Augmented Reality to Enable Users in Learning Case Grammar from Their Real-World Interactions

Fiona Draxler[1], **Audrey Labrie**[2], **Albrecht Schmidt**[1], **Lewis L. Chuang**[1]
[1]LMU Munich, Munich, Germany, {fiona.draxler, albrecht.schmidt, lewis.chuang}@ifi.lmu.de,
[2]Polytechnique Montreal, Montreal, QC, Canada, audrey.labrie@polymtl.ca

ABSTRACT

Augmented Reality (AR) provides a unique opportunity to situate learning content in one's environment. In this work, we investigated how AR could be developed to provide an interactive context-based language learning experience. Specifically, we developed a novel handheld-AR app for learning case grammar by dynamically creating quizzes, based on real-life objects in the learner's surroundings. We compared this to the experience of learning with a non-contextual app that presented the same quizzes with static photographic images. Participants found AR suitable for use in their everyday lives and enjoyed the interactive experience of exploring grammatical relationships in their surroundings. Nonetheless, Bayesian tests provide substantial evidence that the interactive and context-embedded AR app did not improve case grammar skills, vocabulary retention, and usability over the experience with equivalent static images. Based on this, we propose how language learning apps could be designed to combine the benefits of contextual AR and traditional approaches.

Author Keywords

Augmented Reality; Language Learning; Grammar; Contextual Learning; Self-Directed Learning

CCS Concepts

•**Applied computing** → **Interactive learning environments;**
•**Human-centered computing** → *Mixed / augmented reality;*

INTRODUCTION

"If we spoke a different language, we would perceive a somewhat different world." – Ludwig Wittgenstein

Language provides us with the means to create and communicate imagined settings, scenarios, and situations. Thus, it is unsurprising that language learning presents a key application field that could greatly benefit from innovative AR implementations [3, 34]. In particular, AR provides a unique opportunity for unfamiliar instructional material to be directly embedded in the familiar setting that a user interacts with on a daily basis. It offers users a level of interactivity that is believed to foster task motivation and engagement [30]. Thus, the growing ability of consumer mobile computing devices for rendering AR presents the opportunity to make the world a classroom for language learning.

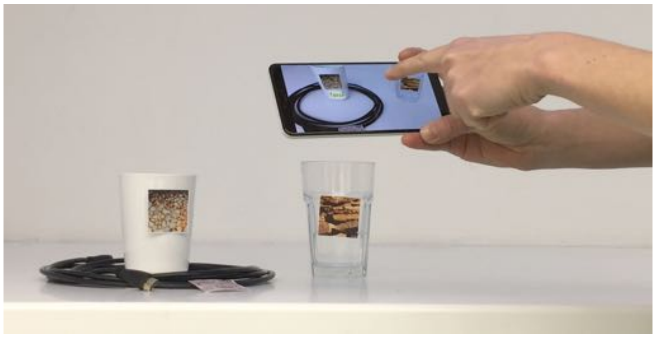


Figure 1. Our Augmented-Reality app for learning case grammar.

Recent years have witnessed notable successes of AR systems that support self-learning in everyday contexts. For example, some systems provide labels of new foreign words with real-life objects [21, 38], while others have provided visualisations to communicate the unseen physical properties of everyday objects (e.g., heat conductivity) and the laws of physics that govern them [22, 36]. In general, the use of AR for contextual learning has been shown to deliver learning benefits over more traditional methods, such as rote learning. Nonetheless, it is unclear if this is necessarily true for all types of learning. While there are many desirable aspects related to the use of AR for learning, these aspects can also be implemented in traditional learning media. For example, flashcards need not be purely text-based; they can also depict a photo-realistic scene wherein objects are labelled with foreign words. Like AR, such learning material can similarly establish strong contextual associations to assist subsequent recall, even if they do not necessarily 'mirror' the user's actual environment.

A unique characteristic of AR is that it enables users to interact and actively discover relationships between objects in the real world. In this work, we evaluate how AR could be implemented to help users learn the influence of physical context on language grammar. Specifically, we address how AR could be developed to help users learn how context modifies the article of gender nouns; this is an aspect of learning certain foreign languages (e.g., German, Russian, Croatian) that is known to be especially challenging to native speakers of languages that do not have cases (e.g., English, French, Mandarin).

CHI '20, April 25–30, 2020, Honolulu, HI, USA.
© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
http://dx.doi.org/10.1145/3313831.3376537

More specifically, we developed an AR app that allowed adult learners to interact with real-world objects and discover how spatial relationships between them affected case grammar (i.e., article + preposition constructs), see Figure 1. As a control, we presented the same high-resolution content, generated and experienced by users of the interactive app, *Snapshot*, presented in a non-interactive static format. Our working hypothesis was that users would be better in learning case grammar from an AR experience that allowed them to see how their own manipulations of real objects modified the sentence itself. In contradiction, Bayesian t-tests analyses of learnt performance, based on testing immediately and one week after self-learning, revealed substantial evidence that there was no difference between learning with the *AR* or *Snapshot* app. Nonetheless, qualitative results revealed that our participants were enthusiastic in making these AR techniques part of their foreign language learning routines, suggesting a high potential for AR-based self-learning applications. By considering the quantitative and qualitative results together, we address the design challenges that AR apps could target to allow them to surpass traditional self-learning methods.

BACKGROUND AND RELATED WORK

The current work investigates the potential of AR in helping adults learn a second language on their own. The following examples will illustrate how the application of AR for learning is tightly interwoven with context and context-based technologies. Therefore, we summarise the role of context in learning and respective teaching strategies from a cognitive point of view. Finally, we provide an overview of the relevance of AR to context-based (language) learning applications.

Learning, Context, and AR

Learning can benefit from the relationship between context and learning content in several ways. Firstly, context determines the immediate relevance and, thus, the motivation to engage with learning content. It has been shown that people are more motivated to learn if (1) they see the value of the content – its importance or utility –, (2) the cost of learning is not too high, (3) they find the content interesting [30]. Perceived interest, in turn, is often rooted in a learner's situation [1]. For example, being in a café in a foreign city is likely to raise the interest in learning phrases necessary for ordering coffee. Secondly, context makes it possible to form associations that ease later retrieval in similar circumstances [18]. For example, new words relevant to the learning context are more likely to be recalled than unrelated words [8, 12]. Thirdly, rich context fosters situated learning, because it is a contributing factor to knowledge building – in addition to social behaviour, activities, and the underlying culture [5]. AR is the preferred technology for contextual learning because it can be used to contribute context-specific, just-in-time information in an interactive manner [11, 35].

Besides context, AR can enhance learning by embedding additional content to facilitate multimedia learning, i.e., learning through associations between verbal and imagery information by drawing on different sensory and working memory channels [28]. For instance, learning new vocabulary with both text and image leads to better acquisition rates than with textual information only [7]. In this light, AR applications, which typically interleave real-life and virtual content, are likely to be especially suitable for language learning and can provide an additional modality for processing and understanding learning materials [34, 35].

Existing Applications for Language Learning

Numerous language apps have been developed for both research and industry (for an overview, see [19]). Below, we present a selection of learning systems that use AR as well as systems that select learning material based on the learner's current context. In particular, we focus on how conceptual knowledge can be taught (as opposed to aspects such as vocabulary, which rely on rote learning).

AR Learning Applications

In language learning, AR has been used to label real-life objects in the foreign language for vocabulary learning. For instance, in one project, a HoloLens setup was used to augment objects in the learner's surroundings by displaying the respective foreign language words and presenting audio samples [21]. Vocabulary acquisition with this AR setup was compared to a web-based flashcard control condition; a 4-day delayed test revealed better recall in the AR condition. A drawback of this setup is the need to manually tag objects. This issue was tackled in a similar project that also used the HoloLens, but automatically generated labels with object-recognition and content-retrieval systems [38]. In addition, object enhancements were extended with definitions, example sentences, and related multimedia content. Here, the automated detection and augmentation processes significantly broadened the range of application in comparison to manual tagging, but, as of yet, there was no formal evaluation of this system.

Language is more than its vocabulary, though, and AR could also be used for teaching structural concepts of a language as it is a suitable technology for visualising spatial and temporal relationships [3, 34]. For example, with the 'Block Talks' system, children train their literacy skills by combining tangible letter objects to create sentences [13]. The app provides corrective feedback, as well as AR animations illustrating the meaning of sentences. As of yet, Block Talks was not tested for performance over a traditional non-AR method. While not designed for foreign language learning as such, similar concepts could be used for training sentence construction, a common issue, for instance, in German [32].

Finally, in science education, AR has been used to simplify experimental setups or to augment experiments with information that cannot be perceived by humans. For example, the thermal flux experiment has been augmented with a HoloLens setup by showing temperature changes of a rod that is heated on one side and cooled on the other [36]. Similarly, Ibáñez et al. displayed electromagnetic forces on top of tangibles representing circuit elements [22]. Their AR condition outperformed their web-based control condition that provided the same learning content but gave no dynamic feedback on circuits constructed with tangible elements. These two approaches allow learners to observe how their changes influence the experiment setups and thus gain an understanding of the overarching concepts.

Overall, it can be seen that AR has been applied for teaching in different domains, although not all projects have been empirically evaluated. Specifically, to the best of our knowledge, the benefits of AR for teaching conceptual language skills such as grammar structures have not been studied so far. Moreover, the web-based systems used as control conditions in the experiments described above used media that were not as rich as in the corresponding AR condition or provided no dynamic feedback. Below, we broaden the scope to also present projects that base learning content on context but do not include AR.

Context-Based Learning Applications

Learning systems can be designed to adapt to the user's context. In particular, sensors on a user's mobile device information can currently provide information on the user's geo-location, activity, and cognitive state [2], which can subsequently be used to shape the learning experience [9]. In this section, we provide examples of previous implementations as well as evidence suggesting that such context-based systems can enhance learning.

For example, it has been shown that vocabulary learning is enhanced when unfamiliar words are presented to learners in a relevant setting, compared to if they were presented in a random context. A traditional flashcard learning application could determine which words to present depending on the user's current location (e.g., "train" when the user is travelling; [12]). This can also be implemented as an implicit learning system whereby the geo-location selects relevant new words for presentation on the lock screen of a user's mobile device [8]. Both implementations have successfully demonstrated that learning words that are relevant to geo-location can result in better retention of presented new words.

Besides geo-location, learning content can also be shaped by the content of any linguistic material that currently engages a user. For example, the application 'WaitChatter' infers the current topic of a user's instant messenger conversation to determine which new words to present during idle moments [6]. Text on user-frequented websites can also be used for content generation. In 'ALOE', first-language words can be replaced with their foreign-language translations [37]. In WaitChatter and ALOE, learners demonstrated an improvement in their vocabulary skills. However, there was no comparison to a non-contextual solution. In another system, the texts were turned into clozes by removing articles, determiners, or prepositions [29]. In this case, the focus was on evaluating system accuracy; the learning effect was not tested.

There has also been work on constructing virtual contexts to aid learners. In one example, students were instructed to apply a set of grammar structures in their chat messages with other students [26]. A supervising teacher corrected mistakes made in the chats. Providing this immediate and relevant context enabled targeted students to achieve a better understanding of grammar structures compared to other students who relied on a coursebook and oral communication exercises instead.

Immersive environments can be specifically developed in Virtual Reality (VR) to convey appropriate contextual presence for taught language learning [27, 33]. For instance, VR environments have been specifically tailoured for vocabulary learning [16] or for medical students practising conversations with virtual patients [41]. Nonetheless, creating bespoke virtual environments can be extremely time-consuming, resulting in stationary setups that cannot be readily re-purposed, and interaction with the environment is commonly limited to handheld controllers instead of direct object manipulation.

To summarise, geo-location has successfully been used to enhance vocabulary learning. The impact of the activity context, on the other hand, still needs further confirmation. As with AR systems, there is again a predominance of projects teaching words rather than grammar structures.

Summary

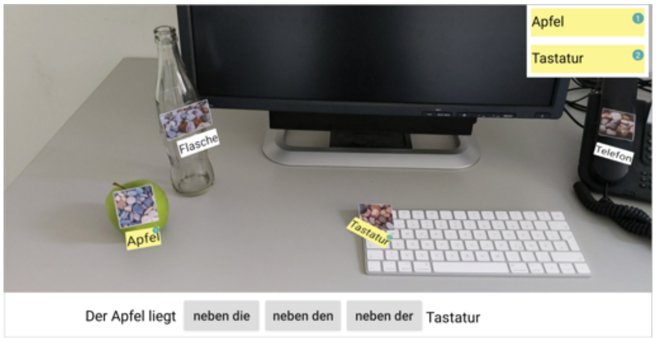
The examples above clearly demonstrate that presenting learning matter in context can improve learning performance. AR provides a unique opportunity of embedding the learning matter in the real-world context itself in a way that is sensitive to a user's activity; it allows users to select their own context and learning matter. Nonetheless, it remains an open question as to whether this is necessarily superior to existing context-relevant and media-enriched learning experiences.

The majority of the language learning tools and empirical studies presented above target vocabulary learning [10, 19]. However, integrating the context could also be helpful when learning conceptual and structural aspects of a language (e.g., grammar or the description of spatial and temporal relations). The current work focuses, not on rote learning but, on how context-based AR could support the *transfer* of concepts from a learnt instance to other situations [17].

APP DEVELOPMENT AND FEATURES

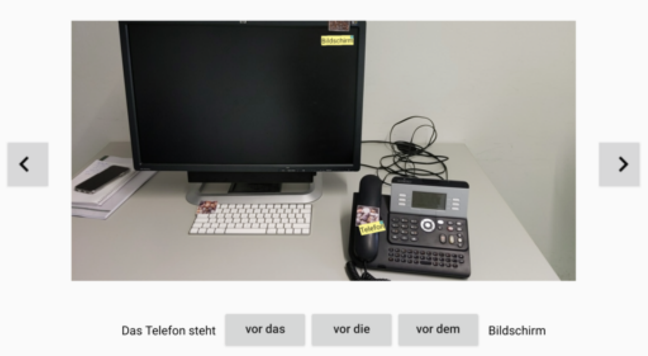
So far, we have shown that strategies for language learning have been widely researched, and applications have been developed based on these strategies. However, there is little empirical evidence that demonstrates the benefits of AR and context-based systems in comparison to traditional non-contextual systems. Where examples exist, they are usually targeted at vocabulary learning and rarely on the self-learning of structural and grammatical concepts. We contribute to filling this gap with a *handheld-AR* app and a corresponding *Snapshot* app for learning *preposition + article* constructs in German using real-life objects in the user's surroundings in the first case and using static images in the latter case (see Figures 2 and 3, respectively). The construct was chosen as an example of a language structure where the underlying rules can then be applied to similar situations once a learner has identified them. In addition, we decided to use AR rather than VR because of the theoretical plausibility that real-world relevance would motivate grammar learning. Furthermore, unlike VR, mobile-based AR offers the opportunity for users to learn on-the-go.

In this section, we explain our content selection process, design decisions, and technology choices. We complement our argument with the findings of an online survey amongst language learners, as detailed below.



Der Apfel liegt | neben die | neben den | neben der | Tastatur

Figure 2. Screenshot of the Augmented-Reality (AR) App, depicting virtual labels of objects in a common office scene.



Das Telefon steht | vor das | vor die | vor dem | Bildschirm

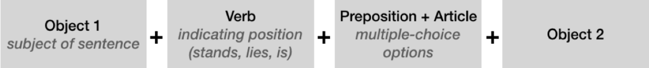
Figure 3. Screenshot of the *Snapshot* app, depicting virtual labels of objects in a common office scene.

Survey on Learner Experiences and Needs

The online survey we conducted polled potential users on their motivation, experiences, and needs for learning languages. It was announced via mailing lists, social media, and through local language course companies. We received responses from 122 participants. Of these, 27 were currently learning German and answered some additional questions specifically addressing issues with German. The questions concerned language learning habits, the efficiency of different strategies, and problems that learners face. Overall, learners stated work (54.9%), studies (44.3%), and travelling (43.4%) as a motivation for them to learn a language (multiple selections were possible). Methods that were considered most helpful were watching movies or TV ($M = 3.81$, $SD = 1.30$) and talking to native speakers ($M = 3.76$, $SD = 1.43$; *1 = not effective at all, 5 = very effective*). Language apps received the second-lowest score of the eight methods we presented ($M = 3.10$, $SD = 1.43$). This suggests that there is still room for improvement of apps that are currently on the market. Pertinently, the survey results guided the design of our two applications that are discussed in more detail in the next subsections.

App Design and Integration of Context

As mentioned above, spatial relationships are suitable for modelling with real-life objects [3, 34]. In several languages like German or Russian, the description of such relationships requires the correct grammatical case of nouns. Grammatical cases indicate the function a word serves in a sentence and typically require modification –inflexion– of determiners or nouns themselves[1]. We selected this concept for our *AR* app, as it lends itself to contextual instruction: there is only a limited number of rules to learn and real objects can be used to create learning scenarios. We further chose German as the target language, where the relationships are described with *preposition + article* constructs[2].



Object 1 *subject of sentence* + Verb *indicating position (stands, lies, is)* + Preposition + Article *multiple-choice options* + Object 2

Figure 4. Description of how sentences would be constructed for quizzes on case grammar, given the object labels and their spatial relationships.

For a better understanding of the general concept, below, we describe the challenges that learners experience and why this is the case. In German, nouns can either be masculine, feminine, or neuter, and there are only a few heuristics that help non-native speakers select the correct gender. In addition, German nouns have four cases. The cases are indicated by modifying the article and sometimes also the suffix of the noun (cf. Table 1). For instance, for the masculine word “Apfel”, “*der* Apfel” (“the apple”, nominative) designates use as a subject, “*den* Apfel” (accusative) defines a direct object, and “*dem* Apfel” (dative) corresponds to an indirect object. Thus, the article changes for each case. The differences may be small, but due to the relatively flexible word order in German sentences, the correct inflexion is sometimes crucial for understanding the meaning of a sentence.

The learning content is presented as quizzes based on two everyday objects, as shown in Figure 4. We chose to use quizzes instead of presenting correct sentences in order to increase engagement and because testing has been shown to improve recall in comparison to passive intake [31].

The *AR* and *Snapshot* apps differ in the way the quizzes are created. In the geo-location app, the users select objects in their actual surroundings, whereas the *Snapshot* app uses a predefined static photographic image that is independent of their environment (see below). Thus, the *Snapshot* app is *less embedded in the context*. Moreover, the *level of context interactivity is lower*. These two aspects are core strengths of using AR in education [34].

The AR Application

With our handheld-AR app, learners scan their environment using the phone camera. Whenever the AR system identifies tagged objects, labels are shown on top of the live camera preview. Learners can select tagged objects by tapping the corresponding labels. Once two objects have been selected, a short quiz is presented. This quiz takes the spatial relationship of the two objects into account. For instance, selecting an apple and a keyboard that are positioned next to one another produces a quiz, as shown in Figure 2[3]. Based on the estimated real-world coordinates, the app computes if an object 1 is *to the right, to the left, in front of* or *behind* an object 2. The

[1] See https://en.wikipedia.org/wiki/Grammatical_case for a more detailed explanation

[2] Research has shown that this topic is indeed an issue for learners, see [14, 32]

[3] English translation: “The apple is next to the keyboard”

Case	Example	Translation
Nominative	*der* Apfel (m), *die* Banane (f), *das* Obst (n), *die* Nüsse (f, plural)	the apple, the banana, the fruit, the nuts
Accusative	Ich lege *den* Apfel/*die* Banane/*das* Obst/*die* Früchte auf *den* Tisch.	"I put the . . . on the table."
Dative	Die Tasse ist neben *dem* Apfel/*der* Banane/*dem* Obst/*den* Nüsse*n*.	"The cup is next to the . . . "

Table 1. Cases in German: examples of how articles change depending on the function in the sentence and grammatical gender. *m*, *f*, and *n* indicate the gender: masculine, feminine, and neuter

case left–right is further varied by adding the neutral option *next to* at random intervals. Learners can move objects as they want, as their relative position is updated automatically. Thus, learners can try out different spatial constellations of object positions and observe how doing so modifies the quiz. By connecting the learning content to surroundings/images to a textual description of a situation, we further aim to increase associative strength by adding a second processing channel (see Related Work).

In this study, we integrated 20 different objects that were commonly found in an office (e.g., a notebook, a cup, a telephone). Even with this limited set, there is already a large number of possible combinations to explore, especially if objects are moved.

The app was implemented as an Android app using ARCore[4] for the AR functionality. The tracking was realised with ARCore's image marker system that also manages the continued tracking of markers even if they are out of sight. Each object was associated with an image label showing a different rock formation. This type of label is easily recognisable by a tracker system but relatively homogeneous in its visual appearance. The (relative) position of objects was based on the position in the real world as calculated by the ARCore toolkit.

The Snapshot Application

The exercises in the *Snapshot* consist of screenshots taken while using the geo-location app. Thus, the quizzes are exactly the same. However, using predefined still images instead of having learners select objects means that learners do not have to interact–and thus actively engage–with their environment. The only interaction we added was navigation between quizzes using arrow buttons on the left and the right side of the screen. The screenshots were generated through participant pairing during the study (see Procedure). Whenever the items used in a quiz were not both visible on the screen, we manually reproduced matching screenshots.

The *Snapshot* app was also implemented as an Android app and could be used with Android versions 4.1 and above.

EVALUATION

We compared our *AR* and *Snapshot* apps in a between-subject study. The between-subject design was chosen in order to compare learning. We collected quantitative data on the participants' usage behaviour and performance as well as qualitative data on their learning strategies, perceived usability, and general comments of the respective app. This study was designed in accordance with the Declaration of Helsinki (2013) for research involving human subjects, and all participants provided informed signed consent prior to participation. Our aim was to investigate (1) whether participants in the *AR* condition would perform better in subsequent tests on vocabulary and *preposition* + *article* constructs and (2) if they would find the *AR* app more enjoyable and useful and why.

Participants

Twenty-five learners of the German language participated in this study (8 males, 15 females, 1 undisclosed). Their ages ranged between 21 and 35 ($M = 27.2$, $SD = 4.1$). Our participants were mostly students or PhD students in different fields (e.g., medicine, business administration, and linguistics), but also included program and sales managers as well as a housekeeper. One additional participant provided only an early qualitative evaluation of the *AR* implementation that guided its final implementation. The next 24 participants were evenly and randomly assigned to either the *AR* or *Snapshot* condition. The participants' self-assessed German level on the Common European Frame of Reference[5] ranged from A1 (beginner) to C1 (advanced) with an almost even distribution in both groups. Eleven participants stated to have no prior experience with Augmented Reality and/or Virtual Reality applications. The others had tried at least one of these technologies. Participants were recruited via a university mailing list and social media. As compensation, they received 10€ in cash or a 10€ Amazon voucher once they had completed the final part of the study.

Apparatus and Materials

The apps for both conditions were run on a Pixel 3 XL running Android 9. The objects used as input for the quizzes in the *AR* condition were distributed freely in the room. We attached image markers to the objects, so they would be identified by the app when the Pixel's rear camera was pointed at them. Through the ARCore system on the phone, the real-world position of the objects was continuously tracked even when the objects disappeared from the camera's viewport. In both conditions, the screen content was captured during app usage. App interactions, such as object selections, were also logged. Questionnaires were answered on a laptop computer.

Before, directly after, and one week after using an app, we administered tests on vocabulary and case grammar skills. In the vocabulary test, participants were asked to translate the names of objects we used in the study from English to German. In the grammar tests before the study, their task was to select the correct choice of prepositions and articles for some example sentences with fictitious words. After the study, we showed a set of images that showed similar situations as the ones constructed in the study, but with fictitious objects and given grammatical gender. These images had to be described

[4] https://developers.google.com/ar, last accessed September 06, 2019

[5] See https://www.coe.int/en/web/language-policy/cefr, last accessed September 11, 2019

in full sentences (cf. Figure 5 for an example). The order of items in the recall and transfer test was randomised.

Procedure

Each session began with a welcome and briefing of the purpose and procedure of the study. We explained what data we would collect and gave participants some time to consult our data protection guidelines (in accordance with European data protection laws) before signing our consent form.

Following this, participants responded to a pre-study questionnaire (*Pre*) on their demographics, prior experience, and prior knowledge of vocabulary and case grammar. These measures were later used as a base for quantifying the learning.

Next, participants were introduced to the app that they would use to learn German case grammar and its features. In the *AR* condition, participants practised using it with objects that were not included in the subsequent learning phase until they were confident in using the app. After the introduction, the experimenter left the room and participants had 15 minutes to use the respective app. In the *AR* condition, they were able to explore the room, select objects, and solve quizzes at their own pace. In the *Snapshot* condition, they could solve quizzes generated from snapshots generated with the AR app by their paired participant. They could navigate between quizzes using pagination arrows and were thus able to choose which quizzes they wanted to answer and how often. All participants were allowed to end the learning phase early if they had already done everything they wanted. The maximum time of fifteen minutes was chosen based on iterative piloting and lies within the range of the duration of a microlearning session as defined in [20].

After using the app, the participants returned to the computer for a post-study questionnaire (*Post1*). They graded their experience using the System Usability Scale (SUS) [4], answered additional app-specific questions, and identified the situations where they would use the self-learning system assigned to them. Finally, they completed the same vocabulary recall test as they had performed before self-learning. In addition, we tested their transfer performance with the image description task described above. We explicitly added the recall and transfer tests after the qualitative feedback in order to introduce a short delay.

Long-term retention was also assessed with an additional test conducted one week after the main part of the study (*Post2*). For comparability, we administered the same type of tests that were presented for *Post1*. We only replaced the fictitious objects. A one-week delay for the post-test was chosen since this is a typical window for memory consolidation [15].

RESULTS

This section first describes the learnt performance and interaction measures. When relevant, qualitative statements of our participants are provided to support our interpreted findings. Next, we present the qualitative results on usability and recommendations for future applications. For the reporting, we assign the participant codes A1–A12 for *AR* and S1–S12 for their paired participants in *Snapshot*. The pilot participant is referred to as A0; we only use their data for qualitative analyses.

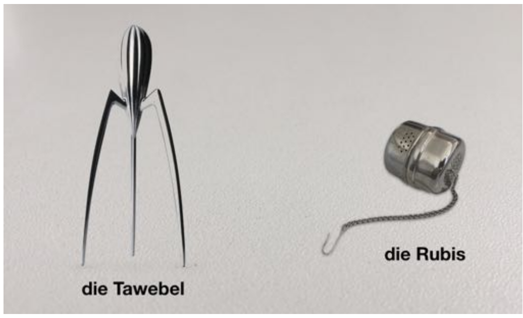


Figure 5. Example image that was used for the transfer test, depicting fictitious objects.

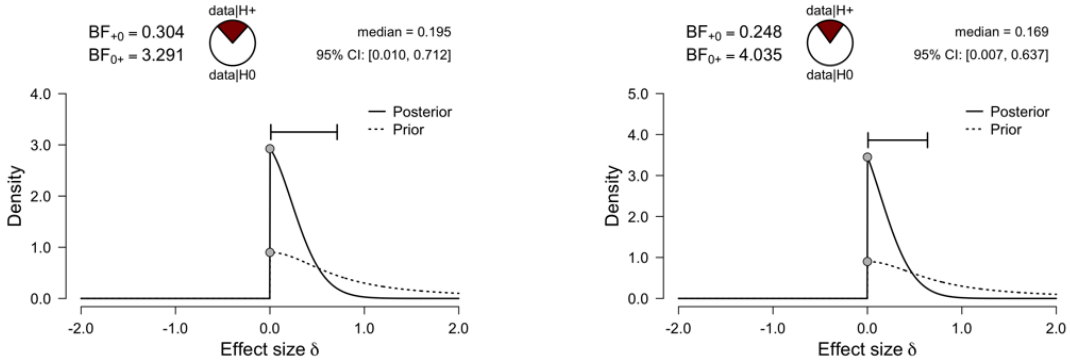
Our null hypothesis H_0 is that "*AR* is no different from *Snapshot*". Given previous work (e.g., [21]), our test hypothesis H_+ was that learnt performance with the *AR* app. In addition to frequentist analyses, we used one-tailed Bayesian t-tests for the comparison between the two apps. This allowed us to evaluate the likelihood of the null hypothesis relative to the test hypothesis, expressed as Bayes Factors (B_0+). A $B_0+ > 1$ indicates that the null hypothesis is more likely than the test hypothesis, while $B_0+ < 1$ would indicate the opposite. We adopted a default Cauchy prior with a width of $r = 0.707$ in our analysis [39] and we describe the resulting B_0+ using discrete terms of evidential strength (e.g., $B_0+ > 3$ is substantial evidence for the null hypothesis over the test hypothesis) [24].

Transfer Performance

Table 2 shows the participants' performance on the transfer tests. A response was considered correct when the article for the dative noun (the word after the preposition) was correct. To recap, test 1 was a multiple-choice test, while tests 2 and 3 were image description tasks and, hence, more difficult than test 1 as they tested grammar case production instead of recognition. In either case, a fictitious object, including its gender, was given for each test. There were no significant differences in the transfer tests for *AR* and *Snapshot*. Similarly, a t-test for independent samples showed no significant difference in the performance change from *Pre* to *Post2* between the conditions ($p = 0.75$). Indeed, a Bayesian t-test for independent samples provided substantial evidence in favour of the null hypothesis, namely that *AR* participants did not outperform *Snapshot* participants $B_{01} = 3.29$ (cf. Figure 6a). This trend was apparent even when we took potential ceiling effects into account by excluding participants that correctly answered all pre-test preposition questions—doing so returns a Bayes factor $B_{01} = 4.10$ for the remaining seven participants in *AR* and six in *Snapshot*.

Recall Performance

We also checked vocabulary recall in the pre-test and the two post-tests. Directly after the study, *AR* participants recalled a mean average of 1.1 more words ($SD = 4.1$) and *Snapshot* participants, 2.4 more words ($SD = 2.5$) than before using the respective app. After one week, the number of additional words in *AR* rose to 1.8 words on average ($SD = 3.3$) and dropped to



(a) Transfer Improvement

(b) Additional Words Recalled

Figure 6. Prior and posterior distributions for the likelihood ratio of H0 and H+, and corresponding Bayes Factors, based on performance improvement on the transfer test and learnt vocabulary after one week compared to performance before the study. Relative likelihoods are illustrated by the pie-chart. Created with JASP [23]

	Test	AR	Snapshot	p	B_{0+}
Ratio of correct sentences in the transfer tests	*Pre* (Recognition)	0.73 (0.33)	0.81 (0.24)	0.49	4.03
	Post1 (Production)	0.68 (0.45)	0.90 (0.24)	0.15	5.57
	Post2 (Production)	0.71 (0.36)	0.80 (0.30)	0.47	4.07
Total number of words recalled before the study	*Pre*	6.59 (4.03)	7.92 (4.96)	0.72	3.35
Additional words recalled in comparison to *Pre*	*Post1*	1.08 (4.14)	2.42 (2.50)	0.93	2.52
	Post2	1.82 (3.30)	2.25 (2.22)	0.49	4.04

Table 2. Mean and standard deviation of prior knowledge and learning scores. Here, B_{0+} shows the relative likelihood of no difference between the conditions compared to *AR* performing better than *Snapshot*

	AR	Snapshot	p	B_{01}
TT (s)	595.2 (202.5)	552.0 (316.0)	0.69	2.53
RT (s)	8.7 (2.6)	10.9 (5.0)	0.19	1.40
#Q	21.6 (13.5)	59.3 (70.5)	0.08	0.84

Table 3. Mean and standard deviation of interaction measures for the *AR* and *Snapshot* app: total usage time (TT) and time to select answer (response time, RT) in seconds; the number of quizzes taken (#Q)

2.3 in *Snapshot* ($SD = 2.2$), see Table 2. In spite of this suggestive trend, namely that *AR* could support better long-term retention, frequentist analyses revealed no significant performance differences (*Post1*, $p = 0.93$; *Post2*, $p = 0.49$). The Bayes factor for additional words recalled immediately after app usage provided anecdotal evidence for the null hypothesis ($B_{0+} = 2.52$). For long-term retention of learnt vocabulary, we obtained substantial evidence in favour of the null hypothesis ($B_{0+} = 4.04$).

Interaction Measures

Table 3 summarises interaction patterns across the two apps. The mean time spent using the app was similar in both conditions: 9 minutes and 55 seconds in *AR* and 9 minutes and 12 seconds in *Snapshot*. Once a quiz was shown, participants in the *AR* condition took around 8.7 seconds to select an answer, and *Snapshot* users around 10.9 seconds. Participants using the *Snapshot* app were able to take the same quiz several times. On mean average, they took 2.75 times more quizzes than in the *AR* condition and their rate of quizzes per minute was 2.65 times higher. We also performed two-tailed Bayes factors B_{01} on these measures, but only found anecdotal evidence in favour of the null hypothesis.

Judging by the interaction logs, A1, A3, A6, and A10 strategically reused one or two objects in subsequent quizzes, i.e., in at least a third of their quizzes. A9 and A11 never used the same objects twice in a row . The remaining participants did so at least once. In the *Snapshot* condition, five participants took quizzes in the order that they were presented, an additional five repeated or skipped single quizzes and two jumped back and forth several times. Based on this, we infer that participants utilised different learning strategies across the two apps, which resulted in comparable learning performance.

Qualitative Evaluation

The post-study questionnaire provided feedback on the functionality and design, the suitability of each app for self-learning, and future usage potential. In the following text that presents key qualitative statements, accompanying numbers in parentheses indicate the number of participants that support the statement.

Usability and Innovation

Both apps obtained similar SUS scores at $M = 80.6$ ($SD = 13.5$) for *AR* and $M = 80.8$ ($SD = 11.0$) for *Snapshot*, although SUS scores were more variable for *AR*. In line with transfer and recall test performance, the Bayes factor $B_{01} = 3.13$ shows substantial evidence for H_0 over H_1, i.e. for the SUS score not being higher for *AR*. The *AR* app was considered to be more cumbersome to use than the *Snapshot* app at $MD = 7.5$ ($SD = 2.39$) versus $MD = 10$ ($SD = 1.99$; lowest score 1, best score 10). In spite of this, participants said they would use

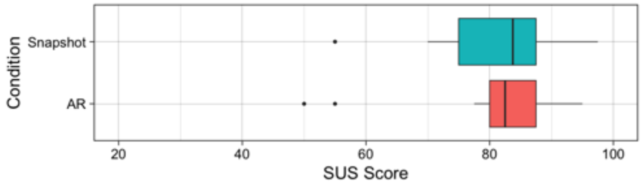


Figure 7. Box plots of SUS Scores, indicating the median, upper and lower quartiles, minimum and maximum values, and outliers.

the *AR* app slightly more frequently at $MD = 7.5$ ($SD = 2.17$) versus $MD = 6.25$ ($SD = 2.70$). This suggests that the form factor of the mobile device limited the usability of *AR*.

Overall, the *AR* app was described as *"fun"* (4), *"easy"* to use (3) *"interactive"* (2), *"interesting"* (2), and *"very innovative"* (1). A1, A2, A5, and A7 appreciated a self-learning experience that was integrated in the environment. Participants also raised several points for improvement in *AR*, namely the precision of relative location (2), and the speed and performance of the marker detection (2).

The *Snapshot* app was generally regarded as being easy to use (5). The participants liked that grammar and vocabulary could be practised at the same time (2), that the images could be used as a support or for learning the words (4), that they received immediate feedback (2), and that they could repeat words with different prepositions (1). However, self-generated images by their paired participants were not always clear enough (4) and symbols, e.g. *"arrows"*, could have helped *"to illustrate the spatial relationship between two objects"* (S7). S9 and S12 would have liked a gamification element in *Snapshot*, e.g., points or levelling up.

Participants in both conditions said that the apps would benefit from presenting a larger number of different prepositions (5) and showing the gender on labels (5). A3, S4, and S10 suggested adding rule explanations.

Interaction Strategies and Learning

We asked participants how they selected objects for quizzes in order to investigate how strategies might have differed across individuals and between the app conditions. In the *AR* condition, four participants reported that they chose objects randomly. Others shared their specific strategies: A3 changed *"the spatial arrangement of objects to see which preposition the app would suggest"*, A10 checked *"every preposition, [. . .] the correct answer and [then checked] that on further labels"*. A0 and A11 explicitly selected objects they could not name or where they did not know the gender.

In the *Snapshot* condition, only a few participants commented on their usage strategies. S7 said she *"read the sentence and then again I looked at the pictures, this way helped me to choose the right answer"*, whereas S10 found *"the picture [. . .] rather irrelevant to the question, I can choose simply based on the choices"*. S11 relied on the feedback for *"[remembering] the particular case and [applying] that knowledge on next similar cases"*.

These free responses on learning strategies suggest that the current participants were familiar with traditional learning methods (i.e., *Snapshot*), which elicited more consistent methods. In contrast, *AR* allows for more diverse self-learning and might have benefited from a more guided experience, at least until users are more familiar with this medium.

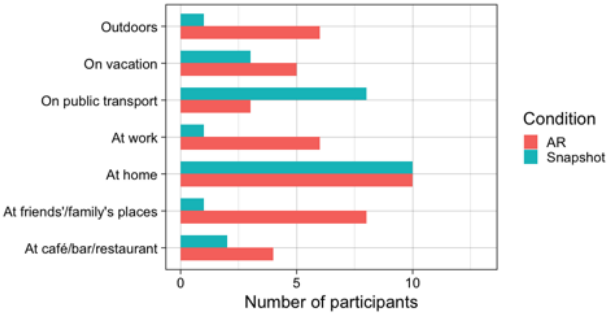


Figure 8. Summary of potential use-locations and the number of participants who indicated they would use either the AR or Snapshot app in these locations.

Usage Situations and Scenarios

As an outlook into future application, we asked participants if and in what situations they would use such a language learning app. A large majority stated they would likely use the respective app a few times a week or every day: 9/13 in the *AR* condition and 7/12 in the *Snapshot* condition. Figure 8 gives an overview of situations where participants would use the apps. A notable difference in the response was the belief that the *AR* app would be used more often outdoors or in social environments, e.g., at a friend's place, whereas the *Snapshot* app seems to be more suitable for usage on a commute. For instance, S2 said she *"can learn and practise in [. . .] fragmented time, like on public transport"*.

Scenarios where the *AR* app would be useful were *"immediately in real-like situations"* (A0) and by *"[making] a game out of being able to identify objects and [getting] instant results"* (A2). A12 added it might help him *"get grips on words and phrases of a language quicker [sic], but only during free time"*.

Several participants considered the *Snapshot* app a useful addition to language courses, e.g., S2: *"I think it is a good complementary [sic] to textbooks. It is more flexible and convenient to use"* and S11: *"not very convenient for learning new material. I would use it to test my knowledge or to revise"*.

DISCUSSION

Context and Learning Performance

In contrast to [21] and [22], our study shows that self-learning performance in the *AR* condition is not necessarily better than in the control condition for self-learning. There are at least two main reasons for this. First, users demonstrated greater usage familiarity with *Snapshot* and more usage diversity with *AR*. It is possible that *AR* generates higher learning performance if it represents a more constrained experience. Nonetheless, this could unnecessarily limit the exploratory and interactive experience that *AR* affords. Second, the *Snapshot* represented self-generated learning material with photo-realistic content, which sets a high bar for the control condition. Thus, providing high-quality media of a familiar context could suffice in providing for an immersive and engaging learning environment, without incurring the cost of constructing highly realistic virtual environments.The high-quality media in *Snapshot* may have enabled the construction of an imagined context, i.e., context may have been utilised for learning across both conditions. Therefore, unlike in [8] and [12], we were not able to observe performance differences between context-based content and non-context-based content.

Interaction Behaviour

Our work highlights that *AR* can come at a certain cost to users. For example, AR can cause a high cognitive load [11]. Although we did not explicitly measure cognitive load, the AR app was rated to be more complex and cumbersome to use than the Snapshot app. Thus, handheld *AR* may indeed have led to slightly higher cognitive load in comparison to *Snapshot*.

Further differences were revealed between the interaction behaviour of the two conditions. For example, slower response time in *Snapshot* and the lower number of quizzes answered per minute in *AR* suggest that our participants applied different learning strategies in the two conditions (besides the requirements imposed by the app design, e.g., the mandatory selection of objects in *AR* before a quiz appears). More specifically, *Snapshot* users obtained similar results even though they answered more quizzes and could concentrate on the exercises instead of exploring the room. This may indicate a more superficial level of processing compared to *AR*.

In both conditions, we also found substantial *individual* differences in usage behaviour, both in the activity logs and qualitative statements. Some participants were more arduous than others, took more quizzes than others and reflected on their choices. We believe that, especially in informal, self-directed learning, such differences will always exist and technology is therefore not the only factor that influences learning outcomes. This also means that there is not a “one-size-fits-all” approach, and certain challenges will need to be addressed before AR becomes universally useful.

Usage and Application Contexts

Our qualitative evaluation showed that learners would be willing to use both the *AR* and the *Snapshot* app on a regular basis, but that the primary use cases would be different. A typical usage scenario for the AR app could be a situation where someone arrives at an unknown place that contains unknown objects. Users might want to explore their surroundings, i.e., get on-demand information on their location. They would do so on their own or together with someone else. Thus, the focus is on getting new information and usage in a social context seems to be acceptable. The *Snapshot* app, on the other hand, would most likely be used when the user is alone, in a familiar environment or on the go, and when learners intend to revise previously learnt material. Since such situations only have a limited overlap, we see a large potential for combining *AR* and *Snapshot* approaches. For instance, content for *Snapshot* learning could be generated through screenshots taken during exploration with *AR* and then revisited at a later moment.

Limitations and Future Work

The results revealed limitations of our implementation and areas that afford further research and development, which relate to both technical aspects as well as the study design.

The current implementations were limited by the number of tagged sample objects. This was a necessary limitation in the current study because it relied on a small marker set that provided a high level of stability and reliability, which allowed for adequate preparation of learning content. In the future, we intend to extend our AR app by using computer vision libraries for markerless object recognition that would allow users to generate content automatically (see, WordSense [38]). A more fine-grained analysis of spatial relationships and dealing with object occlusion in the captured image would make it possible to quiz more complex object relationships, e.g., “between” or “underneath”. The ability to track object movements would also extend the self-learning of case grammar to include accusative statements as well. Besides language self-learning, the current apps are easy to adapt for teaching other concepts that could benefit from a similar instruction technique. Teaching cross-multiplications, for instance, could be generated by arranging objects of different sizes and shapes.

It also has to be mentioned that handheld AR can easily cause fatigue when the phone is held for a longer time. Since we primarily designed our AR app for frequent, but short sessions during idle moments in a busy schedule, we did not expect this to be a major issue. However, some participants did mention that doing so was cumbersome. This issue could be readily resolved by the use of smartglasses as well as other devices that do not require manual manipulation.

From an evaluation perspective, the effect of context and interactivity could be further isolated by (1) restricting the study to a task defined in more detail and with a fixed number of quizzes, and (2) a fixed amount of time to reduce variance between participants. In our study, we decided against this because prescribing an object selection sequence in AR would have conflicted with its potential for exploration and interactivity, and would have made the study setup less realistic by interfering with the participants’ natural learning style (see also [34]).

The current study motivates the need to perform self-learning studies with more participants to uncover diverse individual learning strategies and, also, over a long period of time. Prolonged use could increase usage familiarity, allowing users to generate a more efficient learning style with innovative learning methods. In addition, our study specifically targeted adults who had at least one proficient language, with the interest or need to learn new languages, but with limited spare time; our sample of young adults (age: 21-35) fulfilled these criteria and included participants with different backgrounds and levels of experience with AR technology. This work could be extended to include children and older adults, who can be expected to have different requirements for both system design and content. For instance, users with less experience in using smartphones may find it challenging to select small objects on the screen. Some users might also require more playful approaches to increase motivation.

Finally, we acknowledge that the learning effect could have been higher with different implementations. There is a large design space for AR apps, and there were several aspects that could have (and have been) approached differently. However, we do not expect this to have had an impact on the generalisability of our comparison, as the design of both apps was based upon the same decisions.

DESIGN CONSIDERATIONS FOR AR LEARNING

During this work, we identified several key issues that should be resolved in order to achieve the full potential of AR for self-learning. Moreover, we discussed the situations where AR could be a promising solution and other situations where traditional approaches might be just as effective. Below, we summarise points to consider when designing AR (language) learning applications.

Content for AR Learning – An important hallmark of using AR is that learning is less controlled than when predefined learning content is used instead. Consequently, it cannot be guaranteed that all elements of a curriculum will be covered and it is more difficult to schedule spaced repetitions [25] to improve recall. On the other hand, AR presents the potential for scalable (unlimited) content that is tailoured to the user's personal experience. Even without personalisation algorithms, learners can individually select what they are interested in and leave out what they already know. Therefore, we suggest using AR for learning experiences that are intrinsically motivated and shaped by learners themselves, rather than strictly following a curriculum. For example, besides practising language skills, a potential use case is that of a tourist exploring a city to learn about its history.

Interaction in Context – We have proposed that AR is particularly suitable for learning material that involves the visualisation of a spatial or temporal relationship. However, we assume from current findings that the benefit of AR is greatest when such specific contextual features cannot be equalled in virtual or simulated contexts. AR can also be costly because exploration takes more time than simply presenting prepared exercises. Therefore, an AR solution may only be justified if it contributes an additional quality to an interactive learning experience.

Content Generation and Variation – Motivation typically declines over time [30]. Thus, AR learning systems will need to, just like any other learning technology, provide enough content variation to keep learners interested and avoid high drop-out rates. In combination with content generation algorithms, AR has the potential to provide unlimited content. However, content development must also take into account that the learner's skill level will change over time and exercises need to be adjusted accordingly.

Choice of AR Technology – In this work, we opted for an AR solution that is based on smartphones because they are comparatively cheap and widely available. Moreover, they are suitable for short periods of interaction that can be performed with one hand. If only a small field of view is needed, current smartglasses could serve as an alternative, especially if bimanual interaction is desired. In the future, we expect smartglass technology to improve further and thus become even more relevant for ubiquitous learning.
In addition to the hardware, software for object detection and tracking is required. Marker-based object detection has high accuracy but requires objects to be tagged before they can be used. Markerless object recognition largely increases the potential interaction space but often lacks robustness. Existing services also tend to provide only a small set of pre-trained categories[6] and are optimised for static images instead of dynamic environments. Future systems that allow for real object interaction could rely on state-of-the-art real-time object tracking systems as described in [40].

Comprehensive Solutions – For future designs, we suggest a hybrid system that adapts to the learner's situation: AR would be used to explore new content on-demand and a traditional system like flashcards for restudying the knowledge acquired in AR in quiet moments. Flashcards created using AR could also be extended with curriculum-based information.

CONCLUSION

This paper investigated the impact of interactivity and embeddedness (i.e. two major advantages of AR) on language learning. To this end, we developed a novel AR application teaching German vocabulary and the description of spatial relationships of objects as well as a static app presenting the same learning content as static images. Our in-between study showed that learning performance was not higher when using the contextual and interactive AR app. We discuss several drawbacks and strengths of AR technologies in comparison to traditional learning methods. Nonetheless, we argue that there is still a large potential for AR or hybrid solutions, especially for exploratory learning in unknown environments or social contexts.

ACKNOWLEDGEMENTS

This research is financially supported by the European Union's Horizon 2020 Programme under ERCEA grant no. 683008 AMPLIFY and the German Research Foundation (DFG) within the project C06 of SFB/Transregio 161.

REFERENCES

[1] Mary Ainley, Suzanne Hidi, and Dagmar Berndorff. 2002. Interest, learning, and the psychological processes that mediate their relationship. *Journal of Educational Psychology* 94, 3 (2002), 545–561. DOI: http://dx.doi.org/10.1037//0022-0663.94.3.545

[2] Niels Van Berkel, Denzil Ferreira, and Vassilis Kostakos. 2017. The Experience Sampling Method on Mobile Devices. *Comput. Surveys* 50, 6 (2017), 1–40. DOI: http://dx.doi.org/10.1145/3123988

[3] Mark Billinghurst and Andreas Duenser. 2012. Augmented Reality in the Classroom. *Computer* 45, 7

[6] For example, Tensorflow Lite Object Detection only comes with 80 pre-trained categories (https://www.tensorflow.org/lite/models/object_detection/overview, last accessed December 20, 2019)

(July 2012), 56–63. DOI: http://dx.doi.org/10.1109/MC.2012.111

[4] John Brooke and others. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[5] John Seely Brown, Allan Collins, and Paul Duguid. 1989. Situated cognition and the culture of learning. *Educational researcher* 18, 1 (1989), 32–42. DOI: http://dx.doi.org/10.3102/0013189X018001032

[6] Carrie J. Cai, Anji Ren, and Robert C. Miller. 2017. WaitSuite: Productive Use of Diverse Waiting Moments. *ACM Transactions on Computer-Human Interaction* 24, 1 (March 2017), 1–41. DOI: http://dx.doi.org/10.1145/3044534

[7] Dorothy M. Chun and Jan L. Plass. 1996. Effects of Multimedia Annotations on Vocabulary Acquisition. *The Modern Language Journal* 80, 2 (1996), 183–198. DOI: http://dx.doi.org/10.1111/j.1540-4781.1996.tb01159.x

[8] David Dearman and Khai Truong. 2012. Evaluating the implicit acquisition of second language vocabulary using a live wallpaper. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*. ACM Press, Austin, Texas, USA, 1391. DOI:http://dx.doi.org/10.1145/2207676.2208598

[9] Tilman Dingler, Dominik Weber, Martin Pielot, Jennifer Cooper, Chung-Cheng Chang, and Niels Henze. 2017. Language learning on-the-go: opportune moments and design of mobile microlearning sessions. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '17*. ACM Press, Vienna, Austria, 1–12. DOI:http://dx.doi.org/10.1145/3098279.3098565

[10] Guler Duman, Gunseli Orhon, and Nuray Gedik. 2015. Research trends in mobile assisted language learning from 2000 to 2012. *ReCALL* 27, 02 (May 2015), 197–216. DOI: http://dx.doi.org/10.1017/S0958344014000287

[11] Matt Dunleavy and Chris Dede. 2014. Augmented Reality Teaching and Learning. In *Handbook of Research on Educational Communications and Technology*, J. Michael Spector, M. David Merrill, Jan Elen, and M. J. Bishop (Eds.). Springer New York, New York, NY, 735–745. DOI: http://dx.doi.org/10.1007/978-1-4614-3185-5_59

[12] Darren Edge, Elly Searle, Kevin Chiu, Jing Zhao, and James A. Landay. 2011. MicroMandarin: Mobile Language Learning in Context. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 3169–3178. DOI: http://dx.doi.org/10.1145/1978942.1979413

[13] Min Fan, Uddipana Baishya, Elgin-Skye Mclaren, Alissa N. Antle, Shubhra Sarker, and Amal Vincent. 2018. Block Talks: A Tangible and Augmented Reality Toolkit for Children to Learn Sentence Construction. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, Montreal QC, Canada, 1–6. DOI: http://dx.doi.org/10.1145/3170427.3188576

[14] Helga Fervers. 1983. *Fehlerlinguistik und Zweitsprachenerwerb: Wie Franzosen Deutsch lernen*. Number 62. Librairie Droz.

[15] Paul W. Frankland and Bruno Bontempi. 2005. The organization of recent and remote memories. *Nature Reviews Neuroscience* 6, 2 (Feb. 2005), 119–130. DOI: http://dx.doi.org/10.1038/nrn1607

[16] Sarah Garcia, Ronald Kauer, Denis Laesker, Jason Nguyen, and Marvin Andujar. 2019. A Virtual Reality Experience for Learning Languages. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, Glasgow, Scotland Uk, 1–4. DOI: http://dx.doi.org/10.1145/3290607.3313253

[17] John K. Gilbert, Astrid M.W. Bulte, and Albert Pilot. 2011. Concept Development and Transfer in Context-Based Science Education. *International Journal of Science Education* 33, 6 (April 2011), 817–837. DOI: http://dx.doi.org/10.1080/09500693.2010.493185

[18] Duncan R Godden and Alan D Baddeley. 1975. Context-dependent memory in two natural environments: On land and underwater. *British Journal of psychology* 66, 3 (1975), 325–331. DOI: http://dx.doi.org/10.1111/j.2044-8295.1975.tb01468.x

[19] Catherine Regina Heil, Jason S. Wu, Joey J. Lee, and Torben Schmidt. 2016. A Review of Mobile Language Learning Applications: Trends, Challenges, and Opportunities. *The EuroCALL Review* 24, 2 (Sept. 2016), 32–50. DOI: http://dx.doi.org/10.4995/eurocall.2016.6402

[20] Theo Hug. 2005. Micro learning and narration. In *Fourth Media in Transition conference, May*. 6–8.

[21] Adam Ibrahim, Brandon Huynh, Jonathan Downey, Tobias Hollerer, Dorothy Chun, and John O'donovan. 2018. ARbis Pictus: A Study of Vocabulary Learning with Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics* 24, 11 (Nov. 2018), 2867–2874. DOI: http://dx.doi.org/10.1109/TVCG.2018.2868568

[22] María Blanca Ibáñez, Ángela Di Serio, Diego Villarán, and Carlos Delgado Kloos. 2014. Experimenting with electromagnetism using augmented reality: Impact on flow student experience and educational effectiveness. *Computers & Education* 71 (Feb. 2014), 1–13. DOI: http://dx.doi.org/10.1016/j.compedu.2013.09.004

[23] JASP Team. 2019. JASP (Version 0.11.1)[Computer software]. (2019). https://jasp-stats.org/

[24] Harold Jeffreys. 1961. *The theory of probability*. Oxford University Press.

[25] Nate Kornell. 2009. Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology* 23, 9 (2009), 1297–1317. DOI:http://dx.doi.org/10.1002/acp.1537

[26] Mariusz Kruk. 2015. Practicing the English Present Simple Tense in Active Worlds:. *International Journal of Computer-Assisted Language Learning and Teaching* 5, 4 (Oct. 2015), 52–65. DOI: http://dx.doi.org/10.4018/IJCALLT.2015100104

[27] Tsun-Ju Lin and Yu-Ju Lan. 2015. Language Learning in Virtual Reality Environments: Past, Present, and Future. *Journal of Educational Technology & Society* 18, 4 (2015), 486–497. https://www.jstor.org/stable/jeductechsoci.18.4.486

[28] Richard Mayer and Richard E Mayer. 2005. *The Cambridge Handbook of Multimedia Learning*. Cambridge university press.

[29] Detmar Meurers, Ramon Ziai, Luiz Amaral, Adriane Boyd, Aleksandar Dimitrov, Vanessa Metcalf, and Niels Ott. 2010. Enhancing Authentic Web Pages for Language Learners. In *Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications (IUNLPBEA '10)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 10–18. DOI: http://dx.doi.org/10.5555/1866795.1866797

[30] Paul R. Pintrich. 2003. A Motivational Science Perspective on the Role of Student Motivation in Learning and Teaching Contexts. *Journal of Educational Psychology* 95, 4 (2003), 667–686. DOI: http://dx.doi.org/10.1037/0022-0663.95.4.667

[31] Henry L Roediger III and Jeffrey D Karpicke. 2006. The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science* 1, 3 (2006), 181–210. DOI: http://dx.doi.org/10.1111/j.1745-6916.2006.00012.x

[32] Margaret Rogers. 1984. On major types of written error in advanced students of German. *IRAL-International Review of Applied Linguistics in Language Teaching* 22, 1 (1984), 1–40.

[33] Maria V. Sanchez-Vives and Mel Slater. 2005. From presence to consciousness through virtual reality. *Nature Reviews Neuroscience* 6, 4 (April 2005), 332–339. DOI: http://dx.doi.org/10.1038/nrn1651

[34] Marc Ericson C. Santos, Arno in Wolde Lübke, Takafumi Taketomi, Goshiro Yamamoto, Ma. Mercedes T. Rodrigo, Christian Sandor, and Hirokazu Kato. 2016. Augmented reality as multimedia: the case for situated vocabulary learning. *Research and Practice in Technology Enhanced Learning* 11, 1 (2016). DOI: http://dx.doi.org/10.1186/s41039-016-0028-2

[35] Dov Schafer and David Kaufman. 2018. Augmenting Reality with Intelligent Interfaces. In *Artificial Intelligence - Emerging Trends and Applications*, Marco Antonio Aceves-Fernandez (Ed.). InTech. DOI: http://dx.doi.org/10.5772/intechopen.75751

[36] M P Strzys, S Kapp, M Thees, P Klein, P Lukowicz, P Knierim, A Schmidt, and J Kuhn. 2018. Physics holo.lab learning experience: using smartglasses for augmented reality labwork to foster the concepts of heat conduction. *European Journal of Physics* 39, 3 (mar 2018), 035703. DOI:http://dx.doi.org/10.1088/1361-6404/aaa8fb

[37] Andrew Trusty and Khai N. Truong. 2011. Augmenting the web for second language vocabulary learning. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. ACM Press, Vancouver, BC, Canada, 3179. DOI: http://dx.doi.org/10.1145/1978942.1979414

[38] Christian David Vazquez, Afika Ayanda Nyati, Alexander Luh, Megan Fu, Takako Aikawa, and Pattie Maes. 2017. Serendipitous Language Learning in Mixed Reality. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17*. ACM Press, Denver, Colorado, USA, 2172–2179. DOI: http://dx.doi.org/10.1145/3027063.3053098

[39] Eric-Jan Wagenmakers, Jonathon Love, Maarten Marsman, Tahira Jamil, Alexander Ly, Josine Verhagen, Ravi Selker, Quentin F Gronau, Damian Dropmann, Bruno Boutin, and others. 2018. Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic bulletin & review* 25, 1 (2018), 58–76.

[40] Qiang Wang, Li Zhang, Luca Bertinetto, Weiming Hu, and Philip H.S. Torr. 2019. Fast Online Object Tracking and Segmentation: A Unifying Approach. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1328–1338.

[41] Marjorie A. Zielke, Djakhangir Zakhidov, Gary M. Hardee, Jithin Pradeep, Leonard Evans, Zahra Lodhi, Kevin Zimmer, and Eric Ward. 2018. Exploring medical cyberlearning for work at the human/technology frontier with the mixed-reality emotive virtual human system platform. In *2018 IEEE 6th International Conference on Serious Games and Applications for Health (SeGAH)*. IEEE, Vienna, 1–8. DOI: http://dx.doi.org/10.1109/SeGAH.2018.8401366