals. For the inhibitory control test in *ZooKitchen**, participants had to indicate in line ups of five arrows on a screen in which direction the third arrow pointed.

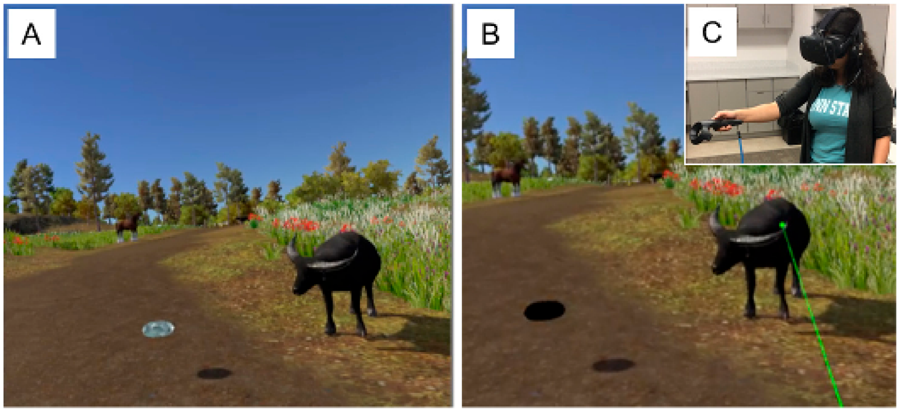


Figure 11: Images of the zoo environment in IVR and a participant [6]. On the left the gem to indicate the target word *cow* is floating above the road. On the right the gem has turned black to indicate that the participant has successfully clicked on the target word.

when looking at the results separately for successful versus less successful learners, it was noted that less successful learners showed a clear benefit of learning in IVR, while the best learners performed equally well in both conditions.

Legault *et al.* also looked with *ZooKitchen** at the influence of manipulation on VA, making the project also have a movement-based approach, which is further discussed in Section 4.5.5.

4.2.6 IrishSuper*

Collins *et al.* [84] created *IrishSuper** after iterating through three case studies of which all three are discussed in [10] and the last iteration is discussed separately and in more detail in [84]. The final version of *IrishSuper** consists of a supermarket VE, where the supermarket is filled with visualised objects of target words. These objects are, contrary to a real supermarket, placed with generous spacing around them on display, where one object type takes up the whole shelf. Collins *et al.* [84] focus with *IrishSuper** on vocabulary retention, and learner's anxiety and motivation on learning, but place their main focus on improving the motivation of participants to see themselves as being capable of becoming Irish speakers in the future. Their participants are ten trainee Irish primary school teachers aged between 18-40 for whom Irish language competency is a requirement, so who already try to become future Irish language speakers.

The participant starts the experiment in IVR by approaching an NPC standing in front of the supermarket, who tells the participant the four items they need to retrieve from the shop. The participant can walk through the supermarket and pick up an object to hear their Irish (i.e. L2) pronunciation and a text hovers over the object with the Irish text equivalent. Participants can put objects in their shopping basket and when they think that they have retrieved the four requested items, participants can go to a checkout desk to place their selected items on the counter for evaluation. The NPC shopkeeper will then shake their head if the participant is still missing items, and will also tell the participant how many items they still need to find, or the NPC shopkeeper will give the participant a thumbs up if they collected all items successfully and the participant will continue on to the next level. There are four levels in total, where the only difference between levels is the difficulty of the L2 items that are requested. The *learning phase* and *quiz phase* are thus combined together in a game-like approach, as the participant can explore the supermarket and choose to learn the text and audio of each word that piques their own interest (i.e. *learning phase*), while also trying to find explicit requested target words (i.e. *quiz phase*).

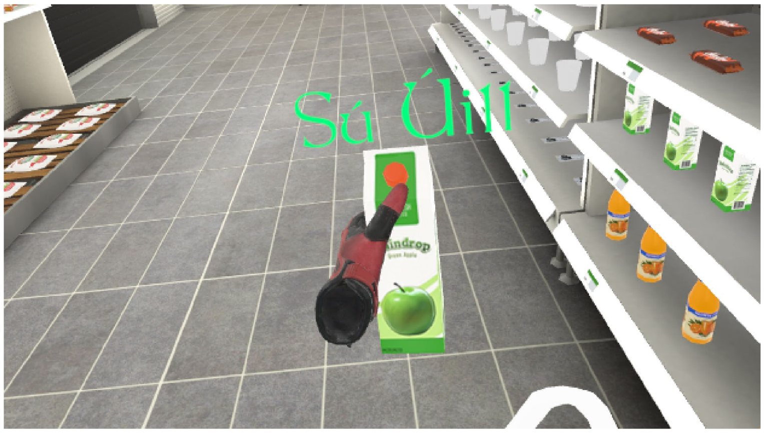


Figure 12: Image of *IrishSuper** from Collins *et al.* [84].

Each participant played through *IrishSuper** three times, where each session had a duration of twenty minutes and took place in a time period of five weeks. There was a pre- and posttest regarding vocabulary, where the posttest was administered a week after the last session, so retention rate would also play a role in the results. Collins *et al.* found a 21% increase on word retention when comparing the results of the pretest with the posttest and conclude that the simulated environment increases the L2 vocabulary of participants. Participants also filled in a pre- and posttest regarding their anxiety for learning Irish and their attitude towards learning Irish. The pretest scores indicated a lack of confidence among participants regarding their ability in Irish, while the posttest indicated more confident participants as a result of interacting with *IrishSuper**. The IVR experience allowed participants to self-assess their ability and motivations through interacting with the Irish language in a naturalised environment instead of through the usual classroom experience.

4.2.7 House of Languages

Alfadil [41] studied the influence of IVR on VA through the commercial game *House of Languages*, which is discussed in more detail in Section 3.4. *House of Languages* is a game in which a raccoon NPC named *Mr. Woo* teaches players vocabulary in different VEs. Players can select an object by gazing at them, which reveals the text and pronunciation of the L2 word associated with the selected object. *House of Languages* starts with *Mr. Woo* requesting specific items which the player must find (i.e. *quiz phase*), but the player can meanwhile also learn about the other objects in the room by gazing at them (i.e. *learning phase*). This setup allows the player to go through the *learning* and *quiz phase* simultaneously, as was similarly done in *IrishSuper** [84]. *House of Languages* then concludes with some mini-games and puzzles to test the player further.

The hypothesis of Alfadil was that participants who use IVR, in contrast to participants using a traditional method, score better on VA. Participants were 64 students from a school where they learned English as L2 and were ages 12-15. Participants in the IVR experimental group used *House of Languages* for twelve school days for a duration of 35-45 minutes with sessions of around 8 minutes, while participants in the traditional control group continued with their usual school learning and made use of books, lectures and worksheets. Results showed that participants in the experimental group scored significantly higher on vocabulary acquisition than participants in the control group.

Alfadil also observed participants during the experimental period and noted that students were clearly excited for participating in *House of Languages*, which kept their motivation for learning high. Alfadil also sees further potential for IVR language learning in the classroom since IVR can provide a personal *teacher*, although in the form of an NPC, for each student, enabling each student to study at their own pace without influencing the overall classroom balance.

4.3 Conversation-based approach

Vocabulary learning can revolve around studying and practicing with individual objects, as with the projects in the exploration-based category, or around learning words from conversation or practicing vocabulary by applying them in conversation. *Crystallize* and *MuseumTour** have such a conversation-based focus, while *VirtualCustoms* applies such a focus in their *quiz phase*, while having an exploration-based focus in their *learning phase*.

4.3.1 MuseumTour*

Monteiro and Ribeiro [82] wanted to explore the potential of IVR for language learning. Therefore they placed their 25 participants, with a median of 23 years of age, in *MuseumTour** in a museum by using Google Cardboard, where they followed a tour through a virtual museum where seventeen text target words were placed in the environment and where each is close to an object that the target word has a connection with. Most of the chosen target words are difficult to capture in a visualised object, like *bedridden*, *hardships* or *strength*. After a participant selected a word a real life teacher, who joined the participants in IVR, would first provide synonyms to the target words and/or ask questions directed to the participant regarding the target word, while always speaking in the target language. Then the teacher would tell a story about the object while using the target words at least five times in their story.

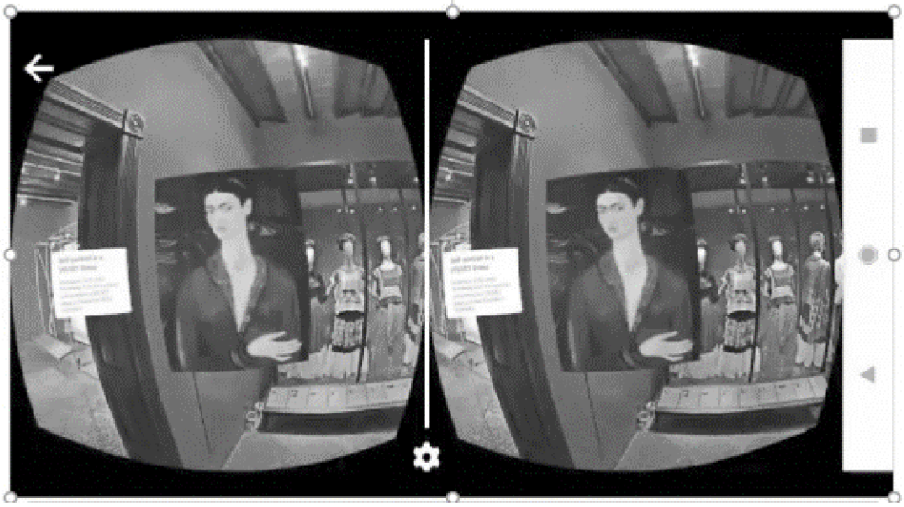


Figure 13: Image of *MuseumTour** [82] depicting the target word *self-portrait* and its accompanying object of the first self-portrait of Frida Kahlo.

It was thus intended that vocabulary was explicitly taught through the context of a story to participants. A pre- and posttest on the seventeen words resulted in an average gain in learning of 43%, but Monteiro and Ribeiro also asked participants about their IVR experience, where participants reported feeling immersed and some mentioned how IVR was able to support their learning style with realistic images and kinaesthetic aspects.

4.3.2 Crystallize

Cheng *et al.* [73] created a 3D video game for learning Japanese called *Crystallize* to increase engagement when learning a language [79]. Cheng *et al.* converted this 3D version to an IVR demo of *Crystallize* to explore the impact of IVR on language learning, and studied if IVR could also teach embodied cultural interaction (i.e. bowing in Japanese greetings). The IVR environment of *Crystallize* is a Japanese teahouse with NPCs that provide different types of dialogue. The player can eavesdrop on an NPC group to obtain new words for their inventory, and can join a conversation to use these words through prompts where they must choose a word from a multiple choice selection, see Figure 14, or where they must put words in order to form a sentence.

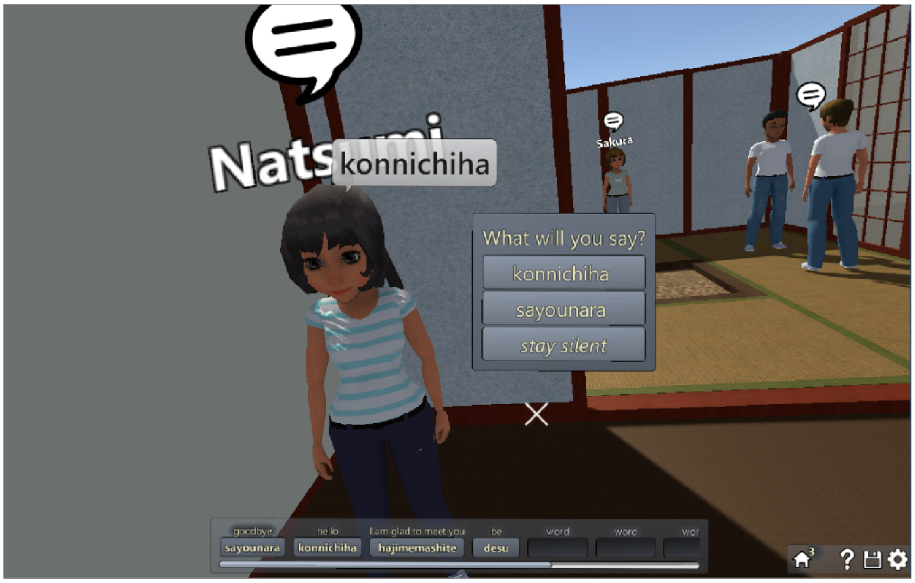


Figure 14: Image of *Crystallize* [73] where a participant can choose a response from a multiple choice selection.

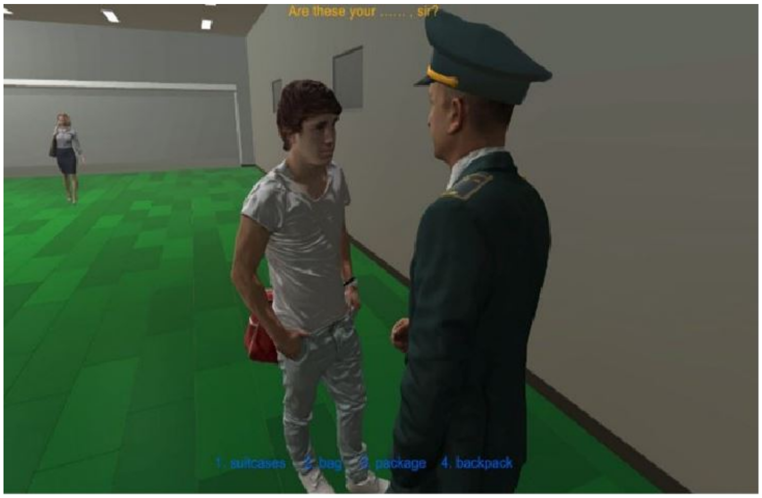
Cheng *et al.* evaluated the impact on learning by comparing the results of 34 participants who learned in the original 3D *Crystallize* to 34 participants who learned in the IVR demo. All 68 participants made a pre-test before the experiment and a post-test after the experiment to determine how many of the eight target words they knew. The differences between the two scores were used to measure language acquisition, where on average the non-IVR group learned 5.39 new words and the IVR group learned 4.77 words, although the difference was not statistically significant. Cheng *et al.* suspect that the difference is caused by participants being unfamiliar with the IVR interface which might contribute to less effective learning. However, projects like *Ogma* [8] and *Words in Motion* [9], which also provided a vocabulary pre- and posttest of target words, found similar results with the IVR group performing worse than the non-IVR group, but administered a second post-test for recall after some time had passed and discovered that the retention rate was significantly higher for IVR groups [8], [9].

4.3.3 Mondly VR

Tai *et al.* [85] used the commercial IVR program *Mondly VR* to study the potentials of mobile-rendered HMDs for vocabulary learning by comparing IVR language learning with walkthrough video language learning. An experimental participant group of 24 participants could look around in five different VEs and interacted with NPCs through voice command, while a control participant of 25 participants group watched *Mondly VR* gameplay through videos on a desktop. Target words in *Mondly VR* were either heard and read during conversations with NPCs, seen as a picture in the speech bubble of an NPC, seen as an object in the VE, or encountered as a combination of these options. No emphasis was placed on target word objects in the VE and it being a target word. All 49 participants, aged 14-15, evaluated their vocabulary acquisition by partaking in a pretest, posttest and an delayed posttest with a delay of one week. Results showed that both participant groups performed significantly better on the posttest in comparison to the pretest, with the IVR group also performing significantly better than the control group. The IVR group also performed significantly better on the delayed posttest in comparison to the pretest, where the retention rate of the IVR group was also significantly better then that of the video watch group.

4.3.4 VirtualCustoms*: conversation

Dobrova *et al.* introduce with *VirtualCustoms** an IVR experience for language learning without further evaluations. Their IVR experience consists of two VEs of which the VE called the *red channel* is exploratory-based and was discussed in Section 4.2.1, and the second VE, called the *green channel*, is conversation-based. In the *green channel*, see Figure 15, the user can interact with a customs officer while the spoken sentences of the customs officer are also displayed in text in the VE. Through conversation with the customs officer the user can learn the names of goods and types of luggage. The user can also test their knowledge in a mode where the customs officer skips some phrases in their dialogue and the user must then select from a multiple choice option which phrase was left out. Input is given by the user by selecting a text option in the VE.


Are these your ...... , sir?

1. suitcases 2. bag 3. package 4. backpack

Figure 15: Image of the *green channel* in *VirtualCustoms** from Dobrova *et al.* [74].

4.4 Location-based approach

Research projects are seen as having a location-based approach if they have a focus on how location plays a role in VA. Only one research project falls under this location-based category, namely *ClassroomVS** from Cho.

4.4.1 ClassroomVS*

In their master thesis Cho [72] compares VA in IVR to VA in a desktop equivalent, with a specific focus on the connection between spatial presence and memory retention. Unlike projects with an exploration-based approach where participants are free to learn target words in any order during the *learning phase*, were participants in *ClassroomVS** guided inside a classroom VE to see all target word objects in a specific order. The classroom VE housed twenty classroom objects as target words, and participants had to follow red arrows placed on the ground, which can also be seen in Figure 16, to go past all target word objects. 64 participants participated in total with ages between 18-65. One half of the participants learned in the IVR VE and the other half learned in the desktop VE. Participants were allowed to take their time for each object and were asked to see each object twice by following the arrows a second time after finishing their first round. Approaching an object would result in the accompanying L2 word appearing in text next to the object. Participants in the desktop VE moved around with a keyboard and mouse while participants in the IVR VE moved around with an Xbox controller. Participants were tested

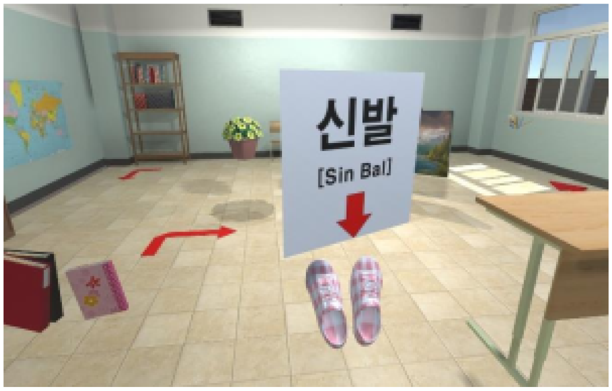


Figure 16: Image of *ClassroomVS** from Cho [72].
The arrows on the ground show the walking route participants must follow.

afterwards not with a vocabulary posttest to test their L2 word knowledge, but with a layout of the classroom and a list with the L2 target words texts combined with a target word object screenshot image, and participants were asked to fill in where they had seen the target word in the classroom. Doing so allowed for testing participants on *Memory of Loci*, where knowledge is connected to a pattern that is created from the order and location something is encountered in [86]. Cho concluded that IVR showed superior results in remembering where objects were compared to the desktop VR equivalent, but also added that they were not able to find the exact mechanism for increased memory retention.

4.5 Movement-based approach

Embodied cognition theory revolves around the concept that cognitive processes are grounded in the body's interaction with the world, and that learning languages as a cognitive process also involves the sensorimotor systems and their sensorimotor representations in the brain [78], [80], as discussed in more detail in Section 2.2.3. Four projects use embodied cognition theory as the basis for their IVRALL+VA research, although interestingly enough each project uses the knowledge of this basis to focus on a different type of movement in IVR. Repetto *et al.* [38] from *LimbVerbs** study the effect on learning an action target word while moving a part of the limb that is connected to that action word (e.g. learning the action word *to throw*, which needs an arm to be put into action, and moving the thumb, which is part of the arm, while learning). Macedonia *et al.* [80] with *CaveGrasp** let participant make a grasp movement around objects in IVR. Fuhrman *et al.* [78] from *ObjectManipulation** let participants make a relevant movement with regards to the target word (e.g. making a movement of stirring food, while learning the target word *spoon*). Lastly, Vázquez *et al.* [9] from *Words in Motion* connects the movement of drawing specific symbols to target words.

4.5.1 LimbVerbs*

Repetto *et al.* [38] mention how action words are better recalled if subjects pantomime the corresponding action during the *learning phase*, which they call the *enactment effect*. Repetto *et al.* hypothesise that if for learning a verb's meaning the motor simulation of the action described by the verb is important, then a synchronous action that involves the same responder limb of the action verb should modulate its recall later. They also wondered if, if indeed involved, motor simulation would be triggered by actual motion or virtual motion. To test their hypotheses they created a VE that looked like a park and chose fifteen target words, of which five are hand action verbs (e.g. to peel, to leaf through), five are foot action verbs (e.g. to kick, to jump) and five abstract verbs (e.g. to undertake, to forget). They tested with 40 participants aged between 19-49 years and with an average age of 33.17. Twenty participants were for the first condition instructed to continuously move around in the park and the other twenty participants were for the base condition told to sit on a bench while they were only allowed to look around. To move around participants used an Xbox controller while standing still themselves. The actual hand motion was therefore the thumb that moved the left analogue stick on the controller, while the virtual motion was the virtual walking in the park. Repetto *et al.* found no differences between the two conditions and conclude in their results that simulation is apparently not involved in verbal learning, but then nuance it by mentioning that their project only provides a shared generic motion for all words, while research that combines each word with a different gesture does find significant results. However, possible explanations not named by Repetto *et al.* might also be that the movement of the thumb on an analogue stick was too small to have an effect, that the corresponding limb movement mixed with the uncorresponding limb movement might negate each other, that the avatar movement through the park, which can more easily cause nausea for people who are sensitive to the side effects of vection, influenced the results, or if the motor simulations acquired for L1 are automatically transferred to L2.

4.5.2 CaveGrasp*

Macedonia *et al.* [80] use for *CaveGrasp** one generic movement for their experimental condition; namely grasping. Macedonia *et al.* explain that neuroimaging studies on words for tools or instruments, so manipulable objects, show a stronger activity in motor brain areas than words that have a low manipulability. Therefore, the more intensive the interaction with an object is the more it will be grounded. So by actually grasping or manipulating an object the brain is able to create strong sensorimotor networks that are connected to that specific word. Hereupon Macedonia *et al.* wondered if grasping virtual objects would lead to better memory performance for L2 words than virtual VA learning without grasping.

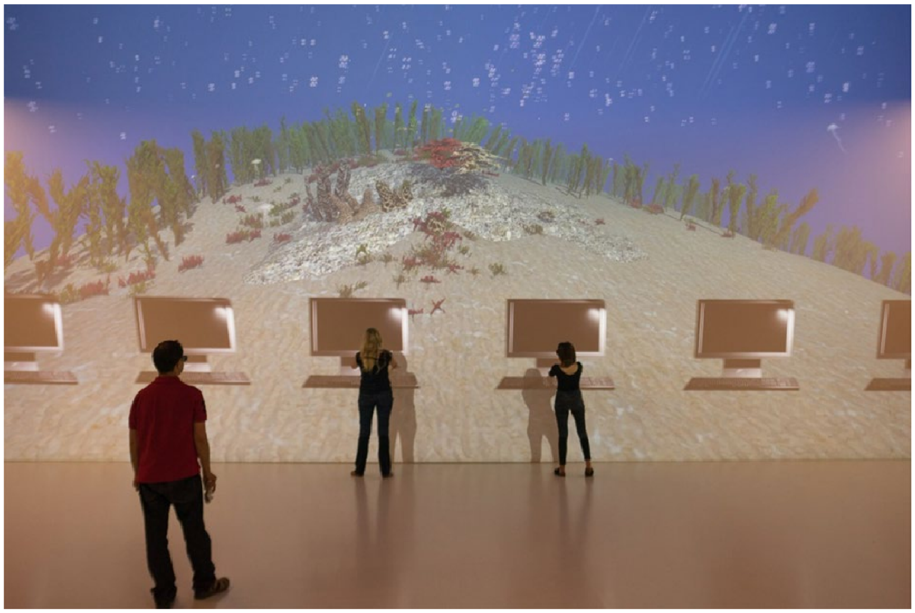


Figure 17: Image of *CaveGrasp** from [80] [80] showing participants making a grasp movement in front of a target word object.

Macedonia *et al.* set up their research project in the Deep Space 8K, which is an IVR cave situated in Linz with a wall and a floor of both 16 by 9 meters which can generate stereoscopic 3D visualizations in 8K. Users wear 3D glasses to be immersed in the projected virtual environment. Macedonia *et al.* used three conditions for their project, where for all conditions participants were placed through Deep Space 8K at the bottom of the ocean floor. For two conditions participants saw multiple instances of the same oversized object being dropped in the ocean, and after a second the objects would reach the ocean floor and stay there for a little while while participants could read the L2 target word belonging to the objects and hear how it is pronounced. After some time the target word would then disappear and the next would follow, until all target words were seen. For the first condition the participant was also asked to grab the object, however, since the object is visual and cannot provide haptic feedback, the participant was essentially asked to position their hands around the virtual object like they were grasping it, while the second condition requested no movement. The last condition provided only the text and audio of a word without a visual representation and without any movement requests. Participants were tested on their recall and recognition immediately after their learning session and also 30 days later. Their results showed a statistically significant difference for word retrieval when comparing the grasp condition to the text and audio only condition, and for word recognition when comparing the grasp condition with the other two conditions. They conclude that learning L2 words in a VE while grasping their visualisations enhances the memorability of L2 words as well as their recognition. They also state that IVR allows for grasping without the use of real objects, making it suitable for embodied learning of L2. However, when translating this to IVR with an HMD instead of an IVR cave, the results might be different, if the program is not able to project the user's hands into the environment.

4.5.3 ObjectManipulation*

Fuhrman *et al.* [78] looks with *ObjectManipulation** at the concept of manipulating objects during VA. They aim to explore in their project the effect of meaningful motor information that is acquired in IVR while learning new words. Interestingly, their VEs are built in a similar manner as how VEs are built for studies with an exploration-based focus, meaning that they built a room and dressed it up with various items like a table and a kitchen surface, while also placing visualisations of target words inside the room.

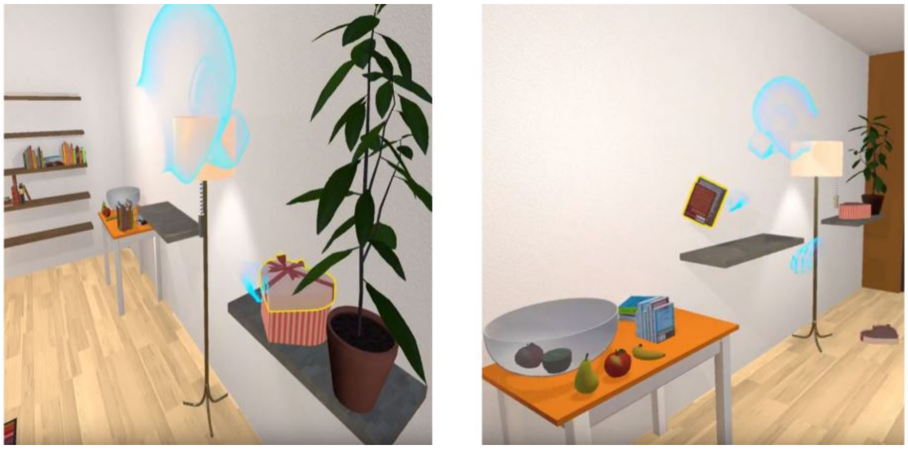


Figure 18: Image of *ObjectManipulation** from Fuhrman *et al.* [78].

However, what makes it different from exploration-based projects is that participants were not free to explore the room to learn the words, but instead one object would have a yellow outline around it and arrows on the ground pointing towards it to indicate that this was the current target word of the participant. Upon locating the object the participant would hear the pronunciation of the word and was required to say it aloud followed by walking towards it and then they would hear the word a second time. From here three different options could be required from the participant, depending on the current condition: (i) the participant had to repeat the word a second time, (ii) the participant had to repeat the word a second time while performing an unrelated non-interactive movement (e.g. drawing half a circle in the air), and (iii) the participant had to repeat the word a second time while performing a meaningful action with the object while mimicking reality (e.g. stirring food with the visualised target word for *spoon*). Participants were tested after each condition with a word-picture matching test and a week after their session with again a word-picture matching test and a verbal free recall test. The relevant manipulation condition always outperformed the non-relevant manipulation condition, where Fuhrman *et al.* speculate that the non-relevant manipulation condition required participants to not only process the word, but to also think about a movement that was separate of the presented word, possibly creating a higher cognitive load which can have a negative effect on the learning process. The watch condition was compared to the manipulation condition only marginally favourable during the word-picture test a week later, with it being the other way around for the first test. Fuhrman *et al.* [78] hypothesise here that IVR technology might have enhanced learning for both the watch-only and manipulation condition, and that both groups performed so high on the posttest that the manipulation effect might have been erased. This effect might have become again visible for the delayed posttest where the manipulation condition performed slightly better, which might indicate that the motor information that was paired with the target word might have worked as a more powerful mnemonic.

4.5.4 Words in Motion

Vázquez *et al.* [9] from *Words in Motion* do not provide participants with a dressed up VE, but use an empty room to allow participants to focus on the connection between their bodily experiences and the target words to enable kinesthetic learning. 57 participants were placed in one of three conditions. Twenty participants in condition (i) first saw a symbol being drawn in the air which they had to redraw with their right hand, see Figure 19. The Spanish target word would then appear with the corresponding English translation. Symbols did not match the target word in any manner. Participants were then asked to make the movement again after which the words would stay visible for fifteen seconds before continuing to the next movement. Twenty words were learned in this manner and each word was encountered twice. Another twenty non-kinesthetic IVR participants in condition (ii) followed the same setup as participants in condition (i), but all the movements were made by the program so the participants only had to watch. The text-only participants in condition (iii), consisting of 17 participants, saw each word for fifteen seconds on a computer screen where each word appeared twice. A post-test was taken by all participants where they had to fill in the English translations to the Spanish target words. A second post-test was repeated a week later. Text-only participants did significantly better on their first post-test than IVR participants.

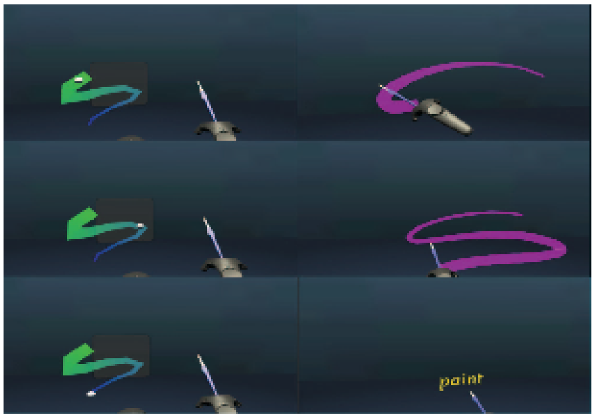


Figure 19: Images on the left show the animation of the movement in *Words in Motion* and the images on the right show the movements of the participant [9].

However, in the second post-test, text-only and kinesthetic IVR participants performed almost the same, while non-kinesthetic IVR participants performed significantly lower. So even though the text-only group was able to remember more words in the first post-test (M=14.6), many of those words were forgotten during the second test (M=7.56), while the kinesthetic IVR could recall not that many words during the first post-test (M=10.8), but of those words they remembered many during the second post-test (M=7.8), meaning that their retention rate was significantly higher than the text-only and non-kinesthetic IVR groups. These results suggest that kinesthetic elements in virtual reality can positively impact language learning, especially for retention. Vázquez *et al.* also noticed that their results showed a strong comparison to the results from Ebert *et al.* [8] and their *Ogma* project, where the performance for the text-only condition was better for the initial test but where participants of the IVR condition performed better on the later recall test.

4.5.5 ZooKitchen*: movement

Legault *et al.* [6] created with *ZooKitchen** an exploration-based kitchen and zoo VE for VA, which is discussed in Section 4.2.5, but also placed an emphasis on movement-based VA by comparing the kitchen VE, where objects can be picked up and moved around and participants can walk around in the VE, to the zoo VE, where target words cannot be manipulated and participants move in the VE through teleportation. Using manipulable objects for VA is also the core element of *ObjectManipulation** [78] from Section 4.5.3, but the difference in *ZooKitchen** is that participants are not specifically instructed to interact with the objects. Instead, the opportunity is built into the VE, but it is up to participants to decide on any interaction, which objects to interact with, what the interaction is and how long the interaction occurs. Target word objects were also not marked to encourage active and self-paced learning in the kitchen VE, unlike the zoo VE where target word objects were indicated with a gem floating next to them. However, arrows appeared next to not yet activated target word objects after some time had passed, to prevent that participants would miss any target words.

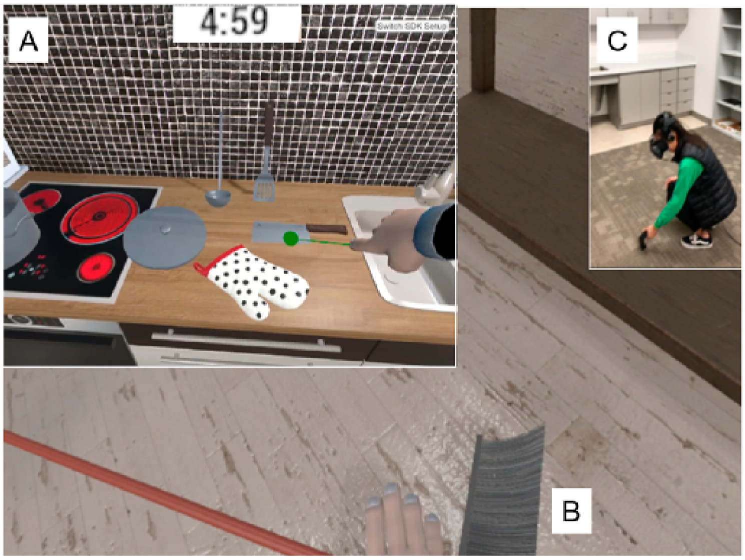


Figure 20: Images of the kitchen environment in IVR and a participant [6]. On the left a participant points towards a knife, and on the right and bottom a participant is picking up a broom.

Legault *et al.* found in their results that kitchen items were more accurately learned as compared to zoo items and refered to earlier research regarding embodied representations for a possible interpretation of their finding. In this earlier research of Martin *et al.* [87] the activation of neural networks was studied while participants looked at 2D drawings of tools and animals, where it was found that looking at tools activated motor regions in the brain, while animal words activated additional visual processing regions. Legault *et al.* conclude that items that can be seen as tools may make more use of embodied networks and may therefore be more effectively learned in comparison to target words from the zoo VE. Most participants also stated that it was the ability to move objects in the kitchen VE that aided their VA learning process.

4.6 Summary

There are a few trends noticeable when looking at all discussed IVRALL+VA studies. So has each study a learning phase, quiz phase or both, are multiple studies looking at retention and do studies include sometimes a non-IVR alternative as control condition. These and other observations relevant for this research are summarised in more detail in this section. An overview of the characteristics of all sixteen IVRALL+VA studies can be found in Appendix A.

4.6.1 Learning phase and quiz phase

Overall two phases are recurring in each of the sixteen discussed IVR projects: (i) a research or *learning phase*, where participants have the opportunity to study the target words in IVR so they can be learned, and (ii) a search or *quiz phase*, where the IVR program requests a specific target word and the participants must identify or provide the corresponding target word. However, IVR *quiz phases* are never used as a test for evaluating participants word knowledge. Instead all tests are made outside of IVR so non-IVR participants can also make the same tests or to easily use the same test for the pretest as well as the (delayed) posttest.

All movement-based and location-based projects only have a *learning phase*, in which the target word is connected to a specific movement, and do not include a *quiz phase* to quiz on vocabulary in IVR. A *quiz phase* is probably excluded because the main interest of researchers here lies in studying the effect of movement or location on language learning, and as long as participants learn the words, which they do in the *learning phase*, then a test outside of IVR is sufficient for evaluating the conditions of the research projects.

All conversation-based projects include a *learning phase*, where participants listen to NPCs to learn vocabulary. Target words in *Crystallize* [73] are indicated by allowing participants to collect them in a dictionary, while target words in *MuseumTour** [82] are provided in text next to the object of interest of the tour guide story. Target words are not specifically emhasised in *Mondly VR* [85]. *Crystallize* and *VirtualCustoms** both also include a *quiz phase*, where the *learning phase* in *VirtualCustoms** is exploration-based instead of conversation-based. Both projects leave out words in NPC sentences during their *quiz phase*, where the participant needs to fill in the empty space by selecting the correct term.

All exploration-based projects include a *learning phase* in their program with the exception of *ProtoQuiz** [83], where participants do get to move around freely to get acquainted with the controls, but there are no words to learn while doing so. One possibility for the decision to not include a *learning phase* is that *ProtoQuiz** [83] is a proof of concept/prototype project with a focus on user experience and without an evaluation on language learning, making it less important if participants have sufficient possibilities to learn the target words first. In all the other projects participants are encouraged to study the target words during the *learning phase*.

Ogma [8], *Words in Kitchen* [77], *House of Languages* [41] and *IrishSuper** [84] also include a *quiz phase* in their design. In *Ogma* [8] the target word indicators in the form of exclamation marks from the *learning phase* are removed and the textual representation of a target word no longer appears when approaching a target word object. Instead, a textual target word will appear in the field of view of the participant and the participant must approach the corresponding target word object until all target words are found. In *House of Languages* [41] and *Words in Kitchen* [77] participants are guided and accompanied by an NPC teacher, who also provides the target words that must be found by participants in the *quiz phase*. In *House of Languages* [41] the NPC teacher is a raccoon character named *Mr. Woo* and in *Words in Kitchen* [77] it is an unnamed small humanoid character named *Tony* with blonde hair and blue eyes to look like a foreigner for its Chinese participants. Lastly, the *IrishSuper** [84] combines both phases and hides them in a game-like approach, where participants receive a grocery list inside a supermarket where they are asked to find the listed items and check them out. While browsing the supermarket, participants can pick up an object and an audio recording will pronounce the word and a text will appear for the written form of the object. This allows the user to explore target words freely (i.e. *learning phase*) while also asking them to look for specific target words (i.e. *quiz phase*).

4.6.2 Retention rate

Five research projects evaluated the retention rate of their IVRALL+VA method by comparing an immediate vocabulary posttest after the experiment with a delayed posttest, where *Ogma* [8], *Words in Motion* [9], *ObjectManipulation** [78] and *Mondly VR* [85] had their delayed posttest one week after the first posttest, and *CaveGrasp** [80] had their delayed posttest 30 days after the first posttest.

*CaveGrasp** and *ObjectManipulation** both only test with IVR conditions, where *CaveGrasp** has three conditions: (i) grasping a visualised object in IVR, (ii) looking at a visualised object in IVR, and lastly (ii) only reading the L2 word in IVR. *ObjectManipulation** has also three conditions: (i) making a meaningful movement when learning the target word in IVR, (ii) making an irrelevant movement in IVR and (iii) a watch-only condition in IVR. In both research projects almost no significant differences were found between the post-test and delayed post-test, with only *CaveGrasp** noting a significant difference for their grasp condition in one of their five test modes, namely a free recall of L1. In *Object-*

*Manipulation** the participants in the irrelevant movement condition performed poorly, but participants in both the watch-only as the manipulation condition performed almost equally well, with watch-only participants performing slightly better in the posttest and manipulation participants performing slightly better in the delayed posttest.

Mondly VR, *Words in Motion* and *Ogma* also include one non-IVR condition. *Mondly VR* worked with two conditions, where (i) one group learned in IVR with the *Mondly VR* application and (ii) the other group watched video on a desktop of *Mondly VR* gameplay. In *Words in Motion* participants tested in one of three conditions: (i) a kinesthetic condition in IVR, (ii) a watch-only condition in IVR, and (iii) a non-IVR text condition in which participants would look at a computer screen and read the L1 and L2 word. *Ogma* tested an IVR condition against a non-IVR condition where participants learned with a word list and flash cards. In all three research projects was a positive significant difference found for the IVR condition when looking at retention.

Collins *et al.* [84] with *IrishSuper** do not compare between conditions or two posttests, but have their only posttest one week after their experiment, so retention rate does play a role in their results, but it is not possible to see how many words participants exactly lost between the experiment and the posttest. However, Collins *et al.* did find a satisfying 21% increase on word retention between the pretest and posttest.

4.6.3 Non-IVR versus IVR

Six research projects compared their IVR condition(s) to a non-IVR condition and tested on VA of L2. (i) *Ogma*, (ii) *ZooKitchen**, (iii) *House of Languages* and (iv) *Words in Motion* used for their non-IVR condition a traditional method of L1-L2 word association, where *ZooKitchen** and *Words in Motion* presented the words on a screen while *Ogma* provided all words on paper. For *House of Languages* the traditional method was provided inside the classroom by using a book, lectures and worksheets. (v) *Crystallize* used a desktop equivalent of their IVR project as their non-IVR condition and (vi) *Mondly VR* showed gameplay videos of their IVR condition for their non-IVR condition.

*ZooKitchen**, *House of Languages* and *Mondly VR* found all three that participants in an IVR condition performed significantly better on the posttest than participants in the non-IVR condition. For *Crystallize* the non-IVR group performed slightly better than the IVR group on the posttest, but not significantly. With *Ogma* and *Words in Motion* the non-IVR group did perform significantly better on the posttest than the IVR group. However, *Ogma* and *Words in Motion*, together with *Mondly VR*, also looked at retention with a delayed posttest, where it was found for all research projects that IVR participants had a significantly higher retention rate than non-IVR participants, meaning that after some

TABLE III

Overview of research projects comparing a non-IVR and IVR condition and their results.

	non-IVR condition type	**posttest score higher for**	**retention rate higher for**
Ogma [8]	trad. (paper)	non-IVR (sign.)	IVR (sign.)
Words in Motion [9]	trad. (computer)	non-IVR (sign.)	IVR (sign.)
*ZooKitchen** [6]	trad. (computer)	IVR (sign)	-
House of Languages [41]	trad. (classroom)	IVR (sign)	-
Crystallize [73]	desktop equivalent	non-IVR	-
Mondly VR [85]	gameplay videos	IVR (sign.)	IVR (sign.)

time had passed there was a larger word loss for non-IVR participants while IVR participants still remembered most of their words.

Cheng *et al.* [73] with *Crystallize* conclude after their posttest that the impact of IVR was inconclusive, as they find no significant difference between their non-IVR and IVR participants with the non-IVR participants also performing slightly better. However, it could be that a possible impact might have been found if participants would also have participated in a delayed posttest, similar to *Words in Motion* and *Ogma*.

Overall IVR participants score slightly more often significantly better on posttests than non-IVR participants, but when a non-IVR group scores significantly better on the posttest, then the retention of the IVR group is significantly higher. Therefore we conclude that a non-IVR condition can be left out if the focus of a research lies on the long term effects of IVRALL+VA, but only if participants are also tested on retention.

4.6.4 Activating words and manipulation

Projects with a conversation approach have no objects for participants to activate and learn, while all projects with only a movement approach present their objects, if they have any, in a predetermined order to their participants, making it unnecessary for participants to activate the objects themselves. However, all projects with an exploration approach enable participants to activate an object so they can hear either the audio, read the text or do both. In *ClassroomVS** participants go past the target words in a predetermined order, similar to most movement approach projects, but participants are here free to decide for themselves when they continue on to the next target word. Target words are then activated by simply stepping into their vicinity. In *House of Languages* and *ProtoQuiz** target words are selected by gazing at the target word, however, Garcia *et al.* from *ProtoQuiz** concluded from participant input that they want to switch to touch controllers for a next iteration. In *Ogma* participants can activate target words by pointing at the target word objects while wearing a Myo bracelet that detects arm movements. In *ZooKitchen** and *Words in Kitchen* participants can also activate objects by pointing at them, but use instead an IVR controller where in the VE a ray is cast from the controller that can be used to point at target word objects, after which clicking on the object then activates it. In *IrishSuper** target word objects are activated by picking them up from their shelves.

*ZooKitchen** took both an exploration as a movement approach, and also looked at the difference between learning words that can be manipulated (i.e. objects in the kitchen) and objects that cannot (i.e. objects in their zoo), and found that objects that can be manipulated in the VE are more accurately learned. Therefore, if conditions that do not involve manipulation are researched, it might be better to let all target word objects be either picked up and moved around or to let all target word objects be unmovable, to not let manipulation create a difference between target words and how well they are learned. An approach like *IrishSuper** is then possible if all items can be picked up, while a pointing towards method can be used while also being able to pick everything up, like in the kitchen from *ZooKitchen**, or for VEs where all target word objects are static. Approaching a target word object to activate it is also a possibility, but worked in *ClassroomVS** because target word objects were presented along a clear walking route with able space between the objects, while using the approach of participants becomes troublesome if target word objects are presented closely next to each other in a VE.

4.6.5 Target word objects

All research projects that work with target word objects include only objects that are logical for the participant to encounter in the environment in which the objects are presented. So are animals placed in a zoo, kitchen utensils in kitchens, types of luggage in a customs point, and living room furniture in living rooms. Collins *et al.* [84] from *IrishSuper** also name the challenge to only use words that belong to a chosen context to ensure that the experience is authentic for users. Placement of target word objects should also be taken into account when placing target word objects in a VE, as a participant from *Words in Kitchen* mentioned that they associated words with locations, while participants from *Ogma* also commented that they could mentally explore the apartment VE and visualise the target word objects during the test.

Two projects indicate which objects are target word objects in the VE right from the start: in *Ogma* target word objects are marked with an exclamation mark and are in the zoo VE from *ZooKitchen** marked with a hovering gem that also changes colour to indicate that an object is successfully activated. Target word objects in the kitchen VE from *ZooKitchen** that have not yet been activated are called to the participants attention by placing arrows next to them, but only after some time has passed.

The total number of words that are learned by participants in the different research projects lies between eight words for *Crystallize* and 64 words for *IrishSuper**, but those 64 words are learned in a time span of five weeks. The most words learned during one session is for *ObjectManipulation** with 40 words, but those words are divided over three different conditions. The highest number of words learned in one session and for one condition are 30 words for *ZooKitchen**, although participants did need to come back later for another session to learn again 30 words. Most research projects also provided participants with a specific time limit in which to learn the words, with *ClassroomVS** and *Words in Kitchen* letting the participant decide when to continue on, while *Ogma* also lets their non-IVR participants decide when they are finished with learning, while IVR participants are tied to a time schedule.

4.6.6 New and learned contexts

Words can be learned in a new or learned contexts, as further detailed in Section 2.2, where new contexts are contexts in which a learner has not yet encountered a specific word, and learned contexts are contexts that a learner has previously encountered an L2 word in. All research projects let participants learn in at least a new context, as all presented VEs are new for all participants. *IrishSuper**, however, also creates the opportunity for participants to learn in a learned context.

In *IrishSuper** the main focus lies on improving the Irish language identity of Irish language learners. To allow for the observation of a gradual increase in presence ratings, three sessions of 20 minutes each were spread along five weeks. During a session a participant received a grocery list in IVR and had to find all target words on the list in the IVR supermarket VE. The supermarket was always the same, but the target words were always different, meaning that if a participant only interacted with the words on their list that all words were learned in a new context, but if they explored the supermarket and also interacted with other items, and repeated this action for the same items in another session, that they were also learning words in a learned context. So depending on the behaviour of the participant, the context in *IrishSuper** could also be a learned context, although this is not further evaluated by Collins *et al.*

4.6.7 Participants

Most studies that compared conditions chose a between-subject approach for testing their conditions (i.e. *Ogma*, *House of Languages*, *ClassroomVS**, *Crystallize*, *Mondly VR* and *Words in Motion*), while *ZooKitchen**, *CaveGrasp** and *ObjectManipulation** chose a within-subject approach. Proof of concept research like *ProtoQuiz** and *Words in Kitchen* used respectively four and five participants to evaluate their project design, while other research varied between ten and 68 participants. Ages of participants vary from eight years old from *Words in Kitchen* to an average of 33.17 years with an SD of 15.95 years from *ObjectManipulation**.

4.6.8 Recycling words in different contexts in IVR

Using IVR for VA seems to be a promising addition to the many different methods of L2 word learning when looking at the results of contemporary IVRALL+VA research. IVR participants often score high on posttests and have a higher retention rate than non-IVR participants. These scores are, with the possible exception of *IrishSuper**, acquired by participants who have interacted during one learning session with target words in one VE. A learning session is here defined as the period that a participant can study target words and which ends when the participant is not able to study the target words further.

Learners need, as explained in Section 2.1.2, multiple encounters with a word to gradually learn more and more word knowledge aspects because it is not possible to learn all word knowledge aspects at once. Therefore it is important that words are recycled during learning and that it is possible to have multiple encounters with a word to increase the depth of word knowledge. Providing multiple encounters with a word, so recycling a word, is also possible in IVR, but if a word is encountered a second time by a learner then this can be either in a new or learned context. Since the discussed sixteen IVRALL+VA research projects did not recycle or evaluate on recycling words, which was also not necessary for their research, the question is raised if and how presenting learners with recycled words in either new or learned contexts might affect learning.

To study the possible effects of learning in either a new or learned context inside IVR we propose a system called Wics that recycles words in multiple learning sessions. To do this, we choose to follow an exploration approach to keep a sole focus on learning words. Projects with a movement approach include always at least one movement variable, while studying specifically the contexts in which learning occurs does not need to be connected to any movement. Moreover, if not all words can be manipulated then it is probably better to remove all possibilities for manipulation to avoid learning benefit differences between words, which was shown in *ZooKitchen** and mentioned in *Ogma* and *Words in Kitchen* by participants. Similar to projects with a movement approach, will a project with a location approach always study the influence of the location of target word objects on learning. Creating an opportunity for learners to also remember the location of objects, which is more easily accomplished in a learned context than in a new context, can also provide learning benefits as Cho showed with *ClassroomVS**, so can better be avoided when comparing contexts.

Projects with a conversation approach work often with fabricated contexts to enable conversations, that take most often place in new contexts, as detailed in Table II, as repeating the same words in an entirely new context can make the conversation feel forced or artificial, making a conversation approach also less suitable for comparing learned with new contexts.

Projects with an exploration approach follow the design behind rote association memory exercises, where traditional rote association memory exercise methods provide the L2 word in text, while IVR projects include also at least the visualisation of an L2 word. Such rote association memory exercises are for the traditional text method most common in learned contexts, however, since IVR can easily switch between different VEs, it becomes possible to also present L2 words in new contexts during rote association memory exercises in IVR. Therefore an exploration approach can be suitable for presenting words in new and learned contexts.

Previous IVRALL+VA research projects have shown that posttests can be made significantly better by both IVR as non-IVR conditions, but that the retention rate is always significantly better for an IVR condition, see Table III. Since language learning is mainly about acquiring a skill for the long term, so where retention rate is more valuable than remembering words for a short period, we choose to not compare learning in new and learned contexts in IVR with a non-IVR equivalent, but to test instead on retention after one week.

5 Experiment design

The focus of the experiment, based on the findings detailed in Section 4.6, is studying possible effects when words are *recycled* in multiple learning sessions and are presented in either new contexts or learned contexts. For the IVR world we provide an exploration-based learning approach, meaning that participants learn target words by activating target word objects in the environment and participants can decide on the order the words are learned. None of the target word objects can be manipulated by participants to prevent a learning disadvantage for target word objects that cannot be manipulated. Target word objects that are presented in learned contexts need a different location in each learned context to prevent a learning advantage for learning in learned contexts. Otherwise learned contexts provide an extra memorability for learning (i.e. location) that new contexts do not. Lastly, the experiment will test for retention after one week has passed. To test for retention the experiment will have a posttest and a delayed posttest.

Testing must, by definition, also occur in either a new context or a learned context. Language learners try to learn, and therefore know, another language. By knowing another language it might be possible to get a job, emigrate to another country, talk to people on a foreign holiday and much more [88]. The conversations, but also texts that are encountered in such situations, are not scripted but occur naturally and are often unexpected, meaning that they take place in new contexts. Therefore, to improve external validity of the study, we choose to let the posttest and delayed posttest also be in a new context to see how learning in either context also prepares for such an unexpected encounter.

In summary, the experiment is set up to answer the following research question:

> RQ: What are the effects of recycling words in IVR in learned or new contexts on retrieving words when encountered in a new context?

5.1 Context specification

The difference between conditions revolves around the concept of *context*. Context can include and exclude different things depending on what is included in the concept. Therefore we will first define the concept of *context* so we can determine what exactly must change to make something a new context, while all other elements stay the same, while also deciding on the exact number of contexts that will be used for our experiment. Lastly, the time allowance for participants is discussed.

5.1.1 Context

A context for a word can be many things, but since IVR can present objects visually to users, here the context will be the representation of the target word, which is the visual object of a word, and the virtual environment in which the word is presented. Therefore, to create a new context, the two things that are changed from the learned context are (i) the visual representation of a target word and (ii) the virtual environment.

5.1.2 Number of contexts

Users must have multiple learning sessions to come across multiple contexts, either new or learned, but the question is *how many*. Having more learning sessions will increase the difference between the two conditions, but after some time people will stop taking in new information. Furthermore, for non-experienced IVR users it might also be problematic to stay for a prolonged time in IVR and each additional new context learning environment will increase the time needed to go through the experiment. Non-experienced IVR users might experience nausea or other IVR related discomforts, which might also hinder learning and give therefore a disadvantage to non-IVR participants. Providing participants with two learning sessions creates a difference of one learning environment between participants, as the first learning environment will always be in a new context. Three learning sessions adds an extra learning session where the context is different for both conditions, so then the learning environments would be two-thirds different from each other and one-third the same. Providing participants with four learning environments might require too much of the concentration of participants and their ability to take in information. Since all learning sessions are shortly after each other it might also not be necessary for participants to see words for a fourth time if they have learned them three times relatively shortly before. It could be that some last and difficult words are finally stored in memory during such a fourth learning session, but the chance for that happening does not outweigh the possibility of participants experiencing discomfort or boredom. Therefore we choose to let participants have three learning sessions to let participants encounter a different learning environment twice in the new context condition. With three learning environments (i.e. one learning environment shared between conditions and two additional learning environments for the new context condition) and two test environments (i.e. one for the posttest and one for the delayed posttest) there are five VEs needed for Wics.

5.1.3 Time allowance

Two IVRALL+VA research projects, *ClassroomVS** and *Words in Kitchen*, allowed participants to learn words until the participant decided that it was time to continue on, where in *Ogma* non-IVR participants were also allowed to learn to their heart's content. Other research projects had a strict time allowance for learning words, including *ZooKitchen** where participants had 20 minutes to learn 30 words and *Ogma* where IVR participants had 5 minutes to learn 10 words.

Our experiment will not exclude participants based on their IVR experience, thus we include both IVR experienced participants as participants who have never been in IVR before or have little experience. It is expected that non-experienced IVR participants are able to move around the VE and interact with target word objects, but that they are slower in doing so. Setting a time limit might thus negatively influence results for non-experienced participants. Participants will therefore all choose themselves when to continue on to the next VE by trying to find the moment that they are content with their learning and they feel that there is not much more to gain by staying in their current VE. Hopefully this also removes any possible pressure a participant might feel with an imposed time limit, since stress might also affect results negatively.

5.2 Experiment setup

The experiment is a between-subject design with the two conditions: (i) learning words by recycling them two times in a new context in IVR and (ii) learning words by recycling them in two learned contexts in IVR after learning them once in a new context. The independent variable of the experiment is the context in which words are repeated for learning a second and third time (i.e. new or learned) and the dependent variables are the test scores for the posttest and delayed posttest, the number of activations of target word objects, and the time spent in the VEs. For the experiment the context is the visuals of the virtual environment and the target word representations. Participants can study the target words until they feel ready to continue. Table IV provides an overview of the experiment setup.

5.3 Hypotheses

Participants in both conditions will have seen all target words in three learning sessions before doing the posttest. Participants are encouraged to learn the words until they feel content in each learning session. Participants can still rely on their short-term memory when doing the posttest and are expected to make use of this memory to retrieve the learned target words. Therefore we hypothesise:

H1: New and learned context participants will perform similarly on

TABLE IV
Setup of the experiment for both contexts

Learned context	**New context**
Learning environment 1	
Environment A Object representations A Object locations A	
Learning environment 2	
Environment A Object representations A Object locations Y	Environment B Object representations B Object locations B
Learning environment 3	
Environment A Object representations A Object locations Z	Environment C Object representations C Object locations C
Posttest	
Environment D Object representations D Object locations D	
Delayed posttest	
Environment E Object representations E Object locations E	

the posttest.

There are 18 word aspects that can be learned for a word [3], as described in Section 2.1.2 and listed in Table I, like what other words it makes us think of or how it is pronounced. It is impossible to learn all word aspects simultaneously so multiple encounters are needed to learn different word aspects, where each encounter with a specific word aspect also deepens the knowledge for that word aspect. If such an encounter adds new information to a specific word aspect then knowledge about this word aspect will deepen further. Participants in both contexts are presented the exact same information regarding word aspects with the exception of the word aspect of *what is included in the concept* which also refers to how a word can look as a visual object. Here participants in the new context condition receive additional information about this word aspect in each learning session, while participants in the learned context condition always receive the same information about this specific word aspect. Because new context participants have received more information regarding how a target word can visually look as an object, we expect that this has deepened their word knowledge in comparison with learned context participants. In the delayed posttest, both groups need to rely on their long-term memory instead of their short-term memory, where we expect that a more deepened word knowledge allows for a better retrieval of this knowledge. Therefore we hypothesise:

> H2: New context participants will have a higher retention than learned context participants.

5.4 Conclusion

The main directions for this thesis, which were established in Section 4.6.8, have been bundled in a research question which places a focus on researching the effects of recycling words in different contexts. A context is defined as the virtual environment and the visual representation of target words. There is no time constraint for learning so both non-experienced IVR users as experienced IVR users can be included in the experiment. For this experiment a system must be built for IVR so it is possible to recycle words in IVR in different contexts. The design of such a system is described in the next chapter.

6 System design

Participants must be able to explore multiple contexts in IVR for the experiment in order to answer the research question introduced in the previous chapter. Therefore, a system is needed to carry out the experiment. The system must have certain requirements in order to do the experiment set up in the previous chapter. The full list of identified requirements is discussed in the next section. In Section 6.2 a platform is chosen to build the system in, namely *Neos VR*. Choices on how to fill in the requirements specifically are explained in Section 6.3. How target words and their objects, together with virtual environments, are selected is discussed in Section 6.4. Putting the virtual environments and the target word objects together is discussed in Section 6.6, while specifics regarding the functionality of the system are discussed in Section 6.7. Lastly, in Section 6.8 is detailed how the system explains its workings to participants through an additionally designed VE that functions as introduction room, and in which is explained how participants can take breaks.

6.1 Requirements of the system

A system is needed to do the experiment, and requirements for the system should enable the system to be used in the experiment. Requirements are separated into functional requirements, which are features that must be implemented in the system to make it function for our purposes, and non-functional requirements which are properties of the system. Those are both further divided into requirements that are needed to learn words for the experiment, and requirements that are needed for the workings of the system.

6.1.1 Functional requirements

The experiment has two conditions, where in both conditions participants learn words in three learning sessions, but in one condition the context is new in each learning session while for the other condition the context is learned for the second and third learning session. Breaks are needed between learning sessions to avoid that participants are overloaded with information and to provide non-experienced IVR users with time outside of IVR between learning sessions. Since the experiment has an exploration-based approach, the target words should be presented as visual object representations. Because it is not possible to learn an L2 word without providing a word form of the L2 word (i.e. written or spoken form), there needs to be a clear connection between a visual object representation and the L2 word form. It is also difficult for learners to learn more word aspects if the link of an L2 word to the equivalent L1 word form has not yet been established [4], so the system should also enable the user to make that link as effortlessly as possible by presenting an L1 word form at the same time as the L2 word form. However, participants can have different L1s from each other for this experiment, making it difficult to present participants with their own L1. We therefore replace the L1 with an English translation and require all participants to speak English as either an L1 or an L2. As the experiment tries to look at the word knowledge of participants when they are confronted with a new context, and participants are tested for retention, a posttest and delayed posttest are needed. With these specifications for the experiment we identify the following functional requirements for the system with a focus on learning words:

- The system must present a user with three new contexts or one new context followed by two learned contexts
- The system must present a user with visual object representations of target words
- The system must provide the user with a form of L2 (i.e. written form or audio form)
- The system must enable the user to make a link between L1 and L2
- The system must allow for multiple learning sessions and include breaks
- The system must test target word knowledge twice in a new context, once for the posttest and once for the delayed posttest

To acquire multimodal representations of an object in the brain, as discussed in Section 2.2.3, an object needs to be close to a person. As we want to take advantage of this benefit which IVR can provide, participants must be able to move around in the VEs so they can get close to the target word objects in it. Moving around in IVR also supports the exploration-based approach of the experiment. To see the L2 word form and English translation belonging to a target word it should be possible to show this connection to the participant. Several VEs are presented for the experiment and participants must be able to reach those different VEs. Participants also should not need to divide their attention by having to figure out how the controls or workings of the system work while trying to learn words. Such additional attention loss would not be beneficial for word learning and it creates a disadvantage for non-experienced IVR users when compared to experienced IVR users. To minimise differences between participants with a different IVR experience level, everything should be optimised for both groups. Some data also needs to be collected regarding the experiment, so time duration, word clicks and word input need to be stored so that data can be retrieved. Participants should also be able to communicate target words that they knew prior to participating, since these words will by default be filled in correctly at both the posttest and the delayed posttest which will skew the results for retention. Thus the following functional requirements are identified for the system, with a focus on the functionality of the experiment:

- The system must enable the user to move around
- The system must enable the user to activate target words
- The system must enable users to become familiar with the locomotion, controls and workings of the system before words are learned
- The system must collect time duration, word clicks and word input of a participant for analysis
- The system must allow the user to go from one VE to another VE
- The system must allow users to indicate that they knew a target word before entering the system
- The system must be optimised for both non-experienced IVR users and experienced IVR users

6.1.2 Non-functional requirements

Manipulation can help with learning a word, just as the location of an object can act as a mnemonic for remembering that word. To prevent such advantages for some words but not for others, or for one condition over the other, it is necessary to not have target word objects that can be manipulated and to always provide target word objects with a new location in learned contexts. However, in both contexts a mnemonic might also be created if target word objects are often grouped together with the same target word objects, so this should also be avoided to prevent a learning advantage for some words. Similarly, a VE might create a more beneficial environment for learning if elements of the VE are moving which might increase immersion. Thus, to keep VEs between conditions as similar as possible to each other, VEs should not have any elements that move in them. Similarly, VEs should roughly have the same walking size area and should invite a similar level of exploration, so if not all target word objects can be seen in one environment then they should also not be shown all at once in another. VEs in the new context condition should also differ from each other as the VE is one of the two elements that is included in the definition of *context* for this research. Therefore a VE should appear clearly different from any other previously encountered VE. Lastly, all words should be seen during a learning session to avoid a disadvantage for some words. With these specifications in mind we discern the following non-functional requirements with a focus on learning words:

- The system must ensure that users see every target word in the learn and test environments
- The system must have target word objects for each learned context in a different location
- The system must not allow any manipulation of target word objects
- The system must not have elements that move and are part of the VE
- The system must have VEs with a roughly equal size walking area
- The system must have VEs in which all words are not visible in one view
- The system must provide new context VEs that appear and feel different from each other
- The system must have VEs in which target word objects are not grouped together with the same target word objects across multiple VEs

The focus of participants should be on learning words, so the composition of VEs in combination with the target word objects that are placed in them should not call for attention by standing out because the aesthetic of objects does not belong or because the objects are illogically placed. Furthermore, participants should be able to deduce if something is a target word or not, so they can purposefully learn words instead of spending time on guessing and trying to understand on what they should focus. Therefore, if a target word is represented by multiple objects in a VE, for example there are multiple coins to depict the target word *coin*, then all the objects belonging to this one target word should be grouped together in the VE, so all coins should be placed close to each other. Then non-target word objects can be used for decoration of the VE as long as these objects are spread out over the environment. Thus participants will hopefully deduce that objects that are bundled together are target words (e.g. coins bundled together), while objects that are strewn about the VE are merely there for decoration (e.g. bushes spread out over a garden). Thus the following non-functional requirements of the system are discerned:

- The system must have target word objects that fit in to the VE that they are in
- The system must have VEs in which objects are logically placed
- The system must have objects that belong to one target word grouped together at the same location in a VE

6.2 Choosing a building platform

Target participants include people who have IVR experience and own a headset of their own. Therefore it is preferable that the experiment can be offered through an already existing program that is free to use so it is not necessary to already own the program. To also allow for participant recruitment from within the program, the program should also have an active player base. Furthermore, the program should allow for interactive world building so it is possible to create a custom made experiment.

Five programs were found that allowed for active world building: (i) *VRChat*, (ii) *Sansar*, (iii) *Neos VR*, (iv) *Rec Room* and (v) *AltspaceVR*. However, the player base of *Sansar* turned out to be almost gone and the player base of *Rec Room* are mostly children, making both programs not suitable for the purposes of this research. *VRChat*, *AltspaceVR* and *Neos VR* all have an active player base, with *VRChat* having the most players.

VRChat and *AltspaceVR* allow for interactive world building but need an external program outside IVR (i.e. *Unity*) to realise this. *Neos VR* offers an all-in-one solution to world building and offers all interactive world building possibilities inside the program itself, making it possible to see at all times how the world looks while in IVR. Therefore we choose to use *Neos VR* for building the experiment.

Neos VR is a free-to-play massively multiplayer online virtual reality metaverse. It provides each user with a *home world* VE in which they always begin when starting up the program. Players can also create a new *world*, which opens up a new and empty VE in which the player can build things. When saving such a world a *world orb* asset is created which represents the saved world. An example of a world orb can be seen in Figure 21. The world orb is an orb roughly the size of a small hand and by interacting with it, it is possible to again open and enter that specific saved world. Players can share these world orbs with each other to enter worlds that others have made. It is possible to share world orbs when players meet each other or to send world orbs through the in-game message system from one player to another player. If multiple players enter the same world, then players can meet up inside that world. However, it also possible to open up a private session of a world to be in that world without others.

Three worlds are created for this experiment: (i) a world which stores the first experiment part for participants in the new context condition, (ii) a world with the first experiment part for participants in the learned context condition, and (iii) a world with the delayed posttest for all participants.

Learning Japanese words in VR - (visit this world to learn more about this experiment)
by Thérèse

Figure 21: A world orb with the title of the world up top and the username of the world creator below.

6.3 Choices for the system

The established requirements from Section 6.1 are requirements that a system must adhere to in order to be able to do the experiment. However, some requirements can be filled in in different ways. Choices made regarding those requirements and their chosen specification are discussed here.

6.3.1 L2 form and translation to English

To provide users with the possibility to learn the L2 form of a target word when presenting it as a visual object, a target word object also needs a written or spoken L2 form connected to it. Since learners have different preferences for learning words, we choose to offer both to users. Thus the written L2 word form is seen each time a target word is activated, and the spoken L2 word is also heard for each activation. However, as learners tend to first create a form-meaning link, the English translation of a target word must also be provided upon activation to minimise their cognitive load, so participants do not have to come up with the meaning themselves. Because there is already an audio fragment playing upon activation, it is undesirable to also play an audio fragment of the English translation at the same time. Therefore we will show the English translations only in their text word form.

6.3.2 Break duration

For the break duration between learning environments and the test environment a minimum duration of three minutes is chosen so participants will at least take a small break. However, since some participants will need a slightly longer break than others before being able to take in more information, the maximum break duration is set on fifteen minutes.

6.3.3 Number of target words

Participants do not have a time restriction for learning and have not one but three learning sessions to learn each word in. Most IVRALL+VA research projects did have a time restriction and one learning session to learn all words in from the IVRALL+VA research project experiments. With 30 target words *ZooKitchen**, has the highest number of words that participants had to learn for one condition in one learning session. For *ZooKitchen** it was also mentioned that some participants scored really well on the posttest, even with 30 words. Since Wics provides participants with more and possibly longer learning sessions, and *ZooKitchen** showed that close to 30 words can be learned by some participants in 20 minutes, we propose to have a minimum of 30 target words that can be learned with Wics. However, we do not want to include many more target words since participants must activate all words at least once for each learning environment, so three times in total, and interact with the words during the test, meaning that participants must interact at least 120 times with words when there is a minimum of 30 target words. To keep people interested and concentrated, we decided to keep the number of target words similiar to the number used in *ZooKitchen**, with a final chosen number of target words of 32, of which the rationale is explained in Section 6.4.3.

6.3.4 Indicating missed words

Participants are allowed to explore the environments without any time pressure. This lack of time pressure removes the necessity to make participants aware at all times of what a target word is, as happens in *Ogma* by placing an exclamation mark next to each target word. There is also a risk of a VE becoming cluttered with target word indications with a minimum of 30 target words for Wics. However, missed target words should still be communicated to the participant to avoid frustration from endless searching because participants cannot continue on to the next VE if they have not yet activated all target word objects at least once in their current VE.

Since participants can search for target words and learn them without a time limit, there is no need to make participants aware of missed target words at a specific time. Instead, participants should decide for themselves when they are convinced that they have seen everything in a learning environment. Therefore participants should be able to activate a button at a time of their choosing after which missed target words are indicated in the VE. As indication for missed target words a small orb that glows brightly yellow was chosen, since it draws attention by shining, while avoiding becoming obtrusive by taking up much space or being hard to ignore. It can also resemble, for those familiar with the franchise, the light of a fairy when seen from afar from *Peter Pan* or a fairy without wings from *The Legend of Zelda*, providing it with a bit of rationale for why its floating (i.e. fairies can fly) or guiding the participant (i.e. fairies from both franchises are known for their offered guidance). Participants should also be able to turn the light orb indicators off if they do not want them to be a part of the VE any longer, by pressing the button again. An example of the small orb is shown in Figure 22.

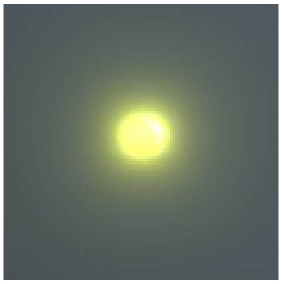


Figure 22: Example of an orb that indicates a missed target word.

6.3.5 Activating objects

To learn the L2 of target word objects the user is provided with the L2 text, L2 audio and L1 text of a target word object when the target word object is activated. These texts should always be directed towards the participant, even when they turn away or turn around, so texts can always be read. To not overwhelm the user with at least 60 words hovering in the air of the environment (i.e. one L1 word and one L2 word for each target word object), and to allow the user to think about the object without other word aspects imposing themselves on the user, the L2 and L1 representations must be activated by the user when they want to see those word aspects, while the word aspects remain invisible and silent at all other times.

There are several methods to activate a target word in IVR, when looking at IVRALL+VA research projects. Words can be activated with gaze, picking the object up, by approaching the object or pointing to the objects and clicking. Selecting target word objects with gaze was not well received by participants from *ProtoQuiz** and will therefore not be used here. Picking objects up defies the requirement that manipulation with objects must not be possible, so is also discarded. There are at least 30 objects in the VE, making it almost unavoidable that objects are next to each other. Using participants' proximity for activation would thus result in multiple audio forms of words activating at the same time. Since a cacophony of sounds will make it difficult to focus on specific words, target word objects should not be activated by using proximity. The point and click method offers participants the possibility to be in control of when a target word object is activated since the participant not only has to point their ray cast on the object, but must also click a button to activate it successfully, making it difficult to activate objects by accident unlike the gaze or proximity activation methods. *Neos VR* also comes with ray cast functionality included. Therefore a point and click method is chosen for target word object activation. To also make the participant in charge of the duration in which they can read the text, we choose to let the participant not only click the button, but to also hold it until they are finished with the text form of a word.

6.3.6 Testing

Wics presents participants with five word aspects, of which all 18 word aspects are discussed in Section 2.1.2 and presented in Table I, in all learning environments. These five word aspects help with the receptive knowledge of a word, so the knowledge that is needed to recognise a word. This is in contrast to productive knowledge, which is needed by the learner to produce a word aspect themselves. For each word aspect that is connected to gaining receptive knowledge there is a word aspect counterpart that is connected to productive knowledge. The five word aspects that participants encounter during the learning environments are (i) *What does the word sound like?* through the audio fragment of the target word which is provided upon activation, (ii) *What does the word look like?* which is shown as L2 text word when activating the target word object, (iii) *What meaning does this word form signal?* which can be deduced from the target word object and which can be confirmed with the English translation text that is also shown upon activation of the target word object, (vi) *Where, when, and how often would we expect to meet this word?* which connects to the environment a word can be found in although in our experiment the *how often* is consistent between VEs, and (v) *What is included in the concept?* which is taught by showing visual representations in the form of an object for each target word. An overview can be seen in Table V. The last word aspect of *What is included in the concept?* is the only word aspect that is different between conditions, as new context participants see a different visual representation for each learning session while learned context participants see the same visual representation in each learning session.

The tests are the first time that participants are specifically asked for their productive knowledge. Therefore we do not want to test them on productive

TABLE V

Five receptive knowledge word aspects that are provided in the learning environments, taken from the 18 word aspect framework provided in Table I

Form	spoken	R	What does the word sound like?
	written	R	What does the word look like?
Meaning	form and meaning	R	What meaning does this word form signal?
	concepts and referents	R	What is included in the concept?
Use	constraints on use	R	Where, when, and how often would we expect to meet this word?

knowledge word aspects if those are connected to a receptive knowledge word aspect which the participant has never seen, as it then becomes increasingly difficult to produce the target word. We therefore want to test on the productive knowledge word aspect counterpart of one of the four receptive knowledge word aspects that are presented in the learning environments. Participants can thus be tested on the word aspects *How is the word pronounced?*, *How is the word written and spelled?*, *What word form can be used to express this meaning?*, *What items can the concept refer to?* or *Where, when, and how often can we use this word?*

Because we want to study if learning in a specific context can help with word retrieval when encountering the words in a new context, we are not interested in testing participants on *What items can the concept refer to?* or *Where, when, and how often can we use this word?*. We are, however, interested in if participants are able to come up with a word form when encountering a visual representation, which relates to *What word form can be used to express this meaning?*. However, we want to define that word form here further and ask participants to specifically come up with the written form of a word (i.e. *How is the word written and spelled?*) as not all participants might be comfortable with leaving voice recordings (i.e. *How is the word pronounced?*). Because this is the first time that a participant is asked to make use of their productive knowledge for the written form of a word, we will not evaluate them on how the word is specifically spelled, but instead we shall evaluate them on if the pronunciation produced by their written entry corresponds to how the target word is pronounced.

6.4 Word and environment selection

Wics needs five VEs and a minimum of 30 target word objects to be able to present participants with three learning virtual environments and two test virtual environments that all have their own target word object visualisation. Those target word objects must match the aesthetic of the VE in which they are placed, and VEs should appear and feel different from each other. This

Table VI: 75 initial object suggestions list

knife	couch	cup	mug	sun
rainbow	squirrel	flower	plant	vase
clock	tree	book	cat	statue
rug	umbrella	glass	bottle	remote
water	stone	television	lamp	bird
bowl	hatstand	frog	cable	outlet
chair	table	shoes	socks	coaster
sunglasses	handkerchief	ball	kettle	eye
rose	tea	teapot	computer	airco
coffee	beer	curtains	window	door
key	magazine	screwdriver	spoon	hand
pot	watering can	tablet	phone	brush
mirror	blanket	headset	music	stereo
pillow	cloud	fox	dragon	button
light	Eiffel Tower	picture	painting	closet

section describes the process of how each VE asset and all target word assets were selected for Wics.

6.4.1 Word suggestions and VE asset selection

To create an idea about what kind of objects the environments could possibly hold, a list was created with 75 object suggestions of which most can easily take on different representations and that do not look out of place in different places, as can be seen in Table VI. A search was performed among existing assets with this list in mind to identify possible environments that could house a multitude of the objects from the list in them. Selected assets were checked on ability of customisation with *Blender*, a 3D computer graphics software, and on being able to open in Neos VR. Environments also should not look like each other, so, for example, including a cafe twice is not desirable.

This search for environments resulted in five assets that depicted different types of environments or a different aesthetic style. The original assets can be seen in Figure 23. The five assets are: (i) A round inhabited island with rocks on the sides, palm trees and a fallen log in the middle, covered with a large sky dome with a sun with clouds. The palm tree leaves became invisible when looking up at the sky from under them, so where removed in Blender. (ii) A fully furnished bedroom connected to a fenced garden in low poly style, so with a block-like, simple and often colourful appearance. (iii) A realistic looking apartment with a living room, kitchen, scullery, hallway, two bedrooms and a children's bedroom, bathroom, toilet, and an office, all filled with furniture. (iv) A theatre in low poly style, filled with red chairs that look out to an empty stage. In the air hovered a drawn elephant which was removed with Blender. (v) A simple barn with a bar in it, with an open door at the front and no walls on the left and right side. There are stairs that lead to a small balcony on the side. Walls on the side were added with Blender to close off the building and the balcony was made wider to allow for a person to walk around comfortably.

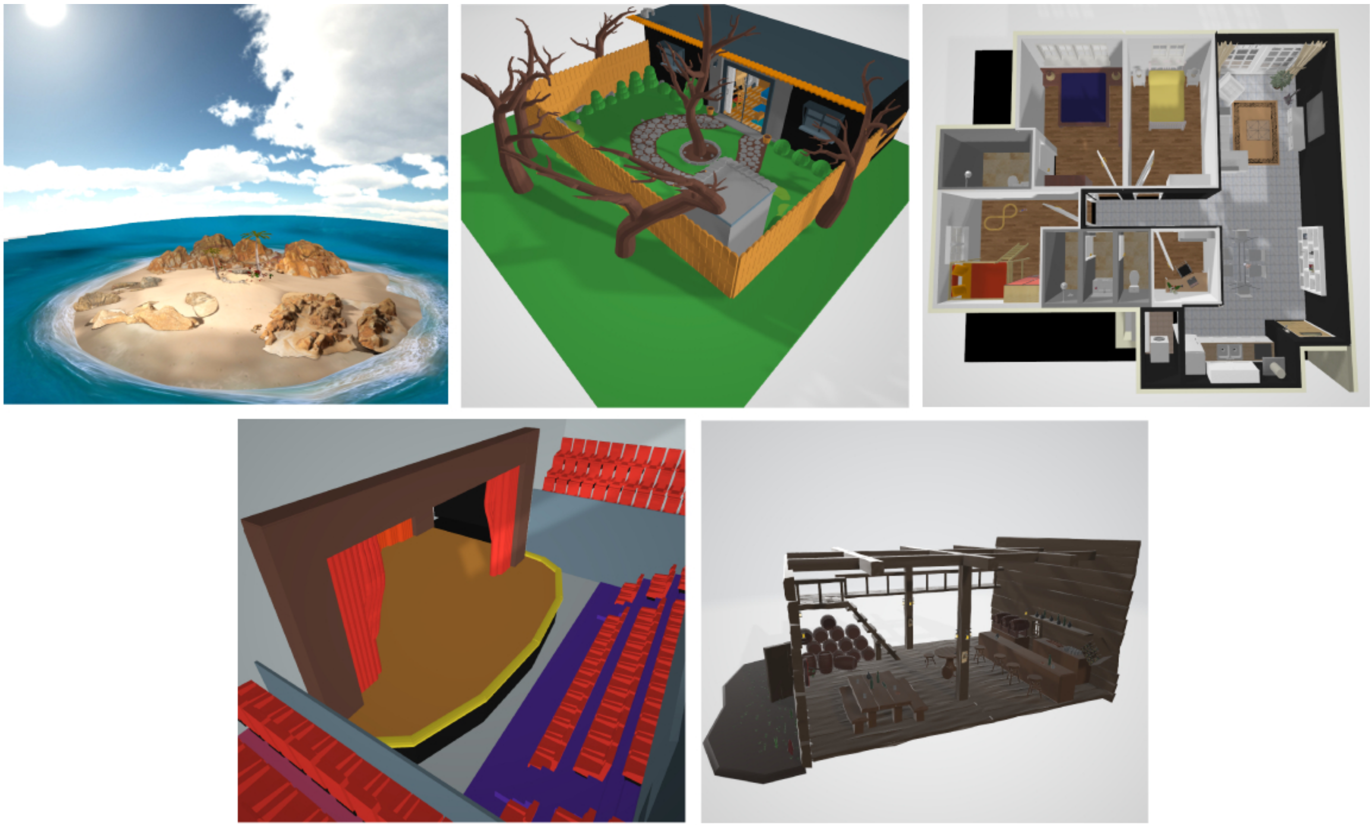


Figure 23: The five original base assets (from left to right, top to bottom): an island, a furnished bedroom with garden, a realistic apartment, a theatre and a simple barn with a bar.

The bedroom with garden environment was selected specifically for its low poly style to present participants with a different aesthetic from some of the other environments. A low poly style creates a simple appearance because there are a minimal amount of sides used to create an object. To maintain the low poly aesthetic of the bedroom with garden, it should be filled with low poly target word objects. Therefore the first search for assets was for low poly objects, as realistic looking assets are more widely available, but if a low poly objects of a proposed target word objects does not exist then that proposed target word cannot be used. A search for low poly assets was therefore performed on *Google Poly*. If a low poly version did not exist for one of the 75 initial object suggestions then it was removed from the list, while new objects that could possibly fit in one of the five chosen VEs were included, resulting in an updated list with 140 possible target word objects.

6.4.2 Narrowing the word suggestion list

More criteria were established in order to reduce those 140 object possibilities. When arriving in the first word learning environment, participants will not know which exact words are target words. If target words objects are mixed with other objects that look like they could be target word objects, then the experience will become a guessing game for the participant. There is also a risk that the participant might get disappointed when they want to learn a word, and expect to be able to learn that word, that it then becomes clear that the object is just mere decoration. It could make the participant doubt all the other objects in the environment. Therefore there should be an intuitive logic behind which objects are target words and which are just for decoration. To create this logic all items from the 140 possible target words list were removed that are part of the essential structure of an environment, like *door* or *window*. Objects of which there are more than one and that are spread out over the environment, like *stone* in the bedroom with garden, were also taken off the possible target word object list. Excluding objects that are repeated throughout the VE allows the VE to be decorated with objects so it does not feel empty, while minimising the affordance of participants to try those decorative objects out as target words.

Environments also need to look logical, so it should be believable that someone outside of the participant's view put everything together and that the environment looks like a naturally occurring environment, instead of participants getting the impression the environment was built especially for them. Therefore all items were removed from the list that would look strange in one or multiple of the selected environments, like *dynamite* in an apartment or only one set of *curtains* in the barn bar which has many small windows. Items were also removed if they needed to be toys or statues in all environments in order to work for the environment, like *hot air balloon* or *lighthouse*, or the participant might become actively aware that everything is specially fitted for their environments, which might diminish their immersion.

Everything should be static inside each VE. Therefore items that would look strange when remaining stationary or unchanged, for example a *fox* or a *squirrel*, but also a *mirror* that is expected to show a reflection, were also removed from the list. Lastly, items of which was expected that it would be difficult to find four other assets of were also removed, like *koala*. This brought the word suggestion list from 140 words to 60 words.

6.4.3 Selecting words on their L2

Next the Japanese equivalents of all 60 remaining possible target words were listed. Then items were selected based on their Japanese equivalent. Items with short (i.e. one or two syllables, e.g. *fune* (*boat*), *hana* (*flower*)), medium (i.e. three syllables, e.g. *kinoko* (*mushroom*), *tsukue* (*desk*)) and long (i.e. four syllables, e.g. *waninashi* (*avocado*), *matsukasa* (*pinecone*)) word lengths were separated, as were words that are clearly loan words from English, like *naifu* from the word *knife*. This last category was also included to provide the participant with words that they can quickly learn to create a feeling of success. Since all remaining words adhered to the aforementioned requirements, a final selection of 40 words was made by choosing words from all categories randomly, while keeping a varied mix of assets with different sizes, but no more than five long words were selected to avoid overwhelming the participant. Words with similar sounds to the words that were already selected were also dismissed, as same sounding words can confuse participants when learning a word for the first time because such words can be easily mixed up. The word for *book*, which is *hon*, and *bookcase*, *hondana*, together with the word for *chair/stool*, *isu*, and *couch*, *nagaisu*,[23] were also purposefully selected, so participants can realise that one word fits in the other. Discovering this connection could provide participants with a feeling of accomplishment, which might motivate them to continue with the experiment.

Since every one of the five environments needs their own representation of a target word, additional assets were sought for every selected word, where target word assets that were already part of the five selected environment assets were also taken into account. For example, there was already a *chair/stool* in the bedroom with garden VE, apartment VE and barn bar VE, so only two more *chair/stool* assets needed to be found. If there was an abundant choice of assets then the different feels of the environment were also taken into account when searching for assets, like an old or broken look for the uninhabited island and a stylish look for the apartment. If less than five different assets were available for a word, then the word was dismissed. This resulted in a final list of 32 target words for the system, which are listed in Table VII.

6.5 VE allocations

With all VE assets and target word object assets chosen it is possible to determine the order in which VEs appear for a participant and their function. Furthermore, one VE should be selected to function as learn VE for the learned context condition, meaning that it is seen as a learning environment three times for participants in the learned context group.

[23] The literal translation of *nagaisu* is *long chair*.

Table VII: Final 32 selected target words

avocado	-	waninashi
bag	-	kaban
ball	-	tama
bed	-	beddo
boat	-	fune
book	-	hon
bookcase	-	hondana
broom	-	houki
butterfly	-	chouchou
camera	-	shashinki
car	-	kuruma
chair/stool	-	isu
chest of drawers	-	tansu
coin	-	kouka
couch	-	nagaisu
desk	-	tsukue
earth	-	chikyuu
fish	-	sakana
flower	-	hana
garbage bin	-	gomibako
glasses	-	megane
key	-	kagi
knife	-	naifu
lamp	-	ranpu
mushroom	-	kinoko
pinecone	-	matsukasa
rug	-	juutan
shoe	-	kutsu
table	-	tebburu
teapot	-	chabin
telescope	-	bouenkyou
television	-	terebi

6.5.1 Learned context VE selection

Of the five selected VEs, two VEs will be used for the two test environments and two VEs will be used once for a learning environment. However, one VE must be used three times as a learning environment for the learned context condition. This VE will also be encountered for both conditions as the first learning environment in Wics to keep conditions the same regarding content for as long as possible.

Of the five VEs, the realistic apartment and the low poly bedroom with garden VEs are least suitable to function as repeated learning environment. Due to the clear function of the rooms and the small room space, many target word objects can only logically appear in a specific area of the VE, making it difficult to really switch the location of objects and to create a total different dynamic for each separate learning VE. There is a bit more freedom to place objects logically inside the barn bar. However, the asset for *car* for the barn bar is a real size car, which stands out in the environment and is difficult to place anywhere except outside where there is limited room for variation. The barn bar is therefore also not the most suitable VE for the learned context conditions.

Remaining are the uninhabited island and the theatre. Both have as advantage that they are almost entirely empty and that they do not have a strong affordance regarding a specific placement for objects. The uninhabited island is located inside a large skybox, where a sphere with an image of a blue sky with clouds and a sun is shown on the inside of the sphere. The theatre is enveloped by walls on stage and chairs at the front, making it a closed box in which the stage is located with no windows, making it also the only VE with no view to the outside or the possibility to step outside. Because participants in the learned context condition must go through the chosen VE three times, and a closed off environment might become oppressive after a while, we choose the uninhabited island as the first and repeated learning environment.

6.5.2 Order and function of environments

There are three learning environments, of which one learning environment is repeated three times for the learned context group, and there are two assessment environments of which one is encountered a week later than the other.

With the uninhabited island chosen as first learning environment, there are four more VEs that must be chosen as either second or third learning environment, posttest environment or delayed posttest environment. The uninhabited island has a realistic style and is seen by participant as the first learning environment. To present the new context participants with also a different style of environment, the bedroom with garden and its low poly style was also selected as a word learning environment for that group. Next, the theatre was selected for the posttest environment. The theatre holds both realistic looking objects as low poly objects, making it an interesting environment to present to participants from the learned context group who only come across realistic looking items and participants from the new context group who have seen both. The barn bar holds also some simple objects and has a simple aesthetic of itself, providing it with a somewhat similar vibe as the theatre, which is why the barn bar was selected as the second assessment environment. That left the apartment as another learning environment for the new context group. New context participants will go through three different learning environments, starting with the uninhabited island. To mix the style of the environments as much as possible, the new context group goes then from the uninhabited island VE to the bedroom with garden VE as second environment, and as third environment to the apartment, to follow an order of realistic, low-poly and realistic.

6.6 Completing the environments

Target word objects must be placed in a VE to complete the VE as a learn or test environment that can be used for the experiment. Therefore the VE assets need to be prepped and target words must be divided over and placed in VEs.

6.6.1 Prepping the environments

After creating the word list and choosing the order of appearance of VEs, it was possible to continue with the creation of the environments. As mentioned before, during the word selection it was established that it should be clear what a target word is in the environment. To accomplish this, all objects that could be mistaken for a target word were removed from environments with Blender. For example, the bedroom with garden has many decorative objects, like frames on the wall, but also a computer, and a lounge chair. Such non-target word objects were all removed, and if there were multiple representations of objects that were target words, like how the apartment has of itself six *chair/stool* assets, then only one was kept per VE. Objects that could remain were objects that were part of the building, like doors and windows, and objects that were in the VE in abundance and spread out, like bushes and stepping stones in the garden. The apartment has many different chambers, but since there are only 32 target words, all rooms were closed off with a door, instead of the door standing open, and only the living room with kitchen and one of the bedrooms remained reachable for the participant.

All five environments should also have roughly the same walking space area size, and should provide a similar view on words to keep the two conditions roughly the same. So words should not all be visible at one glance in one environment.

Therefore, two small sets were added to the open stage to block the view of the participant so it becomes not possible to see all words with one look. The walking area size in the roof barn was larger than the other areas. Therefore, the roof was lowered and no longer making the top of the stairs available to go to the balcony. Lastly, the environment should, just like the target word objects, be stationary to not create a possible unintended favourable learning condition for one environment over another environment. Therefore, no moving water was used to encompass the uninhabited island, but instead a flat plane was used with a water colour to appear like a sea.

6.6.2 Dividing target words over environments

All selected 32 target words have five or more corresponding assets, of which some assets are already part of an environment. For example, the barn bar already had a *knife* and *lamp* and the apartment had among other things already a *couch* and *table* as furniture. However, the assets for the target words that were collected still needed to be divided over the environments that did not yet have that target word. Due to starting the search for assets with only low poly assets, it was possible to provide the bedroom with garden environment with an entire set of low poly objects. Realistic looking items were divided over the uninhabited island, apartment and barn bar, with stylish and new looking items going to the apartment, old-fashioned looking items going to the barn bar, and broken, mended and random items going to the uninhabited island. The assets still left were a mix of low poly and realistic objects, but since the last environment is a theatre, and props can come in many different colours and flavours, a colourful mix of both low poly and realistic was selected for the theatre. The size of objects was also not always consistent with their non-prop counterpart to strengthen the feeling that the theatre is filled with props.

6.6.3 Placing target words in the environment

Target words objects are placed in a VE with the purpose that a participant is able to find a rationale behind its placement. For example, there are three versions of the uninhabited island for the learned context group and the target word objects are placed differently in each environment. Each island tries to convey a different personality for the person who must once have inhabited the island. One island has the telescope in the middle of the island, with a chair and glasses next to it to see clearly through the telescope. A bag is already packed in the boat to leave at any moment as soon as a ship is spotted. Simple objects like a toy car are placed aside as unwanted distractions. Another island tries to show more the spirit of someone who has accepted their fate and who has tried to make the most of it. They have made a chill corner with the bookcase, a sofa, and on a table next to it are a teapot and a lamp for easy reading.

Objects are also mixed up for each environment, for both conditions, to not be next to the same object in another environment. The only exception on this are the *bookcase* and *books*, which can always be found together, just like how their Japanese words are similar and belong together. A complete overview of the visual representations of all target words in each of the environments can be found in Appendix B. An overview of every filled learn and test environment can be found in Figure 24 and additional images of all VEs can be found in Appendix C.

6.7 System functionality

Different functionalities to make the system work are discussed in this section.

6.7.1 Word activation

Pointing a ray cast on an object and holding a button on the controller must activate a word, so show the L2 written word of the object together with its English translation while playing the L2 audio form. This functionality was implemented for all target words in learning environments. If a participant activates a target word object, a black line goes from the object towards the L2 word, as can be seen in Figure 25, to indicate that the object and the word belong together. The black line also underlines the L2 word. Under the underlined L2 word there is an English translation that is smaller than the L2 word so it does not ask for as much attention. The words and the black line will always rotate with the head position of the participant, so the words are always readable regardless of the participant's positions. The audio is played once and the texts disappear if the ray cast is not pointed on the object any longer or if the participant lets go of the button. If participants want to hear the spoken L2 form again or read the texts once more, then they must activate the object again.

6.7.2 Word fill-in

Words need to be filled in by participants as text input in the test environments. One text field was therefore added to each target word object in the test environments. The corner of a text field always touches the surface of the target word object it belongs to so participants know for which target word they need to provide input. A few examples of empty text fields connected to their target word object can be seen in Figure 26. Clicking a button on the controller while pointing the ray cast on a text field *spawns* (i.e. the creation of a character, item or NPC in a VE) a virtual keyboard in front of the partic-

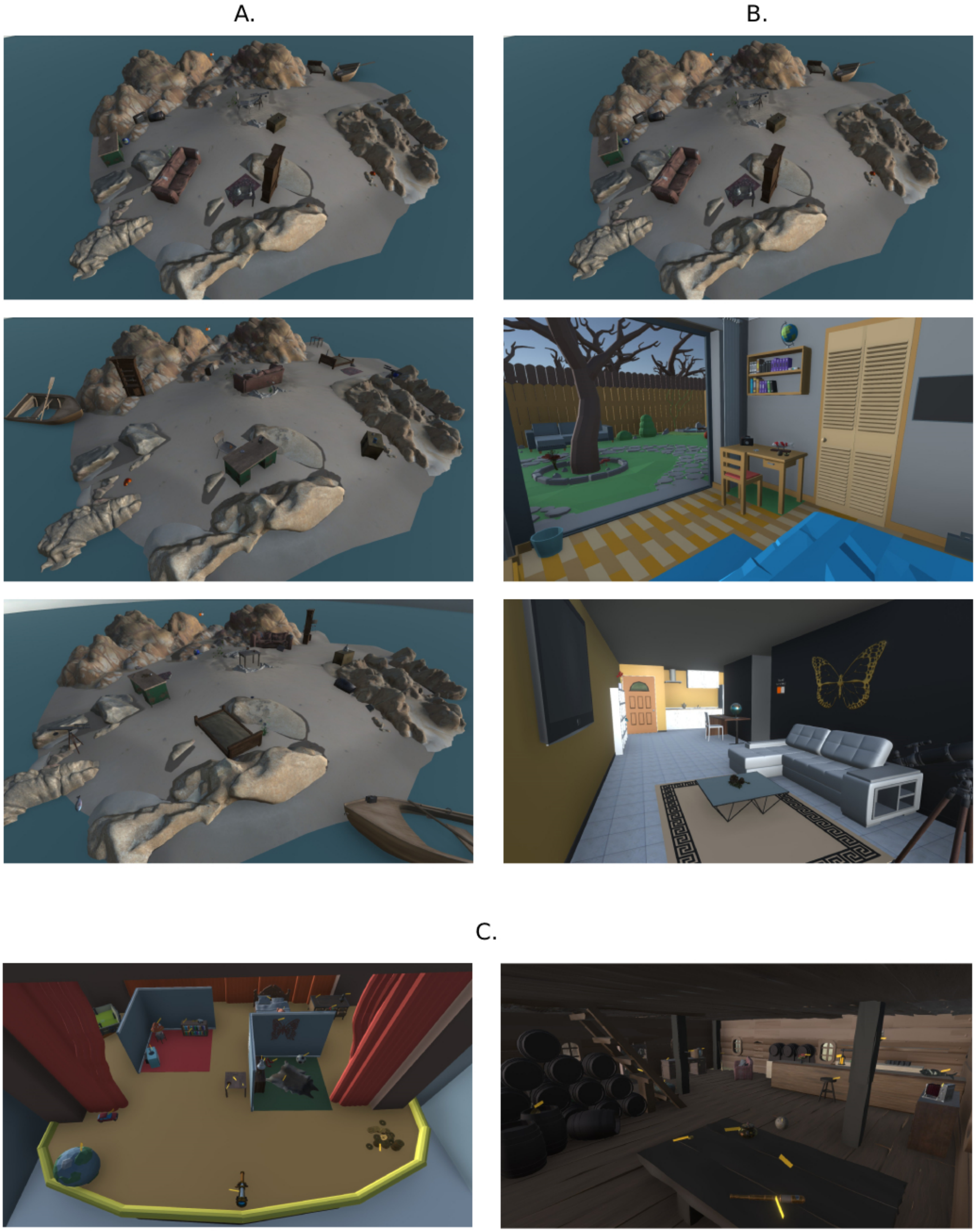


Figure 24: A: The three learned context learn VEs. B: The three new context learn VEs. C: The two test VEs.

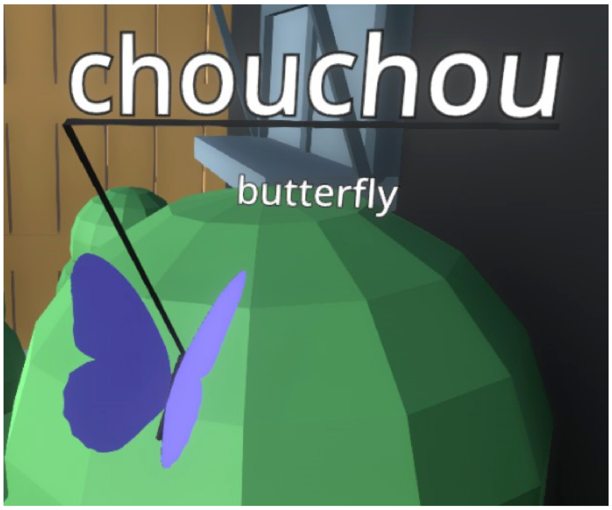


Figure 25: The target word object for *butterfly* which is currently activated.

ipant. Letters can be selected from the keyboard with the ray cast and input appears inside the text field. The keyboard can be removed again by clicking *close* or *enter*. Participants are asked through a text field inside the VE, which also explains the workings of the test, that they fill in an *x* behind an entry if they already knew that word prior to participating. The delayed posttest environment has the same workings as the posttest environment, with the exception that participants do not need to fill in an *x* behind words that they knew prior to participating any longer, because all text fields where an *x* was filled in by the participant in part one are now removed.

6.7.3 Portal system

Users need to be moved to different VEs to go from one to the next. Participants in learning environments are allowed to leave the VE after they have interacted with each target word object at least once. If participants have fulfilled this requirement and use the toggle button to check if they have found everything then a button will spawn below the toggle button with the text *open portal*. Similarly, participants cannot leave the test environment until something has been filled into every text field to ensure that participants have seen all target word objects on which they will be tested. If they do not know a word then they must fill in a dot (i.e. .), so the text field is not seen as being empty. After all text fields have been filled a button will again spawn with the text *open portal*. The participant can choose when they want to open the portal to continue, so it is possible to continue learning or check all filled in words a last time if participants choose to do so.

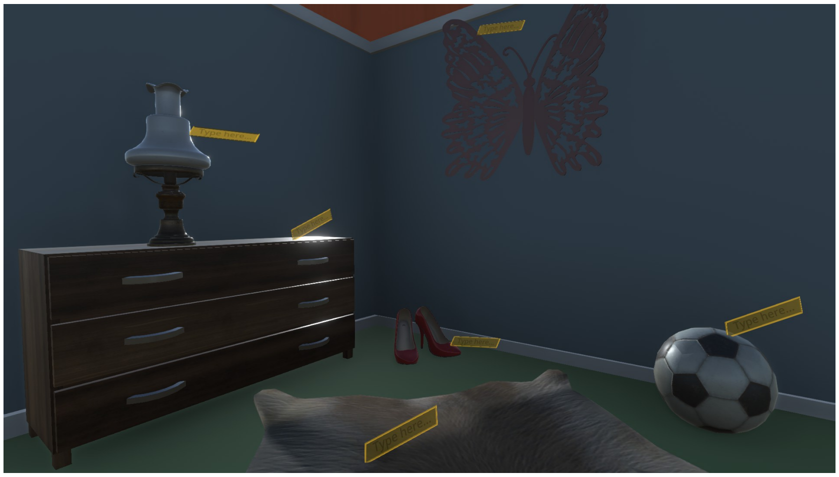


Figure 26: Empty text fields connected to target word objects in the theatre test VE.

To not startle the participant by teleporting them without warning to the next location, the user should make the decision themselves to be teleported. Therefore a clear indication is needed that doing a certain action will teleport the user. To create the idea of something being a portal an asset was used that has many colourful particles going up in a circular shape, as can be seen in Figure 27, providing it with a feel of something otherworldly. Using the *open portal* button would make this portal asset appear in the VE. By stepping into this cocoon of light users are then moved to the next VE.

6.7.4 Restricting the user's movement

The movement of users is restricted through the placement of invisible blocks around the walking area which the participant cannot bypass. In the uninhabited island VE participants cannot enter the water or climb or teleport to high rocks. The garden of the bedroom with garden VE is encompassed with a fence without a garden gate. The bedroom has a door that is closed and cannot open. The apartment also has many closed doors of which none can open. In the theatre it is not possible to move further than the stage, while in the barn bar it is possible to go outside, but participants can only move on black stone slabs that are close to the side of the building.

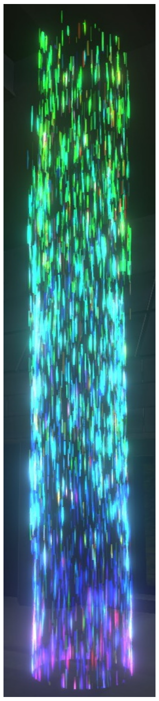


Figure 27: Portal asset.

6.7.5 Collecting data

The input of all filled in text fields, together with the number of clicks for each target word object for each location, are stored in one text asset in a location inside the experiment *world* that is not reachable by the participant. The number of target word objects a participant had already found before using the toggle button to check if there are any words left is also stored. A low number could indicate that the participant did not explore the environment, which would make their entry invalid if their behaviour would differ too much from other participants. The username of a participant is also stored so their participation in the first and second parts of the experiment can be linked to each other. Furthermore, the time duration for each part of the world is stored and the date is stored so it is possible to send reminders to participants for the second part of the experiment a week later.

All data that is collected inside the world is connected to a cube asset in a room that participants reach last when going through the experiment world. The cube is blue in the learned context world, purple in the new context world and yellow in the delayed posttest world to keep conditions easily apart. The cubes are the only assets that a participant can manipulate, so pick up and move around.

By holding the asset the participant can send it to the researcher through the in-game message system of *Neos VR*. Since all data collection assets are linked to the cube, those are also sent along with the cube.

6.8 Introduction and break room specifics

Participants should first become familiar with the controls and the workings of the system so they can focus fully on learning words when being in a learning environment. Therefore, an introduction room is created to explain to the participant the specifics of the system and what is expected of them before beginning the experiment. Here participants are also asked for demographic information. Next to the introduction room a break room is created to provide participants with the possibility to take breaks between learning.

6.8.1 Explaining the system to the user

It is undesirable to put the user immediately into a learning environment, where they need to focus on learning words, without them understanding the controls and workings of the system yet. Therefore an introduction to the workings of the system in a separate VE is necessary before starting the learning process. A historical temple-like room with four elongated windows on one side was chosen to function as the introduction room. The temple-like room has a high ceiling and windows to avoid a feeling of being locked up or closed off, as this is the first VE a participant will see and the VE should feel inviting and safe. The introduction room has three functions: (i) obtaining relevant demographic information from participants, (ii) teaching users about some Japanese language facts that are nice to know for the experiment, and (iii) conveying to participants how to interact with the environment and objects.

Upon entering the experiment *world* the participant starts in the middle of the room at location (1), see Figure 28. At locations (2) to (5) they will fill in demographic information, at location (6) they learn about Japanese as a language, and at locations (7) to (12) they are educated on the use of the system. Showing all this information at once will make it difficult for participants to understand what is expected from them in the introduction room. Therefore everything is still hidden from the participant when they enter the room, except for the information located at (2), shown in Figure 29.

At location (2) the participant is asked to choose between a teleport locomotion or a walking locomotion, and to choose between doing the experiment sitting or standing. The participant is asked to choose on basis of what makes the experience most comfortable for them. Providing answers to both questions spawns a button underneath those questions with the text *Begin*.

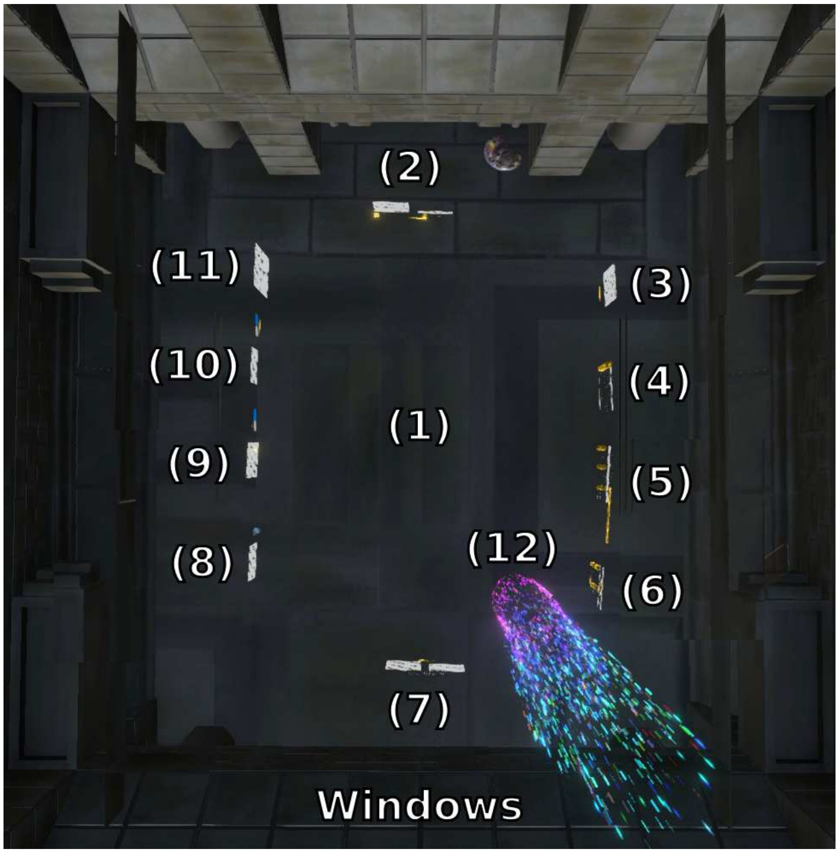


Figure 28: Top-down view of the lay-out of the introduction room with everything visible.

Clicking the *Begin* button from (2) spawns an information text with information regarding what to expect from the introduction room at (3) with a *continue* button underneath. Clicking the *continue* button spawns at location (4) the question what the participant's proficiency in Japanese is, with six levels to choose from. Choosing either *I have none / I know a few words* or *I'm a beginner* will spawn a text underneath that tells the participant they can continue. Choosing one of the four higher levels of proficiency will warn the player that they can continue, but that it is likely that they will not meet the condition that they should not know more than twelve words for their results to be taken into account. The participant is then allowed to choose if they want to run the risk of their results not counting or not. Whatever option they chose, at (5) more questions will spawn. Here participants are asked to fill in their first and optionally their second language and their age range. After answering each question a request will spawn at (6) to perform a sound check by pressing a button with *play sounds* on it. Pushing the button plays a sound fragment of a

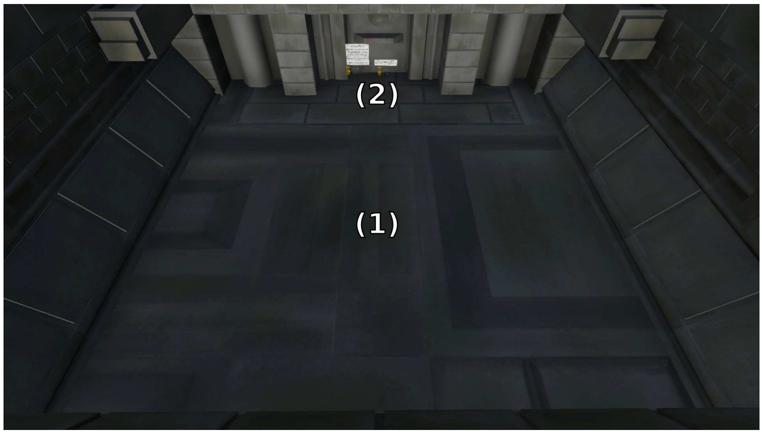


Figure 29: The introduction room as it is seen by the participant upon entering at location (1).

cow mooing and a cat miaowing. Playing the sound will also make a question appear about which sounds they heard, together with ten possible answers of which the participant can select as many as they want, and they can push a button with the text *done* below the options if they are finished.

The button *done* of (6) spawns two Japanese language facts at (7) that are useful to know during the experiment. One, that the experiment will not use a Japanese script but instead will work with a script with Latin letters which is called *romaji* in Japanese. Two, that the same word in Japanese can mean both the single and plural variant of a word, which should explain why some learning environments show a single target word object for a target word while another learning environment can show multiple objects for that same word. For example, the uninhabited island shows one *shoe* while the bedroom with garden after that has two *shoes* while both teach the Japanese word *kutsu*. It is also mentioned that the provided English translation is always in single form, but that they should remember that in Japanese it can be both. There is a button below this information with *continue*, which toggles on the information at (8).

The information at (8) is accompanied with the object of a mug. The information text explains to the user how objects can be activated, and asks the user to activate the mug three times by pressing the primary button of their controller while holding the laser on the mug. As long as the trigger is pressed and the laser is on the object, the textual form of the word mug in Japanese (i.e. *magu*) can be read, while the word can be heard once. Pressing the mug again activates the pronunciation recording again. After activating the mug three times an information text field appears at (9), together with a toggle button. Here the participant can read about what to do to check if they think they are done with all the objects in the scene, and how toggling a toggle button will make glowing orbs of light appear next to missed objects. A new object spawns close to (2) for demonstration purposes when the participant tries this button. After finding the so-called missed object, a cat in the corner of the room, and activating the cat to see its Japanese word (i.e. *neko*), more information will spawn at (10). Now a copy of the toggle button is also spawned next to this new information, but now with an added button beneath that says *open portal*. The information explains that the button will appear after activating all the words in an environment at least once and after using the toggle button to check if everything has been found. The participant is then invited to use the button, upon which a last information text field spawns at (11), together with a portal at (12). The information at (11) conveys to the participant that they should only use the *open portal* button if they think they are finished with an environment. The information also provides the participant with an overview of what to expect from here on, so that they will three times go through a learning and break environment, followed by a fill-in environment and lastly will go to an environment to save their data. The participant is asked to step into the portal if they are ready to begin.

In the second part of the experiment the participant makes the delayed posttest. The introduction room for the second part of the experiment contains much less information as most information from participants is already obtained in the first introduction room and the workings of the system are not repeated in the second introduction room. Instead it is expected from participants that they still remember these, but non-experienced IVR participants are able to ask for a reminder regarding controls from the researcher when doing the second part.

The spawn location of participants is again at (1) in the second introduction room with info already being present at (2). At (2) the locomotion type and standing or sitting position in real life are asked and a small text field informs participant that this experiment part will have one test environment to go through. An *open portal* button below the small text field enables the participant to open up a portal to the delayed posttest VE.

6.8.2 Break rooms

Users receive much new information to process when going through the learning environments. Therefore, a learning environment is always followed by a break room to provide participants with a moment where they can take a rest from learning. Thus the break room should not include any stimulating views that call for attention. Therefore, the break rooms are a simple low poly shape with rising walls on the side to act as natural barriers for the participant. A counter that counts the number of minutes, starting from the moment the participant entered the break room, communicates the current duration of their break to the participant. How the break room is encountered by participants upon entering can be seen in Figure 30. Non-experienced IVR users are expected to go out of IVR at this point so they can take a break away from IVR. However, as IVR experienced participants have less to worry about IVR discomforts [45], there is a chance that they will choose to stay inside IVR, thus inside the break room. Therefore, a video player is included in the break room where participants can watch a video of rabbits playing around. There is no wild action that demands the attention of the viewer in the video and the sound of the video only includes wind rustling and bird song. The original rabbit video,[24] which has a duration of one hour, was cut into three parts of fifteen minutes to provide each of the three break rooms with its own video with rabbits, so there is always something new to watch. After fifteen minutes the video will disappear as the maximum break duration has been reached. The counter stops counting at fifteen minutes and a new message appears to tell the participant that they should open up the portal and continue on.

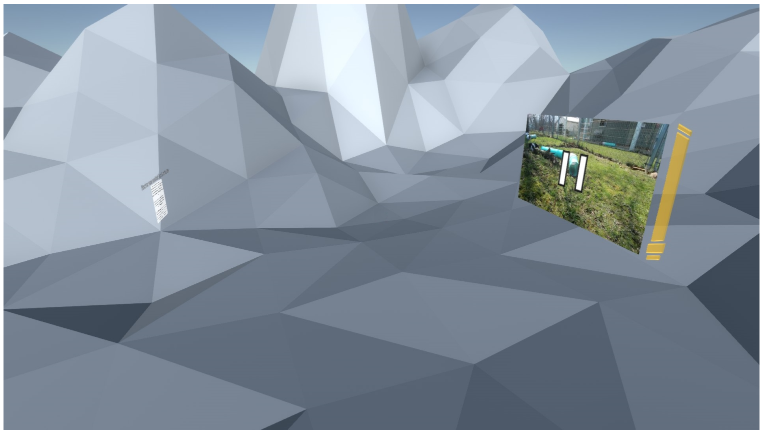


Figure 30: Lay-out of the break room.

6.9 Conclusion

A system called Wics was created based on the identified requirements for a system to be used in the proposed experiment discussed in Chapter 5. Wics first teaches participants about the workings of the system inside an introduction room which is specifically designed for this purpose. After becoming familiar with the workings of the system and the controls, words can be learned throughout three VEs, which are three uninhabited island VEs for the learned context condition and which are an uninhabited island VE, bedroom with garden VE

[24] *Bunnies Playing - 1 HOUR of Relaxing Bunny Cam Video!* by Hook's Hollands on YouTube: https://www.youtube.com/watch?v=Z-lNpn0Le10

and apartment VE for the new context condition. Wics also provides participants with a break opportunity between each learn VE. Target words are tested for the posttest in a theatre VE and for the delayed posttest a week later in a barn bar VE, which are both new contexts for all participants. How participants are recruited for the experiment and how the data from the experiment is analysed is discussed in the next chapter.

7 Method

Participants are recruited with both IVR experience as without any experience to do the experiment. How we reached out to possible participants and what the exclusion criteria for participants are is discussed in the next section. The specific equipment used to run the system and do the experiment is mentioned in Section 7.2. The process a participant went through from the beginning of the experiment until the end is discussed in Section 7.3 and in Section 7.4. How the data received from participants will be analysed is discussed in Section 7.5.

7.1 Recruitment of participants

Non-experienced IVR participants were sought through word-of-mouth and an information brochure. Experienced IVR participants were sought within the *Neos VR* community through the *Neos VR* Discord server and a published *world* for promoting the experiment in *Neos VR* that was accessible for all *Neos VR* users. On the Discord server a call for participation was sent in the *edu-science* channel, where a description of the experiment was provided, as well as specifics regarding the expected duration of an hour and the specifics on how to sign up for participating. In the published promotion world *Neos VR* users were spawned inside the temple interior of the introduction room upon entering the world. The promotion room contained textual information about the experiment, a link to the consent form, three L2 word objects to learn, a portal visualisation and a promotion poster, and can be seen in Figure 31. All persons needed to do, if they wanted to participate, was to open a link that was provided in each message which led to a consent form and an information document with detailed specifics about the experiment. No participants were promised any kind of compensation, and no compensation was given at the end of the experiment. Our study was approved by the ethical committee of Electrical Engineering Mathematics and Computer Science (EEMCS) of the University of Twente (RP 2021-223).

Participants were excluded from participation if they were younger than 16 years of age. Data of participants that met one of the following criteria were not taken into account for the results: (i) the participant could hear no sound in IVR, (ii) the participant toggled on the glowing lights before activating 20 words in any of the learning environments, (iii) the participant knew more than twelve target words from the experiment before entering the first learning environment, and (iv) the participant only completed the first experiment part and did not complete the second experiment part.

The age of participants was checked by ensuring that the box on the consent form, in which participants indicate that they are at least 16 years of age, was ticked. Whether a participant could hear sound in the experiment *world*

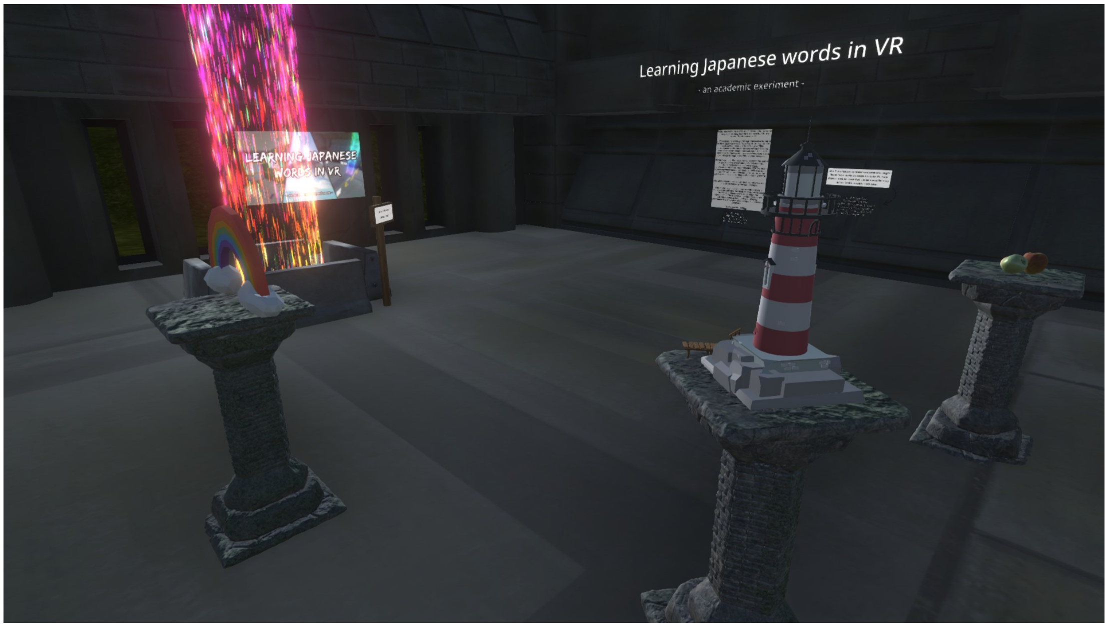

Learning Japanese words in VR
- an academic exeriment -
LEARNING JAPANESE WORDS IN VR

Figure 31: Promotion room for participating in the experiment for *Neos VR* users.

was checked by letting participants hear two animal sounds in IVR and by letting them choose from a list of ten animals which two animals they just heard. How many words were activated before a participant checked if they had found everything was tracked within the experiment *world*. How many words participants knew prior to participating was checked by instructing participants to indicate which words they already knew during the posttest.

7.2 Equipment

Neos VR user participants used their own hardware for Neos VR. Non-experienced IVR participants travelled to a location where a room contained a desktop, monitor, mouse, two VIVE base stations, one HTC VIVE (2016) headset and two VIVE motion controllers. Participants only interacted with the headset and two motion controllers. Participants were asked to bring their own earphones for hygienic reasons. The desktop with monitor and mouse stood in one corner of the room to allow the program to be run and the monitor allowed the researcher to maintain an overview on the happenings in the IVR environment of the participant. A space was cleared in the middle of the room for the IVR play area in which the participant could move freely with their arms around. A chair stood in the middle of the IVR play area for the participant to sit on, as sitting reduces the chance for IVR sickness as opposed to standing. Neos VR was run on Windows 10 with release 2021.11.10.1253 of Neos VR.

7.3 Procedure for part one

Non-experienced IVR participants were welcomed into the prepared physical experiment room and received a short explanation regarding the IVR introduction room they would start in. The researcher then helped them with putting on the headset and providing them with two motion controllers. Participants then went through the introduction room inside IVR while the researcher stayed in the physical experiment room and watched what the participant did and looked at on the monitor. The researcher provided guidance on how to move around inside the VE and provided tips on controller specific interaction possibilities, while the participant could also ask any questions they might have. Once the participant reached the end of the introduction room and opened the portal, and felt like they understood how to move around in IVR and how to use the controllers to manipulate the environment, the researcher left the room to provide the participant with a safe space to do the experiment. This was done because it could be that a participant is shy with repeating the Japanese words out loud with the researcher present, or that they feel watched and judged which might skew the results.

Experienced IVR participants who were *Neos VR* users and own a headset of themselves were already able to make use of *Neos VR* and understand its workings, so did not need a researcher to setup the IVR equipment or to explain the workings of *Neos VR*. Instead, *Neos VR* participants were sent the experiment *world orb* for the first experiment part through the *Neos VR* message system from which they could retrieve the world orb to enter the *world*. They then also started in the world in the middle of the introduction room. The same worlds were used for all participants and the entire experiment took place in IVR, although participants were free to remove the headset during breaks.

The order of VEs which a participant goes through can be seen in Table VIII.

7.3.1 Introduction room - part one

The participant starts in an introduction room, which looks like an historical temple room with four elongated windows on one side. The participant is spawned in the middle of the area with a few text fields where they fill in if they use the teleport or walking locomotion and if they are sitting or standing in real life. They continue on to an information text regarding what to expect from specifically the introduction room they are in, and fill in their proficiency in Japanese. They then fill in their first language and optionally their second

TABLE VIII
The VEs in the experiment and their order of appearance, which matches the order of the experiment setup shown in Table IV.

Part one

Learned context	**New context**
Introduction room - part one	
Uninhabited island v.1	
Break room 1	
Uninhabited island v.2	Bedroom with garden
Break room 2	
Uninhabited island v.3	Apartment
Break room 3	
Theatre	
Save room - part one	

Part two

Introduction room - part two
Barn bar
Save room - part two

language and their age range, which is followed by a sound check. They then read a bit of information about Japanese and practice activating a target word object with a mug asset. Next they receive an example of how a missed target word object looks and how it will have a glowing light next to it if the participant toggles a button which will show all missed words. The participant also reads about how they can open a portal with a button if all words are found and can press the button to spawn a portal. Lastly they read an overview of what to expect from the experiment after which the participant is invited to step through the portal to go to the first learning environment.

7.3.2 First learning environment

The portal brings the participant to the first of the three learning environment, which looks like an uninhabited island with objects strewn about by a previous washed-up island dweller. The participant learns each of the 32 target words that are included in the VE by placing their ray cast on a target word object and clicking and holding a button on the VIVE controller. This action shows the corresponding L2 text word form and English translation belonging to that target word object. After the participant feels like they have seen all target word objects they toggle a button to see if they are right. A text above the button tells them if they have indeed seen all target words or if they need to explore a bit more. If they need to explore a bit more then a glowing orb appears next to each missed target word object. After gaining an indication from the toggle button text that there are still target word objects that have not been activated, the participant toggles off the button if they want to explore further without the provided missed target word indicators. They leave the button toggled on if they want to see which word exactly they have missed. An example of how a room can look where all target word objects are indicated as missed can be seen in Figure 32. After activating all target words the participant sees an *open portal* button appear beneath the toggle button if it is toggled on. All three possible toggle button states are shown in Figure 33. The participant decides if they are finished with learning words in the VE. If they are not finished, they continue learning the words, and if they are finished they press the *open portal* button to spawn a portal. They then enter the portal to go to the first break room.

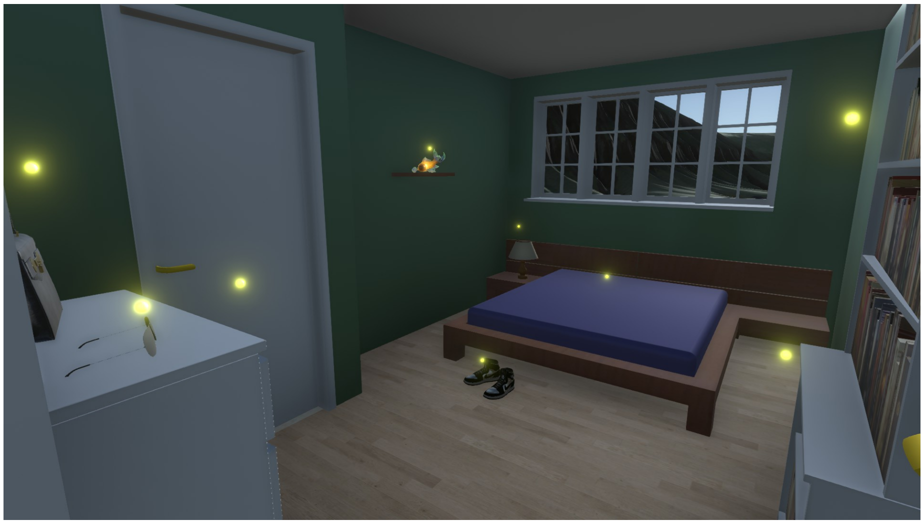


Figure 32: A room where all target word objects are indicated as missed.

7.3.3 First break room

In this first break room participants read about the risk of having to process too much information and the importance of breaks and they are informed that the break duration has a minimum of three minutes and a maximum of fifteen minutes. Participants can read from a counter how many minutes have passed in their break. An *open portal* button appears after three minutes below the informative text. There is also a video in the break room with rabbits hopping around that the participant can turn on if they want to watch it while spending their break in IVR. However, especially non-experienced IVR participants are

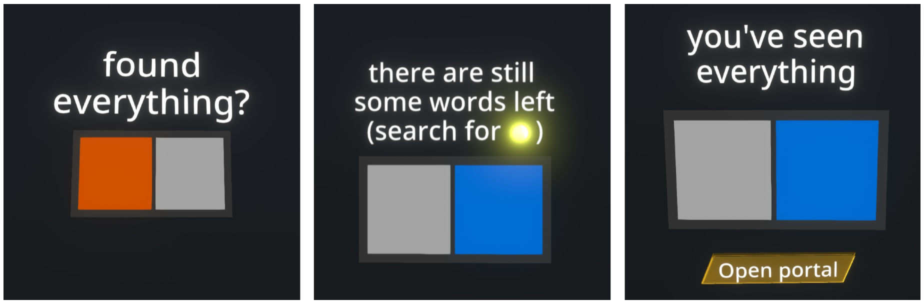


Figure 33: All three possible toggle button states: default state (left), after toggling the button on while missing target words (middle), and after toggling the button on while missing no target words (right).

recommended to take a break outside of IVR to reduce the risk of negative effects of IVR occurring. To do this, they can take off the headset by themselves, which was practised while the researcher was still present at the start of the experiment, and they can also put the headset back on by themselves when they want to continue. However, it is always possible to call the researcher if help is needed in any way. After 15 minutes of break time have passed the video disappears inside the IVR environment and a new text appears below the informative text with the request to please open and enter the portal. After clicking on the *open portal* button a portal will again open which the participant enters to continue on to the second learning environment.

7.3.4 Second learning environment and break room

The second learning environment different for each of the participant groups. The new context group spawns in low-poly bedroom with garden VE while the learned context group spawns on a second version of the uninhabited island from the first learning environment where all target word objects now have a different location. However, participants go through the environment in exactly the same manner as in the first learning environment. Once all target words have again been activated at least once and the participant checks if they have found everything, participants choose to open the portal when they feel ready to continue to the next break room. After entering the portal they see a second break room that is almost exactly the same as the first break room, with the only differences being that there is a different rabbit video and the information text now communicates that they are in the second break room. The participant enters again the portal when they feel like their break has been long enough and between three and fifteen minutes, which brings them to the third learning environment.

7.3.5 Third learning environment and break room

The third learning environment also differs for each participant group. The new context group spawns in a realistic looking apartment VE while the learned context group spawns in a third version of the uninhabited island where again all the locations of target word objects are different from the previous island versions. Everything in the third learning environment is again repeated for a third and final time and participants go to a third break room after entering the portal. They can see a new rabbit video during their break and open a portal after feeling finished with their break. Stepping into this portal brings them to the next VE.

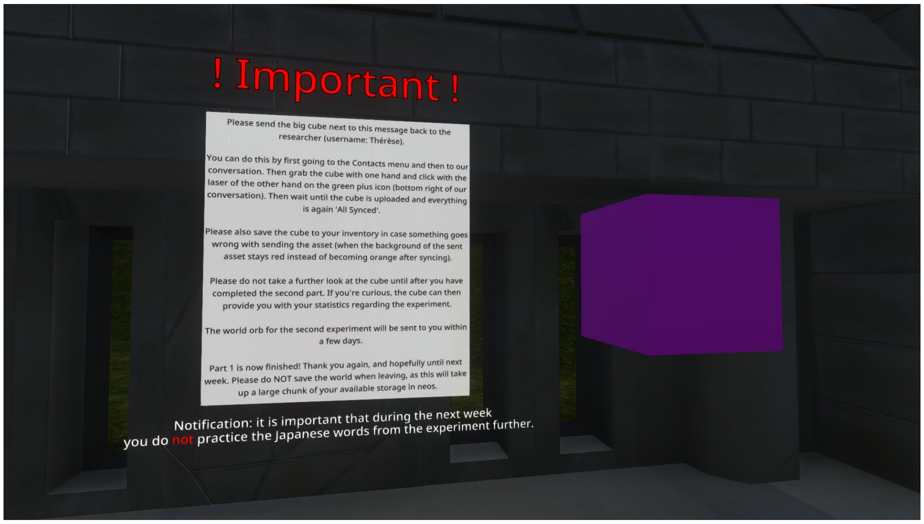
7.3.6 Test environment - part one

Both participant groups teleport after the third break room to the same test environment. The spawn point of the participant is in the middle of the stage with a clear view to a text placed on a wall that explains to the participant the purpose of the environment. The participant reads that it is expected of them to fill in all 32 text fields that are present in the environment, where each text field is connected to the target word object that it belongs to. If the participant does not know a word then they should fill in a dot (i.e. .), and if they already knew a word they should fill in a letter *x* after it. Participants are also informed that the *open portal* button will not appear until all text fields have been filled. If participants have filled in something in all 32 text fields, and also feel satisfied with what they filled in, they can use the *open portal* button to open a portal to continue on.

7.3.7 Saving room - part one

Participants always end the experiment in a saving room that has the same interior as the introduction room, however, now it only has one wall filled with information in two parts. The first part asks the participant if they want to get reminders about the second part of the experiment that must be done a week later. There is a toggle button that they can turn on if they want reminders through the Neos VR message system, and there is an empty text field where they can fill in their Discord username or email if they want reminders through Discord or email. Reminders are promised to be sent on both day six and day seven after participating, so on the day before the second experiment and on the day itself.

The other information part has large and red letters with *!Important!* written above it to draw the attention of the participant towards it. Here an informative text explains to the participant that they first need to send an asset, namely a large cube that is placed next to the cube, to the researcher before closing the experiment world. There are also instruction on how to send the asset. The warning text, instructions and cube asset for saving can be seen in Figure 34. After sending the asset the participant closes the experiment world without saving, but before leaving there is a last message underneath the information text that reminds the participant that they should no longer actively try to learn or repeat the words in their head.



! Important !

Please send the big cube next to this message back to the researcher (username: Thérèse).

You can do this by first going to the Contacts menu and then to our conversation. Then grab the cube with one hand and click with the laser of the other hand on the green plus icon (bottom right of our conversation). Then wait until the cube is uploaded and everything is again 'All Synced'.

Please also save the cube to your inventory in case something goes wrong with sending the asset (when the background of the sent asset stays red instead of becoming orange after syncing).

Please do not take a further look at the cube until after you have completed the second part. If you're curious, the cube can then provide you with your statistics regarding the experiment.

The world orb for the second experiment will be sent to you within a few days.

Part 1 is now finished! Thank you again, and hopefully until next week. Please do NOT save the world when leaving, as this will take up a large chunk of your available storage in neos.

Notification: it is important that during the next week you do not practice the Japanese words from the experiment further.

Figure 34: The warning text, instructions and cube asset for saving in the saving room as encountered by participants in the new context condition.

7.4 Procedure for part two

A week after participating in the first part participants are asked to go through the second part of the experiment in which the main activity is going through a second test environment. The pre-IVR process is the same as for the first part of the experiment, where the researcher is present while the participant goes through the introduction room so questions can be answered, but where the researcher leaves when the participant goes to the test environment to allow for privacy while testing.

7.4.1 Introduction room - part two

The introduction room of the second part of the experiment has the same interior of the ancient temple. Here the participant starts again with questions regarding their locomotion and standing or sitting position, and answering those will reveal an informative text about what to expect from this second part of the experiment, namely going through another test environment. Beneath this text is an *open portal* button which the participant presses to open a portal to go to the next VE of the experiment.

7.4.2 Test environment - part two

The environment of this second test environment is a barn with a bar inside it that is located in the middle of a desert canyon environment. Participants are spawned close to a text field so they read that they should fill in all text fields and type a dot (i.e. .) if they do not know a word. Looking around they see a text field attached to each target word object that they have learned in the previous week. After filling in all text fields an *open portal* button will once again spawn which the participant pushes if they feel ready, which opens a portal to continue on to the last environment.

7.4.3 Saving room - part two

Similar to the saving room from the first experiment world, the participant is asked to send an asset to the researcher, which contains all necessary data. The participant is then thanked again for participating and is told through text that from that moment on they can take a look at the sent assets, if they would like to do so, to see the data collected in the experiment.

7.5 Analysis

Words filled in for the posttest and delayed posttest will be scored with 1 point if the answer is phonetically correct, 0.5 points if the answer is almost correct (i.e. one syllable is wrong or missing), and 0 points if the answer is incorrect, similarly to *Ogma* [8] and *ObjectManipulation** [78]. Points are assigned by the main researcher. We are only interested in the improvement of participants, so if participants already know words prior to participating, we correct the score to not count those prior known words in the score. We call this corrected score the *performance score* of a participant. To calculate the performance score we transform the initial score into a performance score percentage by dividing the initial score with 32 minus N words that are known prior to participating. For example, if a participant knew 6 words beforehand, and has scored 18 points on the words they did not know beforehand, then their performance score percentage is $18/(32-6)*100\% = 69.23\%$, whereas their score would be $24/32*100\% = 75\%$ without the correction.

The performance and retention scores seems to be close to normally distributed continuous data, however, is expected to break parametric assumptions due to a skewed data set.[25] Therefore, the performance score is dealt with as ordinal data and we use a non-parametric test to compare between the independent new and learned context conditions using the *Mann-Whitney U test* [89].

The time spent in total for the learning environments, the time spent in total for the break rooms, and performance scores for the posttest are compared between the new and the learned context group using also a Mann-Whitney U test for determining if there are significant differences between the two groups for these topics.

With the performance scores for the posttest and the delayed posttest the word retention for each participant is determined by calculating *delayed posttest performance score/posttest performance score.* Again we compare this retention between learned and new group using a Mann-Whitney U test to determine if there is a significant difference in scores between both groups in the delayed posttest and thus whether learning with new contexts increases retention.

[25] Consider the cut-off at 100% and as participation is voluntary, participants are likely to be more motivated, hence we expect practically no cases where no words at all are memorized.

8 Results

There are in total 26 participants who started with the experiment. All 26 participants completed the first part of the experiment, while 22 participants also completed the second part. The four participants who did not complete the second part of the experiment were excluded from the results, leaving 22 participants in total. There were 6 participants in the learned context condition and 9 participants in the new context condition who are experienced IVR users and 5 participants in the learned context condition and 2 participants in the new context condition who are non-experienced IVR users with no or little prior IVR experience. The age range for most participants is the age range of 26-30 with 11 participants, followed by the age range of 21-25 with 8 participants and 2 participants with an age range of 56-60 and one participant with an age range of 61-65. The number of participants in each age range can be seen in Table IX. There are no participants that did not wish to disclose their age range.

TABLE IX
Division of age range of participants.

Age range	**Learned**	**New**
21-25	4	4
26-30	5	6
56-60	2	-
61-65	-	1

Two of the 22 participants experienced a *world* crash in Neos, one during their second learning environment and one during their second break. The first occurred for an experienced IVR *Neos VR* participant in the learned context condition, where the participant quickly went through the previous VEs a second time until reaching the second learn VE again, and the other occurred for a non-experienced IVR participant where the researcher went through the VEs a second time until the second break area was reached again. The times inside learn VEs and break rooms of these two participants are left out of the results, while their test scores are included.

Table X shows the max score, score, and performance percentage per participant per condition, as well as the median and standard deviation between participants for both conditions. The max score refers to the maximum score a participant could get given the number of words they knew beforehand. Performance of participants in the learned context condition (Mdn = 89.06) did not differ significantly from participants in the new context condition (Mdn = 79.69) on the posttest, U = 58.5, z = -0.1286, p = 0.8977.

Table XI shows the max score, score, performance percentage, and retention percentage per participant per condition, as well as the median and standard deviation between participants for both conditions. Figure 35 shows the box

TABLE X
Posttest scores and performance of participants per condition

Learned				New			
Participant	Max score	Score	Perf. (%)	Participant	Max score	Score	Perf. (%)
L1	32	14	43.75	N1	25	15.5	62.00
L2	25	24	96.00	N2	32	25	78.13
L3	32	28.5	89.06	N3	32	25.5	79.69
L4	25	19.5	78.00	N4	31	30	96.77
L5	32	16.5	51.56	N5	32	30.5	95.31
L6	32	28.5	89.06	N6	20	19	95.00
L7	32	29.5	92.19	N7	32	31.5	98.44
L8	32	30	93.75	N8	32	7	21.88
L9	32	30,5	95.31	N9	28	24	85.71
L10	32	6.5	20.31	N10	32	17.5	54.69
L11	32	30	93.75	N11	28	16.5	58.93
Median	32	28.5	89.06	Median	32	24	79.69
Mean	30.73	23.41	76.61	Mean	29.45	22.00	75.14

TABLE XI
Delayed posttest scores and performance of participants per condition

Learned					New				
	Max score	Score	Performance (%)	Retention (%)		Max score	Score	Performance (%)	Retention (%)
L1	32	14	43.75	100.00	N1	25	13	52.00	83.87
L2	25	19	76.00	79.17	N2	32	12	37.50	48.00
L3	32	22	68.75	77.19	N3	32	14.5	45.31	56.86
L4	25	18.5	74.00	94.87	N4	31	22.5	72.58	75.00
L5	32	9	28.13	54.55	N5	32	23.5	73.44	77.05
L6	32	18	56.25	63.16	N6	20	16	80.00	84.21
L7	32	22.5	70.31	76.27	N7	32	30	93.75	95.24
L8	32	22	68.75	73.33	N8	32	3.5	10.94	50.00
L9	32	30	93.75	98.36	N9	28	16	57.14	66.67
L10	32	5	15.63	76.92	N10	32	9.5	29.69	54.29
L11	32	10.5	32.81	35.00	N11	28	12	42.86	72.73
Median	32	18.5	68.75	76.92	Median	32	14.5	52.00	72.73
Mean	30.73	17.32	57.10	75.35	Mean	29.45	15.68	54.11	69.45

TABLE XII
Total time spent in the learn and break environments of participants per condition

	Learned			New	
	Learn	Break		Learn	Break
L1	-	-	N1	45	35
L2	59	27	N2	31	13
L3	32	19	N3	20	9
L4	19	19	N4	50	30
L5	71	27	N5	38	23
L6	52	29	N6	25	17
L7	65	20	N7	24	15
L8	54	40	N8	-	-
L9	43	18	N9	24	12
L10	9	20	N10	18	11
L11	36	37	N11	27	22
Total	440	256	Total	302	187
Median	47.5	23.5	Median	26	16

plots of both the performance and the retention percentages of both conditions. Retention of participants in the learned context condition (Mdn = 76.92) did not differ significantly from participants in the new context condition (Mdn = 72.73) on the delayed posttest, U = 75, $z = 0.9053$, $p = 0.3653$

Table XII show the total time spent in rounded down minutes in the learn and break environments per participant per condition, as well as the median and standard deviation between participants for both conditions. Time spent in the learning environments of participants in the learned context condition (Mdn = 47.5) did not differ significantly from participants in the new context condition (Mdn = 26) during the experiment, U = 73, $z = 1.6996$, $p = 0.08921$. Time spent in the break environments of participants in the learned context condition (Mdn = 23.5) also did not differ significantly from participants in the new context condition (Mdn = 16) during the experiment, U = 74, $z = 1.7789$, $p = 0.07526$.

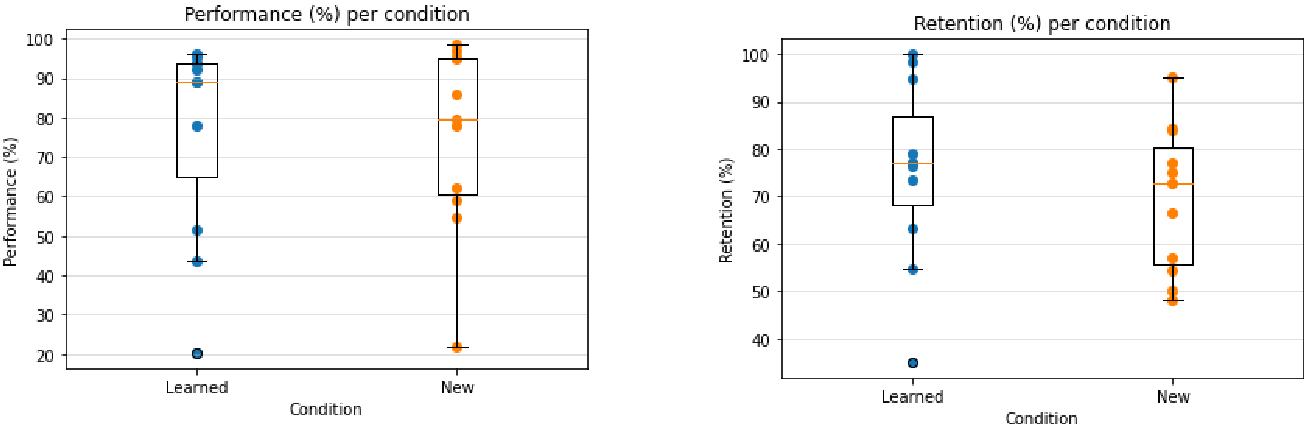


Figure 35: Box plot of both the performance (left) and retention (right) percentages of both conditions.

9 Discussion and conclusion

We build a dynamic IVR system called Wics where participants learned 32 Japanese words in three learning sessions and were tested in IVR in order to answer the research question of this thesis:

> RQ: What are the effects of recycling words in IVR in learned or new contexts for retrieving words when encountered in a new context?

With as hypotheses:

> H1: New and learned context participants will perform similarly on the posttest.

> H2: New context participants will have a higher retention than learned context participants.

For H1 we tested the performance of both groups on the posttest when compared with each other and predicted that both groups would perform similarly. The new context group performed marginally better ($z = 0.1286$) on the posttest than the learned context group, with no significant difference ($p = 0.8977$). Therefore, H1 is supported.

For H2 we tested the retention between the two conditions groups by comparing the retention percentages and predicted that the retention of the new context participants would be higher. The learned context group even performed a bit better ($z = -0.9053$) than the new context group, but with no significant difference ($p = 0.3653$). Thus, H2 is not supported.

To answer the research question we set up requirements such as that the system must provide users with multiple learning sessions, target word visualisations, and an L2 word form, see Section 6.1 for a full list, where we managed to fulfil all requirements.

9.1 Learn duration and results

There is no significant difference between contexts on the performance of the posttest or on their retention rate, with the new context group performing marginally better on the posttest and the learned context group scoring slightly better on retention. We identified several possible reasons for how these results might have come about.

TABLE XIII

All participants ordered by total time spent in learning environments

Participant ((L)earned/ (N)ew context #)	Total time spent in learn environment (min)	Indicated non-experienced IVR user
L5	71	**X**
L7	65	**X**
L2	59	
L8	54	**X**
L6	52	**X**
N4	50	
N1	45	
L9	43	**X**
N5	38	**X**
L11	36	
L3	32	
N2	31	
N11	27	
N6	25	
N7	24	
N9	24	
N3	20	
L4	19	
N10	18	
L10	9	
L1	-	
N8	-	**X**

Firstly, when looking at the total time spent by participants in both the learn and break environments, we see that the learned context group spent substantially more time in the learning environments, but also the break environments. Although no significant difference in time spent by the learned context group could be identified with a 95% confidence, we can see a significant difference with a 90% confidence. Participants were allowed to pinpoint the moment where they felt ready to continue themselves and no time limit was imposed on them during the learning environment.

Participants with higher learn times were most often non-experienced IVR participants, of which 2 participants were in the new condition and 5 participants in the learned condition. We miss the times from 1 non-experienced IVR participant in the new context condition. All the participants ordered by their total time in the learning environments reveals that all non-experienced IVR participants belong in the half that took the most time learning. One possible explanation could be that non-experienced IVR participants need more time in IVR due to unfamiliarity with controls and locomotion inside IVR. However, most non-experienced IVR participants noted that they felt familiar with moving around and activating words somewhere throughout the first learning environment. Non-experienced IVR participants came across as being more excited due to the novelty of IVR, where some participants noticed their virtual hands and were impressed by them while others expressed a clear excitement when seeing the portal asset with moving particles open up in the introduction room. This excitement might have resulted in more motivation to learn. To follow this conjecture we use *Spearman's Rho* to informally investigate if there could be a correlation between the total learn time and performance on the posttest ($r_s = 0.33434$, $p = 0.14965$) and the total learn time and retention ($r_s = -0.03009$, $p = 0.8998$). If at all of influence it seems that if the total learn duration would have had an influence on the results, then this could have been on the performance of the posttest, but not on the retention.

There is an indication that there might be a correlation between learning time and performance on the posttest ($p = 0.14965$). Most participants in the learned context condition were also participants with a long learning time. However, there is almost no difference between the performance of the groups on the posttest ($p = 0.8977$). Therefore it is possible that learning in the new context condition prepared participants better for retrieving words in a new context posttest, but that this benefit was erased due to participants in the learned context condition learning for a longer period of time. This is, at first glance, the opposite of what we expected to see with both hypotheses, as this seems to indicate that learning with new contexts is more beneficial for word retrieval shortly after learning, while its advantage diminishes in the long run. However, only so many word aspects can be learned at a time and learners first create a form-meaning link with their preferred word form before adding more word aspects to their word knowledge, as discussed in Section 2.1.2. This form from the form-meaning link is the textual form or audio form from the L2 word. If after establishing the form-meaning link there is still room to process more information, other word aspects as the textual or audio form that was not the favourite of the learner, or the visual representation of an object, can still be learned. The strength of these extra learned word aspects depends on learner's ability to still take in new information after creating the form-meaning link, but is expected to be weaker than the form-meaning link. How weaker the link to information is, how more difficult it is to retrieve, and how easier it is to forget after time passes, but shortly after learning it is expected to still be in short-term memory. So if new context participants did have a benefit during the posttest as opposed to learned context participant, then we expect that it was this weak but additional word knowledge regarding how a word can be visually represented that provided new context participants with an extra benefit.

9.2 Test observations

All non-*Neos VR* users, so users who did the experiment with the researcher present, and some *Neos VR* users who communicated through *Discord*, appeared excited to go back into IVR to do the test because they were curious to see how much they would be able to remember. Also, participants from both groups indicated that they wanted to see the next environment and were curious as to what it would be. This might indicate that testing in IVR, in contrast to text-based testing, could also increase motivation for wanting to evaluate existing word knowledge. Something we advise interested researchers to further explore.

Some participants filled in a word on the posttest that was awarded 0 points due to missing more than one syllable or having more than one syllable wrong. Some of these words were filled in written exactly the same on the delayed posttest where the words were awarded again 0 points. However, there is a form of retention happening here. Participants have stored the words incorrectly into their memory, and since there is no moment between the posttest and the delayed test to verify answers, there is no manner for participants to be corrected on the word they stored for L2. So remembering the incorrect form of L2 a week later again is also similar to remembering the correct form of L2 a week later. However, since we did not anticipate for this occurrence, we did not include this form of retention in our calculations, so now only retention on correct word is calculated, and not on purely retaining stored word forms. However, as this research focuses on how learners might perform in an unexpected new context as might be found during total immersion, a better representation is produced by only looking at retention of correct words.

Another unexpected occurrence is that words were sometimes filled in correctly on the delayed posttest after making a mistake in the posttest, and even words that were filled in the delayed posttest after filling in no word in the posttest. N5 explained after the delayed posttest an occurrence of the latter, where they told after the delayed posttest:

> I didn't remember the word for *avocado* during the posttest and was confused with the *Pokémon* called *Wishiwashi*. It was during the bike ride back home that I suddenly thought *Nooo, that's not it! It was* waninashi*!*, and that information has never left my brain from that point on.

while N6 wrote:

> Occasionally memories of the learning world came to mind unprompted, but I wasn't deliberately rehearsing anything.

L1 even had an 100% retention score, while not remembering all the words they had filled in the week before, but where their forgotten words were countered with words that were now remembered correctly.

Participants had to go through the second part of the experiment one week after the first part, so 7 days later. However, in the learned context L3 took the delayed posttest 8 days later, while from the new context P9 took the delayed posttest 8 days, P3 9 days later and P5 10 days later. Their results do not look particularly different from other participants, but the longer waiting time before testing could have influenced retention.

9.3 Learning observations

Context was defined as (i) the virtual environment and (ii) the visual representation of a target word. All other aspects were kept the same between conditions to only let the difference be in the context in which was learned. However, some participants, indicated that they remembered on the delayed posttest the words they had learned by searching for the sounds of the word and still hearing the words as they were said by the audio cues. N6 from the new context group noted:

> I can still visualise and recall the spatial layout and arrangement of objects (and the environment itself) in the first learning world pretty clearly. I spent a lot more time in the first one, so it's not too surprising to me that my memory of that one is stronger than the subsequent 2. I can recall the approximate layout of the other two, and placement of some of the objects, but it's not as clear. I only have a strong memory of the visual aspect of some objects (e.g. the rowing boat from the first world) - for most, the strongest memory is of the words being spoken. For example, for the knife & rug I'd probably not be able to pick out exactly which model was used in that world, but I remember the audio cues very clearly. I guess because the audio cues were shared across all learning worlds. When I was trying to learn the words I did quite a lot of sub-vocal practice and would try to learn a few new ones before going back to practice the previous few.

N6 thus noticed that they most strongly remembered the audio cues, as those were repeated exactly the same over the three learning environments and the participant was sub-vocally practising the words by repeating the audio cues. Including the audio cues in the definition of context, so the audio cues are different for each context, might diminish the strong effect that audio now seems to have had on learning for at least some participants.

P5 indicated that their ability to visualise objects in their head was low and that imaginations appeared only dim and vague, which resulted in them almost not remembering any visual representations of words, but where they remembered the sounds instead. They wrote:

> I think I had a vague recollection of the first boat while I was trying to recall *fune*, but I'm generally not a very visual thinker, so for most of them I was just trying to recall the sounds of the words. *Bouenkyou* was one of the difficult ones to remember, and I had no imagery while remembering it - just tried out mouth sounds until it felt right.

For them this did not diminish their experience in the new contexts they went through, as they named it was fun to go through and they wanted to continue with learning and exploring and seeing what the next environment would bring.

That there are 32 words to learn for each environment was communicated to participants beforehand. However, as there was much information to take in, some participants might have thought that it was possible that after the first learning environment they would have to learn new words that they had not seen before. Thus some participants indicated that they had spent much time in the first learning environment to learn everything really well, only to discover that everything was the same in the next learning environment and that they had already learned most words well at that point. This expectation and resulting approach might have resulted in quite a different learning outcome than participants who knew they had still two more learning environments to learn the words in.

Some participants also communicated their opinion regarding the experience of going through Wics. N9 wrote:

> This was super fun! I loved that the worlds kept changing.

while N5 wrote:

> Was a really fun way to learn new words.

and N6 wrote:

> It was a very interesting experience - certainly richer than simply trying to learn a list of words!

Unlike non-experienced IVR participants, experienced IVR participants also tended to sometimes name possible improvements for the system as they seemed to understand more clearly the possibilities IVR has to offer, and therefore to see missed opportunities. Most often the advise consisted of adding the possibility to manipulate objects, which was purposefully not included in Wics, but adding a social aspect to learning or letting objects appear at unexpected places were also named.

9.4 Limitations

We chose for a between subjects design, as a within subject design would ask for a large commitment of people, where they go through learning, waiting a week, delayed posttest, waiting a bit more, learning, waiting a week and another delayed posttest. However, a within subject design will counter people's individual learning differences much more. Because our sample size is small, there is more influence of individual learn preferences on the results. We divided participants randomly over groups to avoid bias, but this can have resulted in an unequal division of participants with specific learning differences.

There was no check on participation of *Neos VR* participants, as everything occurred outside of the view of the researcher. There is known how much time they should have been in an VE, the number of activation of objects, after how many activated objects it was checked if everything was found and on which dates the participant participated. But there was, for example, no control on if someone cheated on the test, if someone took a break outside of the break rooms or if someone was talking with somebody else during learning or was otherwise occupied.

Wics can only really depict a large variety of material objects easily, so it would be difficult to depict words like *impression* or *appearance*. Meaning that not all words can be learned in the current system.

9.5 Future work

There might be a possible advantage for studying in new contexts if learners have ample opportunity to also add the visual representations to their word knowledge. Future research looking into recycling words and different contexts would be advised to either take more time between learning sessions so knowledge regarding word aspects can be processed before continuing with the next learning sessions. To prevent some participants learning to their heart's content and going far beyond other participants regarding learning effort, some form of a time restriction is also recommended.

It is also recommended to test for visual imagery vividness to divide or exclude participants based on their visual imagery possibilities. Learners with a low ability to create visual images in their mind are expected to not experience an advantage for learning with visual object representations for the long term, although they can still benefit for the enjoyment of going through environments in IVR to learn target words. A well-known test for testing the strength of visual imagery of people is the *vividness of visual imagery questionnaire*. [90]

To avoid creating learned contexts for the word aspect for which a participant has a learning preference, so to enable participants to create a really strong connection to one word aspect, it is recommended to also change the context for other word aspects. By changing the voice and intonation of audio cues, participants cannot continue strengthening one specific voice instance for each target word. Perhaps changing the font would do something similar for participants with a learning preference for the textual form.

We have been quite mild with using the possibilities of IVR for this research to focus specifically on recycling words in different contexts. However, IVR has many possibilities to show objects or virtual environments. Other research might study possible effects when switching contexts much more quickly, letting a teacher change the content of a context live, or to continue on into the realm of impossibilities, where some objects behave or are placed as they would normally not.

9.6 Conclusion

In this research we looked at recycling words in IVR in a new context, where for each learning session each of three VEs and the visual representations of objects within was changed, and recycling words in a learned context, where the context was kept the same each learning session. Two tests also took place within IVR and happened in a new context so learning in both contexts could be tested for an unexpected situation as can be encountered in real life. One test was a posttest taken after the three learning sessions and the second test was a delayed posttest taken one week after the first posttest.

A system called Wics was built so participants could go through either three new contexts learning environments or one new context and two learned context learning environments. Each learning environment contained the same 32 target words as visual representations. Participants could explore the environment and activate target word objects to hear and read their L2 and read an English translation. Participants were tested in a similar environment, but where target word objects could not be activated. Instead, target word objects had text fields connected to them where participants could fill in the L2 of the target word object.

Participants were scored on performance for the posttest and retention was determined by also looking at the delayed posttest. There was no significant difference for performance on the posttest between conditions, nor on retention. There was also no 95% significant difference for total time duration spent in learning environments, nor for break environments, but there was a 90% significant difference seen where participants in the learned context condition learned longer and took longer breaks. Participants were allowed to pinpoint the moment they felt ready with learning the presented target words, to allow for differences in IVR familiarity and learn differences, which made it possible for a difference in learn duration to occur.

When taking a closer look at possible correlations based on informal investigation between learn time and performance on the posttest and retention then there are indications that a longer learn time could have benefited the performance score, but not the retention. Therefore there is a chance that learning in a new context was more beneficial in Wics for retrieving words during the posttest in comparison to learning in learned contexts. This could possibly be caused by providing too many new word aspects at once, where the visual representation word aspect, being not the first word aspect to be stored into memory for learners, is only stored minimally and is mostly lost over time. Therefore, it might be worthwhile to provide learners with ample opportunity to process words between learning sessions so multiple word aspects have a similar opportunity to be stored in memory. However, future research is needed to support or disprove such speculations.

For as far as we are aware, Wics is the first IVR system to provide learners with multiple learning sessions where words are recycled, where learners have control over their own learning by being able to activate indicators for missed words, and where the posttests are also inside IVR. Participants showed enthusiasm for going through the learning environments, discovering which virtual environment would wait for them next and looked forward to testing themselves in a new IVR environment a week later. The possibility of IVR to change VEs with ease should not be overlooked by language learning researchers as it can keep motivation up while providing learners with many contexts to learn, whether they switch contexts consecutively or repeat one context a few times. Furthermore, as participants showed a clear enthusiasm for going back in IVR to do the posttest, studies might consider to do their posttest(s) also in IVR, to base the results on a test method that compliments the learning method that is studied.

References

[1] P. Meara, “Vocabulary acquisition: A neglected aspect of language learning,” in *Language teaching and linguistics: Abstracts*, vol. 13, 1980, pp. 221–246.

[2] E. A. Levenston, “Second language acquisition: Issues and problems,” *Interlanguage studies bulletin*, pp. 147–160, 1979.

[3] I. S. P. Nation, *Learning vocabulary in another language*. Cambridge University Press, 2013.

[4] S. Loewen and M. Sato, *The Routledge handbook of instructed second language acquisition*. Taylor & Francis, 2017.

[5] R. Godwin-Jones, “Contextualized vocabulary learning,” *Language Learning & Technology*, vol. 22, no. 3, pp. 1–19, 2018.

[6] J. Legault, J. Zhao, Y.-A. Chi, W. Chen, A. Klippel, and P. Li, “Immersive virtual reality as an effective tool for second language vocabulary learning,” *Languages*, vol. 4, no. 13, 2019.

[7] N. P. Holmes and C. Spence, “The body schema and multisensory representation(s) of peripersonal space,” *Cognitive processing*, vol. 5, no. 2, pp. 94–105, 2004.

[8] D. Ebert, S. Gupta, and F. Makedon, “Ogma: A virtual reality language acquisition system,” in *Proceedings of the 9th acm international conference on pervasive technologies related to assistive environments*, 2016, pp. 1–5.

[9] C. Vázquez, L. Xia, T. Aikawa, and P. Maes, “Words in motion: Kinesthetic language learning in virtual reality,” in *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*, IEEE, 2018, pp. 272–276.

[10] N. Collins, B. Vaughan, and C. Cullen, “Motivation in situated immersive games for irish language learning, a dbr approach,” Academic Conferences limited, 2020, p. 96.

[11] P. Li, J. Legault, A. Klippel, and J. Zhao, “Virtual reality for student learning: Understanding individual differences,” *Human Behaviour and Brain*, vol. 1, no. 1, pp. 28–36, 2020.

[12] L. Freina and M. Ott, “A literature review on immersive virtual reality in education: State of the art and perspectives,” in *The international scientific conference elearning and software for education*, vol. 1, 2015, pp. 10–1007.

[13] J. C. Richards, “The role of vocabulary teaching,” *TESOL quarterly*, pp. 77–89, 1976.

[14] W. Marton, “Foreign vocabulary learning as problem no. 1 of language teaching at the advanced level,” *Interlanguage studies bulletin*, pp. 33–57, 1977.

[15] M. Thomas, *Universal grammar in second-language acquisition: A history*. Routledge, 2004.

[16] N. Schmitt, *Vocabulary in language teaching*. Cambridge university press, 2000.

[17] B. Laufer, “A factor of difficulty in vocabulary learning: Deceptive transparency,” *AILA review*, vol. 6, no. 1, pp. 10–20, 1989.

[18] I. S. P. Nation, *Teaching and Learning Vocabulary.* Newbury House, 1990.

[19] K. S. Folse, “Myths about teaching and learning second language vocabulary: What recent research says,” *TESL reporter*, vol. 37, no. 2, pp. 1–13, 2004.

[20] N. Schmitt, “Instructed second language vocabulary learning,” *Language teaching research*, vol. 12, no. 3, pp. 329–363, 2008.

[21] C. Larrotta, “Second language vocabulary learning and teaching: Still a hot topic.,” *Journal of Adult Education*, vol. 40, no. 1, n1, 2011.

[22] R. Ramos and F. Dario, “Incidental vocabulary learning in second language acquisition: A literature review,” *Profile Issues in TeachersProfessional Development*, vol. 17, no. 1, pp. 157–166, 2015.

[23] N. C. Ellis, “Implicit and explicit language learning,” *Implicit and explicit learning of languages*, pp. 79–114, 1994.

[24] T. Saragi, I. S. P. Nation, and G. F. Meister, “Vocabulary learning and reading.,” *System*, vol. 6, no. 2, pp. 72–8, 1978.

[25] H. Jeong, M. Sugiura, Y. Sassa, *et al.*, “Learning second language vocabulary: neural dissociation of situation-based learning and text-based learning,” *Neuroimage*, vol. 50, no. 2, pp. 802–809, 2010.

[26] J. S. Brown, A. Collins, and P. Duguid, “Situated cognition and the culture of learning,” *Educational researcher*, vol. 18, no. 1, pp. 32–42, 1989.

[27] Y.-F. Yang, “Engaging students in an online situated language learning environment,” *Computer Assisted Language Learning*, vol. 24, no. 2, pp. 181–198, 2011.

[28] V. G. Mendelson, “ “hindsight is 20/20:” student perceptions of language learning and the study abroad experience,” *Frontiers: The interdisciplinary journal of study abroad*, vol. 10, no. 1, pp. 43–63, 2004.

[29] C. Kinginger, “Language learning in study abroad: Case studies of americans in france,” *The Modern Language Journal*, vol. 92, pp. i–131, 2008.

[30] R. Oxford and J. Shearin, “Language learning motivation: Expanding the theoretical framework,” *The modern language journal*, vol. 78, no. 1, pp. 12–28, 1994.

[31] B. L. Savage and H. Z. Hughes, “How does Short-term Foreign Language Immersion Stimulate Language Learning?” *Frontiers: The Interdisciplinary Journal of Study Abroad*, vol. 24, no. 1, pp. 103–120, 2014.

[32] V. Pellegrino Aveni, “Speak for yourself: Second language use and self-construction during study abroad,” 2006.

[33] S. Wilkinson, “The omnipresent classroom during summer study abroad: American students in conversation with their french hosts,” *The Modern Language Journal*, vol. 86, no. 2, pp. 157–173, 2002.

[34] V. Gallese and G. Lakoff, "The brain's concepts: The role of the sensory-motor system in conceptual knowledge," *Cognitive neuropsychology*, vol. 22, no. 3-4, pp. 455–479, 2005.

[35] L. W. Barsalou, "Grounded cognition," *Annu. Rev. Psychol.*, vol. 59, pp. 617–645, 2008.

[36] J. Legault, S.-Y. Fang, Y.-J. Lan, and P. Li, "Structural brain changes as a function of second language vocabulary training: Effects of learning context," *Brain and cognition*, vol. 134, pp. 90–102, 2019.

[37] K. Ladendorf, D. Schneider, and Y. Xie, "Mobile-based virtual reality: Why and how does it support learning," *Handbook of Mobile Teaching and Learning. Springer: New York*, 2019.

[38] C. Repetto, B. Colombo, and G. Riva, "Is motor simulation involved during foreign language learning? a virtual reality experiment," *SAGE Open*, vol. 5, no. 4, p. 2 158 244 015 609 964, 2015.

[39] H. A. Curtain, "The immersion approach: Principle and practice.," 1986.

[40] C. Hastings and J. Brunotte, "Total immersion: VR headsets in language learning," *The 2016 PanSIG Journal*, pp. 101–110, 2017.

[41] M. Alfadil, "Effectiveness of virtual reality game in foreign language vocabulary acquisition," *Computers & Education*, vol. 153, p. 103 893, 2020.

[42] E. G. Q. Palmeira, V. B. Saint Martin, V. B. Gonçalves, Í. A. Moraes, E. A. L. Júnior, and A. Cardoso, "The use of immersive virtual reality for vocabulary acquisition: A systematic literature review," in *Anais do XXXI Simpósio Brasileiro de Informática na Educação*, SBC, 2020, pp. 532–541.

[43] P. D. MacIntyre and L. Vincze, "Positive and negative emotions underlie motivation for l2 learning," *Studies in Second Language Learning and Teaching*, vol. 7, no. 1, pp. 61–88, 2017.

[44] Y. Lan, "Immersion, interaction and experience-oriented learning: Bringing vr into fl learning," *Language Learning & Technology*, vol. 24, no. 1, pp. 1–15, 2020.

[45] E. Chang, H. T. Kim, and B. Yoo, "Virtual reality sickness: A review of causes and measurements," *International Journal of Human–Computer Interaction*, vol. 36, no. 17, pp. 1658–1682, 2020.

[46] R. Schroeder, "Defining virtual worlds and virtual environments," *Journal For Virtual Worlds Research*, vol. 1, no. 1, 2008.

[47] S. Warburton, "Second Life in higher education: Assessing the potential for and the barriers to deploying virtual worlds in learning and teaching," *British journal of educational technology*, vol. 40, no. 3, pp. 414–426, 2009.

[48] A. A. Madini and D. Alshaikhi, "Virtual reality for teaching esp vocabulary: A myth or a possibility," *International Journal of English Language Education*, vol. 5, no. 2, pp. 111–126, 2017.

[49] K. Schwienhorst, “The state of VR: A meta-analysis of virtual reality tools in second language acquisition,” *Computer Assisted Language Learning*, vol. 15, no. 3, pp. 221–239, 2002.

[50] J. Falsetti, “What the Heck is a MOO? And What’s the Story with All Those Cows?,” 1995.

[51] B. Sanchez, “MOOving to a new frontier in language learning,” *Telecollaboration in foreign language learning*, pp. 145–163, 1996.

[52] K. Schwienhorst, “Why virtual, why environments? Implementing virtual reality concepts in computer-assisted language learning,” *Simulation & gaming*, vol. 33, no. 2, pp. 196–209, 2002.

[53] H. Brammerts, “Language learning in tandem using the Internet,” *Telecollaboration in foreign language learning*, pp. 121–130, 1996.

[54] K. Schwienhorst, “The ‘third place’–virtual reality applications for second language learning,” *ReCALL*, vol. 10, no. 1, pp. 118–126, 1998.

[55] E. Reid, “Cultural formations in text-based virtual realities,” 1994.

[56] T.-J. Lin and Y.-J. Lan, “Language learning in virtual reality environments: Past, present, and future,” *Journal of Educational Technology & Society*, vol. 18, no. 4, pp. 486–497, 2015.

[57] J. M. Sykes, A. Oskoz, and S. L. Thorne, “Web 2.0, synthetic immersive environments, and mobile resources for language education,” *Calico Journal*, 2008.

[58] J. C. Chen, “The interplay of tasks, strategies and negotiations in Second Life,” *Computer Assisted Language Learning*, vol. 31, no. 8, pp. 960–986, 2018.

[59] Y. Alshumaimeri, A. Gashan, and E. Bamanger, “Virtual worlds for collaborative learning: Arab EFL learners’ attitudes toward Second Life,” *World Journal on Educational Technology: Current Issues*, vol. 11, no. 3, pp. 198–204, 2019.

[60] M. Kruk, “Dynamicity of perceived willingness to communicate, motivation, boredom and anxiety in Second Life: the case of two advanced learners of English,” *Computer Assisted Language Learning*, pp. 1–27, 2019.

[61] D. Kastoudi, “Using a Quest in a 3D Virtual Environment for Student Interaction and Vocabulary Acquisition in Foreign Language Learning.,” *European Association for Computer-Assisted Language Learning (EUROCALL)*, 2011.

[62] N. Levak and J.-B. Son, “Facilitating second language learners’ listening comprehension with Second Life and Skype,” *ReCALL: the Journal of EUROCALL*, vol. 29, no. 2, pp. 200–218, 2017.

[63] S. Melchor-Couto, “Virtual world anonymity and foreign language oral interaction,” *ReCALL: the Journal of EUROCALL*, vol. 30, no. 2, pp. 232–249, 2018.

[64] K. Newgarden, D. Zheng, and M. Liu, “An eco-dialogical study of second language learners’ World of Warcraft (WoW) gameplay,” *Language Sciences*, vol. 48, pp. 22–41, 2015.

[65] L. K. Sylvén and P. Sundqvist, “Similarities between playing world of warcraft and clil,” *Apples-Journal of Applied Language Studies*, vol. 6, no. 2, pp. 113–130, 2012.

[66] K. Scholz, “Encouraging free play: Extramural digital game-based language learning as a complex adaptive system,” *calico journal*, vol. 34, no. 1, pp. 39–57, 2017.

[67] P. S. Rama, R. W. Black, E. Van Es, and M. Warschauer, “Affordances for second language learning in world of warcraft,” *ReCALL: the Journal of EUROCALL*, vol. 24, no. 3, p. 322, 2012.

[68] F. O’Brolcháin, T. Jacquemard, D. Monaghan, N. O’Connor, P. Novitzky, and B. Gordijn, “The convergence of virtual reality and social networks: threats to privacy and autonomy,” *Science and engineering ethics*, vol. 22, no. 1, pp. 1–29, 2016.

[69] H. Rose and M. Billinghurst, “Zengo Sayu: An immersive educational environment for learning Japanese,” *University of Washington, Human Interface Technology Laboratory, Report No. r-95-4*, vol. 199, no. 5, 1995.

[70] T.-Y. Tai and H. H.-J. Chen, “The impact of immersive virtual reality on efl learners’ listening comprehension,” *Journal of Educational Computing Research*, p. 0 735 633 121 994 291, 2021.

[71] R. Kaplan-Rakowski and T. Wojdynski, “Students’ attitudes toward high-immersion virtual reality assisted language learning,” *Future-Proof CALL: language learning as exploration and encounters–short Papers from EUROCALL*, vol. 2018, pp. 124–129, 2018.

[72] Y. Cho, “How spatial presence in vr affects memory retention and motivation on second language learning: A comparison of desktop and immersive vr-based learning,” 2018.

[73] A. Cheng, L. Yang, and E. Andersen, “Teaching language and culture with a virtual reality game,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 541–549.

[74] V. Dobrova, P. Labzina, N. Ageenko, L. Nurtdinova, and E. Elizarova, “Virtual and augmented reality in language acquisition,” in *International Conference on the Theory and Practice of Personality Formation in Modern Society (ICTPPFMS 2018)*, Atlantis Press, 2018, pp. 218–223.

[75] Y. Xie, Y. Chen, and L. H. Ryder, “Effects of using mobile-based virtual reality on Chinese L2 students’ oral proficiency,” *Computer Assisted Language Learning*, pp. 1–21, 2019.

[76] Y.-j. XU, S.-j. ZHENG, Q.-r. CHEN, S.-r. OU, L. Cong, and X. Yao, “The design and implementation of chinese vocabulary learning case based on mobile vr for “the belt and road”,” *Computational Modeling, Simulation and Applied Mathematics*, pp. 263–266, 2017.

[77] T. Jia and Y. Liu, "Words in kitchen: An instance of leveraging virtual reality technology to learn vocabulary," in *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, IEEE, 2019, pp. 150–155.

[78] O. Fuhrman, A. Eckerling, N. Friedmann, R. Tarrasch, and G. Raz, "The moving learner: Object manipulation in virtual reality improves vocabulary learning," *Journal of Computer Assisted Learning*, 2020.

[79] G. Culbertson, E. Andersen, W. White, D. Zhang, and M. Jung, "Crystallize: An immersive, collaborative game for second language learning," in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 2016, pp. 636–647.

[80] M. Macedonia, A. Lehner, and C. Repetto, "Positive effects of grasping virtual objects on memory for novel words in a second language," *Scientific Reports*, vol. 10, no. 1, pp. 1–13, 2020.

[81] M. Macedonia, K. Müller, and A. D. Friederici, "The impact of iconic gestures on foreign language word learning and its neural substrate," *Human brain mapping*, vol. 32, no. 6, pp. 982–998, 2011.

[82] A. M. V. Monteiro and P. N. d. S. Ribeiro, "Virtual reality in english vocabulary teaching: An exploratory study on affect in the use of technology," *Trabalhos em Linguística Aplicada*, vol. 59, no. 2, pp. 1310–1338, 2020.

[83] S. Garcia, D. Laesker, D. Caprio, R. Kauer, J. Nguyen, and M. Andujar, "An immersive virtual reality experience for learning Spanish," in *International Conference on Human-Computer Interaction*, Springer, 2019, pp. 151–161.

[84] N. Collins, B. Vaughan, and C. Cullen, "Designing contextually: An investigation of design-based research to promote situated irish language identity through virtual reality," in *2020 6th International Conference of the Immersive Learning Research Network (iLRN)*, IEEE, 2020, pp. 147–154.

[85] T.-Y. Tai, H. H.-J. Chen, and G. Todd, "The impact of a virtual reality app on adolescent efl learners' vocabulary learning," *Computer Assisted Language Learning*, pp. 1–26, 2020.

[86] G. Lea, "Chronometric analysis of the method of loci.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 1, no. 2, p. 95, 1975.

[87] A. Martin, C. L. Wiggs, L. G. Ungerleider, and J. V. Haxby, "Neural correlates of category-specific knowledge," *Nature*, vol. 379, no. 6566, pp. 649–652, 1996.

[88] V. Cook, *Second language learning and language teaching.* Routledge, 2013.

[89] C. J. Morgan, “Use of proper statistical techniques for research studies with small samples,” *American Journal of Physiology-Lung Cellular and Molecular Physiology*, vol. 313, no. 5, pp. L873–L877, 2017.

[90] D. F. Marks, “Visual imagery differences in the recall of pictures,” *British journal of Psychology*, vol. 64, no. 1, pp. 17–24, 1973.

A Overview of IVRALL+VA studies

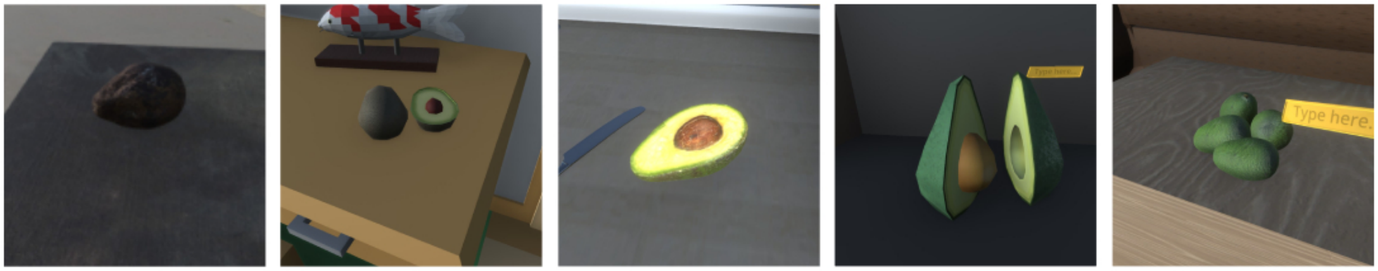
The following table contains an overview of characteristics of the sixteen IVRALL+VA studies that were considered during this research.

	visual target words	audio target words	text target words	T,C,L,M categories	learning phase	quiz phase	non-VR variant vs. IVR	vocabulary evaluation	number of target words	retention evaluation	evaluation on project design	proof of concept	existing program	language L2	participants recruited (number)	participants (age, range or mean)
Zengo Sayu Rose and Billinghurst [69] (1995)	X	X		E,C	X	X								JAP		
*LimbVerbs** Repetto *et al.* [38] (2015)		X		M	X			X	15					CZE	42	33.17
Ogma Ebert *et al.* [8] (2016)	X	X	X	E	X	X	X	X	10	1 week				SWE	20	
Crystallize Cheng *et al.* [73] (2017)			X	C	X	X	X	X	8					JAP	68	18-65
*ClassroomVS** [thesis] Cho [72] (2018)	X		X	L	X		X	X	20					KOR	64	27.28
*VirtualCustoms** Dobrova *et al.* [74] (2018)	X		X	E,C	X	X						X		ENG		
Words in Motion Vázquez *et al.* [9] (2018)			X	M	X		X	X	20	1 week				SPA	60	
*ProtoQuiz** Garcia *et al.* [83] (2019)	X	X	X	E		X					X	X		SPA	4	18-24
Words in Kitchen Jia and Liu [77] (2019)	X	X	X	E	X	X		/	12		X	X		ENG	5	8-10, 24
*ZooKitchen** Legault *et al.* [6] (2019)	X	X		E,M	X		X	X	60					CHI	64	19.05
House of Languages Alfadil [41] (2020)	X	X	X	E	X	X	X	X	30				X	ENG	64	12-15
*IrishSuper** Collins *et al.* [84] (2020)	X	X	X	E	X	X		X	64					IRI	10	18-40
*ObjectManipulation** Fuhrman *et al.* [78] (2020)	X	X		M	X			X	40	1 week				FIN	46	28.41
*CaveGrasp** Macedonia *et al.* [80] (2020)	X	X	X	M	X			X	18	30 days				VIM	46	36.61
*MuseumTour** Monteiro and Ribeiro [82] (2020)			X	C	X			X	17					ENG	25	23
Mondly VR Tai *et al.* [85] (2020)		X	X	C	X		X	X	25	1 week			X	ENG	49	14-15

B Overview visual representation of target words

An overview of the visual representations of all target words per environment are shown in the image grid on the pages below. Each row in the grid corresponds to a specific target word of which the name and Japanese translation is shown to the left above each row. Each column in the grid corresponds to a specific learning or test environment. From left to right, the columns represent: uninhabited island, bedroom with garden, apartment, theatre, and barn bar. Note that the yellow labels attached to the visual representations in the theatre and barn bar environments are text fields and were used to fill in the words in the posttest and delayed posttest respectively.

avocado - waninashi



bag - kaban

ball - tama

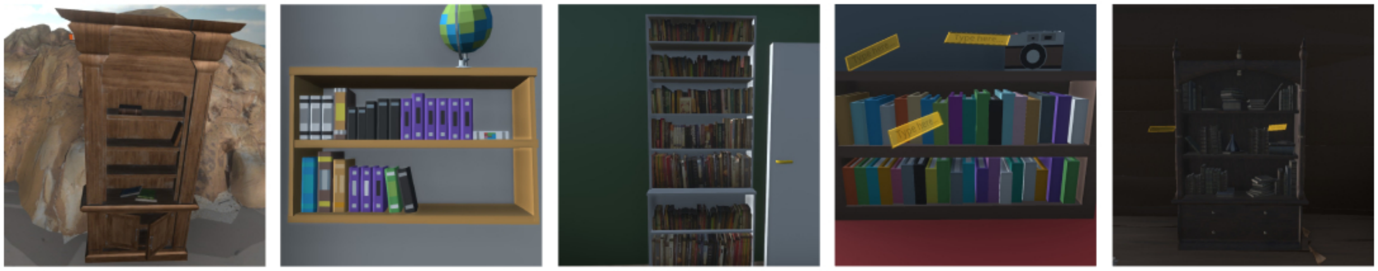
bed - beddo

boat - fune

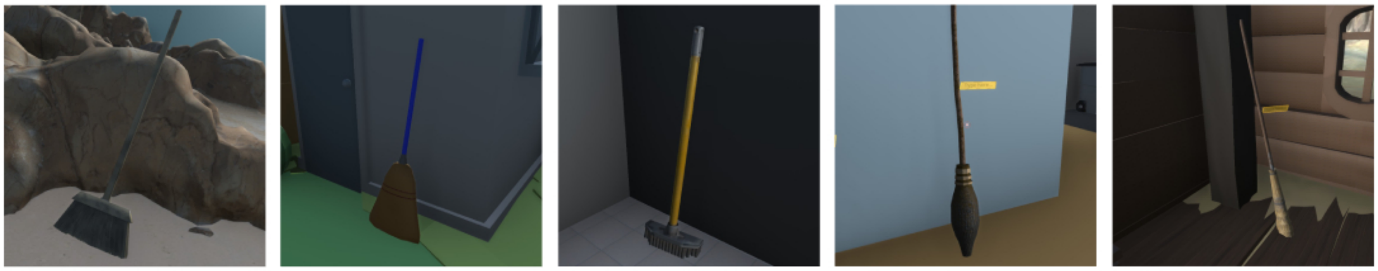
book -hon

pe here...
Type here...

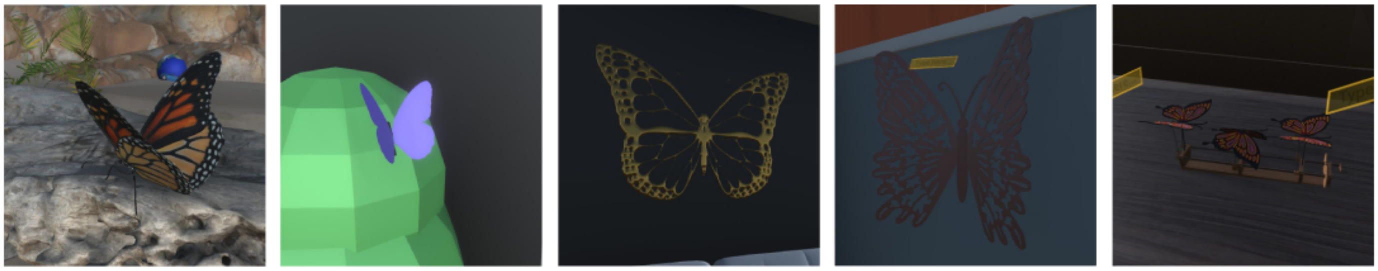
bookcase - hondana



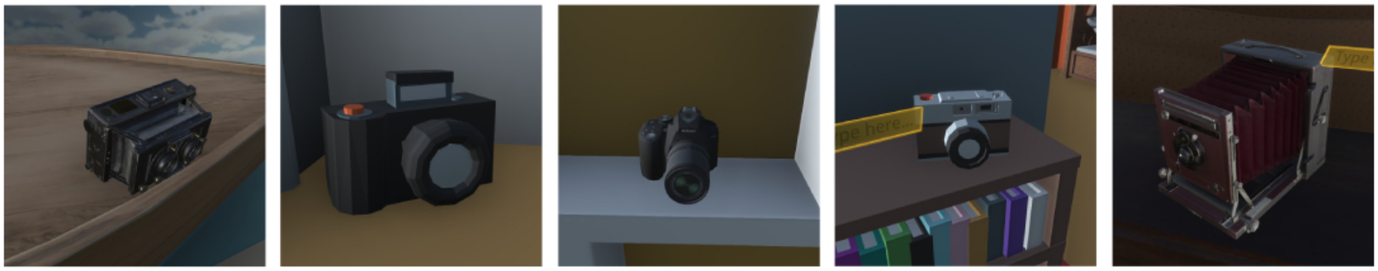
broom - houki



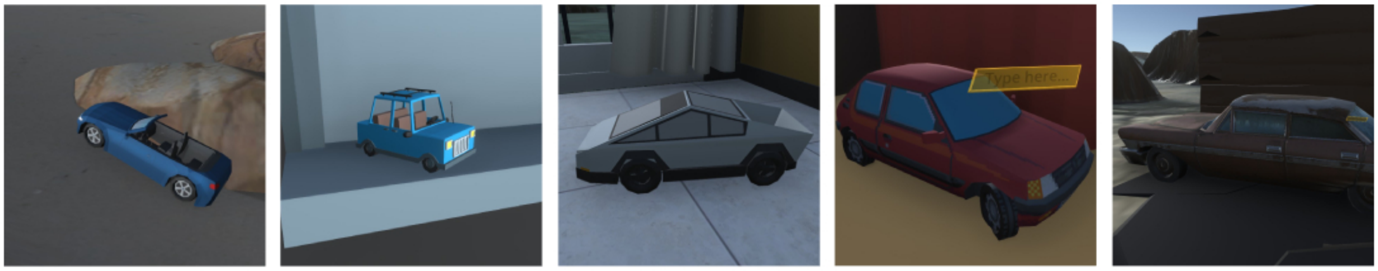
butterfly - chouchou



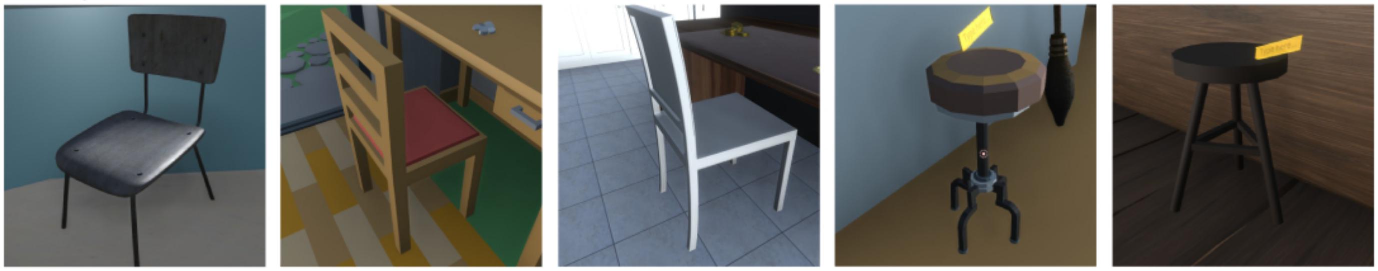
camera - shashinki



car - kuruma



chair/stool - isu



chest of drawers - tansu

coin - kouka

Type here...
Type here...

couch - nagaisu

found everything!

desk - tsukue

earth - chikyuu

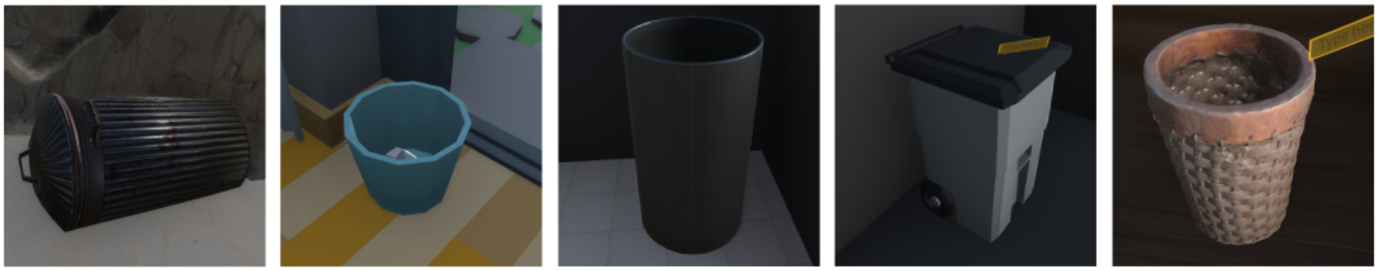
fish - sakana

Type here...
Type here...

flower - hana

Type here...
Type here...

garbage bin - gomibako



glasses - megane

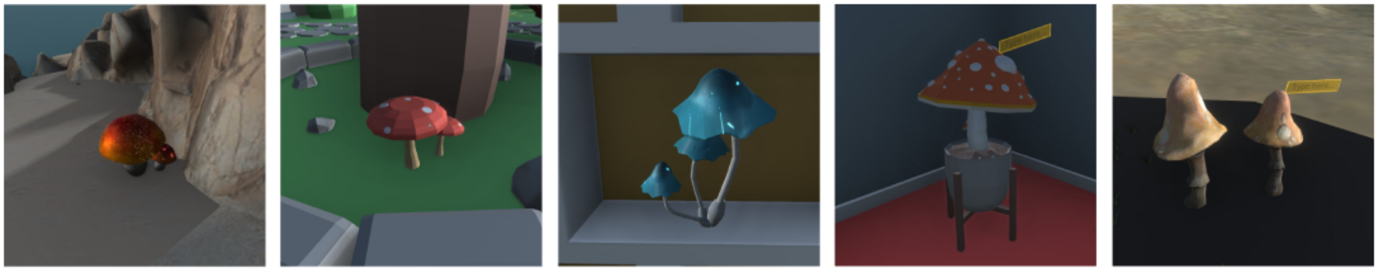
key - kagi

knife - naifu

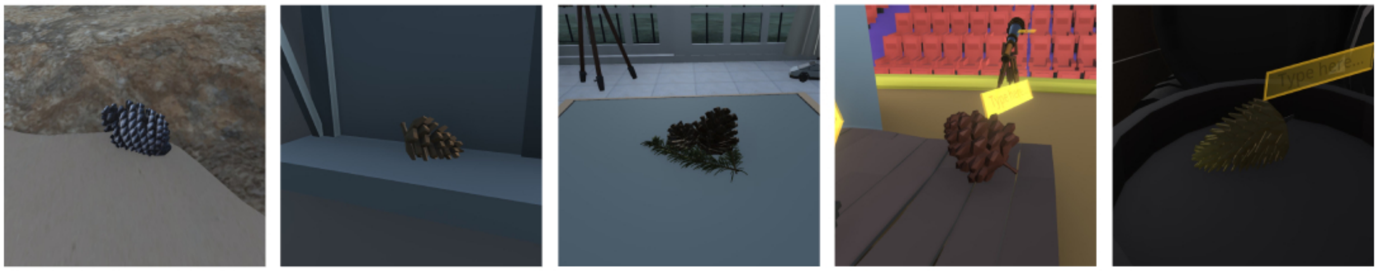
lamp - ranpu

Type here...
Type here...

mushroom - kinoko



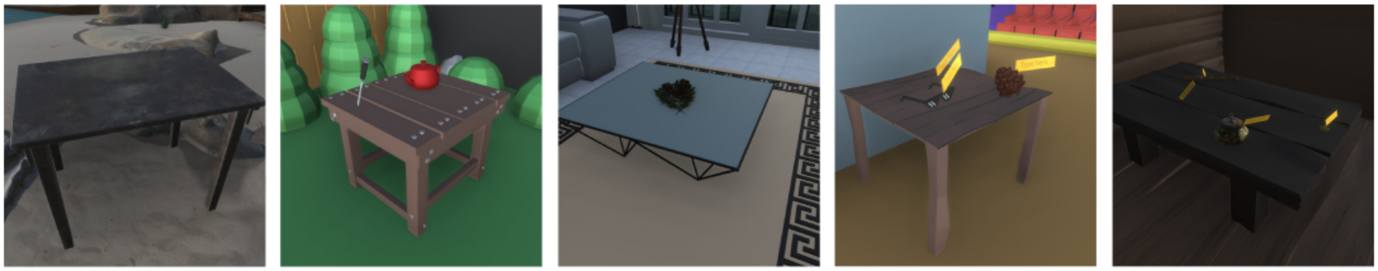
pinecone - matsukasa



rug - juutan

shoe - kutsu

table - tebburu



teapot - chabin

Type here...
Type here...

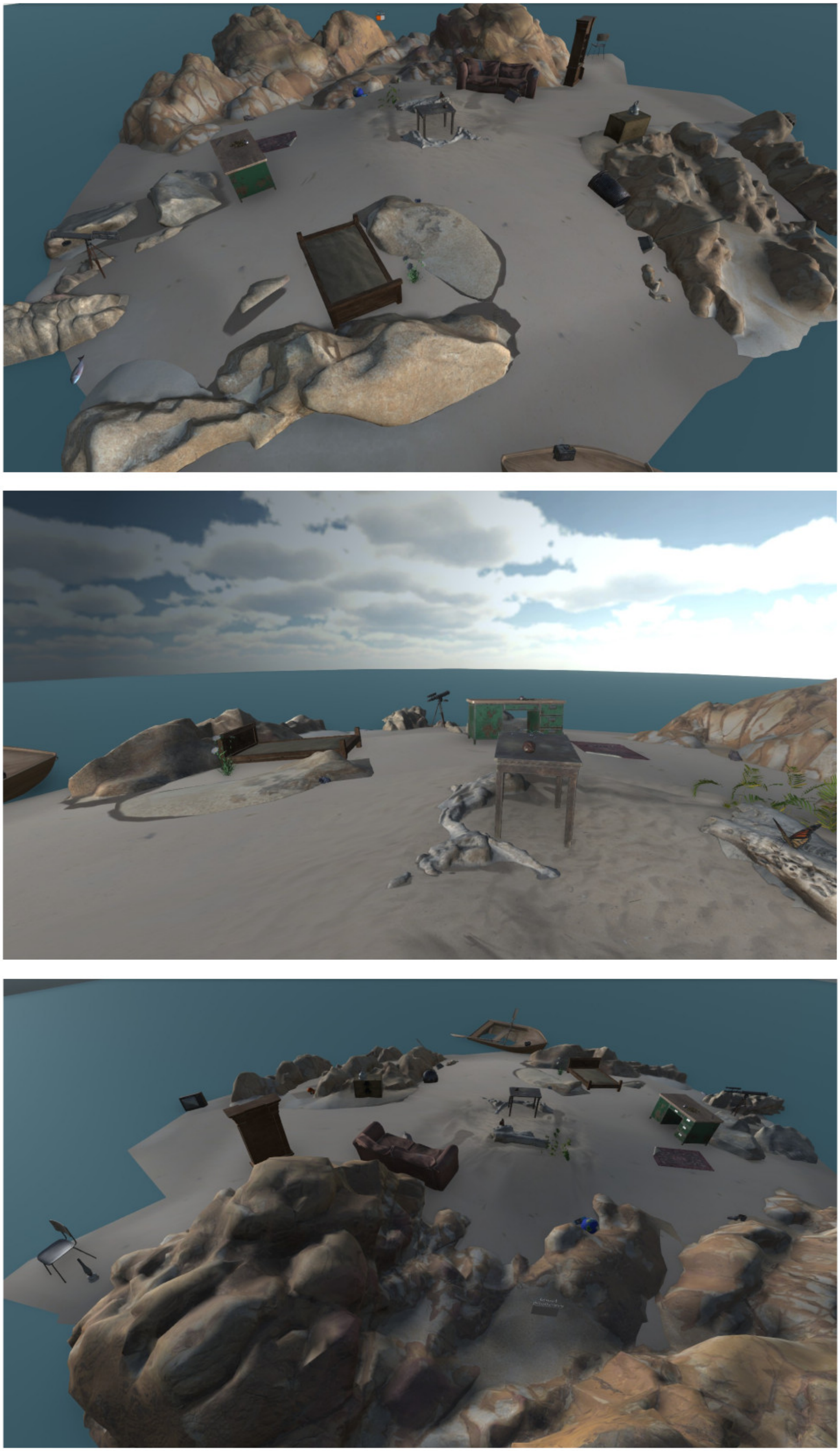
telescope - bouenkyou

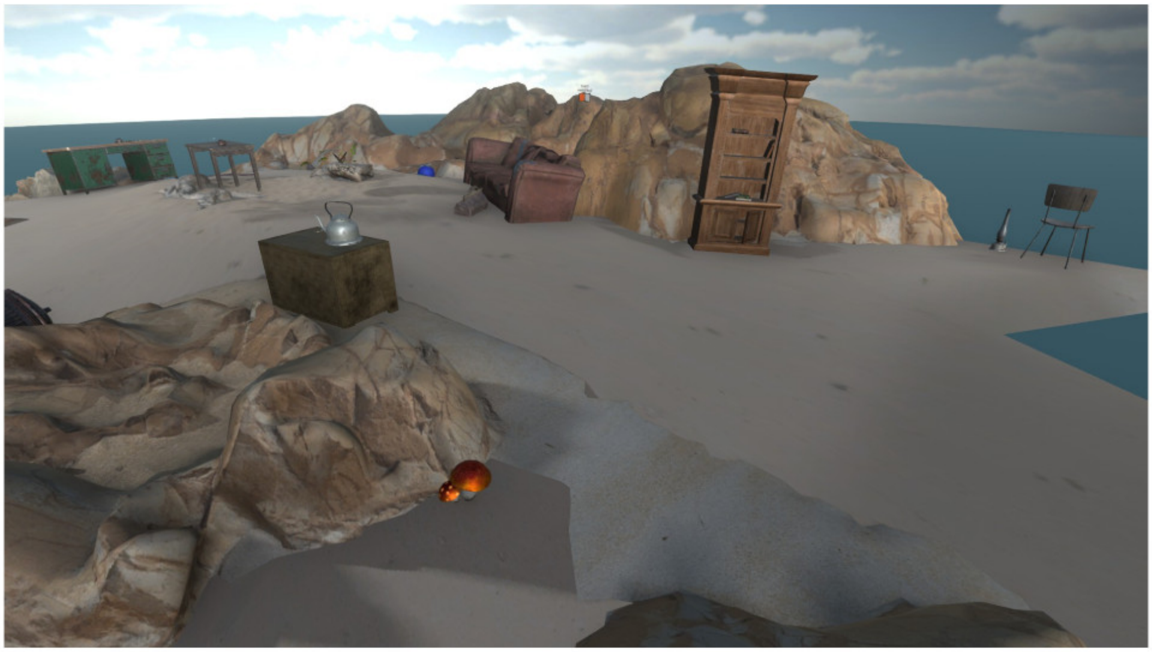
television - terebi

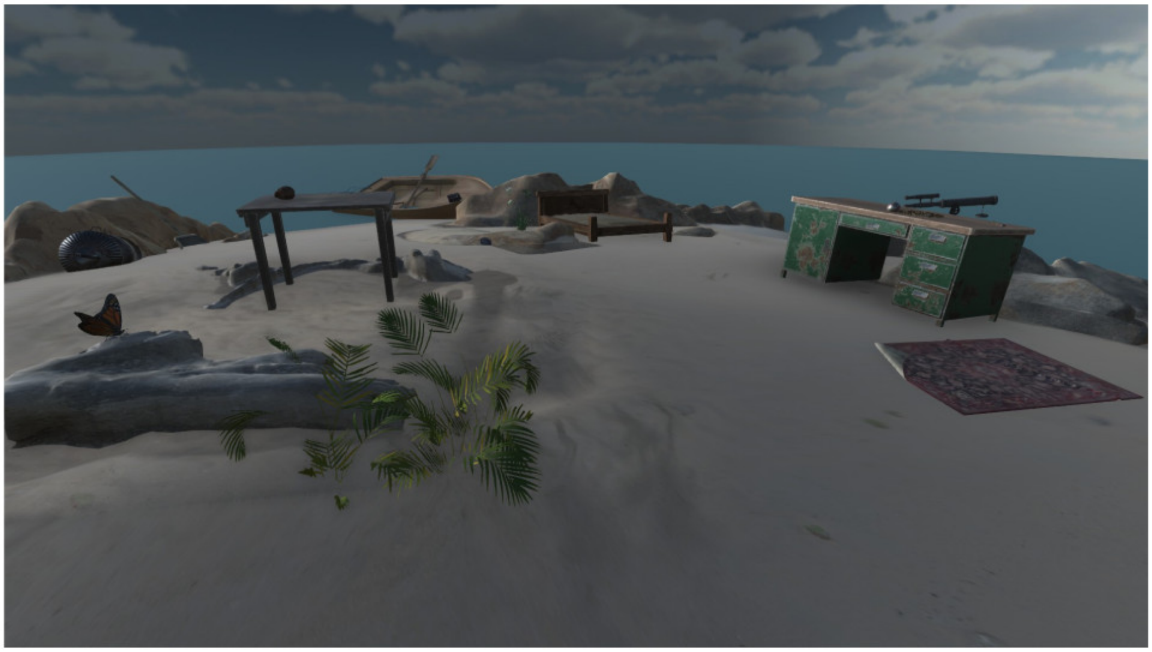
C Overview VEs

Images of respectively the uninhabited island as seen by all participants, the bedroom with garden, apartment, theatre, barn bar and the second and third island of the learned context condition.

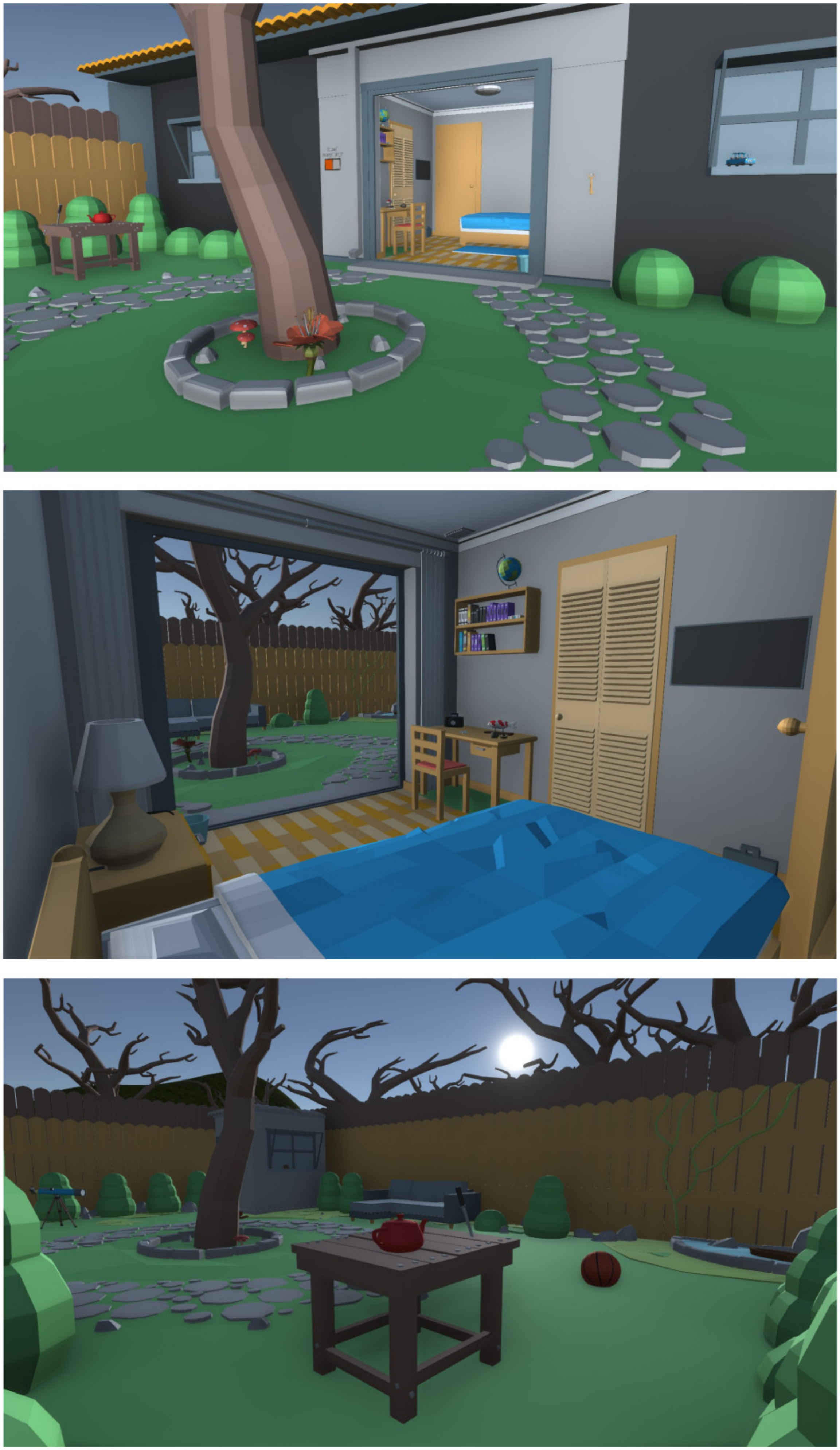
Uninhabited Island 1

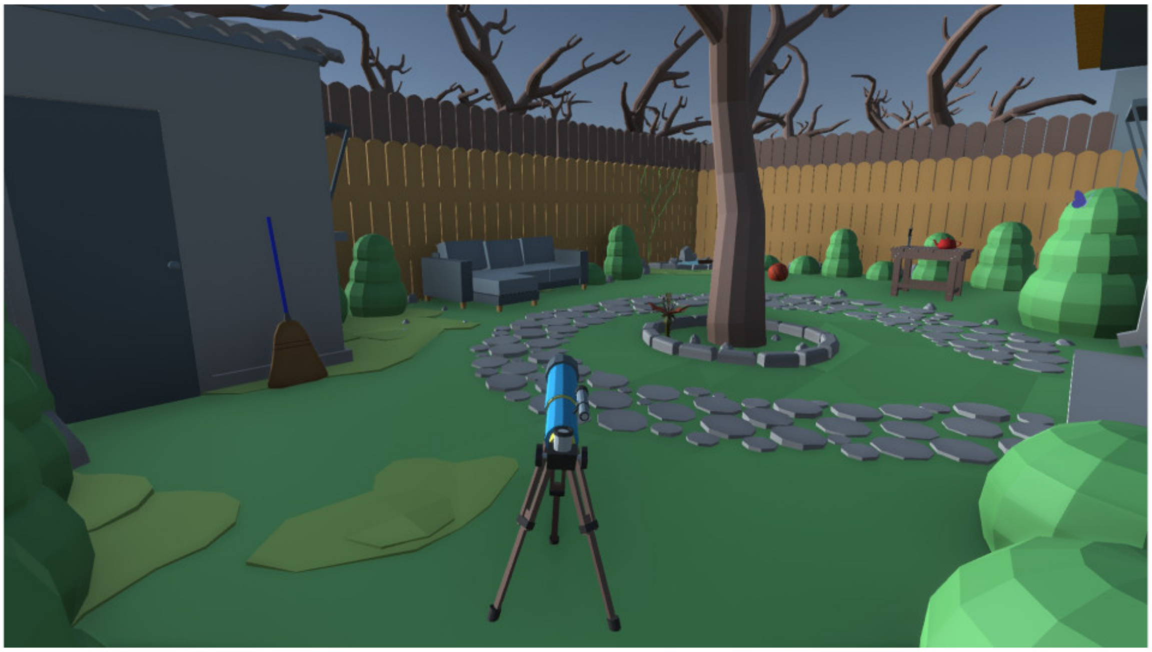




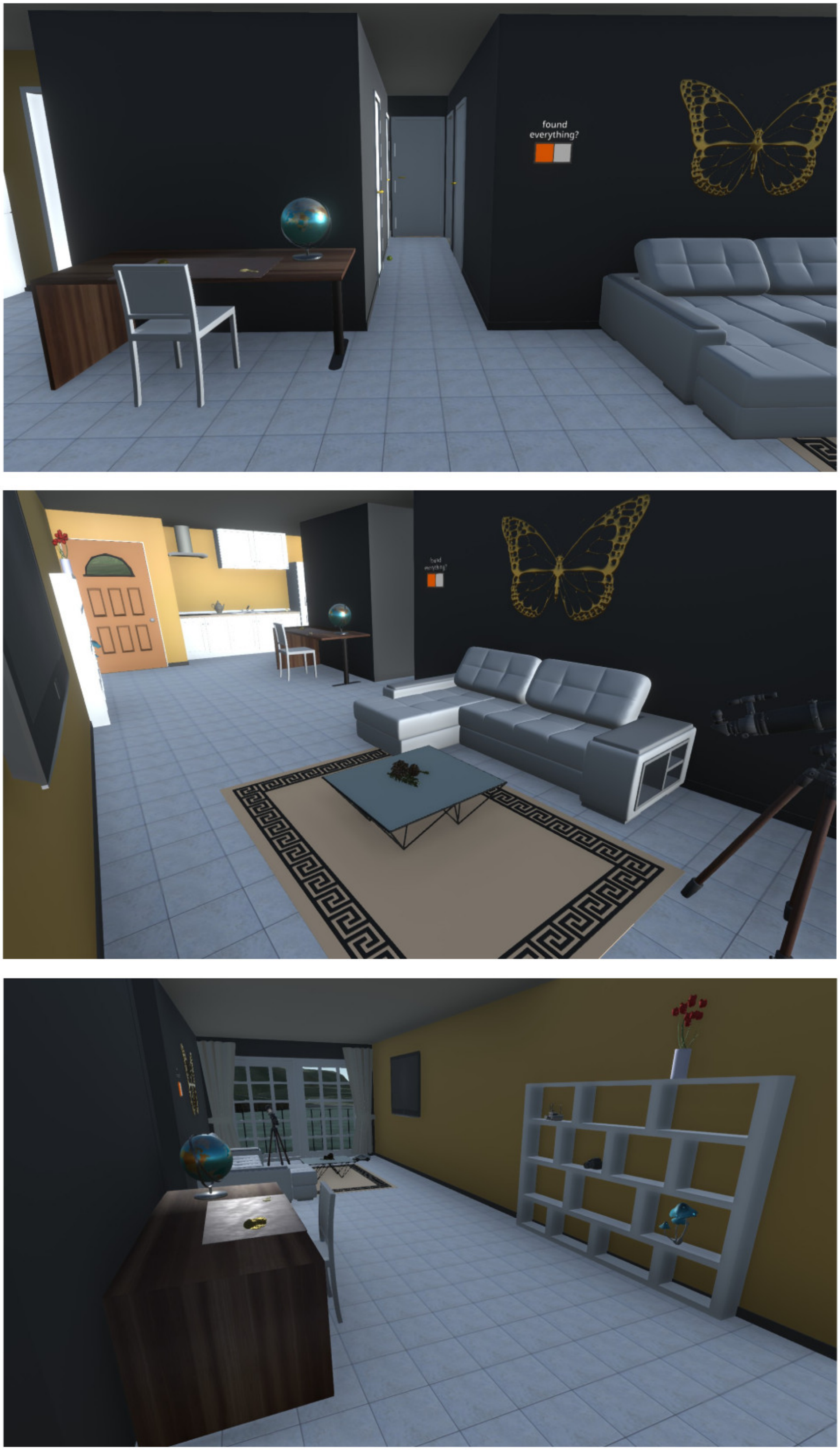


Bedroom with garden

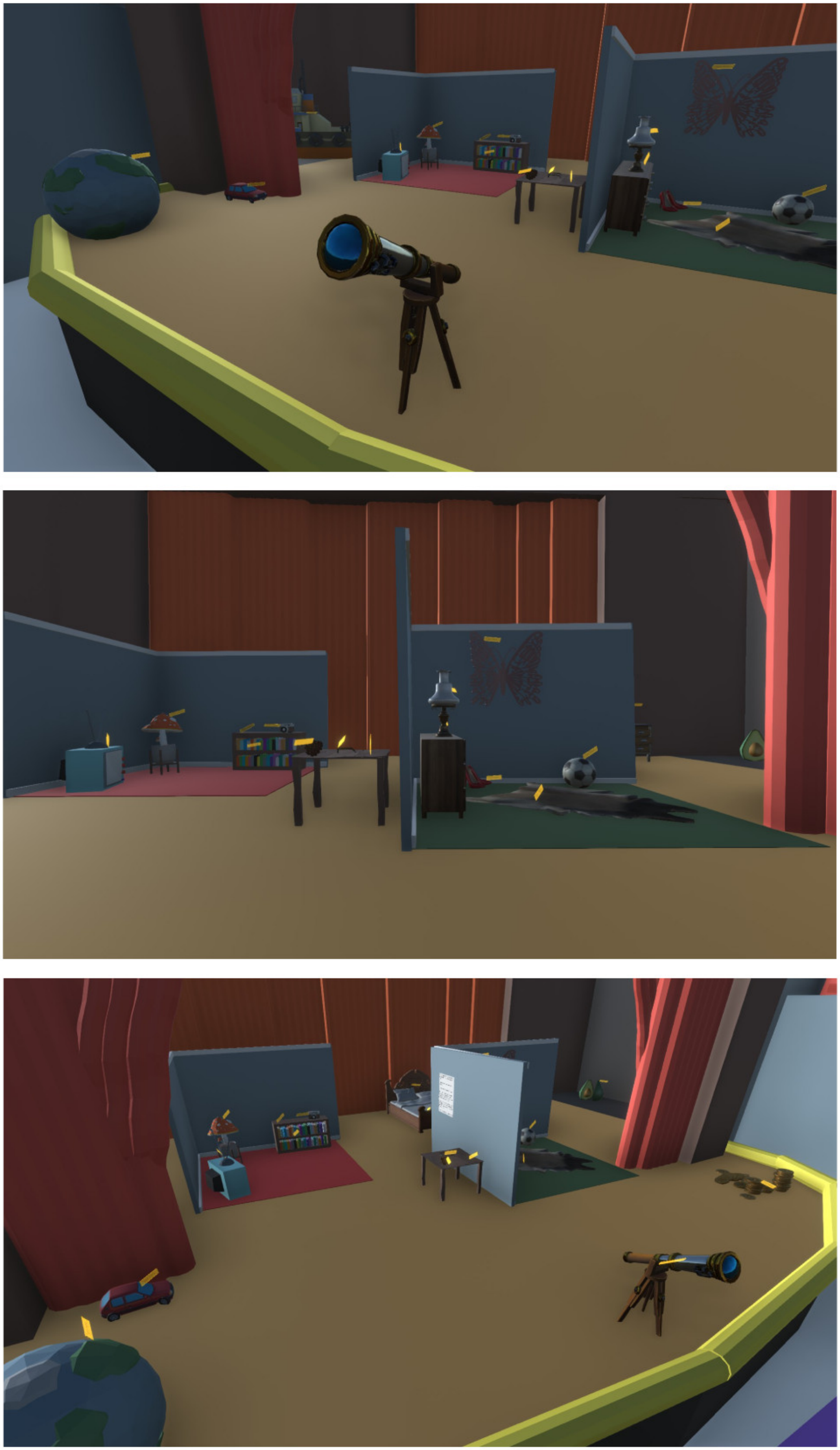


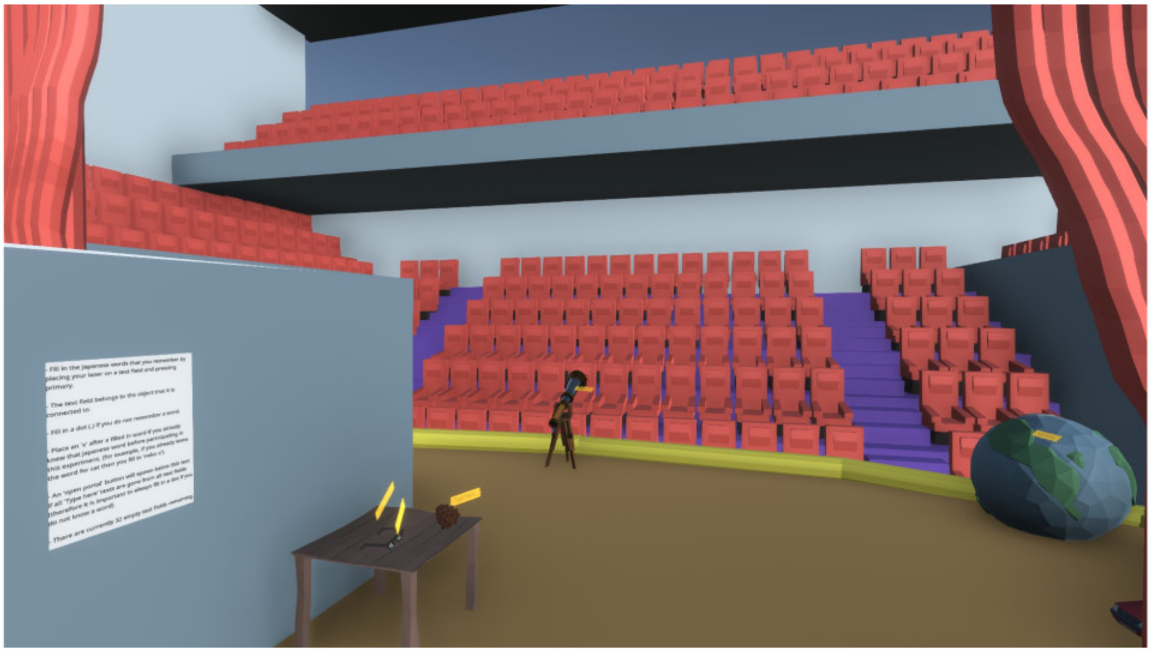


Apartment



Theatre





- Fill in the Japanese words that you remember by placing your laser on a text field and pressing primary.

- The text field belongs to the object that it is connected to.

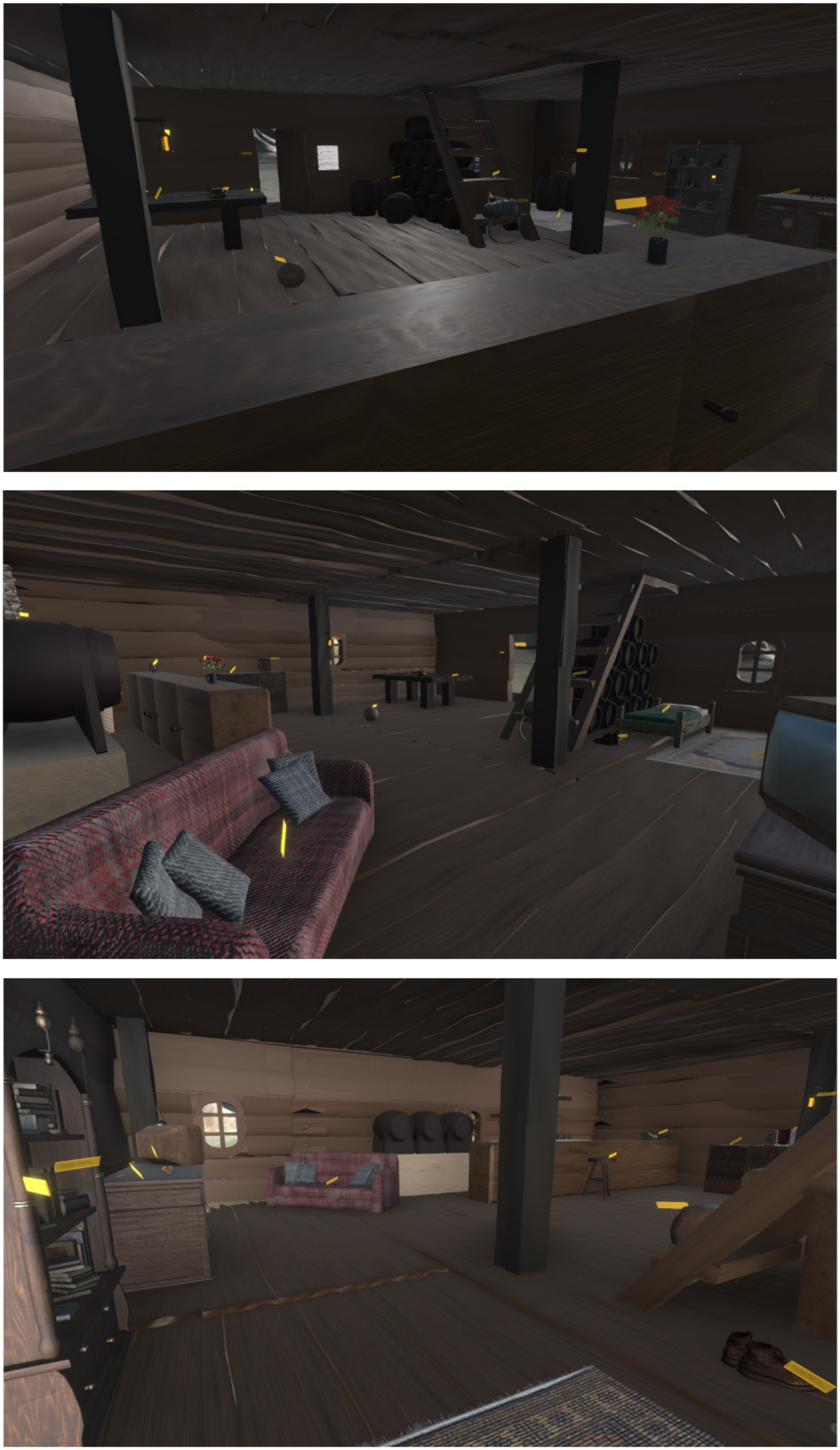
- Fill in a dot (.) if you do not remember a word.

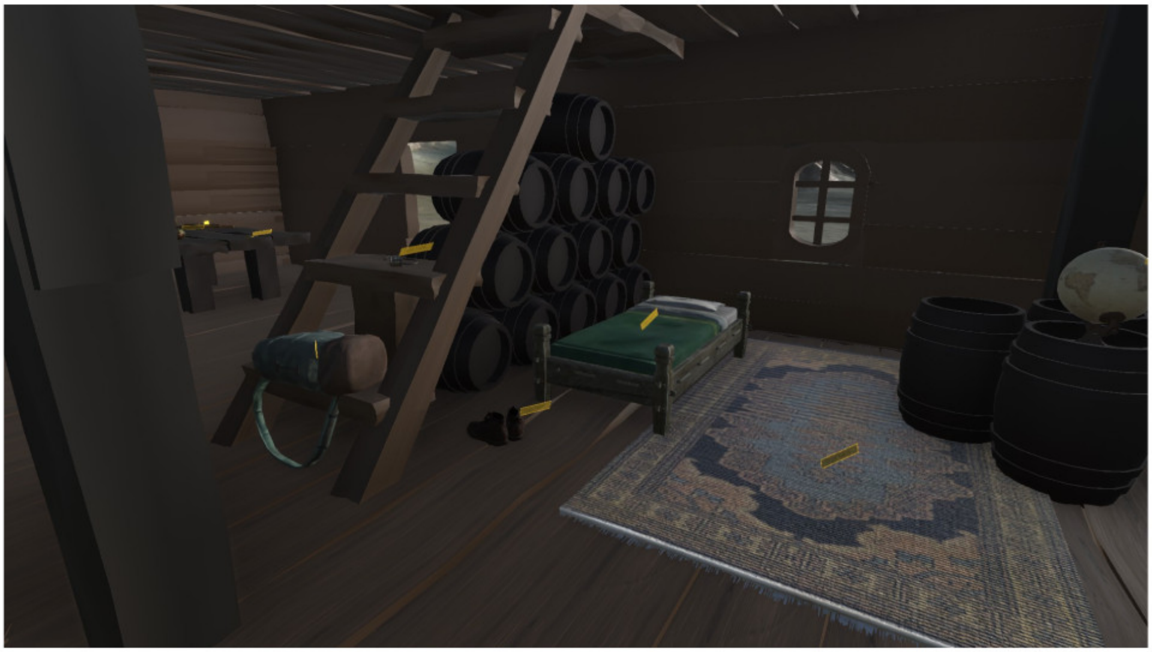
- Place an 'x' after a filled in word if you already knew that Japanese word before participating in this experiment. (for example, if you already knew the word for cat then you fill in 'neko x').

- An 'open portal' button will spawn below this text if all 'Type here' texts are gone from all text fields (therefore it is important to always fill in a dot if you do not know a word).

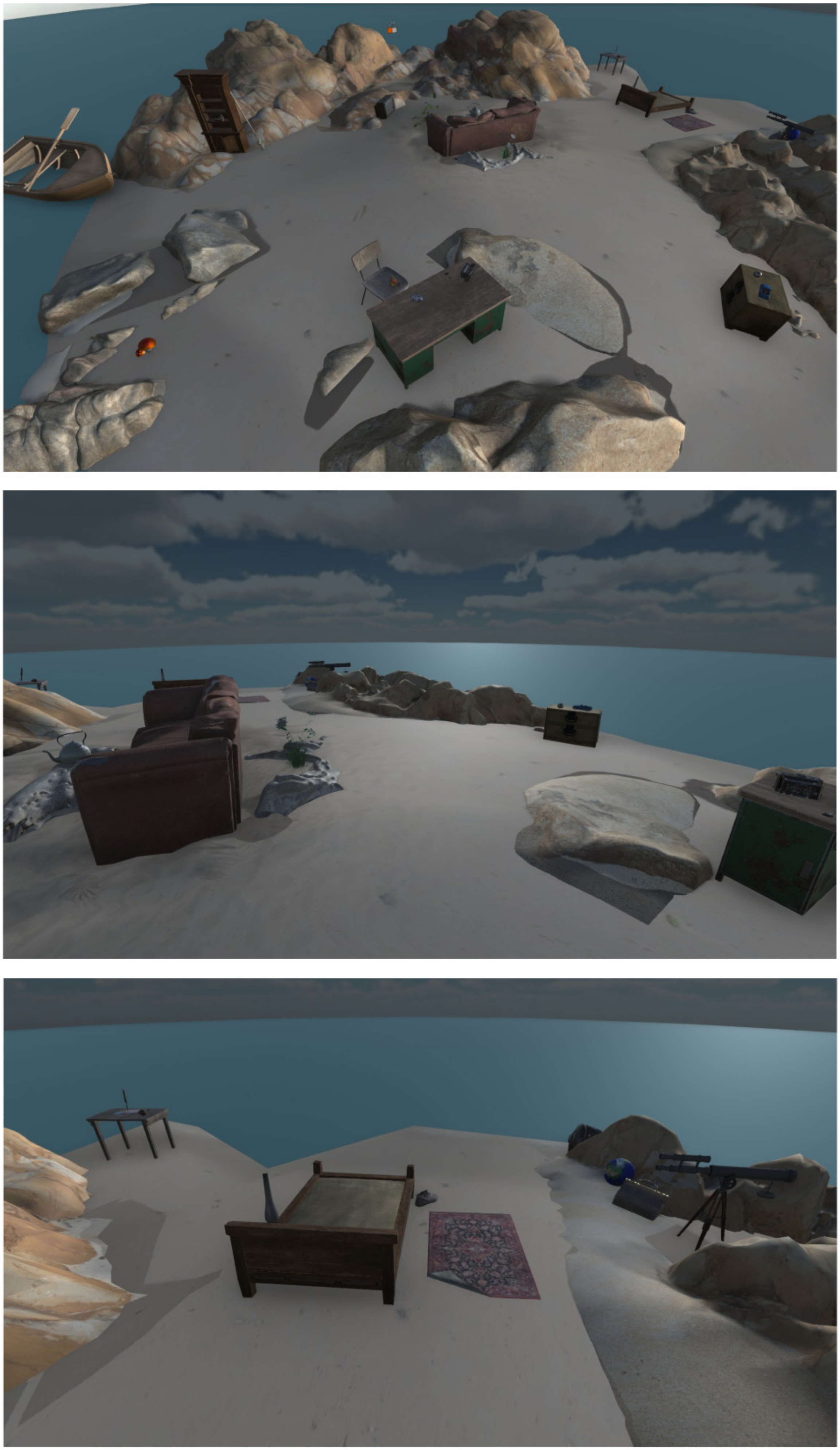
- There are currently 32 empty text fields remaining

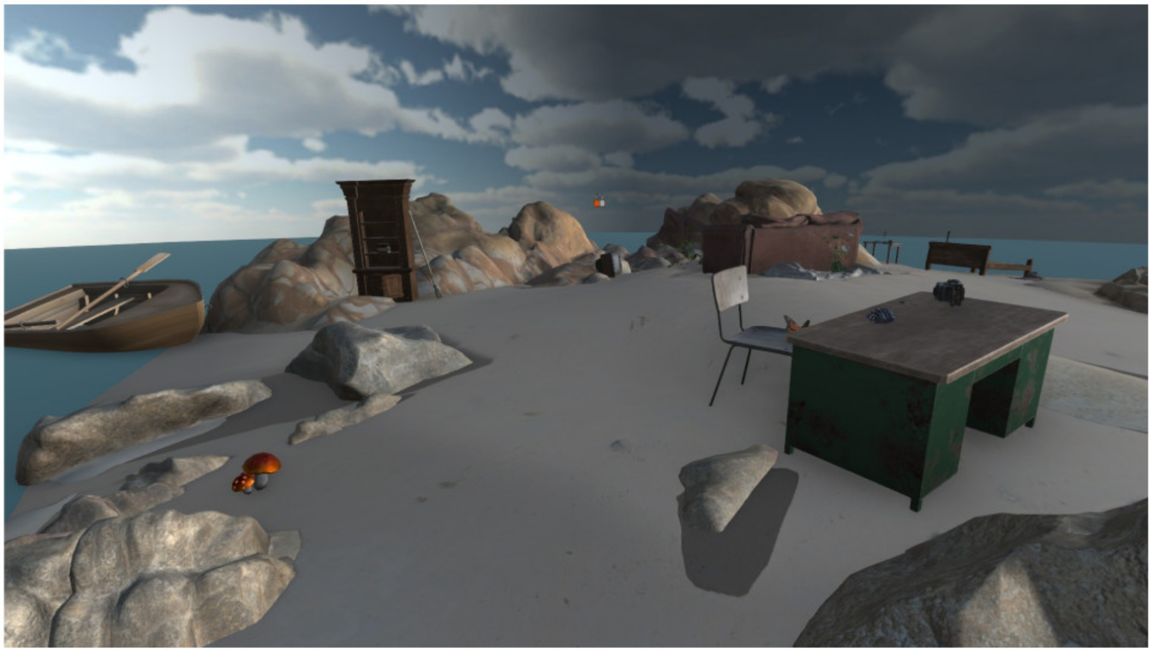
Barn Bar

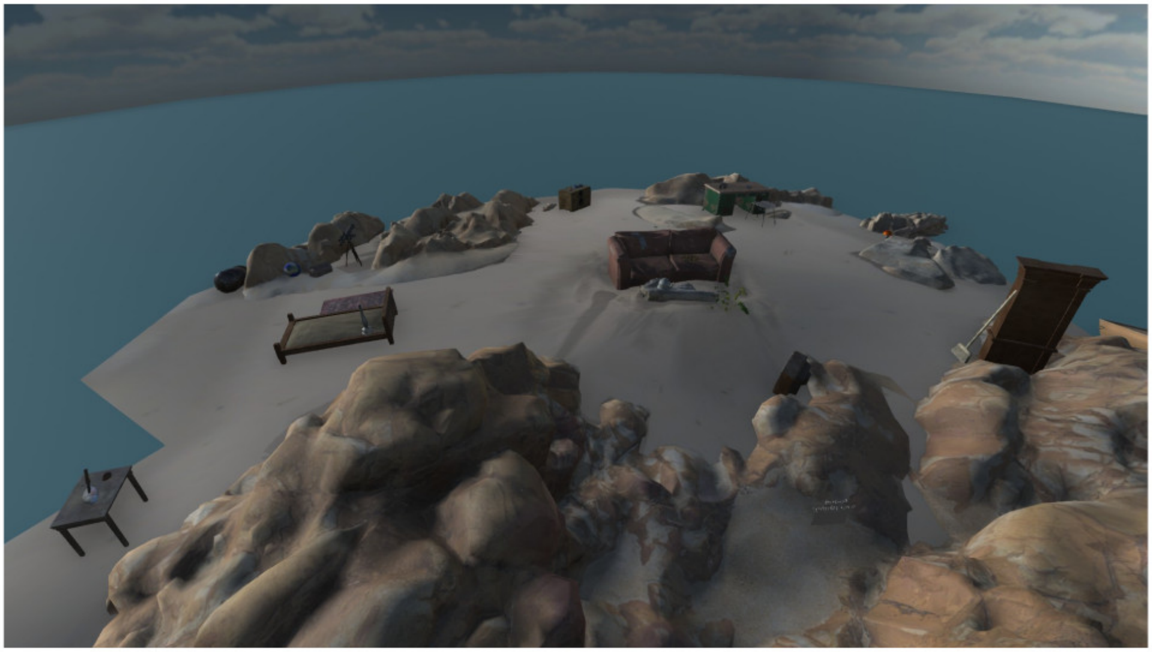




Uninhabited Island 2







Uninhabited Island 3