PROMOTING AUTONOMY FOR LANGUAGE LEARNING POWERED BY DATA-DRIVEN METHODS AND LEARNER-CENTRED DESIGN

MARIANA AKEMI SHIMABUKURO

A thesis submitted to the
School of Graduate and Postdoctoral Studies in partial
fulfillment of the requirements for the degree of

Doctorate in Philosophy
in
Computer Science

Faculty of Science
University of Ontario Institute of Technology (Ontario Tech University)

December, 2024

Thesis Examination Information

Subimtted by: Mariana Akemi Shimabukuro

Doctorate in **Philosophy**

Thesis Title: Promoting Autonomy for Language Learning Powered by Data-driven Methods and Learner-centred Design

An oral defense of this thesis took place on September 30^{th}, 2024 in front of the following examining committee:

Examining Committee:

Chair of Examining Committee	Dr. Shahram Heydari
Research Supervisor	Dr. Christopher Collins
Research Committee Member	Dr. Michael Glueck
Examining Committee Member	Dr. Jeremy Bradbury
Non-voting Advisor	Dr. Michael Glueck
Internal Examiner	Dr. Pejman Mirza-Babaei
External Examiner	Dr. Carrie Demmans Epp, University of Alberta

The above committee determined that the thesis is acceptable in form and content and that a satisfactory knowledge of the field covered by the thesis was demonstrated by the candidate during an oral examination. A signed copy of the Certificate of Approval is available from the School of Graduate and Postdoctoral Studies.

ABSTRACT

This thesis explores innovative, data-driven methods to enhance autonomy in language learning. While it presents interdisciplinary work, the main focus is exploring the learner-centred design of three applications aiming to promote self-regulated behaviours among language learners.

The following projects are presented: (1) H-Matrix offers a visualization tool for analyzing cross-linguistic features in large learner corpora. This tool supports users in identifying transfer effects and other patterns in learner data, aiding in developing targeted instructional strategies and promoting linguistic awareness for learners. (2) Card-it is a dynamic flashcard generator leveraging finite-state morphology to support the acquisition of Italian verb morphology. This application incorporates expert feedback and learner evaluations to refine its design, ensuring it effectively supports Italian verb conjugation training and morphological awareness. (3) LangEye captures the learner's environment to create situated, personalized, context-aware vocabulary learning experiences. It leverages learner-curated curriculum and micro-learning opportunities, and it integrates generative AI to enable context personalization to enhance the relevance and retention of new vocabulary in the target language.

Through user studies, including expert evaluations and learner interviews, this research indicates the efficacy of these applications in fostering autonomous, self-regulated language learning. It provides practical, data-driven solutions that empower learners to take control of their language acquisition journey. These findings contribute to technology-enhanced language learning (TELL) and human-computer interaction (HCI) for education.

Keywords: learner-centred design, language learning autonomy, data-driven methods

DECLARATION

I hereby declare that this thesis consists of original work of which I have authored. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize the University of Ontario Institute of Technology (Ontario Tech University) to lend this thesis to other institutions or individuals for the purpose of scholarly research. I further authorize University of Ontario Institute of Technology (Ontario Tech University) to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research. I understand that my thesis will be made electronically available to the public.

The research work in this thesis that was performed in compliance with the regulations of Research Ethics Board under REB Certificate numbers #14426, #17317, and #17447.

Oshawa, Ontario, Canada, December, 2024

Mariana Akemi Shimabukuro

STATEMENT OF CONTRIBUTIONS

I declare that this dissertation has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where states otherwise by reference or acknowledgement, the work presented is entirely my own.

I acknowledge that the prototypes presented in Chapters 5 and 6 were primarily implemented by undergraduate researcher assistants Shawn Yama and Deval Panchal. In both cases, I managed and designed the application development while the research assistants were the primary software developers.

I also acknowledge that the work on Card-it, Chapter 5, resulted from collaborating with the Computational Linguistics PhD Candidate, Jessica Zipf. Zipf created the Finite-State Morphology (FSM) analyzer used in the project, and both as a linguist and an Italian speaker, they actively informed the design of Card-it. The user studies were designed and executed in our joint efforts.

Some ideas and figures have appeared previously in the following publications:

[1] Mariana Shimabukuro and Christopher Collins. "Cross- Linguistic Word Frequency Visualization for PT and EN." In: *Proceedings of the IEEE Conference on Information Visualization (Poster)*. 2019.

[2] Mariana Shimabukuro, Jessica Zipf, Mennatallah El-Assady, and Christopher Collins. "H-Matrix: Hierarchical Matrix for Visual Analysis of Cross-Linguistic Features in Large Learner Corpora." In: *Proceedings of the IEEE Conference on Information Visualization (short papers)*. 2019.

[3] Mariana Shimabukuro, Jessica Zipf, Shawn Yama, and Christopher Collins. "Evaluating Classroom Potential for Card-it: Digital Flashcards for Studying and Learning Italian Morphology." In: *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA) (Short Paper + Demo)*. 2023.

[4] Jessica Zipf, Mariana Shimabukuro, and Christopher Collins. "Card-IT: a Dynamic FSM-based Flashcard Generator for Learning Italian Verb Morphology." Talk in European Association of Computer Assisted Language Learning - EUROCALL 2021, online. 2021. URL: https://whova.com/web/euroc1_202108/Agenda.

ACKNOWLEDGEMENTS

Learning and education have always been a priority for my family, and my job growing up was to study, which was a great privilege. I first have my parents, Ezio Seisi Shimabukuro and Marilda Serrano Shimabukuro, to thank, they made uncountable sacrifices so I could focus on my education, *muito obrigada*. I want to express my deepest gratitude to Eva and Victor Sawal, when I moved to Canada, they took me in their home as well as part of their family.

I would likely to thank all my teachers, professors, and mentors for inspiring me to learn how to learn. I thank my examiners, Dr. Pejman Mirza-Babaei and Dr. Carrie Demmans Epp, for their honest and valuable feedback on my work. I thank the members of my committee, Jeremy Bradbury, Allyson Eamer, and Michael Glueck for the support and guidance along my journey. Particularly, I thank my supervisor Chris Collins, who I always say is a rockstar researcher, but is also an extremely caring and supportive supervisor. Chris, I have seen you raise your standards for excellence, and rediscover your passion for research and mentorship many times throughout the years; the brightness you have radiates and inspires everyone — I can only hope, that even if for a moment, to shine as bright as you do.

I thank my colleagues for sharing words of support and for cheering for my success, specially Randy Fortier. Randy mentored me and helped me find balance between my doctoral and teaching workload. I am very grateful for the opportunity Greg Crawford gave me to focus on my doctoral research over teaching duties, this professional development time was crucial to finalizing my work.

Along the many years of research to complete my dissertation, I have been luck to have worked with great labmates to share the sorrows and the joys of graduate studies. Hrim Mehta, Riley Weagant, Kevin Desousa, Matt Chan, Zixin Zhao, Feiyang Wang, Gaby Dias, Mohammed Ahmed, Ben Leung, and my sister Tania Sanai Shimabukuro, are some of the many who are part of the Vialab and have shared the burden of graduate school with me. I thank you. Rafael Veras was also instrumental on my journey as researcher, he mentored and inspired me to chase after his footsteps. Beyond grad-

uate school, I really appreciate Rafa and Brit Kondo, who are dear family-like friends to me. I also thank my dear friend and collaborator, Jessica Zipf, for all the work and meaningful moments we shared.

At last, I would like to thank my spouse, Jay, for all the years he has stood by me, he provided me with the support I needed. Jay was patient and supportive, he provided feedback on my work, he listened to all my (too many) rants, he helped me pick up the pieces after rejections, he cheered me through milestones, he shared the burden of graduate studies with me in all aspects. Jay is literally part of my dissertation. Jay, I could not have made this far without you, or even if I could, truthfully, I simply would not want to. Thank you.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AI Artificial Intelligence

API Application Programming Interface

AR Augmented Reality

CALL Computer-Assisted Language Learning

CV Computer Vision

DDL Data-Driven Learning

ESL English as Second Language

FL Foreign Language

FSM Finite-State Morphology

GALL Google-Assisted Language Learning

GPT Generative Pre-trained Transformer

HiVR and High-immersion Virtual Reality

ICLE International Corpora of Learner of English

LLM Large Language Model

MALL Mobile-Assisted Language Learning

ML Machine Learning

MT Machine Translation

NLP Natural Language Processing

L1 First Language

L2 Second Language

LiVR Low-immersion Virtual Reality

REB Research Ethics Board

SDT Self-Determination Theory

SLA Second Language Acquisition

TELL Technology-Enhanced Language Learning

TL Target Language

UI User Interface

UOIT University of Ontario Institute of Technology

UX User eXperience

VR Virtual Reality

XR Extended Reality

Part I

WHAT I NEEDED TO KNOW

As a language learner, I know that learning a new language requires effort. I know the daily struggles to communicate in a foreign language – the embarrassment, frustration, and disappointment in oneself while trying to master a foreign language are all well-known to me. While working on this thesis and learning about how languages are learned and the different ways to support learners, I've come to accept, embrace, and respect myself as a non-native English speaker. In this part of the thesis, I share what I learned from the literature and introduce what I have identified as my contributions to the field so that others can confidently embrace and take control over their language learning journey.

1

INTRODUCTION

Language and communication are key to personal, academic, and professional success. Let's take English as an example; English is known as the "global language" because it is widely adopted by many countries even when it is not the country's native language. On top of that, one-quarter of the world's population is already fluent or competent in English [33]. For many, speaking English as a Second Language (L2) increases their chances of accessing career and education opportunities. This indicates the need for understanding the Second Language Acquisition (SLA) process and developing tools to support L2 learners in different facets and stages of their journey.

While the term L2 can refer to one's literal second language acquired or in acquisition after their First Language (L1) or mother tongue, in the context of this document, L2 refers to any additional language or Foreign Language (FL) being learned as an adult. The reason is that the process of early language acquisition (when we learn our L1, which includes all languages we learn as children in the case of bilingual upbringings) differs from how we learn languages later in life. For example, I grew up in a monolingual household in Brazil, and my L1 is Brazilian Portuguese; when I was around 16 years old, I started learning communicative English as my L2. Meanwhile, my partner, Jay grew up in a bilingual household in Canada, and he equally used Tagalog and English to communicate with his family and schoolmates, respectively. We met during his senior year at university, and Jay started learning to speak Brazilian Portuguese. In this second example, Jay has two L1s, Tagalog and English, acquired simultaneously in his childhood; Brazilian Portuguese is his L2. Independently, the process of acquiring a L2 for Jay and I differed greatly from when we learned our L1s.

Typically, learning a L2 requires intentional effort from learners who are investing their time, energy, and in many cases, money into language courses and resources. For example, when learning English as my L2, I mainly attended an English course that took me from pre-intermediate (A2) to upper-intermediate (B2) proficiency level

in approximate 1600 hours of in-class time plus self-studying and homework completion time. Brazil is a predominantly monolingual country, and there were few opportunities to practice English outside the classroom. Therefore, the experience that really helped me to solidify and improve my communicative skills was 12 months of exchange studies in Canada. On the other hand, Jay learned Brazilian Portuguese as his L2 mainly from taking daily Duolingo lessons, listening and singing to Brazilian songs, and practicing speaking with me for 12 months until he could spend five weeks in Brazil. While his L2 was not well developed to participate in complex conversations, he was exposed to the culture and different native speakers of Brazilian Portuguese, which was fundamental to increasing his vocabulary and listening skills. After his trip, Jay added watching Brazilian TV shows such as soap operas and reality shows (i.e., MasterChef *Brasil*) with Brazilian Portuguese subtitles to his study routine. It took him approximately three years to achieve communicative fluency [25], where he developed a level of linguistic competence that allows him to somewhat express feelings, explain plans, and paraphrase stories using his L2. In his case, being able to write and pursue academic skills in Brazilian Portuguese was not a goal, so his study routine was based on informal and communicative language. Yet, his method was learner-centered, allowing him to study and learn relevant and functional language competence [85, 118]. From these two examples alone, we can affirm that one's journey toward their L2 acquisition can differ depending on their goals and set of tools available, which can be tuned depending on the situation; either way, it requires discipline, grit, and consistency.

Thus, motivation is the key factor for L2 acquisition. What motivates the individual can vary from cultural appreciation for the language and its natives to utilitarian reasons such as educational or professional requirements. In my case, learning English was first an academic and professional requirement to increase my career opportunities; and secondly, it was needed to pursue exchange studies opportunities during undergraduate school. Those goals motivated me to be consistent in my English course throughout the years. However, for Jay, learning Brazilian Portuguese was not part of a lifelong or professional goal. Due to his relationships, Jay's goal was to communicate with monolingual native speakers of Brazilian Portuguese in practical and leisurely conversations and to become well-versed in

their culture — it is possible that Jay's motivation to pursue his second language studies would have been affected if his relationships had been severed, but not necessarily. Either way, motivated learners are self-determined to succeed in studying and learning their L2 to their desired level of proficiency or fluency. These self-determined individuals seek strategies that allow them to trail their way down a path leading to being an autonomous language learner, where being autonomous means that one takes control over their learning [114, 115]. Autonomous learners aim to plan, implement strategies and tasks, and monitor and assess their learning goals [91].

Fortunately, there are many software applications and resources, both free and paid, which can support second language learners and speakers. Among the most popular applications are Rosetta Stone[1], Duolingo[2], italki[3] and Grammarly[4]. Even though this wide range of applications serves different types of learners, based on a preliminary investigation, the existing tools do not well support more proficient speakers or advanced learners. I interviewed 10 non-native English speakers (see Chapter 2), who were mostly advanced learners who had already acquired the second language. From the interviews, I identified that L2 immersion via both explicit (i.e., talking to native speakers) and implicit (i.e., watching TV or movies) practice is key to language learning success. In their opinion, the major issue with existing Computer-Assisted Language Learning (CALL) and Mobile-Assisted Language Learning (MALL) software applications alike is the lack of flexibility and control over the curriculum. This finding aligns with the needs of autonomous language learners. As learners progress towards full autonomy, their degree of autonomy becomes more granular and varies depending on their language competencies. Advanced learners, who are further along the autonomy continuum, require greater flexibility from software applications to choose their study content and track their progress, leading to increased control over their learning process. It is important to consider that learner autonomy is not absolute and develops over time, resulting in varying needs for different learners [67].

Data-driven methods are a way to make software applications more adaptable for learners is by using machine learning techniques like

1 https://www.rosettastone.com/
2 https://www.duolingo.com/
3 https://www.italki.com/
4 https://www.grammarly.com/

personalized content recommendations. This approach uses a student's previous interactions and experiences to tailor the content's topic, difficulty level, and presentation style to their needs. While this is a viable solution for flexibility, some may view it as a passive way of controlling one's learning. On the other hand, learner-driven methods, such as choosing what to learn next and, having the ability of creating or curating suggested lessons based on personal interests, accessing external data and resources that provide digestable evidence for informed choices, are active ways of curating content. The mix of the data-driven and learner-driven methods can empower learners and help them progress in their autonomy as learners.

This thesis presents a collection of learner-centred projects that explore and address language learners' different aspects and needs.

1.1 UNDERSTANDING SECOND LANGUAGE LEARNING

SLA is the process of learning a language other than your native language(s). In the process of acquiring other languages, a learner often exhibits cross-linguistic effects. *Cross-linguistics Effects*, also called *Transfer Effects*, can be structural, grammatical, semantic, or other linguistic rules from the learner's first language being applied to the second language [102, 103]. For instance, a native Portuguese speaker can misuse the English verbs *lost* and *miss* because both translate into the Portuguese verb: *perder*.

Some transfer effects can be identified by comparing language produced by native and non-native speakers. If we have, for example, writing produced by native speakers as the baseline for English, distributional information such as the frequency of words used in non-native writing samples can indicate differences between the two groups of writers. More specifically, word frequency is used to investigate overproduced and underproduced words, word choice, word form, and collocations between two languages [38, 70, 79, 139]. James (2013) classifies transfer effects into two broad categories: grammatical and acceptability errors.

Grammatical errors: where clear grammatical rules are broken. Example:

"I will *driven* to the airport next week".

Acceptability errors: satisfies grammatical rules, but it is uncommon. Usually, the sentence would sound odd, and after a split second, a native speaker would understand that it is acceptable. This category includes errors related to word choice, word form, or collocation. Examples:

"This a strong/*powerful* tea."
"I feel very *confusing* this morning."
"I *ate* a medicine pill."

Thus, language learning experts and linguists know that we are affected by the languages we speak. From more noticeable and recognized traits such as pronunciation and grammar, to cultural aspects that influence personality and sense of humour. Not all linguistic influence is negative. For example, is commonly accepted that if one speaks a romance language such as Italian, learning a second romance language such as Spanish might be "simpler" due to their shared origin and typology. While L2 learners are used to many challenges in language acquisition, self-determination can be affected when they lack competence. In other words, they do not have the necessary tools or strategies to continue improving towards their desired-self. Therefore, learners who understand how their L1 may affect their L2 acquisition and production are better prepared to find suitable strategies to accept and overcome them.

Chapter 4 reports on H-Matrix, a visualization designed to increase cross-linguistic awareness of non-native English speakers. H-Matrix allows people to understand how their L1 influences their English (L2), by showing the most common mistakes someone with a certain L1 is likely to make when learning English as L2. Awareness of cross-linguistic effects empowers learners to properly plan and implement mitigation strategies suitable to their L1 towards language learning autonomy [6, 94].

To help learners to overcome negative transfer effects, H-Matrix, a hierarchical matrix visualization for linguists, language teachers, and learners to analyze transfer effects using a collection of English learners' essays (learner corpus). For example, Korean (L1) learners of English (L2) are likely to make mistakes regarding pronoun-verb conjugation; that happens because Korean verb forms are independent of the pronoun. On the one hand, this observation can either guide English teachers to focus their lessons on pronoun-verb agreement

when they have Korean (L1) learners. On the other hand, Korean (L1) learners of English (L2) would become aware of this transfer effect which can help them mitigate it appropriately.

1.2 TOWARDS LEARNER-CENTRED METHODOLOGY

In the past, learning a language was deemed as something natural that we have innate abilities to do; later, this was proven to be true for our first language(s) [43, 45]. Currently, we know that learning a new language is no easy task. It requires countless hours of study and meaningful practice – similarly to the "10,000 hours" rule presented by Gladwell [55]. This realization put to the test the traditional models of language learning and teaching, which started to shift from teacher-centred to learner-centred [43, 69, 123].

Ultimately, the contextual combination of teacher-centred and learner-centred practices is the most desirable, where the teacher may be responsible for controlling the outline, monitoring and assessing progress whilst constantly responding to the learners' needs and interests; meanwhile, the learner is responsible for communicating their needs to teachers and looking for meaningful practice and opportunities for implicit learning outside the classroom or curriculum [2, 43]. For that reason, CALL and MALL applications grew in popularity propelled by increased computer literacy and Internet and mobile device availability[43]. Besides the proven beneficial multimedia approach to language learning extending and supplementing the classroom experience, CALL also promotes autonomy to learners — which is essential to learner-centred language learning.

The aforementioned traditional CALL applications (i.e., Duolingo) provide learners with structured learning but little control over the content or order of the presented materials. Chapter 5 describes the design, implementation, and evaluation of Card-it — a digital flashcard application for studying Italian verb morphology. Card-it can be used by learners of Italian in both traditional and self-taught use cases. It was designed first to support individual learners of Italian independently of frequenting a class. But, to enable classroom adoption it provides support to teachers to control the content and to track their students progress using quizzes. All users can access functionalities that allow them to plan which linguistic features (i.e., verb, tense, mood, etc) to focus on. Therefore, learners can browse public

flashcard collections (not just those created by their teacher) and create their own collections for personalized study plans. Card-it also supports self-assessment using two different types of quizzes. Thus, Card-it can provide both freedom and structure for learning Italian verb morphology, which promote autonomy for language learners.

1.3 LANGUAGE LEARNER AUTONOMY WITH SELF-CURATED CONTENT

While intuitively, we may think of *autonomy* as a synonym of independence, that is not how we should interpret it in the context of CALL. In their Ph.D. work, Blin [15] has an in-depth discussion around the definition of learner autonomy from language teaching literature where an autonomous learner takes responsibility and control over their own learning without teacher intervention — this could include choosing what to learn, how to learn, practice, monitor, and how to evaluate success in learning [91]. While independent of teacher intervention, the success of an autonomous learner depends on situational and psychological aspects, such as appropriate tools and resources, and motivation and self-discipline, respectively. Therefore, the relationship of applications, methods, or solutions built to promote autonomy and learners seeking autonomy should be interdependent on each other. In CALL and MALL, this interdependence is enabled via interactivity, which allows not only freedom and choice over materials but also structure and control of the learners' experience [98].

The popular language learning apps such as Duolingo and Memrise are very sucessful at the application of adaptive and personalized learning technology (i.e., recommendation algorithms, machine learning models, etc), which means they can suggest and present exercise types or lessons based on the learner's progress and engagement with the apps. Furthermore, these apps offer some L2 immersive content that embed dialogue between native speakers in the form of shorts clips or animations. On the one hand, their content and experience is curated and personalized. On the other hand, the learner has little to no active control over the content itself.

In addition, the professionally curated content presented in these apps is accurate and well informed in terms of pedagogical practices such as program map, and curriculum building. However, more often than not, learners do not have the power to choose and pick their own

content which will make their learning experience more relatable and modular.

From this perspective, Chapter 6 describes the design, implementation, and evaluation of LangEye, a mobile application to support autonomous learners with L2 immersion and vocabulary acquisition. LangEye uses the mobile camera to identify and label objects in the learner's real environment, it saves the labels and photos of those objects collected in the learner's 'memories'. Then, it uses the 'memories' to generate mini-activities where the learner can assess their vocabulary retention. By using the learner interaction with the real world to generate the study content, LangEye provides freedom and control over the study sessions while storing and facilitating self-assessment opportunities for learners.

LangEye leverages a data-driven architecture using methods to automatically extract objects and label them from images. It also applies a Large Language Models (LLMs) to provide word definition and usage sample sentences. This data-driven architecture enables the implementation of many source and target language pairs. Considering that all the content in LangEye is learner-generated and curated, this method relies heavily in the data quality provided by the state-of-the-art Application Programming Interfaces (APIs) employed in the implementation. Thus, the trade-off between learner self-curated content is the difficulty to gauge the accuracy and quality of the content itself which can inadvertently teach learners inaccurate knowledge.

1.4 SUMMARY

In summary, this thesis presents the following contributions:

- H-Matrix — An application to visualize and analyze linguistic features in a large learner corpus to identify transfer effects (Chapter 4).
- Card-it — A digital flashcard application to enable teachers and learners to teach and study Italian verb morphology (Chapter 5).
- LangEye — A mobile application to promote L2 immersion, contextual learning, and vocabulary acquisition (Chapter 6).

In Chapter 2, I present and discuss key interdisciplinary concepts related to language learning autonomy and preliminary interview with non-native English speakers which leads to the identification for flexibility in supportive tools to empower language learners toward autonomy. In Chapter 3, I discuss the different aspects of how CALL can support learner autonomy, and in Chapter 7 I conclude this dissertation by outlining the contributions, their impact in the current literature, and the opportunities they create for future research.

2

LANGUAGE LEARNERS, AUTONOMY, AND SELF-REGULATED BEHAVIOUR

Due to the interdisciplinary nature of this dissertation, the sections in this chapter introduce the necessary background to draw out autonomous language learners' goals explored in future chapters. Section 2.1 defines and presents the main concepts related to autonomous language learning, and Section 2.2 discusses self-regulated behaviour displayed by autonomous language learners.

In the early stages of my doctoral studies, I wanted to understand other non-native language learners; including their views, needs, practices, and struggles regarding acquiring English as Second Language (L2) and the current technology-enhanced language learning landscape. Thus, I designed and ran a semi-structured interview with 10 participants to survey methods and applications to support non-native English speakers. Section 2.3 reports on the analysis and results of these interviews. The results helped me to scope my work towards empowering language learners in their search for autonomy.

2.1 AUTONOMOUS LANGUAGE LEARNING

In different contexts, *autonomy* can be used as a synonym for *independence*. However, the two terms do not have the same definition. Independence can be defined as "the freedom from control, influence, support, aid, or the like, of others" [29]. In contrast, autonomy is defined as "the ability to make your own decisions about what to do rather than being influenced by someone else or told what to do" [28]. Thus, autonomy does not require independence. One can be autonomous by taking actions and responsibility based on their own values with free will and still accept support or depend on others. Another related context is in psychology's Self-Determination Theory (SDT), where autonomy is a psychological need tied to self-motivation and mental health. The self-determination theory outlines three inherent human needs: autonomy, competence, and relatedness [114, 115]. This theory poses that we, as humans, want to improve and achieve high levels

of proficiency in what we do; we want to attain certain skills that will allow us to develop a sense of belonging and control over our lives — leading to more self-motivation and better well-being (mental health). While SDT refers to human or individual needs in general, it helps us to understand the root of motivation or the display of self-regulated behaviour also displayed in language learning, specifically.

Little [91] defined *learner autonomy* as the ability to take charge of one's own learning; learners can achieve that by taking control over planning, implementing, monitoring, and evaluating their own learning. Similarly, Littlewood [92] defined the core attributes of learners' autonomy as their ability and willingness to make choices independently; these attributes were the base of a task-driven framework for language teaching methodologies to increase competence as communicators, learners, and individuals. For example, a person might be able to perform a task by either having access to the correct resources or having an adequate set of experiences; on the other hand, if they are not willing (motivated), the task is inefficient.

Therefore, humans' self-determination is directly related to autonomous language learning and one's ability to self-regulate behaviours for learning. Ushioda [134] explains that motivation can be intrinsic (internal or "within" oneself) or extrinsic (external factors). For example, intrinsically motivated learners may participate in an activity for pure interest, enjoyment, and satisfaction, whereas extrinsically motivated learners act for utilitarian benefits such as passing an exam. Thus, autonomous language learners are, by definition, motivated learners and/or individuals [93, 133].

Naturally, L2 teaching, classroom, and pedagogical methodologies evolved to adopt concepts derived from autonomous learning, such as autonomous communication and L2 production, learner confidence, and learning strategies to motivate learners to extend their studies beyond the classroom and their teacher's guidance; therefore, a shift from teacher-centred to learner-centred methodologies toward autonomy [2, 43, 69, 123].

According to Dörnyei and Ushioda [41], L2 motivation lies in improving oneself, where we desire to become better, by acquiring a new language in this case. This will drive us toward planning and implementing strategies that bring us closer to our ideals-selves or our idealized future selves. Dörnyei and Ushioda [41] believes that this effect is strongly related to globalization and the effects of the

Internet and social media. Their views align with the current perception of English as a global language tied to travelling, business, social status, and professional success [33]. For example, this can motivate learners to associate their ambition of being a world traveller or a successful businessperson with learning English. More recently, the proliferation of Korean as a foreign language also demonstrates the power of globalization, where the enhanced contact with the popular culture (i.e. K-pop and K-drama) of the target language significantly increased the motivation to learn Korean as L2 [87]. In addition, a study by Chan and Chi [26] showed that young adults have five motivational factors, pop culture, career, personal achievement, academic exchange, and interest in foreign languages and cultures.

Thus, exposure to the target language's culture causes learners to envision or create a version of themselves that belongs to the L2-speaking group. Motivation involves the desire to reduce the gap between one's actual self and the projected behavioural standards of the ideal/ought selves. This is what led Dörnyei and Ushioda [41] to conclude that

> ...future self-guides provide incentive, direction and impetus for action, and the sufficient discrepancy between these and the actual self initiates distinctive self-regulatory strategies with the aim to reduce the discrepancy.

In other words, our future ideal self-guides, or the person we want and imagine us to be in the future, give us a baseline of comparison to be matched through behaviour changes such as self-regulatory (autonomous) learning strategies [65]. For example, the behaviour change may include setting up a study routine, signing up for a foreign language course, researching a set of online resources and applications, or a combination of those that may help the learner advance toward their ideal selves. In a different context, someone who wants to have a healthier lifestyle will likely list goals such as a target weight, an aesthetic inspiration, or general health targets; the idealistic vision of their relatively healthier version will motivate them to set actionable plans (i.e., enroll in a gym, reduce white sugar on a diet, etc).

Therefore, one can say that the first step toward language learning autonomy is motivation (intrinsic or extrinsic) to learn a language. Additionally, while a learner has not achieved their ideal self or desired level of proficiency, they will likely need appropriate means to

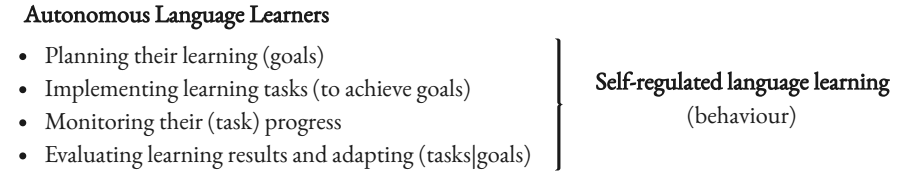


Figure 2.1: Autonomous language learners strategies are considered self-regulated behaviours. Their strategies can be described as planning and implementing learning goals, and assessing and tracking their learning progress.

increase their competence. This indicates an opportunity to provide language learning applications and tools to empower language learners at different steps in their journeys.

2.2 SELF-REGULATED LANGUAGE LEARNING

As seen in Figure 2.1, autonomous learners have the ability to self-regulate. They implement strategies that allow them to plan, implement, monitor and evaluate their learning. However, language learners do not begin their journey as completely autonomous learners, utilizing all strategies simultaneously. Instead, they gradually develop autonomy by incorporating one or more of these strategies over time. This was shown in a framework *learner autonomy continuum* proposed by Hu and Zhang [67] that describes the different levels of dependence and autonomy in a continuum instead of categories (Table 2.1).

Table 2.1: The *learner autonomy continuum* proposed by Hu and Zhang [67].

Innate needs	Needs thwarted	→	→	Needs satisfied
Motivation/behavior	Non-self-determined	→	→	Self-determined
Types of behavioral regulation	External regulation	Introjected regulation	Identified regulation	Integrated regulation
Levels of learner autonomy	Dependence	Relative dependence	Relative autonomy	Autonomy

As displayed in Table 2.1, the *learner autonomy continuum* has four levels: (1) dependence, (2) relative dependence, (3) relative autonomy, (4) autonomy; each related to the SDT's four types of behavioural regulation that allow learners to internalize extrinsic motivation, which is aligned with the varied degrees of need for satisfaction and self-determination. The four behaviour regulation types are (1) external regulation, (2) introjection, (3) identification, and (4) integration. Thus, the more basic needs (autonomy, competence, and relatedness) are satisfied, the more self-determined and motivated one is; consequently, they will produce more self-regulated behaviours. To visualize this re-

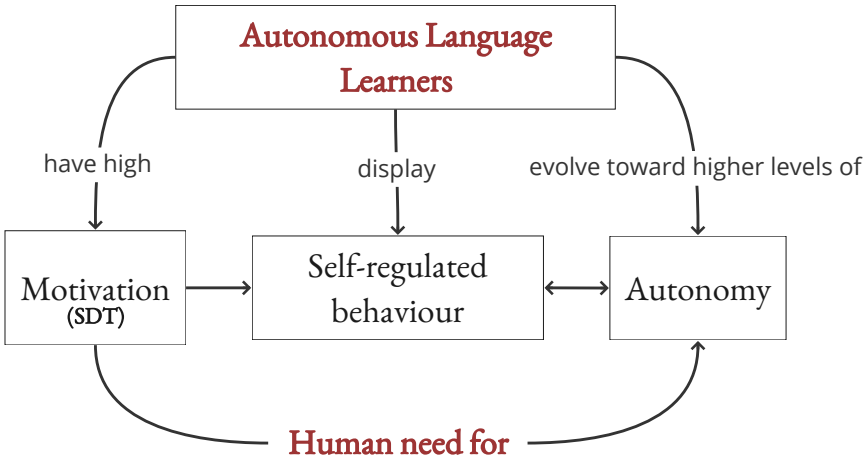


Figure 2.2: Autonomous language learners are motivated individuals who direct themselves to higher levels of autonomy. The higher the autonomy or motivation, the more self-regulated behaviour is displayed by learners. These relationships naturally maintain themselves largely due to the presence of motivation and SDT dictating humans' need for autonomy.

lationship, I present Figure 2.2. It illustrates the autonomous learners' ability to self-regulate strong dependency on motivation.

While motivation is undoubtedly a key part of autonomous language learning, the work presented in this thesis is not focused on language learning motivation. Rather, motivation is assumed and required from language learners as the stakeholders. The goal of the work presented is to empower language learners as they traverse their way in the *learner autonomy continuum* and look for tools for self-regulation strategies (Figure 2.1). Thus, various methods, strategies, tools, and resources are needed to support learners in traversing the *learner autonomous continuum* toward fully autonomous learning. Wherever the learner is in the continuum, they should receive appropriate support. Table 2.2 lists each project in this dissertation and how they support the different self-regulated behaviours displayed by autonomous learners.

Table 2.2: Self-regulated language learning strategies supported by the different thesis projects.

Self-regulated behaviour	**H-Matrix (Chapter 4)**	**Card-it (Chapter 5)**	**LangEye (Chapter 6)**
Plan their learning (goals)	Increase cross-linguistic awareness.	Teacher- or Self-curation of cards, decks, and collections; Easy selection of different verbs, tenses, and other morphological features	Self-curated vocabulary collection.
Implement learning tasks to achieve goal.	—	Study mode with easy and quick feedback; Quiz mode for self-evaluation.	Exploration of contextual vocabulary and sample sentences; Review memories activity composed of three phases with varied levels of support.
Monitor progress toward learning goals.	—	Analytics features of quiz scores via classroom.	Analytic features via the Achievements.
Evaluate results and adapt tasks | goals	Understand cross-linguistic effects.	Cards and decks can be re-organized or expanded to support adaptation of goals.	Pseudo-spaced-repetition algorithm to help with memorization; Endless curation of vocabulary.

2.3 INTERVIEW WITH NON-NATIVE ENGLISH SPEAKERS

I interviewed 10 people to understand common challenges they faced as advanced or fluent speakers of English as a second language. These people use English in their daily and academic routines as residents of Canada. I observed that those with more second language immersion also have higher self-perception of fluency. Participants with higher fluency also reported preferring English over their mother tongue, except when communicating with family members.

The participants' main English-related issues can be classified under first language attrition, acceptability mistakes, and cross-linguistic effects. Even though these issues do not impede them from communicating using English, they may cause discomfort and a loss of communication confidence. Pronunciation was unanimously mentioned as an important factor for fluency and also as one of the hardest skills to master as a non-native English speaker.

Most participants were interested in supportive technologies for both acquiring and supporting language. However, most of them did not currently use any applications for either writing or learning support. When questioned about the reasons why, participants said that the lack of flexibility in the current applications was the main reason. They would like to pick and choose specific skill sets to practice because they have already acquired the main language features as advanced or fluent speakers.

As reviewed by the Research Ethics Board (REB) on file #14426, Appendix A shows the interview questions, which were tabulated in 3 sections: **language background**, **language usage**, and **language support**. I took notes while performing the interviews which were audio recorded and manually transcribed.

DATA ANALYSIS METHODOLOGY The semi-structured interview open-ended questions were coded into categories following commonly mentioned themes in the participants' answers. The original supporting quotes were linked to these categories. For some questions, I used structured codes or open coding depending on the question type. For close-ended questions such as "How often English is used in your home country?", I used a fixed code set of "rarely", "occasionally", "frequently", and "very frequently". For open-ended questions where participants described their experience, tips, and methods for learn-

ing, I used an open coding thematic analysis approach. I aggregated these based on codes and extracted quotes to support the discussion of results.

2.3.1 *Participants*

The participants were recruited via printed posters spread around the university campus. I interviewed a total of 10 participants in the period of January 23rd to February 15th, 2018. They were from 8 different home countries, varying between undergraduate and graduate levels across multiple fields of study. The next sections describe more information about the participants' background, language competency, and adopted technologies.

2.3.2 *Language Background*

The majority of the participants were female (7 out of 10). The participants were university students, where 7 were undergraduate students and 3 were graduate students. They were pursuing degrees in diverse areas such as Computer Science, Engineering, and Health Sciences.

Table 2.3 shows the home countries of the participants, the first language spoken, and, if applicable, any other languages they know besides English. Participants from India and Ethiopia mentioned the other languages they know are regional languages spoken in the local community where they are from; it is common to have different languages and dialects in many countries.

In addition to home country, Table 2.4 shows the amount of time the participants had been living in Canada at the moment of the interview. Half of the participants have lived in Canada for a long time (5 or more years). Meanwhile, the other half had been in Canada for a shorter period (1 month and 3 years). Table 2.4 also shows the ages participants started to learn English, their age at the moment of the interview, and if they reported themselves as fluent in English, the age at which they became fluent. Participants' age varied between 20 and 26 years old, and the methodology or environment where they primarily studied English were at bilingual or immersive schools, English as Second Language (ESL) Schools, or in Canada when enrolled in the Canadian elementary or middle school.

Table 2.3: Countries participants are originally from, their first languages and any languages they can speak besides English.

Home Country	**First Language**	**Other Languages***
Argentina	Spanish	–
Brazil	Portuguese	Spanish
Colombia	Spanish	French
Ethiopia**	Amharic	Urumic and Haramic
India**	Hindi	Bengali
India**	Hindi	Oriya and Asamese
Indonesia	Indonesian	–
Sri Lanka**	Tamil	–
Sri Lanka**	Tamil	–
Vietnam	Vietnamese	–

* Languages other than English.

** English is either also an official language or, it was declared as daily used by participants.

Table 2.4: Tabulated background data from study participants showing their ages at the moment of interview, at the start of their English acquisition studies, at which age (if) they considered themselves fluent. We can also see the country they came from, how long they lived in Canada, and the method they used to study English.

Age start learning	Age fluent	Home Country	Study method	Time in Canada (# of years)	Age
3	18	Indonesia	Bilingual school	2	21
6	13	Vietnam	Bilingual school	3	20
11	15	India	Bilingual school	9	20
3	15	Ethiopia	Bilingual school	2	20
6	20	Colombia	Bilingual school	5	23
8	10	Sri Lanka	In Canada	10	21
8	13	Sri Lanka	In Canada	10	20
7	13	India	In Canada	10	21
8	N/A	Argentina	ESL School	0	26
18	N/A	Brazil	ESL School	0	24

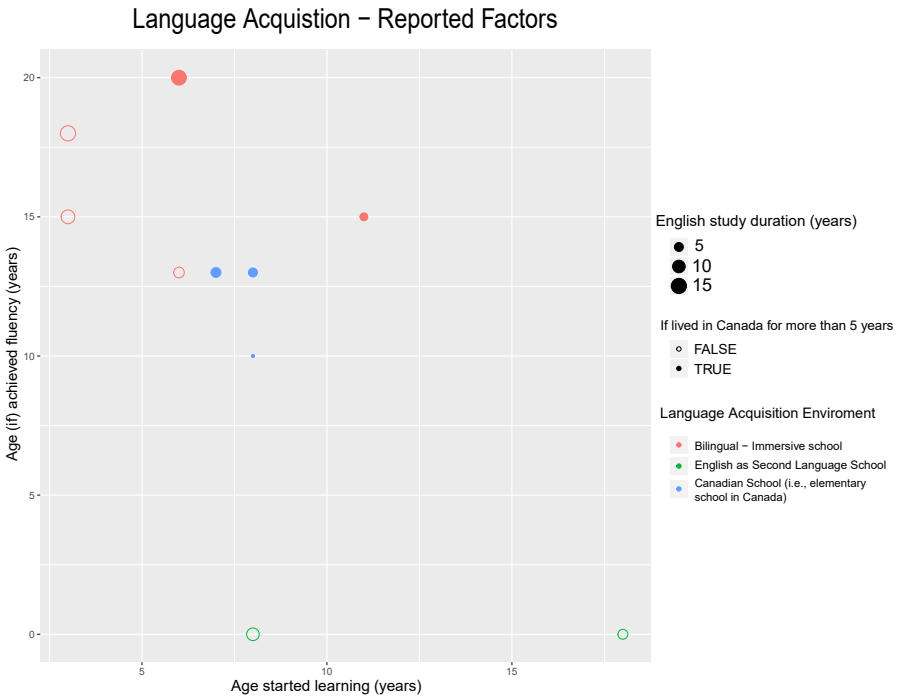


Figure 2.3: This chart shows some reported factors related to the participant's background and fluency. We can observe the age of participants at the start, and when (if) they became fluent in English. If participants did not consider themselves fluent, their age is represented as 0. The colours indicate the reported English learning environment. The size is the number of years they studied or had been studying for. The filled circles are participants who had lived in Canada for at least five years.

Figure 2.3 is a visualization of some of the reported factors from Table 2.4. The factors encoded are age at the start of studies and at fluency, method of study, amount of time studying, and stay in Canada. In the chart, we can see that those who have lived in Canada for more than 5 years (5 out of 10) have all self-declared as fluent despite not having studied English for as long as the contrasting group. This might indicate that language immersion is a strong factor in language learning and/or self-perception of fluency.

2.3.3 *Language Usage*

Table 2.3 shows that half of the participants came from a country where English is also an official language, or they mentioned English is used daily, often mixed with their mother tongue. The other half of the participants declared that English is a commonly learned second

Table 2.5: The common resources and tips mentioned by participants.

Tips for learning English	# participants
Consume English media	6
Observe native speakers and copying	2
Practice with native speakers, immersion	9

language, but it is not used outside of specific companies or academic setups.

Besides the two participants who did not consider themselves fluent in English, all stated English as their preferred language unless when communicating with their family members. First language attrition is characterized when a second language learner loses fluency in their first language. It does not mean the learner can no longer communicate in their native language, but it could cause a loss of vocabulary, among other effects. When asked if they noticed any first language attrition effects, again, all the participants who reported being fluent in English also answered they noticed first language attrition effects.

Table 2.5 shows the tips given by participants when asked if they had any particular methods used to learn and improve their English outside school. They also made the following specific comments regarding immersing themselves in North American culture and in English using digital media.

CONSUME MEDIA IN ENGLISH. Participants advised learners to take advantage of what they like to do, read books in English, watch TV shows with and without subtitles and try to copy the characters, the way they talk and the words they use.

"I find you need to do something you like, you enjoy. So, I like reading books. So, I try to read English books. I like watching TV, at first with subtitles, so I understand what they are saying into my language... And then, I start watching it without subtitles[...]I watch and learn they way they talk. But, sometimes they say some bad words... I ended up picking up bad words too... hmm... because that's the way they speak English, not the way I speak. But I want to learn how natives speak English, so I pick it up."

OBSERVE NATIVE SPEAKERS AND COPYING. Still related to consuming English media, one can specifically use media to observe and copy how natives communicate. Participants listen to music and sing to acquire native pronunciation and enunciation.

"Sing songs, English songs, karaoke... so I did... and try to match what you listen... the language has enunciation of sentences is different from my first language... but for speaking... you have to shed calories, you have to feel tired... I used to spend half an hour alone with music, repeating the words... I was so tired... like when you exercise, but it works."

Participants stressed the importance of learning the native speakers' culture to copy how they express themselves through language. They advised that one should try to learn the culture of English-speaking countries by talking to natives and watching TV from those places.

"I had to learn the culture with the language... because you can say a perfectly correct... grammatically correct sentence... but people can tell when it's not a native speaker's sentence, that sentence is too perfect... when you learn a new language, you have to learn the culture that comes with the sentence... like... the word Google... that's not correctly, but we know the culture... so the sentence it will make sense..."

PRACTICE WITH NATIVE SPEAKERS, IMMERSION. In general, participants mentioned the importance of exposure and immersion in English, as well as practicing with native speakers as much as possible. They said that one should immerse oneself in an English-speaking environment whenever they can; avoid translation; and try to think in English.

"... think in English and not your language, and translate... cuz it doesn't work... it was hard for to understand you can't literally translate word by word... I tried to think in English, watch movies, or listen to music... talk to natives if you can.".

One should get exposed to different native accents, to learn other ways to say things, and understand the language.

"... in English... the accents is very different. So, it's very important you have contact with all natives and non-natives... it's very important... for learning and for communicating..."

To learn how to express one's feelings and emotions in English, one should write a diary.

"...I write a diary...what happens in my day, my struggles, so it helps me to improve my English...I have difficulty to express myself in English, as in my feelings and moods which is very natural in my mother tongue. So one can write a diary to learn self-expression."

Lastly, they advise learners not to limit their exposure to their own native group of friends.

"...you should make an effort...on campus, I see people who cannot speak proper English...change yourself...if only speak to your friends from the same language you cannot improve...they will make the same mistakes you make...speak to other friends even if you are made fun [of]."

2.3.4 *Language Support*

I asked participants about their main difficulties when learning and currently using English (Table 2.6). Regarding writing, the main issues mentioned were punctuation, spelling, and "non-sense" sentences, which participants defined as sentences that sound unnatural to natives despite being grammatically correct. Accents were the most frequent issue related to listening, participants complained that real accents were not introduced when they were learning English. For speaking in English, pronunciation was the biggest issue pointed out by the participants. They also remarked on how pronunciation can affect their confidence level as fluent speakers.

As shown in Table 2.6, vocabulary expansion, literal and idiomatic translations, expressions, and slang usage were also mentioned as difficulties related to writing, listening, and speaking. However, these issues seem more related to cultural references and culture assimilation than language acquisition.

Table 2.7 shows what software, methods, or resources participants use to support writing; we can also observe their comments about idiomatic translations and expressions. Only 4 participants have used dedicated software for writing support. Most of them said to only use default spell checkers offered by whichever writing editor they might use. The mentioned applications were Grammarly Premium (1), Grammarly Free (2), and online websites that offer human tutors for proofreading (1). According to participants, they do not need as much support with writing because in school, they are focused on it, and writing allows more time to think and correct mistakes. For

instance, a participant said *"...when I speak it, it's harder, but when I write it, it's not so hard. Maybe because in school we only learn it [English] theoretically...about grammar."*, another participant said *"when I write I have time, I can catch myself...of course when speaking too...but it's harder [compared to writing] because you realize [you made a mistake] after you said it..."*.

Meanwhile, when asked if they used any software for acquiring English, only three participants answered positively. The language learning application mentioned was Peak (peak.net), which is not specialized in language learning but also offers modules that can be used to practice language skills. For instance, the participants used this application to spot errors in essays, then they were asked to correct them; Duolingo, which is a language learning platform. However, the participant said the application is not as helpful after the beginning level of proficiency; and italki (italki.com), this application matches learners around the globe to exchange and practice each other's language.

2.3.5 *Discussion*

In this section, I list some observations and comments regarding the interview results described so far.

2.3.5.1 *Study Observations*

Language immersion and perception of fluency: Participants who learned English in a dedicated school did not consider themselves fluent at the moment of the interview; they are also the same ones who have been in Canada for the shortest period. The same participants declared a significant preference for their mother tongue versus the others who mostly prefer using English, reserving their mother tongue when they are around family. Another factor that can be co-related regarding this group of participants is that in their home country, English is not commonly used, even though it is commonly learned and taught.

Practice makes perfect: Most of the participants (6/10) indicated the use of TV series and cartoons as their method of learning and improving their English skills. Meanwhile, all participants mentioned that practicing the L2 with natives or just practicing, in general, is among the best ways to learn a second language.

Table 2.6: The main difficulties each participant faces regarding speaking and learning English as a second language grouped by 3 skills (writing, speaking, and listening).

Language Difficulties			
Writing	**Speaking**	**Listening**	**Comments**
Punctuation	–	–	–
Spelling	Expressions, idioms, lack of non-work vocabulary	Difficulty with different accents	Doesn't feel comfortable to engage in learning activities
Vocabulary	Non-sense sentences, auxiliary verbs can be troublesome (order)	–	Feels like confidence is a huge factor into improving language skills
Difficult, always gets friends to proof-read; Writes as is spoken.	–	–	Verb tenses are hard, because there is no equivalent in 1st language;
Literal translations and vocabulary	Vocabulary	Difficulty with different accents	Adj-Noun order is different from 1st language
Run-on sentences, connectors (moreover, further more...), mixes verb tenses	Pronunciation (knife, island); Translation issues (We eat water) from equivalent in the 1st language	–	She has had several private tutors
–	–	Sound alike words (they're / their)	She feels pretty confident about her language skills
Vocabulary and punctuation	Pronunciation	–	Often asks friends to proof-read her essays
Reads a lot to expand vocabulary	Pronunciation, sentence forming and conjugation	Listen to music to learn the pronunciation	He has had several private tutors
Vocabulary	Pronunciation, mainly vowels because of 1st language	–	Feels like French is harder to learn than English because of its similarity to Spanish (participant's 1st language)

Table 2.7: Participants use of software and resources to support writing. We can also see their main goals when adopting these tools or methods: to learn or acquire language, to get writing samples proofread, to improve language and proficiency skills, and to improve confidence in order to not need support in the future.

Usage of support software for writing						
		Goals				
Software	**Comments**	Learn	Get it proof-read	Improve skills	Improve confidence	**Learning Apps**
Regular spell-checker	She heard of Grammarly but doesn't know if it affects the Turn-it-in system.	Yes	–	Yes	Yes	–
Regular spell-checker	Hard time with idiomatic vs literal translations.	–	Yes	–	–	–
Free Grammarly	Grammarly only spots grammar mistakes, it lacks an explanation of why something is wrong; She feels the need for something that can fix "non-sense" sentences that are grammatically correct.	Yes	–	Yes	Yes	Peak and Evaluate - it shows an essay asks to spot a mistake and fix it. Negative point is that there is no option to choose which skill to practice.
Multiple	He never studied grammar, always focused on conversations.	–	Yes	–	–	Several websites and online tools; Also, he has hired tutors to proofread.
Regular spell-checker	–	Yes	–	Yes	Yes	Duolingo, it was hard to improve past the beginner level.
Premium Grammarly	She appreciates that she has to fix the mistakes herself with Grammarly.	Yes	–	Yes	Yes	–
Regular Spell-checker	She doesn't rely on software for proof-reading.	–	–	–	–	–
Free Grammarly	She would want it to show words she overuses and to suggest synonyms or rephrasing.	Yes	–	Yes	Yes	–
Regular Spell-checker	–	Yes	–	Yes	Yes	italki, matches you with another person and you can exchange language skills.
Regular Spell-checker	She would want help with expanding vocabulary, expressions and slangs.	–	Yes	–	–	–

Non-native speakers do not seek supportive writing technology: Most of the participants (6/10) do not use dedicated software for writing support and grammar checking. Most of them said they only use the default spell-checker that the text editor offers. The mentioned used apps are Grammarly Premium (1) and Free (2), and online websites that offer human tutors for proofreading (1).

Acceptability errors are a common concern: Participants expressed that they have difficulties related to non-grammatical issues such as idiomatic and literal translations, "non-sense" sentences that are grammatically correct, overuse of words, need for vocabulary expansion, expressions, and slang.

Advanced learners need more autonomy on learning apps: Only a few participants (3/10) indicated using applications to learn English. The mentioned apps. are Peek and Evaluate, Duolingo, and italk. The main complaints with the available software are "*. . . it [Duolingo] doesn't offer much support past beginner level.*" and "*. . . [in Peak] there is no option to choose which skills I want to practice; I'm always forced to follow the order of the modules*".

Pronunciation affects communication confidence: Participants were generally more concerned with improving their speaking skills because they think writing gives you more time to rephrase and correct yourself. Pronunciation was a common factor among participants regarding speaking and confidence.

Improvement seekers versus settlers: There seem to have 2 groups of participants: (1) the ones that are aware of their language impediments and are willing to use software or other people to improve to feel more confident about their English. And (2) the group that is aware they could improve their English skills and would not mind if they do, but they are confident they do not have to and are unwilling to look for help. I decided to label them "seekers" vs. "settlers" learners.

2.3.5.2 *Study Remarks*

Table 2.8 shows a comparison between fluent and non-fluent participants; it indicates that self-perception of fluency is related to the level of immersion in the second language. This information is aligned with the **language immersion and perception of fluency** and the

Table 2.8: Comparison between self-declared fluent and non-fluent participants. Second language immersion seems to be related to the participants' communication confidence and perception of fluency in non-native language.

	Not Fluent (2)	**Fluent (8)**
Period living in Canada	< month	1 to 10+ years
Learning method	ESL school	Bilingual School (5) or in Canada (3)
Language usage	English for work	English is preferred unless talking to family
English usage (home country)	Not used	Not used (3) Used (5)
First language attrition	None	Present

practice makes perfect observations. On top of that, the **pronunciation affects communication confidence** observation also contributes to this claim that participants living in Canada for longer are more confident in their fluency when speaking to native speakers, considering they already had enough time to adapt and tune their pronunciation. Therefore, **second language immersion directly influences non-native speakers' level of confidence and fluency**.

From the participants, it seems that non-native speakers are **not** actively searching for a tool for writing support. From the following observations, **non-native speakers do not seek supportive writing technology**, **acceptability errors are a common concern**, and **advanced learners need more autonomy on learning apps**. I understand the participants who were advanced learners also could effectively communicate using English as their second language. Therefore, their need for supportive technologies for writing might be similar to the perceived need of an English native speaker — further research is required to confirm this. Each participant claimed to have particular English acceptability-related difficulties. However, the technology they have access to does not help with fine-tuning or nitpicking on specific issues. Thus, **there is a need for developing modular and specific applications that can support acceptability errors such as word choice, word overuse, word collocation, contextual vocabulary, and expressions**.

Regarding the **improvement seekers versus settlers** observations, I can claim two broad groups of non-native English speakers: seekers (6) and settlers (4). Seeker speakers, independently of perceiving themselves as fluent, are actively looking for points to improve their language skills. Meanwhile, settler speakers also, independently of perceived fluency level, already feel comfortable with their language skills. Even though both types acknowledge their weaknesses, the difference between the two groups is in comfort rather than confidence level. Considering this was a small set of participants, one can not lean on this finding for generalization. However, it hints at general personality traits and other personal and situational factors. For instance, if settlers are placed into an environment where their communication and language skills are crucial for their personal success, it might incentivize them to focus on improving their English skills.

2.4 CHAPTER REMARKS

From the interview with non-native English speakers presented in Section 2.3, I identified some aspects that the existing software applications are lacking, such as flexibility in choosing and customizing the topic or skill for learning or practicing. Interviewees also mentioned that L2 immersion, self-confidence, and pronunciation are the main factors for fluency or self-perceived fluency.

In addition, I observed differences aligned with the SDT when interviewing non-native speakers of English; in Section 2.3, I discuss how a group of the interviewees can be categorized as improvement "seekers" and "settlers". The main difference between the two groups is their level of satisfaction with their current foreign language proficiency, independently of perceiving themselves as fluent. Contextualizing this finding in this section, what was observed is that settlers' present selves are fully aligned with their future selves in terms of English as a foreign language. Yet, seekers are not — therefore, seekers are still looking into ways to improve their English proficiency. In other words, seekers self-regulate their behaviour toward becoming their ideal selves as English speakers.

Furthermore, those interviews helped me to identify a lack of computer or mobile-based applications that can support those "seekers"; who can also be called self-motivated or autonomous language learners. As defined earlier, autonomous learning happens when one has

the ability to be in charge of their own learning [91]. This means autonomous learners plan what and how to learn, implement their plans, and monitor and evaluate their progress. While these tasks are not specific to have or not a teacher or tutor involved, nor the adoption of specific software applications, these tasks are related to SDT factors: autonomy, competence, and relatedness. For autonomy, they will be responsible for their learning.

Regarding competence, the learner needs the appropriate skills and means, such as the correct tools and strategies. And for relatedness, they rely on their sense of current and ideal or future selves who belong to the target language culture.

Therefore, there is a need for creating and exploring different tools that empower learners to achieve adequate competence at various aspects as they work toward language learning autonomy. In Chapter 3, I discuss technology-enhanced learning and their relationship regarding supporting autonomous language learners.

2.5 SUMMARY

This chapter introduced the main concepts of language learning autonomy and self-regulated behaviour needed to motivate and drive my proposed contributions. In addition, I discussed the results and findings of my interview with non-native English speakers, which helped me to better understand language learners. Lastly, I contextualized the interview observations using the learner autonomy concepts presented.

Thus, this chapter's contributions are as follows:

- Background on Psychology's SDT, language learner autonomy, learner autonomy continuum, and self-regulated behaviour.
- Results and finding of an interview with non-native English speakers.
- Identification of autonomous language learners' need for flexible support tools that empower learners to traverse through the learner autonomy continuum.

3

LANGUAGE LEARNING APPLICATIONS FOR PROMOTING AUTONOMY

This chapter discusses the core concepts for understanding technology for language learning, including technology aspects to support language learner autonomy, such as Technology-Enhanced Language Learning (TELL).

Furthermore, I present background information to contextualize each project included in this dissertation. Section 3.2 presents Natural Language Processing (NLP) techniques and learner corpora associated with H-Matrix. Section 3.3 introduces the process of Italian verb morphology acquisition and computational linguistic tools used in Card-it. Section 3.4 describes micro-learning, generative AI for personalized learning, and MALL for L2 vocabulary acquisition tackled by LangEye.

3.1 TELL: TECHNOLOGY-ENHANCED LANGUAGE LEARNING

Zhou and Wei [147] define TELL as any language learning activity that uses technological means and/or tools for efficiency, motivation, and learning flexibility. Traditionally, TELL was examined as a learner-centred methodology to help teachers to bridge students and technology — digital media artifacts including video, instant messaging, audio recording for pronunciation feedback, audio listening, or even the Internet itself (Carr et al. [24], Kranthi [81], and Zainuddin [144]). As Bush and Terry [19] indicated, the computer has become more ubiquitous; therefore, TELL focuses on the media itself and how it delivers the learning instructions. In addition, by definition, TELL is not limited to classroom or teacher-instructed support. Meanwhile, CALL focuses on computer applications as the object of study; MALL aims at understanding mobile-based applications, specifically. Data-Driven Learning (DDL) is concerned about employing data resources as a means to learn a L2 or FL; similarly, Google-Assisted Language Learning (GALL) refers to using APIs and other cloud-based services provided by Google, specifically.

Digital media consumed via the Internet, computers and/or mobile devices can provide an accurate portrayal of the target language (i.e., accessing media produced by L2 natives on YouTube) and provide learners with control and feedback (i.e., learner analytics and adaptive learning algorithms). Bush and Terry [19] states that learning resources should support continuous learning by the individual and be relevant to the learners' needs. Furthermore, TELL is more flexible than traditional teaching styles. Yet, it can be used alongside classroom instruction to supplement and improve the learning experience. In other words, TELL can promote and support learner-centred approaches where students can select the order of the material presented to them and control the pace. For example, choosing to learn grammar first, followed by vocabulary. Lastly, the learning activities are not limited to the classroom, and the learning environment is expanded [81]. This means students can study and learn their Target Language (TL) at home or during their commute, and practice language in class, for example.

Macaro [96] and Bush and Terry [19] defend the adoption of the TL in the classroom and in the instructions displayed in the software to help learners develop vocabulary, comprehension, morphology, spelling and even pronunciation for software that includes voice. According to them, this will help students to increase their L2 immersion, improve their communication skills, and promote autonomy. Learners' attitudes and motivation (which is linked to autonomy) tend to be better if learners can be immersed in learning the TL's culture and control the amount and sequence of their studying activities [19, 81, 96, 121, 147].

In contrast, TELL presents some disadvantages or limitations, including cost and accessibility of technology, technical literacy and training, the teacher or instructor must be comfortable with the technology, technical issues or difficulties, inaccurate data or language models can introduce mistakes into the learning materials, and learners' over-reliance on the technology [19, 81, 96]. The TELL system should guide learners through exercises and activities, but they must make a conscious effort to engage with the technology for a better learning experience.

The work described in this thesis is a collection of software projects related to different aspects of TELL. In Section 3.1.1, I present some

popular software applications for language learning and further explain CALL, MALL, DDL, and GALL.

3.1.1 *Related TELL Applications*

The existing software for language support or learning does not differentiate their curriculum or lessons, considering the learners' L1. Heil et al. [62] reviewed the top 50 most popular language learning applications from a joint list of Apple and Google Play stores. The top 5 ranked apps are Duolingo, Rosetta Stone, Memrise, PenyoPal, and Learn English (by Anspear). Besides technical and accessibility factors, such as language, platform, input, and output supported, they evaluate the teaching models applied, corrective feedback, and skills (listening, reading, and writing) supported. They found three main common trends: (1) vocabulary teaching in isolation instead of in context; (2) software auto-adaptation to suit the learner's skill; and (3) lack of explanatory content for corrective feedback.

Grammarly and other default spell and grammar checkers from text editors are specialized applications for supporting writing. Students have widely adopted Grammarly to help with writing, and non-native English speakers are no exception. Karyuatry [74] verified through a study that Grammarly helps to improve the students' writing quality; because the tool handles most of the spelling and grammatical errors, students were able to focus on the writing task itself. Karyuatry [74] also noticed a decrease in the number of errors over time. The authors attributed it to the instantaneous feedback Grammarly provides, which increases students' grammatical awareness.

GALL, corpus-based or DDL have been gained acceptance as language learning tool [30, 32, 40, 95, 132]. GALL defines the suite of applications provided by Google as an informative (access to lexicons and linguistic corpora, search queries for definitions, typo correction, even concordance in its queries, and also translation), productive (Google Docs, Blogger, and Jotspot), collaborative (Google calendar and groups), and communicative (Gmail) tool which is used daily by language learners [27].

In a practical sense, DDL aims to eliminate the researcher or mediator between the learner and the data resource (i.e., corpora, parsers, etc.), increasing accessibility [40]. Data-driven learning happens when

learners take advantage of online access, storage capacity, and text processing power to use web-based corpus tools, such as dictionaries and linguistic corpora. These tools are built based on growing resources that can provide common language usage guidelines, contextual definition, and lexical and concordance queries [30, 140]. Conroy [30] showed that language learners are already competent in web-based tools. Therefore, DDL takes advantage of this familiarity, proving itself effective with non-native English speakers in university.

Similarly, Yoon [140] verified that DDL was an effective cognitive tool for helping participants with their lexical and grammatical problems while dealing with concordance tasks. However, he suggests that some available resources are not user-friendly enough and difficult to use, such as functions for linguistic resources applied for stemming. More recently, a meta analysis of DDL-related publications between 2014–2020 presented by Ueno and Takeuchi [132] showed that DDL influences L2 learners positively. Nonetheless, both GALL and DDL are approaches that are seen as powerful tools for language learning, which suggests that all three proposed projects are a good fit for this scenario of DDL.

Based on the reviewed literature, on my interview with non-native English speakers (Section 2.3), and on my personal experience with language learning and technology, the current language learning apps lack flexibility on modules and lessons. They usually do not allow for manual tuning regarding the learners' personal interests — e.g., pick or skip vocabulary modules, or practice one specific skill such as talking. The software adaptation in the apps regards only the accuracy of tests; the apps also do not allow specific L1 effects targeting, which has been shown to affect learning positively [12].

Even though GALL and DDL are approaches in demand for language learning and the volume of web-based resources is vast, there is a need for applications that can help users to interface with such unstructured knowledge [27, 30, 140]. Even more so for learners who are not proficient with linguistic terms and tools [140]. The rise of these technologies shows the power of independent learning and self-motivation towards language acquisition [30]. Therefore, the intended contribution of H-Matrix (Chapter 4), Card-it (Chapter 5), and Lang-Eye (Chapter 6) fit very well in this gap.

3.2 ERRORS ANALYSIS AND LEARNER CORPORA

An obvious way to explore learner corpora is to find or correct errors. Kochmar's (2016) work uses a collection of learners' essays to detect word choice errors, more specifically, in adjective-noun (AN) and verb-object (VO) combinations. The concept behind her work is that the word paired choices AN, or VO, not only occur more often than expected (native writing), but the distribution will also vary across different first languages. She performed error detection using compositional distributional semantic models for content word combinations (AN and VO) in the learner corpus (Cambridge Learner Corpora).

Similarly, Bestgen, Granger, and Thewissen [12] showed that using the ICLE dataset, it is possible to find common error patterns in English essays that can discriminate native writers of three first languages: French, German, and Spanish. For instance, some of the reported results were that Spanish speakers are strongly associated with spelling errors and lexical errors on single words and phrases. German speakers are more likely to make errors involving the order of adjectives, and French speakers make punctuation errors of the QL type (i.e. use of a conjunction of coordination instead of a punctuation marker or of a punctuation marker instead of a conjunction of coordination). The motivation of his work is to identify the native language of the writer. However, the L1 influence found in the distribution and types of errors made in English (L2) matches the findings by [79].

In a later work, Kochmar and Shutova [80] investigated the extent of transfer effects of lexico-semantic from the native language (L1) to the second language (L2). They chose three typologically diverse languages: Russian, Spanish, and English, and they found that using an L1-specific statistical semantic model from L1 can improve automatic error detection in L2 speakers. These projects show how learner corpora can be used to extract L1-specific language models and also proves that leveraging L1 information can be beneficial to error correction within SLA [12, 79, 80].

A more specific case related to the distributional analysis of errors and direct application into SLA is the work by Zafar [143]. She ran a study on undergraduate students from a Business Communication course at a university in Pakistan. The study collected an initial course essay in English, and the ESL experts from the university identified

the most common English mistakes in the class. They noticed that L1 transfer errors were the most popular category, with the exception of past and present tense errors. The ESL teachers then intervened with English lessons during the semester by focusing on the most common errors found. In the second essay, which was collected by the end of the term, they noticed a significant improvement in the students' writing and communication skills. This study and others show that leveraging L1-specific transfer effects can be useful to ESL teachers and students themselves [110, 111, 143].

Beyond that, one of the main goals of data-driven learning is to be able to remove the researcher or mediator between the learner and the data (i.e. corpora) [40]. For example, Crosthwaite and Baisa [32] proposed CorpusMate a user-friendly corpus explorer designed to teachers and L2 learners to explore a variety of learner corpora. Thus, I present H-Matrix (Chapter 4), tackling the linguistic level, which enables visual analysis of distributional cross-linguistic features in large learner corpora. H-Matrix builds on previous work of Bestgen, Granger, and Thewissen [12], Kochmar [79], and Kochmar and Shutova [80], and it contributes by facilitating the analytics process of language experts and learners themselves who are interested in transfer effects. Furthermore, it may help language tutors to be more aware of transfer effects and potentially modify their lesson plans to help students to improve, as shown by Bestgen, Granger, and Thewissen [12].

3.3 ACQUISITION OF ITALIAN VERB MORPHOLOGY

In Italian, verbs never appear as unmarked forms but always bear inflectional affixes instead. This leads to learners of Italian acquiring a rich system of verbal inflections (e.g.,Pizzuto and Caselli [108]). The verbal inflections express the morphological features of tense, aspect, mood, person, and number. Slabakova [124] argues that inflectional morphology plays an important role in language learning of morphologically rich languages. Past studies on the acquisition of tense have shown that learners tend to first express tense by means of lexical items (today, tomorrow, etc.) and only subsequently use verbal inflections [9].

Previous research on morphological processing indicated a large gap between the processing of regular versus irregular verb forms,

Table 3.1: Irregular versus regular conjugation pattern.

		Irregular *andare* in Present Indicative	**Regular *amare* in Present indicative**
Singular	1st	**vado**	am-o
	2nd	**vai**	am-i
	3rd	**va**	am-a
Plural	1st	and-iamo	am-iamo
	2nd	and-ate	am-ate
	3rd	**vanno**	am-ano

with irregularities in past tense being difficult for L2 learners. (e.g., Larsen-Freeman [84]). This has been linked to the fact that, in contrast to regular past tense, irregular forms have a smaller form overlap with the verb's base form. Another complicating factor is that irregular forms cannot be decomposed morphologically in their constituent morphemes [127]. In Italian, irregular verb forms appear not only in past tense but may show stem irregularities throughout their entire conjugation paradigm. In contrast, regular verbs follow a strict conjugation pattern defined by their conjugation class:

Table 3.1 exemplifies the stem irregularities observed in Italian irregular verbs. While the regular verb *amare* retains the stem throughout all forms, maintaining a strong morphological overlap (consistency), the irregular verb *andare* retains the stem only for 1st and 2nd plural. In all other person and number configurations, the verb form cannot be decomposed into its constituent morphemes and has no form overlap. The capability for learners to understand how words can be structured with morphemes, i.e., their morphological awareness, has been shown to have a positive impact on their reading comprehension, pseudoword reading, and spelling [35, 99, 112]. Thus, morphology plays a major role in different aspects of reading.

In a study on the processing of regular and irregular verbs, De Zeeuw, Schreuder, and Verhoeven [34] showed that L2 learners of Dutch had more difficulties in identifying irregular verbs than regular verbs in 3rd grade —- compared to native speakers. The difference was not observable for 6th-grade students. De Zeeuw, Schreuder, and Verhoeven [34] attributed this to target language exposure. Another contributing factor for supporting the acquisition of Italian verb morphology is that Italian can be regarded as a head-marking language

[17], meaning the morphological makeup of the verb may incorporate its arguments, categorizing Italian as a so-called null-subject language. Studies with highly proficient, near-native speakers of L2 Italian have shown that their production skills with respect to overt pronouns contrast with those of native Italian speakers. This suggests that this is a challenging aspect of the language to acquire for learners whose native language does not exhibit this property.

All these studies, combined with language-specific properties of Italian, highlight the important role of verb morphology in foreign language acquisition. In Chapter 5, I present Card-it. Card-it's design choices regarding the inclusion of linguistic terms, corrective feedback, and explicit training of morphology aim to reflect this role.

3.3.1 *A Large-Scale Finite-State Morphological Analyzer*

FSMs are usually part of a text processing pipeline within NLP tools. In Card-it, I leveraged a FSM as a form generator and analyzer in a language-learning context. The FSM ties a verb form to its linguistic analysis: it may analyze a verb form and return its linguistic tags (analysis) or generate a verb form given its linguistic tags (generation) – see Figure 3.1. The Italian FSM is a core component of the Card-it, containing over 5,000 verb lemmata and their conjugations [10]. It was created by extracting verb roots from free resources, the *Morph-it!* lexicon by Zanchetta and Baroni [145] and the online dictionary provided by one of Italy's leading news magazines, *Corriere della sera*[1]. The FSM was developed by my collaborator and co-author, Jessica Zipf, a computational linguist and Ph.D. candidate from the University of Konstanz [122, 148].

In this case, the FSM consists of a lexicon that contains verb stems, their inflectional paradigms and the appropriate morphological analysis. The lexicon of the FSM creates all verb forms following the regular pattern of concatenating stems with their respective inflectional endings. With the use of regular expressions the FSM is able to manipulate those regular forms of the lexicon on the basis of phonological rules. For example, some forms require the insertion of an *-h* to retain certain pronunciation patterns. Consider the verb *mancare* ("to miss"): the regular inflection paradigm in the lexicon creates the incorrect form *manci* ("you miss"), for the second person singular present

1 https://dizionari.corriere.it/dizionario_italiano/. Accessed 05-2023

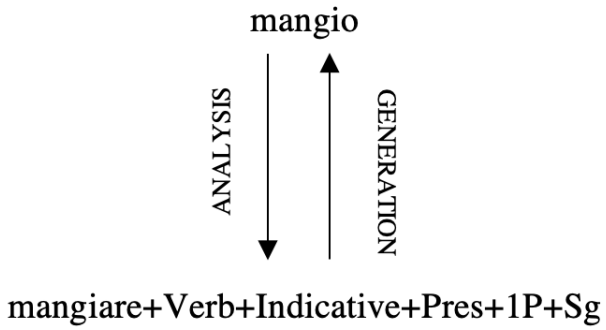


Figure 3.1: Example of FSM Analysis and Generation for the verb *mangio* ("I eat") [122].

indicative. However, to retain the correct pronunciation, the correct form is *manchi*. Whenever the FSM is run, it first creates all forms in the lexicon and then applies regular expressions to manipulate these forms based on phonological rules of the language. This architecture enables a powerful and large morphological resource since it automatically creates verb forms on the basis of their stems. If new verbs are added to it, it simply requires to manually add verb stems into the FSM lexicon. Card-it is powered by the said FSM and it offers DDL experience to learners via a simple user interface; leveraging a web application architecture and flashcards for studying and training Italian verb conjugation. More details on implementation are provided in Chapter 5.

3.4 SITUATED VOCABULARY AND LANGUAGE LEARNING

Situated language learning refers to learning that occurs in a sociocultural relevant or situated context [3, 86, 90]. For example, learning about the history of a building while visiting on site; or learning the names of vegetables in a foreign language while grocery shopping.

To support situated vocabulary learning, many explore immersive technologies such as mixed (Augmented Reality (AR) and Extended Reality (XR)) and virtual reality platforms [22, 44, 64, 66, 138, 150]. An option for building immersive solutions is via Virtual Reality (VR) technology enabled with device screens, such as smartphones, or using head-mounted devices — respectively classified as Low-immersion Virtual Reality (LiVR) and and High-immersion Virtual Reality (HiVR) [73]. Papin and Kaplan-Rakowski [106] explained that both LiVR and HiVR language learning applications were proven to

promote L2 immersion and have a positive impact on learners' L2 vocabulary acquisition due to the contextualized experiences, the dual coding theory (visualization-verbalization) [105], the method loci (memory palaces for memorization) [82], and the embodiment theory [52].

While the AR and XR technology has evolved rapidly in the past five years, most applications and frameworks in the language learning domain are conceptual or experimental due to their limitations, including device prices, human resistance to novel technologies, and the lack of unified design, development, and adoption practices [18, 137]. On the other hand, smartphones are present in everyone's daily life and it is a powerful mediator for situated learning [3]. Smartphones are ubiquitous devices that, combined with cloud computing, have improved our quality of life. MALL is one of the examples of applications that have greatly benefited from this platform (i.e., Duolingo, Ankii, etc.). Lightweight applications are accessible any time of the day and always within reach, making language learning achievable to language learners, particularly autonomous learners, who are self-motivated individuals.

3.4.1 *Learner-curated Vocabulary with MALL Applications*

Mobile devices are a gateway to many services and information. Their high accessibility has contributed to the increased popularity and adoption of MALL applications. Compared to CALL applications, MALL differentiates itself due to the portability of mobile devices which offers opportunities for 'anytime' learning, micro-learning, and situated-learning [8, 21, 36, 39, 54, 60, 129]. Nonetheless, MALL apps showed to be effective, useful, and suitable for studying and practicing new foreign vocabulary [4, 78, 141].

Micro-learning happens via short periods of time (i.e., 5 to 15 minutes while waiting for the bus) using targeted activities [89]. While Dearman and Truong [36] implicitly applied micro-learning by displaying a target word on one's mobile wallpaper, micro-learning usually requires explicit interaction from the learners themselves. When focusing on vocabulary acquisition, combining micro- and situated-learning offers contextual relevance to learners. Tran et al. [129] presented MiniHongo for Japanese as L2 that uses location- and physical-activity-based data to compose and recommend lessons. Their main findings were that "Learning vocabulary via a contextual location-

relevant lesson is more effective" and that micro-learning for vocabulary can be effective when detecting patterns of the learner's free time for triggering micro-learning activities, although this was only partially supported.

While it does not apply micro-learning, VocaBura is also a situated-learning MALL app which combines audio learning with location-relevant L1-L2 word pairs that present new vocabulary items while learners are walking past buildings, shops and other locations. On the other hand, VocabEncounter a CALL application applies contextual- and micro-learning for vocabulary as the learner browsers the web Arakawa et al. [8]. VocabEncounter has a collections of target words being learned and it identifies them in the web content using NLP and Machine Translation (MT) algorithms, it then displays either the word alone or the full sentence in the L2. For example, a Japanese learner of English (L2) will see an English target word in the middle of a news article or the full sentence containing the target word in English; the remaining of the article will be displayed in the original language, Japanese in this case. Similarly, others have studied equivalent techniques for embedding micro-learning vocabulary opportunities in audiovisual content [5, 50, 109].

In contrast, VocabNomad teaches vocabulary in a visual dictionary organized by expert-, user-, and system-generated collections of words [131]. It displays the target word, its definition, pronunciation, image, and a contextual example sentence. Learners can browse through collections, interact with the vocabulary, insert their own vocabulary entries, and record their own pronunciation. Thus, VocabNomad offered micro- and contextual-learning opportunities along. However, it also provided a highly personalized experience with progress tracking, contextual vocabulary recommendation (based on the learners' previous interactions), the embedding of the learners' own pronunciation audio, and learner-curated vocabulary entries. The user study also indicated that VocabNomad promotes situated learning as learner can insert and search for entries as they need to use the vocabulary in real-life. These studies imply the importance of creating relevant and contextual opportunities for vocabulary learning into our daily lives.

3.4.2 *AI-Enabled Context Personalization for Vocabulary Learning*

The introduction of LLM starting from ChatGPT[2] enables the production of natural language in textual format. The generative-AI content can be dynamically produced by a engineering prompts (arguments) to the models, which has the ability to synthesize examples and output definitions. Generative-AI also presents its challenges including ecological and ethical considerations, and the accuracy of its output — which often is indistinguishable from human text [23]. Nonetheless, generative-AI has led many to explore personalized educational applications including language and vocabulary learning. For example, Storyfier aims to help learners to acquire new vocabulary while dynamically creating AI-assisted stories using their target words [107]. While their study did not show better learning outcomes with this method, results show that learners generally favor the generated stories for associating the target words within writing production. Leong et al. [88] explored variations of generative AI vocabulary learning app and they also observed similar results where generative AI-driven context personalization positively affected learning motivation, but not the learning outcome.

In 2017, Vazquez et al. [137] proposed a framework for contextual learning in mixed-reality, WordSense, which relied in a resource-intensive pipeline for automatic object recognition, MT, and Dynamically Linked Content (DLC) as referenced multimedia information where the learning material is not generated or manipulated by the application itself, but queried using context. This implied on combining several other NLP resources such as dictionaries, corpora parsing for example sentences, etc. However, currently, generative-AI is able to somewhat fill the pipeline to enable DLC. For example, FluencyAR (2023) uses AR technology alongside a simulation (Wizard of Oz) of object recognition and ChatGPT to recommend situational prompts about the objects detected in the scene [63]. To further motivate the learner, FluencyAR provides keywords for their self-talk. The system processes the audio prompt and provides them with a transcript and translation of what they said as feedback. However, this can be considered a micro- and situated-learning where the self-talk conversation prompts are rooted in the learner's reality. More recently, in 2024, CuriosityXR was introduced. It is a mixed-reality application which acts

2 https://openai.com/research/overview. Accessed in May 2024.

as a platform for educators to develop contextual and multi-modal interactive mini-lessons. Learners can engage with these lessons and also benefit from AI-supported learning content. Their study revealed greater engagement levels, increased curiosity to learn, and improved visual content retention [136].

LangEye (Chapter 6) is a MALL application for vocabulary learning which implements micro-, situated-, and contextual-learning. Using their mobile camera, learners can discover new vocabulary by taking photos of their real-life objects, promoting situated learning. From the app, learners can explore collected photos, the names of objects detected in them, contextually-related words, listen to pronunciation, see definitions, and example sentences. LangEye focuses on learner-curated vocabulary and short review sessions which are generated using Artificial Intelligence (AI) generative text, MT, and text-to-speech API.

3.5 SUMMARY

This chapter presented key related works and background information to motivate and outline contributions of each of the individual chapters.

Thus, this chapter's contributions are as follows:

- TELL background and related work to this thesis;
- Learner corpora and error analysis research related to Chapter 4;
- Italian verb morphology acquisition including the introduction of FSM analyzer related to Chapter 5; and
- Concepts characterized to MALL applications for vocabulary acquisition and generative-AI for contextual learning related to Chapter 6.

Part II

WHAT I CAME TO KNOW

This part of the thesis presents a collection of projects implemented using data-driven and learner-driven approaches, tackling some of the issues outline in Part I, to promote autonomy in language learning. Furthermore, I reflect on the impact of the presented research and possibilities for future work, concluding this dissertation.

4

H-MATRIX

Hierarchical Matrix for Visual Analysis of Cross-Linguistic Features in Large Learner Corpora

Second Language Acquisition (SLA) is the research field that studies the process of learning a language other than your native language. When learning a different language, there are language transfer effects (cross-linguistic effects) which can be structural, grammatical or semantic rules from the learner's first language being applied to the second language [103]. For instance, a native Portuguese speaker can misuse the English verbs *lost* and *miss*, because both translate into only one verb in Portuguese, *perder*. I am interested in systematically identifying such transfer effects, and do so using learner corpora, International Corpora of Learner of English (ICLE) [59] and TOEFL11 [14], which are collections of written essays of English learners from different native language backgrounds.

My user study showed that linguists, and tutors have similar goals towards understanding characteristics one can improve while producing text in a second language. Tutors are focused on a text level analysis to understand performance of learners overtime using smaller sets of essays to tailor their teaching goals. However, linguists are more interested in higher level analysis in linguistic groups, transfer effects and different levels of linguistic features.

A self-determined learner's goals are similar to the tutors', but they are focused on personal performance and self-improvement. H-Matrix is a data-driven learning (DDL) tool that provides visual representation and data filtering of real essay examples to allow learners to increase linguistic awareness. Understanding the types of errors or grammatical concepts one might be particularly worse or better at due to their linguistic background (L1) can help learners to improve linguistic awareness and by focused practice and managing frustration as they evolve in the autonomy continuum.

On the other hand, part of a linguist's approach to answer a research question involves processing and extracting a list of features from real-world text, then applying statistical analysis to discover pat-

terns in the data. One can take advantage of existing visualization techniques and tailor them to the specific needs of linguists. Besides analyzing higher level patterns from linguistic features, linguists often rely on having direct access of the text and context of the extracted features.

While my ultimate goal is to help learners in their journey throughout the autonomy continuum, I took inspiration from the linguistic visualization subfield to help learners, tutors, and researchers interested in transfer effects to identify linguistic patterns in different language groups. The H-Matrix visualization tool allows users to perform visual analytics tasks on hierarchical feature sets extracted from large learners datasets. In order to support the different analytics tasks, I designed two views.

The **Hierarchical Matrix View** (Figure 4.1) supports high level analysis of a large collection of essays by displaying a matrix with calculated scores based on the observed and expected frequency of features in the corpus. The matrix rows are linked using a hierarchical structure representing the linguistic features. Through the hierarchy, we can perform tasks such as aggregating sub-categories and customizing weights, which encode feature importance.

The **Essay View** (Figure 4.2) enables lower level tasks, such as retrieving specific examples from the text, exploring other essays with similar features and validating feature extraction. It displays a table of essays with their meta-information, which allows one to investigate other dimensions of the data (see columns and filters in Figure 4.2). A side panel shows the selected essay text with its tagged features.

4.1 RELATED WORK

Below, I discuss the background on transfer effects in language learning, and visualizations related to H-Matrix's design.

4.1.1 *Transfer Effects*

Errors can be used to understand language transfer effects [79, 80] as well as to detect the writers' first language [12, 71]. The former can be seen in Kochmar's work where in compositional distributional semantic models were used to detect errors in word combinations in learner data, more specifically in adjective-noun and verb-object.

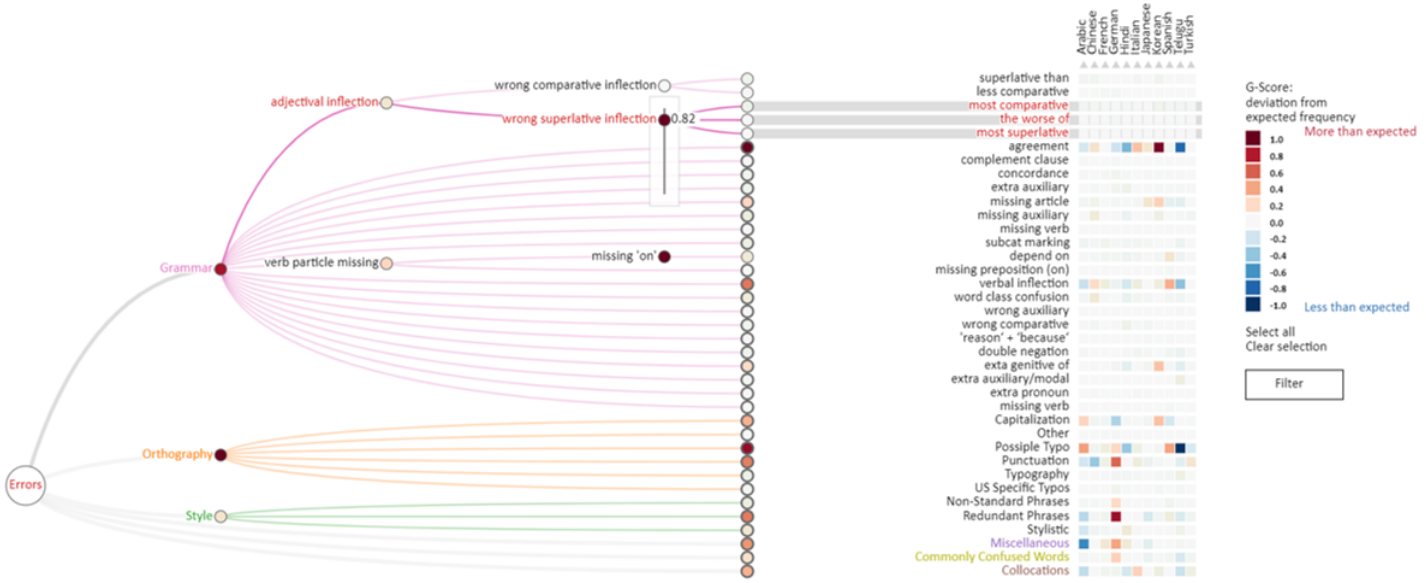


Figure 4.1: H-Matrix combines a hierarchical view of error categories (left) with a matrix view of errors by language (right). The matrix reveals which errors are more common for each language where the hierarchy summarizes the impact of error categories overall. The matrix rows and the categorical tree branches are highlighted on hover. When hovering on a categorical node, it displays a slider to re-weigh the visual impact of a feature (sub)tree in the matrix.

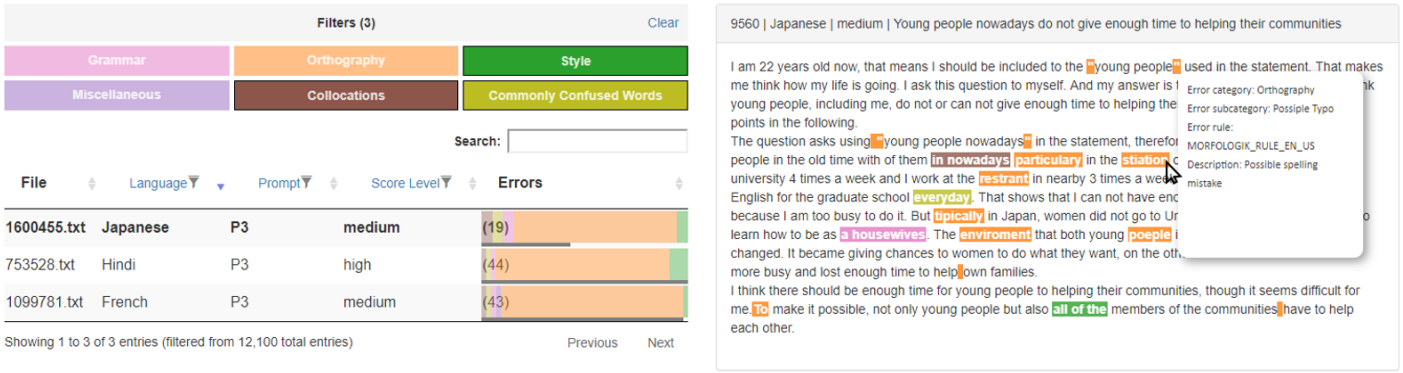


Figure 4.2: The essay view combines a table of essays with a panel where essay text is displayed. We can perform filtering and sorting operations when interacting with the table; it also shows overview information about the tagged errors using an in-cell visualization for each row. On the essay panel, the content of a selected essay can be loaded showing its tagged errors. On text hover, it shows the error categorization and description.

Research on Native Language Identification already uses learner corpora to extract features to distinguish between native languages based on English as second language essays [12, 71]. Bestgen et al. showed that using the ICLE dataset we can find common error patterns in English essays that can discriminate native writers of three languages: French, German and Spanish [12].

Errors made by learners are thus realistic and interesting linguistic features related to transfer effect analysis [12, 71, 79, 80]. Similar to the work of Bestgen et al. [12], H-Matrix allows linguists to identify and distinguish features to create linguistic models to deal with specific transfer effects from each first language.

4.1.2 *Hierarchical Matrix Visualization*

Matrix based visualizations are widely adopted for linguistic analysis. Mayer et al. used a quadratic matrix (matrix of 4 matrices) to investigate cross-linguistic usage frequency of paired vowels [97]. Sacha et al. combined matrices and line chart to highlight intonation patterns [116]. Keim and Oelke proposed a pixel-based heatmap for literary and authorship analysis [76]. Rohrdantz et al. combined a sunburst to encode hierarchical language families with a heatmap for the linguistic features in a set of outer rings [113]. Similarly, in our work we combine a hierarchy with a matrix in order to enrich the data exploration. However, our hierarchical data is the list of linguistic features instead of the language genealogy.

Hybrid visualization techniques are commonly found when dealing with social networks, multi-variate and high dimensional data, Our main data view combines a node-link tree with a matrix visualization. We took as visual and technique references Dendrogramix [13] and Clustergrammer [49]. Dendrogramix combines a dendrogram and a similarity matrix to represent large sets of clusters and data patterns to understand clustering algorithms. Clustergrammer is a heatmap visualization used for hierarchical cluster exploration of high-dimensional biological data, such as, genes and cells.

Lineage [100] and Juniper [101] are visually similar to our approach by combining tree, matrix and table visualizations. However, unlike H-Matrix, which is specifically for analyzing the statistical distribution of hierarchical features, their techniques are to represent attributes of multivariate graphs. Similar to Clustergrammer [49], we designed

an interactive heatmap with operations, such as, ordering, filtering and row aggregation to support dynamic changes on the data for analysis. Dendrogramix [13] differs from our design by using the matrix to represent cluster relationships. However, our categories are pre-defined by an expert. Thus, we combined a node-link tree with categories along with a matrix. This combination aims to enrich the exploration using the feature categorical abstraction to perform aggregation and filtering of the rows rather than analyze the hierarchical structure itself.

4.2 DATA PROCESSING

During design and expert evaluation, the TOEFL11 was used as the large learner corpus due to its well balanced sampling methodology.

DATASET The **TOEFL11** contains 12,100 English learners' essays submitted to the TOEFL language proficiency test [14]. It has 11 L1's (Arabic, Chinese, French, German, Hindi, Italian, Japanese, Korean, Spanish, Telugu, and Turkish), along with information on the score level (low, medium and advanced). They were also evenly sampled from 8 retired TOEFL independent prompts (topics).

FEATURE EXTRACTION For the feature extraction, I used the LanguageTool API which is an open source proofreading software for English and other languages [149]. This API uses a rule-based model to extract errors. The rules are also maintained by the community. It offers a broad categorization of these rules used as a base for the prototype. The rule categorization was later on customized by our expert collaborators. This chapter describes the final and improved version.

4.3 H-MATRIX VISUALIZATION DESIGN

H-Matrix contains three main coordinated components for the matrix, hierarchy, and essays; each is described below.

4.3.1 *Matrix Visualization Component*

In order to support large-scale statistical analysis of the frequency of the extracted features, we built a matrix visualization where the rows are the linguistic features (error rules) and the columns represent the learner data grouped by first languages (right side in Figure 4.1). A matrix visualization is well known and adopted by linguists when performing visual analysis [97, 113, 116]. Therefore, H-Matrix takes advantage of it by having the main component as a familiar tool to the linguistic experts.

4.3.1.1 *Visual Encoding*

For the **encoding values** in the matrix cells, we have 2 options: raw frequency (Figure 4.3) and the g^2-score (Figure 4.4). The raw frequency is the actual number of times a feature or an aggregated set of features (tree leaf) occurs in the dataset per first language. The g^2-score, or Dunning Log Likelihood value, is an expectation measure based on the frequency of the occurrence of a feature in a language group relative to the whole dataset [42]. The g^2-score will be further from zero the more the frequency differs from an expected uniform distribution. For example, a feature observed 100 times within a dataset of 10 first languages, it is expected to appear 10 times in each language. When a language's actual frequency is much lower or higher than 10 it will have a high g^2-score, indicating the deviation from expectancy. We indicate positive or negative scores for cells producing more or less errors than expected, respectively. Table 4.1 shows the error frequencies required to calculate the g^2-score and Equation 4.1 shows its formula.

Table 4.1: Variables used to calculate the g^2-score using the frequency of errors found in the corpora.

	L1	All other L1s	Total
Frequency of error	a	b	$a+b$
Frequency of other errors	$c-a$	$d-b$	$c+d-a-b$
Total	c	d	$c+d$

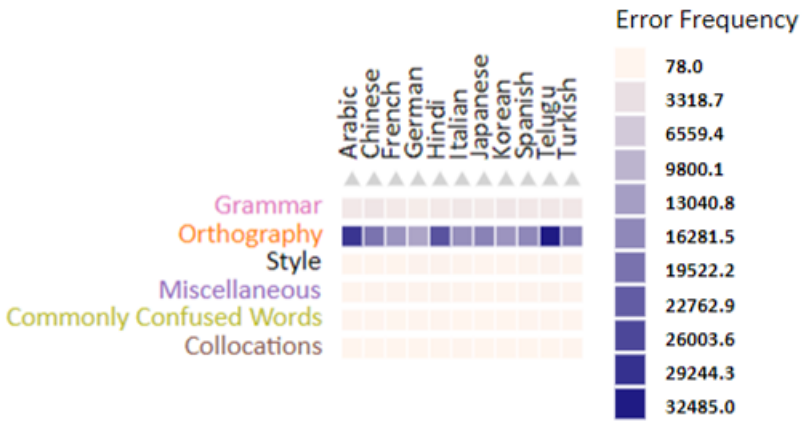


Figure 4.3: Highest level of error categories using the raw frequency without normalization.

$$
\begin{aligned}
E1 &= c * (a + b)/(c + d) \\
E2 &= d * (a + b)/(c + d) \\
g^2 &= 2 * ((a * \ln(a/E1)) + (b * \ln(b/E2)))
\end{aligned}
\tag{4.1}
$$

For **normalizing the displayed values**, H-Matrix offers 4 options: global, row, column, and none (no normalization). The global normalization can help with an overview analysis of the corpora — see Figure 4.4. Row normalization can allow analysis across language groups per feature. The column option helps the analysis of features within a language. See the Figure 4.5 for an example of row and column normalization. Naturally, one also has an option of display the raw frequencies in case they are interested in absolute values — Figure 4.3.

H-Matrix offers two options for **colour scales**: linear and divergent to facilitate the visual analysis for different cases. The linear scale can be applied to both types of scores. However, the divergent scale is only available for the g^2-score as it can be positive or negative.

Another common operation for working with a matrix is **sorting**. H-Matrix supports sorting the rows by their summed score, or by a selected column. One can also sort the columns by their summed scores. These operations support tasks, such as finding high or low aggregated scores per feature or language. Figure 4.6 shows an example of the matrix sorted by the *Korean* column.

Each cell in the matrix provides a tooltip pop-up with extra information on its values (Figure 4.6). We provide the raw and the calcu-

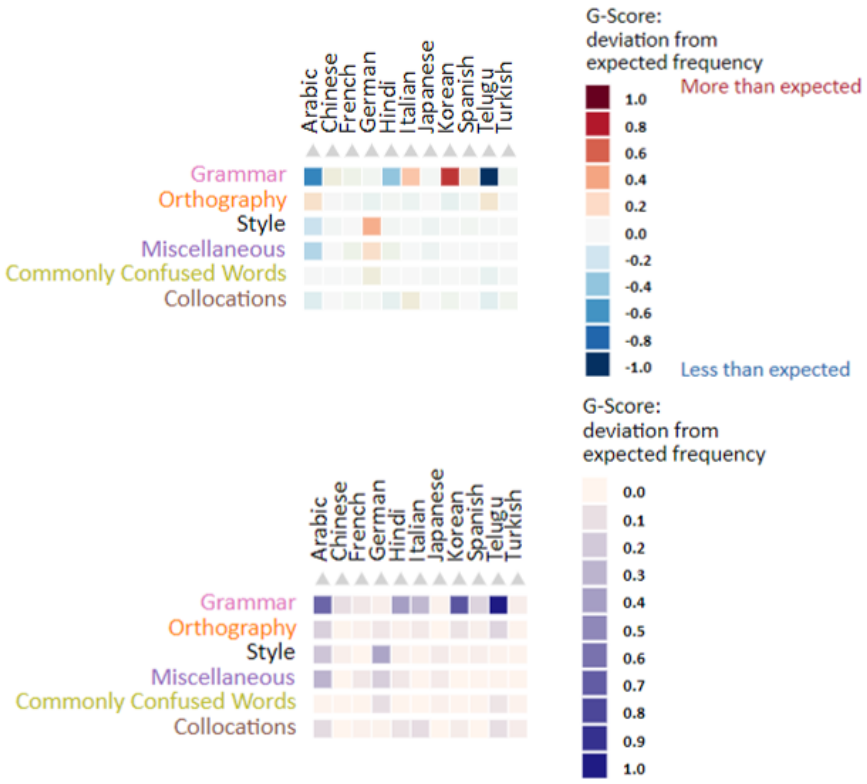


Figure 4.4: Highest level of error categories using the global normalization of the g^2-score. Here the g^2-score is displayed in two colour scales: divergent and sequential.

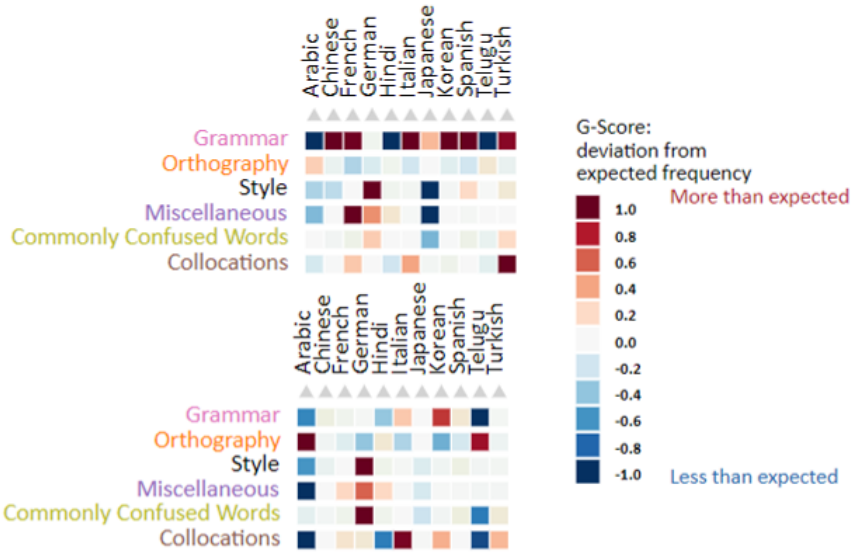


Figure 4.5: Highest level of error categories using the column and row normalization of the g^2-score at the top and bottom of the figure, respectively.

Figure 4.6: Screenshot of the H-Matrix view sorted by the *Korean* column. It also shows the tooltip of when hovering on a cell with the highest g^2-score. The enlarged crop of the screenshot shares a closer view of it.

lated measures. Cells, rows and columns can also be selected in order to filter the **essay level view** with essays containing the selection of errors by clicking in the *Filter* button.

In Figure 4.1 and Figure 4.6, the subcategory *agreement* has the highest aggregated g^2-score (darkest node in the *Grammar* subtree). The *agreement* cell colour in the *Korean* column shows that native Korean learners produce more agreement errors in English than the other languages. This English transfer effect happens because the Korean language does not have agreement rules, such as, verb and subject agreement.

4.3.2 *Tree Component — Hierarchical Features*

In order to perform linguistic analysis, an expert can extract different features from different classes to understand multiple facets of a text. Therefore, experts usually work with different sets of extracted features which can be grouped together in a personalized hierarchical structure. In this project, I worked with linguists to label and categorize the extracted errors from essays into six main categories: grammar, orthography, style, commonly confused words, collocations and miscellaneous (left side in Figure 4.1). Naturally, this categorization is arbitrary and can vary depending on how the features were extracted and the expert's affinities and interests.

I decided to implement aggregation and categorization on the low level features due to large scale of the data which this tool intends to support. For broader categories, such as *grammar*, they can have subcategories and sub-sub-categories containing the rule-level features. For instance, the rule "many kinds of," which signals the plural disagreement of the word *many* and the followed noun, is under the classification "grammar/agreement/nominal agreement (plural)/many kinds of."

Through **interaction** the H-Matrix enables two main operations on demand: dynamic aggregation and re-weighting of (sub)categories nodes. These operations allow experts to change the granularity and importance of the data on-demand.

Without needing to go back into the data processing step, one can investigate different levels of abstraction regarding the error rules being analyzed when **collapsing or expanding nodes** with a left click. Right click opens a menu that offers options to expand/collapse all

levels in a node. When a group of nodes is collapsed all the features under it are aggregated. Dynamically, these changes will reflect in the matrix cells normalization and colour scale accordingly.

When hovering on the nodes, a slider component is shown. This slider enables **dynamic changes in weight** for all the features under the selected tree node. This slider can be used to reduce or neutralize branches or leaves of the tree, therefore, it removes unimportant features on-demand depending on the data analyst's needs. When the weights are changed the appropriate weight distribution is reflected to siblings and children of the node, all the way to the matrix cells scores, depending on the colour scale and normalization used.

4.3.3 *Essay Level Component*

This view is initially populated by all the essays in the dataset. However, the hierarchical matrix visualization view can also be used to filter the essays according to their language and/or type of error feature. The essays are shown in a table which can be used to load a specific essay in the text view (Figure 4.2).

It displays a table where each row is an essay, and the columns can vary based on the dataset's meta-information. In the case of the TOEFL, the columns are file name, language, prompt (similar to topic, is the question learners were prompted when writing the test), proficiency level and the amount of errors. The information in these columns is straight forward, except the errors column.

The data value in the errors column is the number of errors contained in the essay, while the cell contains a visual representation of the number of errors and the proportion of each error category contained in the essay — see Figure 4.7. We render two horizontal bar charts, the top bar represents the distribution of the errors in each category; the lower bar encodes the number of errors which is also displayed in for of text. Using this **in-cell visualization**, one can visually compare the different error categories and their proportions included in each essay. The bar encoding the number of errors helps to visually compare the amount of errors in one essay versus the whole selection listed.

The essays table allows **sort** to be performed in all the columns. However, **filter** can be applied only in the language, prompt (topic) and proficiency columns. Along with the table, this view also has an

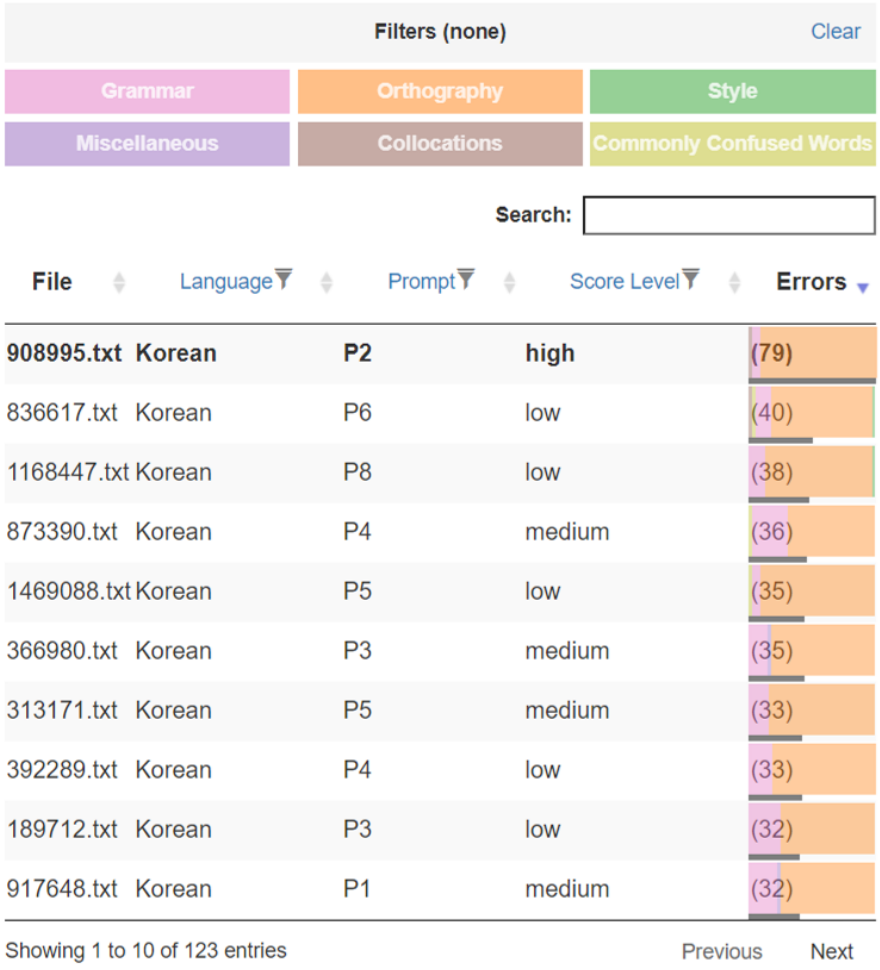


Figure 4.7: The populated Korean essays containing 'verb agreement' errors in the Essay View sorted by the column *Errors*. The Errors column displays the total number of errors in the essay. The number of total errors is encoded in the grey bars at the bottom of each error cell while the remaining of the cell represents the proportion of those errors' categories. For example, the highlighted essay row has a total of 79 errors where majority belongs to the 'Orthography' category (in orange) while fewer belong to the 'Grammar' category (in pink).

error category filtering panel. The **category filtering** options, when used, are applied to the essay table, facilitating the exploration of essays containing a specific error category.

When hovering on an essay row, a **tooltip** pops up with a preview of tagged errors from the essay (Figure 4.8). The tooltip lists each of the tagged errors in a snippet with surrounding context words. If a row is clicked, the **full text** of the essay is displayed on the text panel along with their tagged errors. When hovering on the tagged error, more information, such as, error category, sub-categories if applicable and the rule is shown. This view can be used to verify hypotheses about the data exploration, and to validate the feature extraction step.

4.4 EXPERT CASE STUDIES

I designed a semi-structured interview and an exploratory analysis session, where participants had hands-on experience with the tool. They were asked to describe the methods and processes they would use to identify error patterns in large learner corpora and describe their expectations for a tool that would support their processes.

4.4.1 *Participants*

The study had six participants. The experts who volunteered for the study were computational linguists (N = 3), linguists (N = 1), and language tutors (N = 2). They all had relevant experience with language and second language acquisition. The experts' fields of interest included: writing support, pedagogy, psycho-linguistics, regressive transfer effects, early language acquisition, and bilingualism.

4.4.2 *Exploration Session*

Following the interview and presentation of H-Matrix functions and scheme of the TOEFL11, participants performed a pair analytic session where they had full control of the tool. Meanwhile, the investigator was available to clarify any of the system functionalities, and if needed, to suggest interesting tasks to the participant, *e.g.*, find the highest occurred grammatical error for a specific language and seeking text examples of it.

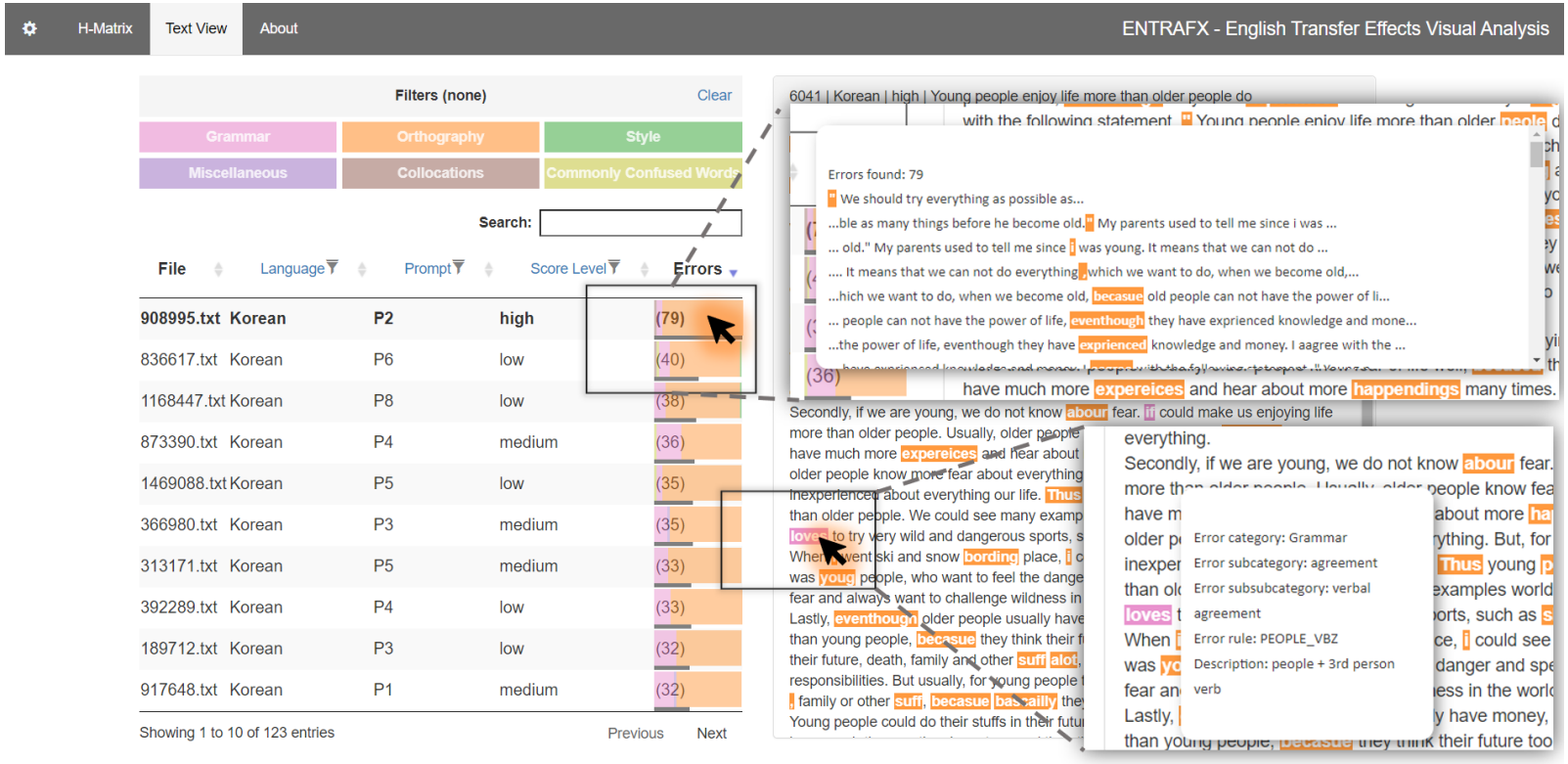


Figure 4.8: Screenshot shows two hover interactions in this view *Errors* cell and tagged error text on-hover tooltips. The tooltip shown when hovering on the error cell enables a quick glance at the tagged error in the essay without loading its full text helping with high level analysis and comparison between essays. The tooltip on the tagged error text displays the error's category hierarchy, rule, and description. It helps analysing and understanding the error itself.

4.4.3 *Results*

The results are described within and between each expert group: language tutors (N = 2) and linguists (N = 4). The **Language tutors** described their work to be focused on personal or small groups of students. They also explicitly stated their pedagogic stance not to point out errors, but rather motivate learners to produce *anything* in their second language. Cohesion and vagueness of arguments were also interests. They seemed overwhelmed when presented with H-Matrix and showed resistance to most of the hierarchical matrix functions. However, they showed interest in the text view for displaying the tagged errors, but suggested some changes in order to make the view more useful for them.

Some of the suggested changes were to add another text panel for comparing two different essays side-to-side, to add more detailed explanations for the tagged errors and functions for editing, grading and commenting on each essay as a tutor. They also suggested a mechanism similar to a file version manager to track student development.

Regarding the hierarchical matrix view, most of their suggestions were related to removing functions, such as, re-ordering of the rows, different levels for the error categories (they preferred high-level only), weighting of the nodes, and the g^2 score. They also would want to work with much smaller groups, if not only individual essays.

They were generally confused with the interactions and scale of the data presented. For instance, they think that the granularity and categorization of the hierarchical errors were too detailed and specific. On the other hand, they mentioned that most of the features they have judged unnecessary could be interesting for researchers to understand language patterns in general. They thought this view would be unsuitable for students and learners to use directly.

The **linguists** were all researchers in language acquisition. Most stated they process and analyze large corpora as part of their pipeline, and some of the methods mentioned to perform this analysis were feature extraction, distributional statistics analysis, spreadsheets and heatmaps. Similar to the language tutors, they initially seemed overwhelmed when presented with H-Matrix for exploration, but after interacting with it, they said it feels easy to use.

For the text view, linguists suggested fading out tagged errors from categories which were filtered out. Another suggestion was to display

a visual representation summarizing the errors in one essay (similar to the details shown on hover, but permanently on screen); this can help to spot interesting essays to analyze. They also were pleased that this view allowed them to validate the feature extraction.

Regarding the hierarchical matrix view, they found it useful to compare groups of learners. One participant mentioned: *"It [matrix view] would be interesting to compare naturalistic learners with instructed learners."* On the other hand, they said the category names seem arbitrary and specific, which requires the linguist to be familiar with them to fully take advantage of the tool. In addition, most linguists wanted to use their own set of features and documents in H-Matrix. The tasks supported were appropriate and they saw it being helpful in their research and data analysis.

They suggested some interactions for the nodes, such as, filtering the essays by clicking at any level of the sub-trees, and to be able to expand or collapse all the children under a node.

In general, they were excited for adopting the tool in their pipeline to look at their own favorite feature phenomena and data patterns. They were particularly fond of the essay view as a mean to validate their feature extraction. They also mentioned that besides linguists, students can benefit from it by having their language teachers use the tool to explore and understand transfer effects.

4.5 DISCUSSION

My initial design challenge was to combine hierarchical categorization of linguistic features with distributional statistical analysis across learner groups. The result was to integrate a tree and a matrix visualization. For on-demand changes on the abstraction level in the feature categories, H-Matrix allows users to aggregate and re-weight rows dynamically in the matrix without having to reprocess the data. The essay view enables lower-level exploration, where users can extract text-level examples, as well as validate the feature extraction itself.

From the expert evaluation, we realized that linguists and language tutors have similar goals to understand and identify transfer effects to help second language learners. However, their methodologies greatly differ. Language tutors are mainly interested in smaller groups of essays. A way to address this would be to design a third view for a smaller set of essays. Perhaps, create profiles for learners display-

ing extracted errors history and similarity to larger learner groups, *e.g.* first language, proficiency. Regarding the granularity of the error categories, we can allow tutors to categorize and extract their own textual features using regular expressions.

In contrast, linguists had a positive response and seemed genuinely interested in using H-Matrix to explore their own corpora. I compiled a list of general functionality improvements from both groups of experts, such as, better labeling and colour legends, and from linguists regarding the tree interaction.

To tackle some of the feedback, a number of changes were applied to the tool. Initially, the matrix and essay view were on the same page. However, I noticed that participants rarely interacted with both at the same time. They preferred to explore each view separately as part of their personal workflow. For this reason, the views were separated into different tabs. The matrix view can still be used to filter and navigate to the essay view, but they are on two separate tabs, which gives them both more screen space and reduces the visual clutter (as shown in the figures). Hopefully, this change will help with the feeling of being overwhelmed the participants mentioned during the study.

In the matrix view, the textual explanations for the colour scale legends were improved, the right click menu options on the tree component for filtering the essay level view and options to expand/collapse all levels of a node to improve feature exploration was also created. On the essay view, I added all the current filter options and the error column as well as its in-cell visualization. The category filter in particular, addresses the complaints of the categorical colour encoding not being memorable on the essay tagged errors.

4.6 SUMMARY

This chapter presented H-Matrix, a visualization tool created to support cross-linguistic features exploration in large learner corpora. H-Matrix is the result of an interactive design approach where expert collaborators were consulted during development and evaluated a prototype with expert participants. The expert evaluation indicated some limitations, such as, the different expectations between language tutors and linguists, as well as some visual and interactive issues addressed in our discussion. Even though tutors, like linguists, are try-

ing to analyze and understand transfer effects, their method focuses on individual or a much smaller group of learners. H-Matrix was designed to deal with large learner corpora and it is over-powered for language tutors. On the other hand, linguists were excited with the functionalities of the tool. Their valuable feedback was used to improve the usability of H-Matrix in the presented version (Figure 4.1 and Figure 4.2), as mentioned in the discussion section.

As future work, one plan would be to evaluate this version of H-Matrix by using linguists own corpora and extracted features; in order to fairly evaluate the usefulness of H-Matrix with real use cases and application. A third view for managing essays of individuals, or smaller groups of learners should be designed to better accommodate tutors needs based on their feedback and our discussion. In addition, one can leverage the visuals and filters from H-Matrix combined with more learner-friendly explanations of linguistic annotations for a more learner-centred use case. Similarly, Generative Pre-trained Transformer (GPT) models can be trained to assist learners with a natural language interface that allows the to input naturalistic queries instead of building hypothesis using the visualization directly.

5

CARD-IT

An Application for Generating, Studying, and Testing Digital Flashcards for Learning Italian Verb Morphology

The verbal conjugation system is an important part of the overall acquisition process of a foreign language [47]. This holds especially true for so-called morphologically rich languages [124], i.e. languages in which significant information concerning syntactic units and relations is expressed at word-level [130]. Italian, in particular, is a highly inflected Romance language, as its verb inflection presents a rich variety of inflection endings for tense, mood, person and number [128]. In addition, Italian morphology is stem-based, which is significant from a processing perspective since Italian stems do not correspond to full words and must always bear some inflectional ending [37].

Foreign language learners are thus confronted with acquiring a rich system of verbal inflections (e.g., Pizzuto and Caselli [108]). While foreign language teaching practices focus on conversational skills and language production for fluency, it is known that linguistic awareness has a positive impact on foreign language learners, but it might be challenging to develop this awareness without foreign language immersion or explicit training. This leads many learners to look for alternative methods and tools to supplement their language learning curriculum. This behaviour is commonly displayed by self-determined or self-taught individuals, who are referred to as "autonomous learners" in this chapter.

Across different degrees of autonomy, autonomous language learners can plan, control, implement, and evaluate their learning goals. Autonomous learners require a wide variety of tools to achieve their different needs throughout their language-learning journey toward their desired level of fluency.

To support autonomous learners of Italian, Card-it was designed, implemented, and evaluated with the help of learners, experts, and teachers of Italian as a foreign language. Card-it is an online application that allows non-native learners to create digital flashcards — based on a semi-automatic approach — with which they can study

and test themselves on verb morphology. The digital flashcard design was chosen for two reasons: (1) it reflects a traditional way of learning vocabulary, and (2) digital flashcards have been shown to be a successful learning tool that students perceive well [20, 142].

Thus, Card-it supports both autonomous learners of Italian as well as learners in an active Italian course. Within a classroom environment, teachers can recommend and moderate learning material for practice outside the classroom. This would help learners to further develop their autonomy, which is one of the goals of language instructors' production-driven practices [2, 43, 69, 123].

5.1 EXISTING CALL APPLICATIONS FOR ITALIAN

An early CALL learning program for Italian was the Rosetta Stone Language Library *Italiano* I, the Italian version of the multimedia CD-Rom package Rosetta Stone Language Library that offered materials for several languages. The tool supported three skills: Listening Comprehension, Reading and Listening, and Reading Comprehension. The program was divided into 8 units, each containing 11 or 12 chapters. One major strength was its high-quality audio materials, which were recorded by native speakers [83].

Italian for Everyone is a tutorial for travellers at a beginner level, taking a situational approach. The program creates a virtual learning environment containing videos and dialogues from common holiday situations, aiming to immerse the learner in authentic learning contexts. Interactions are based on video, text, and audio. It comprises six different sections, each presenting another activity. Graphically, the user is situated in a lobby and can move around in different rooms, browsing shelves and books —- metaphors for Grammar, Dictionary, and Help [104].

Italian Encounters [120] is part of a learning software also available for French, German, Portuguese, and Spanish, targeting beginners and intermediate-level learners of Italian. It is interactive, including a mystery game, *Chi è il ladro?* (Who is the thief?) and twelve different thematic units. Like in Italian for Everyone, the units target real-life situations such as being in a restaurant, making a phone call, or asking for and giving directions. The learners are free to pick a unit from the graphical user interface. In dialogues, the system allows learners to record their voices, assuming the role of one of the speakers. After

recording, learners can replay the entire dialogue and listen to their performance.

The Italian Literacy Tutor (ILT) [31], the Italian counterpart of the Colorado Literacy Tutor (CLT), is a project to support reading in children whose second language is Italian. It integrates two sets of literacy tools, one based on speech and animation technology and one based on language comprehension technology which uses NLP methods. It focuses on reading vocabulary using a 3D animated talking head synchronized with recorded or synthesized speech. This is paired with illustrations. "Children will look at the images, label the item they see, listen to the words, see their spelling, pronounce them, hear how they pronounce them, and receive feedback." [31].

Summary Street [77, 126], a tool based on an NLP-statistical theory of meaning (Latent Semantic Analysis), supports comprehension. Here, children write summaries of chapters, and Summary Street automatically compares their input with the original chapter and provides feedback [31].

In 2015, Bernardinis et al. [11] introduced ICALL-IT, an ICALL learning system for Italian using three existing NLP tools: Morph-it! as lexicon [146], TUT as an annotated corpus [16] and TreeTagger as a part-of-speech tagger [119]. The tool has a user-friendly graphical interface and five exercises, including "Complete the Sentence." Here, the system picks a random sentence from the TUT corpus and removes one word, prompting the user to select the correct missing word from a list of possible answers. Other exercises aim to improve learners' vocabulary and verb conjugation skills, e.g., by having learners identify morphological features.

In recent years, Finite-State Morphologies (FSM) have been increasingly adapted for language learning environments. Kaya and Eryiğit [75] use a Finite-State Transducer for a Turkish word synthesis system and a word-level translation system between Turkish and English. The system targets both Turkish for native English speakers and English for native Turkish speakers. In an evaluation study, the developed system outperformed popular translation tools Google, Yandex, Bing, and Tureng for word-level translation with respect to grammatically correct outputs. Later said the system was used as a base for a mobile application that uses gamification to learn grammar in Turkish [48]; Turkish, like Italian, is also a morphologically rich language.

Another example is an ICALL system for two Saami languages based on Finite-State Transducers [7]. Besides the FSM, their system consists of a lexicon containing grammatical and semantic information and templates for question-answering drill generation. Yet, the FSM is the core component. The exercises include fill-in-the-blank, where the learners are given an infinitive and a personal pronoun for which they must fill in a blank containing the verb in the correct form —- based on a given tense. Furthermore, in 2020, a study compared two approaches to vocabulary learning: digital flashcards *vs.* wordlists [142]. With a total of 57 undergraduate pharmacy students, they conducted two technical vocabulary tests and a post-treatment survey. The results indicated that students learned more vocabulary through digital flashcards, and their perceptions about using digital flashcards to study were also positive. Card-it combines both approaches using an FSM to provide fill-in-the-blank and multiple-choice exercises via digital flashcards.

While there are CALL and ICALL systems that address verb morphology and vocabulary, some of which use similar morphological NLP tools as ours, our system is unique. In ICALL-IT [11], the vocabulary exercises are designed to ask the learner to select the correct answer from a list of suggestions. For example, in exercise 2, "Guessing Morphological features of the word," the system presents a specific word and asks the learner to pick matching morphological features from a finite set of possible answers. In contrast, Card-it has exercises where learners must type the answer without being presented with a list of options. Additionally, our system is designed for inclusion in classroom activities, allowing teachers to include Card-it in their curriculum. It also leverages digital flashcards, as it is a classic method of studying vocabulary and memorization.

5.2 CARD-IT DESIGN

Card-it's design and implementation followed an iterative approach, where a small preliminary expert evaluation with teachers of Italian with an earlier version of the application. This allowed us to obtain and implement the expert feedback leading to an overall informed design approach to Card-it's final version, as described in this section.

5.2.1 *Expert Feedback on Card-it's Early Design Iteration*

For the **preliminary evaluation** at early design stages, a small expert evaluation was conducted with two teachers of Italian at the Institute of Speech and Language at the University of Konstanz, which is the home university of one of the project collaborators. Where the early version of the Card-it system was demonstrated to teachers. While this version had all major learning and studying features, it did not have classroom management features yet. The focus of this preliminary evaluation was to determine the fitness of the flashcards and the quiz formats.

To obtain the teachers' feedback, Card-it and its features were presented to them in an online video call. During the call, they could ask follow-up questions and share comments and feedback with the investigator. After the call, the teachers could access Card-it to test and explore it on their own terms. Lastly, the teachers answered a questionnaire about their experience.

The **first preliminary evaluation results** showed that compared to the teachers' traditional and current ways of verb morphology practice with their students (methods such as worksheets and repetition drills), Card-it offers equivalent, more interactive and engaging activities. Another positive point mentioned by participants is the informative feedback available during a quiz. For instance, if a student submits a valid verb form that does not correspond to the prompted conjugation, Card-it shows the conjugation(s) of that incorrect verb form. When identifying the tense of a verb, if the incorrect tense is chosen, Card-it shows what the verb form would be with the chosen tense. On the other hand, one of the participants expressed concerns regarding their preference to write answers down on paper instead of typing them into the computer screen, which, in their opinion, helps with memorization of the verb forms themselves.

As non-native Italian speakers, both participants highlighted that Card-it would be a great tool for autonomous learners and that they would picture themselves using it occasionally to refresh their knowledge. As teachers, they liked the "Verb Search" view, where verb forms can be looked up in the system. Some of the suggestions they made to improve Card-it are now implemented in the presented version, which were:

- filtering within a deck based on the verb or conjugation;

- presentation of one card at a time —- the presentation view (Section 5.2.4.1);
- addition of compound tenses —- the most popular ones were incorporated;
- review of the verb forms —- teachers found some incorrect ones;
- classroom features to help teachers share collections with specific groups of students (Section 5.2.5).

Other suggestions, such as adding sample sentences for verb forms, a team quiz to be used during a live lecture, or new quiz types, including fill-in-the-blank, were valuable but not within the immediate scope of Card-it.

5.2.2 *Semi-automated Creation of Digital Flashcards & Finite-State Morphological Analyzer*

Instead of manually curated and created content, this application leverages a large-scale finite-state morphological analyzer which can both analyze a user's input and dynamically generate forms. On the one hand, FSMs are widely used in NLP tools as part of a system's text preprocessing pipeline. On the other hand, such linguistic tools are rarely directly used by language learners and teachers due to the technical complexity they require. Bernardinis et al. [11]'s approach was followed when incorporating their NLP and morphological information in a structured database to implement their application (ICALL-IT); according to them, this design choice made the data access faster while increasing the system's scalability. Thus, a script that exhausted all possible word forms generated by the FSM was created, and that output was then restructured into a database. The database then returns forms given specific parametrized features or returns a specific form's morphological analysis. This FSM-generated data is the core component for generating parameterized flashcards in Card-it. Therefore, Card-it can be seen as a way for learners and teachers to benefit from the FSM by using it in a user-friendly application.

Besides learners being familiar with flashcards as a common way of studying vocabulary and other concepts that require memorization and repeated practice, cards are two-sided, representing both func-

tions or concepts of the FSM: **analysis** and **generation**. For instance, when given the verb *finirò* for analysis, the FSM returns that *finirò* is the verb *finire* in the first person singular, future tense, and indicative mood. For generation, one can request the verb form for *finire* given the morphological attributes above, and the FSM returns *finirò*. Therefore, when facing a verb on a card (Figure 5.1, left-panel), the learner sees its morphological description (Figure 5.1, center-panel) by flipping the flashcard, and vice-versa.

Figure 5.1: Both sides of a flashcard can be seen where side 1 (left) shows the verb form – *finirò*; and side 2 (center) shows the verb conjugation (morphology). The card on the right shows the 'spell conjugation' card format, which gives the user feedback on the typed answer, similar to the 'Conjugate' quiz type shown in Figure 5.2.

CREATING A FLASHCARD. Because Card-it's flashcard generation is powered by the FSM analyzer, it can be used to quickly analyze a form or to look up entire verb paradigms where the users can define parameters for features, such as tense, mood, etc. Optionally, they can choose to leave any feature unspecified. Depending on the user's selection, the application returns the corresponding flashcards —- in this context, each flashcard represents a verb form resulting from the user's selection.

ORGANIZING FLASHCARDS. The presented cards can be saved and organized into a new or an existing deck for studying. The organization and sorting of decks and cards allow learners to study verbs based on specific topics or forms, e.g., when one might choose to focus on subjunctive forms or past tense only. Decks can be organized into collections, which can have their visibility set as either public or private; public collections can be seen by any Card-it user, while private ones are only visible to the user who created the collection.

5.2.3 *Meta-linguistic Awareness*

As studies have shown, knowledge of the underlying linguistic concepts benefits the acquisition of a new language (e.g., Ellis [46] and Heift [61]), which led us to enrich Card-it with explanations of linguistic terms such as mood, tense, conditionals, and others using appropriate visual components, to allow learners to identify linguistic patterns and raise meta-linguistic awareness over time. The linguistic explanations are embedded within the Card-it interface.

While the main interface and instructions are available in English or German, Card-it describes linguistic terms in the Italian language (i.e., *Passato Remoto*) to support second language immersion. The teachers who joined our preliminary evaluation saw it as beneficial to students.

5.2.4 *Studying and Testing Mode*

Analogous to physical flashcards, Card-it comprises two different phases: training and testing. The training phase allows the user to study and learn verbs and their inflectional forms; the testing phase quizzes and helps the learner measure their progress. During training, learners can gradually "move" cards from the training set into the test set once they feel comfortable with being tested on them.

5.2.4.1 *Studying Mode*

In training or studying mode, learners have a few options for interacting and visualizing the cards. Firstly, Card-it has two flashcard formats for training: 'flip cards' and 'spell conjugation.' The **flip cards format** is a read-only format where the information on the verbs is displayed, as shown in Figure 5.1, in the left and center panels. The **spell conjugation format** displays the conjugation plus the verb's infinitive form. It has a text box for the learner to type the verb form for the given verb/conjugation —- Figure 5.1 (right-panel) shows an example of this flashcard format. While using the spell conjugation format, learners can easily assess themselves without their performance being tracked.

VIEW MODES. Card-it has two view modes: grid and presentation. In the **grid view**, cards are organized in rows and columns, forming a matrix on the screen, where learners can scroll up and down if the number of cards overflows their screen. While in the **presentation view**, cards are displayed one at a time on the screen, and the learners can navigate among them (using the 'previous' and 'next' buttons).

INTERACTING WITH CARDS. Learners can choose how they interact with the cards while using the 'flip cards' flashcard format (in either grid or presentation view). They can display one side at a time or both sides (conjugation and verb form) at once. For example, side 1 (Figure 5.1, left-panel) is displayed, and when the learner hovers over the card, it flips to side 2 (Figure 5.1, center-panel), or vice-versa; in the second card display option, both sides of the card are displayed at the same time.

To enable searching for a specific form within a deck, learners can filter through them by linguistic attribute(s). Furthermore, learners can activate an edit mode, which allows them to easily select one or multiple cards that can be moved or copied into different decks. This can help learners track which cards they want to build a particular deck with for testing. Alternatively, learners can just quiz themselves on their full original decks.

5.2.4.2 *Testing Mode*

To help learners evaluate themselves, two types of quizzes were designed that can be activated based on a deck of cards: 'Identify Tense' and 'Conjugate.' Alternatively, learners can choose 'Mixture,' which randomly chooses between the quiz types for each card tested. Once a deck is chosen, the three quiz types are available to start.

For each card in the chosen deck, the 'Identify Tense' quiz type presents learners with a verb form and prompts them to pick from three choices of tenses (Figure 5.2's top-panel). The system then displays if learners are correct; if not, learners can pick a different answer until they make the right choice. However, whenever the learner picks an incorrect choice, Card-it presents the correct card (verb form) that corresponds to their selection as feedback. This way, learners are offered a chance to reflect on their choices — an example is shown in Figure 5.2 (bottom panel).

Figure 5.2: On the top panel, the identify tense quiz where the learner is tasked to select, among three choices, the tense of the verb *verrò*, the verb *venire* in the first person singular (*io*). On the bottom panel, see the app's response to the learners choosing the incorrect option. It presents the learner with the card (verb form) of *venire* in the chosen (incorrect) tense providing the learner with feedback on their mistake. The learner can continue choosing until they find the correct choice among the three options.

During the 'Conjugate' quiz, learners are presented with the infinitive form and a specific conjugation and are expected to type in the resulting verb form. Tense and mood are specifically spelled out, whereas number and person are expressed by the respective personal pronoun (here *io*) (Figure 5.3's top panel).

In this quiz type, learners can request a hint, as shown in the right panel of Figure 5.3, which presents three choices, including the correct answer that can be chosen to answer the quiz. The "hint" feature allows learners not to be discouraged if they cannot think of the answer at the time or if they forget the spelling of the verb form itself. In the case of submitting an incorrect answer, either by typing or selecting from the hints, learners receive feedback. Similarly to the Identify

Tense quiz, Card-it displays the cards (conjugations) that are equivalent to their incorrect choice (verb form); an example can be seen in Figure 5.3 (bottom-panel).

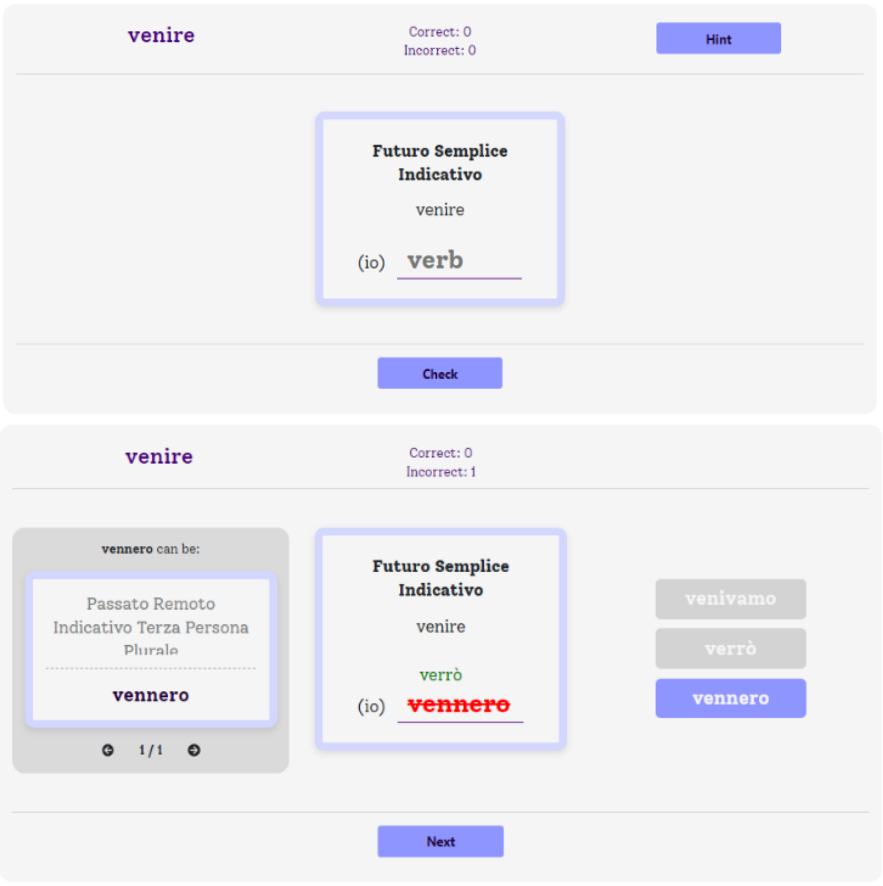


Figure 5.3: On the top panel, see the conjugate quiz where the learner is asked to type the verb form *verrò*, which is the *Futuro Semplice Indicativo* of the verb *venire* in the first person singular; it also shows the "hint" button on the top right corner. On the bottom-panel, see the app's response to activating the "hint" feature, where learners are offered multiple choices to the right of the quiz task; the screen shows the learner has chosen the incorrect answer (*vennero*), the correct one (*verrò*) is displayed in green above the incorrect one; to the left of the quiz task, the app displays other cards that are corresponding the to incorrect (*vennero*) verb form.

5.2.5 *Classroom Management*

To support the adoption of Card-it in a classroom setting, a set of classroom management features was designed for teachers. The tasks supported in this module are: (1) creating classrooms and generating a unique code that can be shared with students allowing them to

join it, (2) sharing specific collections to one or multiple classrooms, and (3) visualizing the students enrolled in the classroom and their progress.

5.2.5.1 *Teacher Features*

From Card-it's "Classroom" tab, teachers can see their existing classrooms or create new ones. Teachers can add collections to their classrooms for enrolled students to access. Each classroom is associated with a system-generated unique code (i.e., "i41xgs"). Teachers can share this code with their students so they can join the classroom.

Besides the collections shared with a class, the teacher can see the list of students enrolled. Furthermore, Card-it presents a set of basic statistical information on the students' progression. For example, the teacher can select a specific deck and analyze each student's attempt (overall score). To understand the class's progress over time, teachers can view the average scores per quiz attempt. Lastly, Card-it shows each card's number of correct attempts in a quiz. This enables teachers to pinpoint specific cards students had the most trouble with.

These filter selections also include the quiz type and their specific statistics —- this can show if students are having more trouble recognizing a tense or conjugating certain verbs. Figure 5.4 shows the statistical overview of students' performance in a quiz. Teachers may filter for a specific collection (here: *Presente Indicativo*), deck (here: Regular Verbs) and quiz type (here: Conjugate). Additionally, teachers see the score for each student. And, Figure 5.5 illustrates how teachers can check on the groups' performance on every single card, sorted from the least correct to the most correct.

5.2.5.2 *Student Features*

For students, Card-it's "Classroom" tab displays all their enrolled classrooms and allows them to join a new classroom using the class code provided by their teacher. From a specific classroom, students can see the collections and decks shared by their teacher. Students are free to study and quiz themselves on all classroom material. At this point, they might or might not have explicit instructions on which collections/decks to work on, as it depends on their teacher's curriculum plan. However, it is presumed that teachers would provide guidelines for their students.

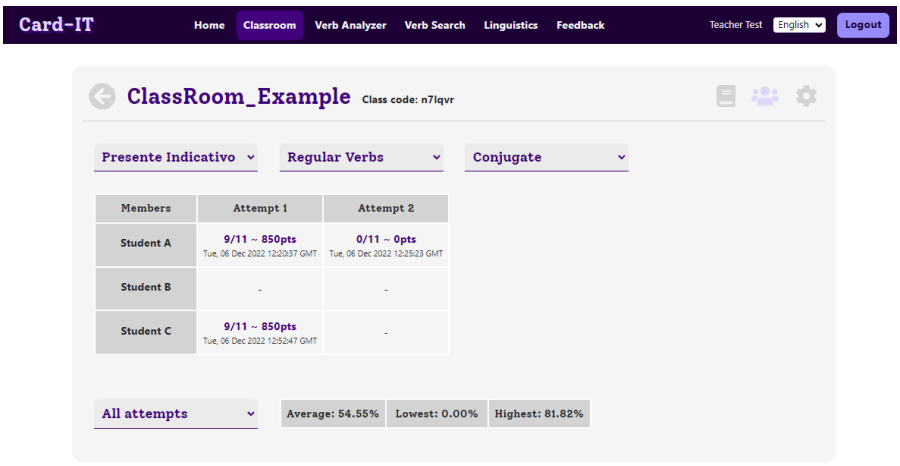


Figure 5.4: Example of a statistical performance overview.

Similar to the student's personal collections/decks, they can access both studying and testing modes for each deck in the classroom. Students can also copy entire collections from a classroom set to their personal set. This enables them to save the study materials for posterity (even after their classroom becomes unavailable) and gives them a chance to practice the material without their progress being tracked. Students are tracked on the material only if they do activities within the classroom environment.

5.3 CARD-IT IMPLEMENTATION

Card-it is a web-based desktop application. It does not require additional software installation besides a web browser, allowing users to access it across different platforms. The main User Interface (UI) components were designed for a desktop experience.

Card-it's system can be divided into two parts: back-end and front-end components. The front-end (UI) was developed using ReactJS, while the back-end (server) was built using Flask, based on the Python programming language. The application's server side also contains the database (MySQL), which stores usage data such as user accounts and the verbs generated by the FSM.

As explained in Section 5.2.2, the FSM is a powerful computational linguistic tool leveraged to create our extensive set of verb morphology. However, for Card-it, the system needed to be scaled for the usage of multiple simultaneous users, and using the FSM in its natural interface without adding a request pool system imposed limitations

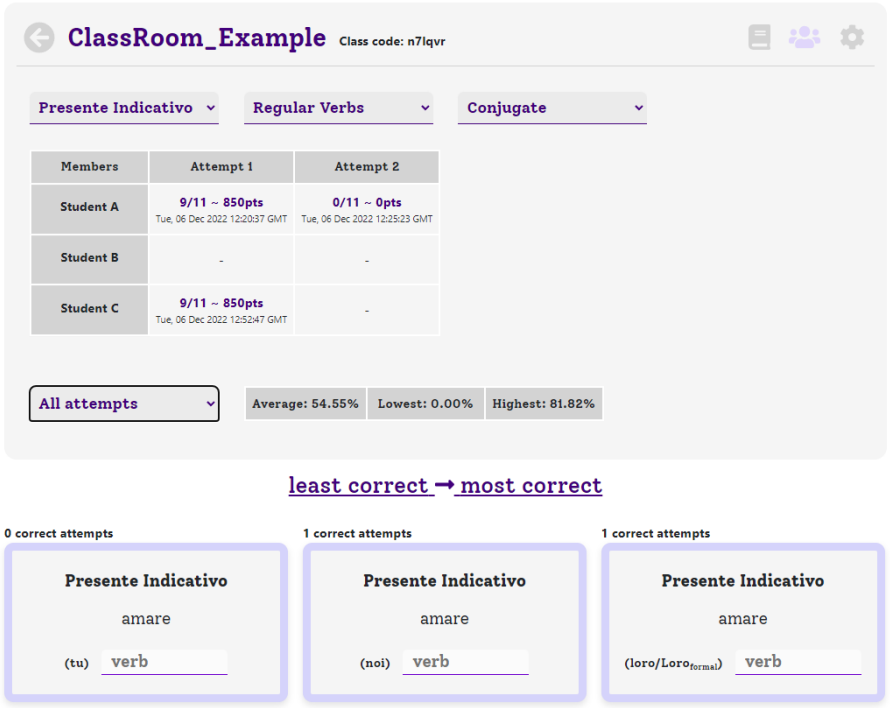


Figure 5.5: Example of a statistical performance overview.

on this non-functional requirement. That said, similarly to ICALL-IT [11], scripts were run on the server side that can synchronize and keep Card-it's database up-to-date with the knowledge generated by the FSM.

In its current state, Card-it can support the following main features: semi-automatic flashcard generation, organization of flashcards in private or public collections and decks, visualization of flashcards using specific display formats for studying cards from a deck, different quiz types to assess learning progress with a deck, and basic teacher and classroom features to manage groups of learners.

5.4 CARD-IT EVALUATION

After implementing changes to reflect the feedback from the first preliminary evaluation (REB #14426); we ran a second expert feedback round (REB #14426) with teachers of Italian, and a longitudinal (three weeks) learner's evaluation (REB #17317). All evaluations were run remotely.

5.4.1 *Expert Feedback*

We invited individuals who were teachers of Italian via our professional networks; in total, five teachers from two institutions in Germany signed up. Of those, three volunteers completed the study. After receiving the study instructions, they had 14 days to complete all steps remotely, then were compensated with a $20 Amazon gift card.

5.4.1.1 *Methodology*

We ran our second preliminary evaluation with experts remotely. This evaluation aimed to obtain expert feedback on the final design implemented for Card-it. Participants were asked to follow three steps to complete the study: (1) Watch a pre-recorded video demo of Card-it's main features; (2) Explore Card-it on their own using both teacher and student account types; (3) Answer a survey about their experience using Card-it.

In the survey (Appendix B), participants were asked 5-point Likert Scale questions on general usability, the potential for classroom adoption, and specific questions on different features such as studying and testing modes. Experts were also asked to answer a section where they give their opinions from a student perspective. All teachers of Italian (experts) recruited were non-native Italian speakers.

5.4.1.2 *Results and Discussion*

The system's usability was rated positively, with two experts selecting easy and one very easy. All experts rated both quiz types, "Conjugate" and "Identify Tense," as "appropriate" or "very appropriate." One expert mentioned that the quizzes were their favourite features. When asked to rate the classroom management usability, two chose "good" and one "very good." As a follow-up, theyr were asked regarding how to create a classroom: one expert found it "difficult," elaborating "not easy to find the different features" and the others found it "easy." They all mentioned that they could foresee themselves using Card-it for homework in their classes or as a tool for students to self-study at home. When asked to take on a student's perspective, they all rated the quiz and verb look-up features of Card-it "most useful." To improve Card-it, they suggested including translations and exam-

ple sentences containing the individual verbs as it would be useful for both students and teachers.

5.4.2 *Learner Evaluation*

Learners of Italian as a foreign language were contacted using a social media community with more than 30,000 members, namely "Let's Learn Italian" on Facebook. Volunteers filled out a screening questionnaire with their contact information and Italian proficiency level when signing up for the study. Using the inclusion criteria (must be a learner of Italian as a second language) and their self-reported Italian proficiency, 31 participants were recruited and finished the study. Participants were compensated with $10 (of their local currency) per session, up to a total compensation of $30 for all the 3 sessions.

5.4.2.1 *Methodology*

Considering this was a longitudinal (three weeks) remote study, only the "Conjugate" quiz type was enabled on Card-it. This simplified the instructions and control over the evaluation. Using Card-it's Classroom feature, specific decks were set up for the study, and each participant was assigned a specific one to use in all sessions. Each deck contained 4 different verbs in first, second and third person singular (*io*, *tu*, *lui/lei*) conjugations of the following three tenses: *Presente Indicativo*, *Futuro Semplice Indicativo*, and *Presente Condizionale*. This resulted in 36 cards (verb forms) per deck.

STUDY STRUCTURE. The study had three remote sessions, spaced by 7-day intervals. Before starting the first session, all participants took an Italian conjugation test ($\texttt{entry}_{\texttt{test}}$) and filled out a background survey. The same conjugation test ($\texttt{exit}_{\texttt{test}}$) was taken again after the last session (Session 3) of the study. The conjugation test and Card-it quiz scores to measure their knowledge pre- and post-study and as the study sessions progressed.

After completing the $\texttt{entry}_{\texttt{test}}$, participants had access to video instructions on how to use Card-it for the study tasks. After exploring Card-it as much as they wished, participants submitted their "Conjugate" quiz on the app; lastly, they submitted a short usability survey based on their first impressions and experience in study session

1. The Card-it quiz completed by participants was built upon an assigned deck with 36 cards.

In between sessions, participants were instructed to use Card-it as often as they would like. After seven days of submitting their session quiz and usability survey, participants were instructed to take the "Conjugate" quiz on the app and submit a usability survey ($usability_2$ and $usability_3$, for study sessions 2 and 3, respectively). After completing study session 3, participants took the $exit_{test}$.

STUDY VERBS. The four different verbs included in the decks were chosen from the 12 verbs shown in Table 5.1. To expose learners to a diverse morphology, the 12 verbs were split considering their (1) regularity (regular and irregular), (2) commonality in regards to usage frequency in the Italian language[1], and (3) conjugation pattern; this resulted in four pools A, B, C, and D, respectively:

- $regular \times common$,
- $regular \times uncommon$,
- $irregular \times common$, and
- $irregular \times uncommon$.

The three verbs in each pool were chosen to represent the three conjugation patterns in Italian determined by the endings of the infinitive verb (*-are*, *-ere*, *-ire*). Each learner was assigned a combination of verbs, and each verb was from a different pool. It was also balanced regarding the conjugation classes — a conjugation pattern is always presented twice, but it was balanced overall. Some sample sets are *trovare*, *suggere*, *dire*, *dare* and *trovare*, *riputare*, *dovere*, *erudire*, see Appendix C for the full combination list.

For the $entry_{test}$ and $exit_{test}$ tests, all learners encountered a subset of eight out of the 12 verbs shown in Table 5.1, resulting in 72 verb forms (eight verbs $\times$ [*io*, *tu*, *lui/lei*] $\times$ three tenses). To evaluate the learning of verbs without prior exposure to them, four verbs were excluded from the $entry_{test}$ and $exit_{test}$. Thus, these forms can only be seen by some of the participants in the three Card-it quiz sessions for comparison ($exposed \times missing$). This resulted in 13 out of

1 Frequency of usage of the verbs was determined based on the frequency lists from WaCky corpus for Italian, published under CC Attribution-Noncommercial-Share Alike 4.0 International. (Link: https://wacky.sslmit.unibo.it/doku.php?id=frequency_lists)

Table 5.1: Set of 12 verbs used in the study split into their morphological pools and conjugation pattern. verbs* were only present in the Card-it and not in the $\texttt{entry}_{\texttt{test}}$ and $\texttt{exit}_{\texttt{test}}$ tests.

	Regular		Irregular	
	A (common)	**B (uncommon)**	**C (common)**	**D (uncommon)**
-are	*trovare (to find)*	*riputare (to ripute)*	*fare* (to do)*	*dare (to give)*
-ere	*mettere (to put)*	*suggere (to suggest)*	*dovere (to have to)*	*parere* (to seem)*
-ire	*sentire* (to feel)*	*guaire* (to yelp)*	*dire (to say)*	*erudire (to teach)*

the 31 participants studying only exposed verbs and 18 participants with one or two missing verbs, as any combinations containing three or more missing verbs were not used $\texttt{entry}_{\texttt{test}}$, which resulted in 31 unique combinations. We decided to exclude the missing verbs to get an indication of whether learners can improve on those verbs despite them not having been studied and to reduce the number of possible combinations.

5.4.2.2 *Participants Learning Background*

All participants declared themselves as non-native speakers of Italian as it was a participation eligibility criteria for the study. Additionally, they self-reported their proficiency level in Italian using the Common European Framework of Reference for Languages (CEFR): beginner (A1), pre-intermediate (A2), intermediate (B1), and upper-intermediate (B2). None of the 31 participants stated to be at advanced levels (C1 or C2), instead the majority self-declared as A1, followed by A2 , B1, and B2 — Table 5.2 shows the distribution.

When asked about their main learning method for Italian, the majority reported to be "self-taught" (N = 18), in contrast to "university" (N = 13), "language institute" (N = 8) and "school" (N = 6). Note that participants could select more than one category for their language learning context. Additionally, an equal number of participants have been learning Italian for "less than a year" (N = 12) or "more than 1, but less than 2 years" (N = 12), with only a few reporting "more than 2, but less than 3 years" (N = 3) and "more than 3, but less than 5 years" (N = 4). This data lines up with their self-reported language proficiency. As Card-it is a digital application, when asked about their experience with other online or digital learning applications for Italian. The five most listed learning applications sorted by

Table 5.2: Tabulation of the number of participants per proficiency level as well as overall. This table also shows the score distribution of the Italian conjugation tests ($entry_{test}$ and $exit_{test}$) along with the Card-it quiz scores in each study session.

Proficiency Level	N	$entry_{test}$ (out of 72)		$exit_{test}$ (out of 72)		Session 1 (out of 36)		Session 2 (out of 36)		Session 3 (out of 36)	
		MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD
A1	12	41.8	20.0	54.4	13.7	25.4	6.5	25.3	3.1	27.1	5.2
A2	8	63.3	3.3	65.0	4.2	30.9	1.9	31.0	1.9	32.0	2.1
B1	6	64.2	2.0	66.2	1.9	32.2	0.8	31.5	2.4	34.8	1.8
B2	5	64.4	2.1	67.4	2.1	30.0	1.2	31.0	1.9	33.2	1.1
All (overall)	31	55.3	16.4	61.5	10.4	28.9	4.9	29.0	3.8	30.7	4.7

number of mentions are Babbel (29%), Duolingo (26%), Rosetta Stone (26%), FluentU (19%) and Italki (16%).

5.4.2.3 *Test Score Results*

Table 5.2 shows the distribution of participants per language level and their average test scores. Kruskal Wallis H tests showed that the proficiency level has an effect on $entry_{test}$ score distribution ($p < 0.05$); however, no effect was found on their $exit_{test}$ scores ($p > 0.05$).

The Paired Wilcoxon signed-rank test showed that the overall Card-it score between sessions 1 and 3, and 2 and 3 resulted in a statistically significant improvement for learners ($p < .05$ and $p < .005$, respectively); as well as the improvement between $entry_{test}$ and $exit_{test}$ ($p < .001$). Furthermore, the score improvement between $entry_{test}$ and $exit_{test}$ for A1 and B2 learners was statistically significant ($p < .05$). In addition, a paired $t-test$ (normally distributed data) showed a statistically significant improvement between quiz sessions 1 and 3 and sessions 2 and 3 for B2 learners ($p < .01$ and $p < .05$, respectively). Figure 5.6 shows the test score distribution per Italian language proficiency level self-reported by the participants; statistically significant pairs are annotated (dotted lines).

CARD-IT USAGE FREQUENCY AND TEST SCORE IMPROVEMENTS. When categorizing the participants by their self-reported usage of Card-it between sessions. For the $entry_{test}$ and the $exit_{test}$ improvement (Table 5.3); those who reported using Card-it at least once between each pair of study sessions improved by 53% on $exit_{test}$ in comparison to $entry_{test}$, and those who used Card-it at least once

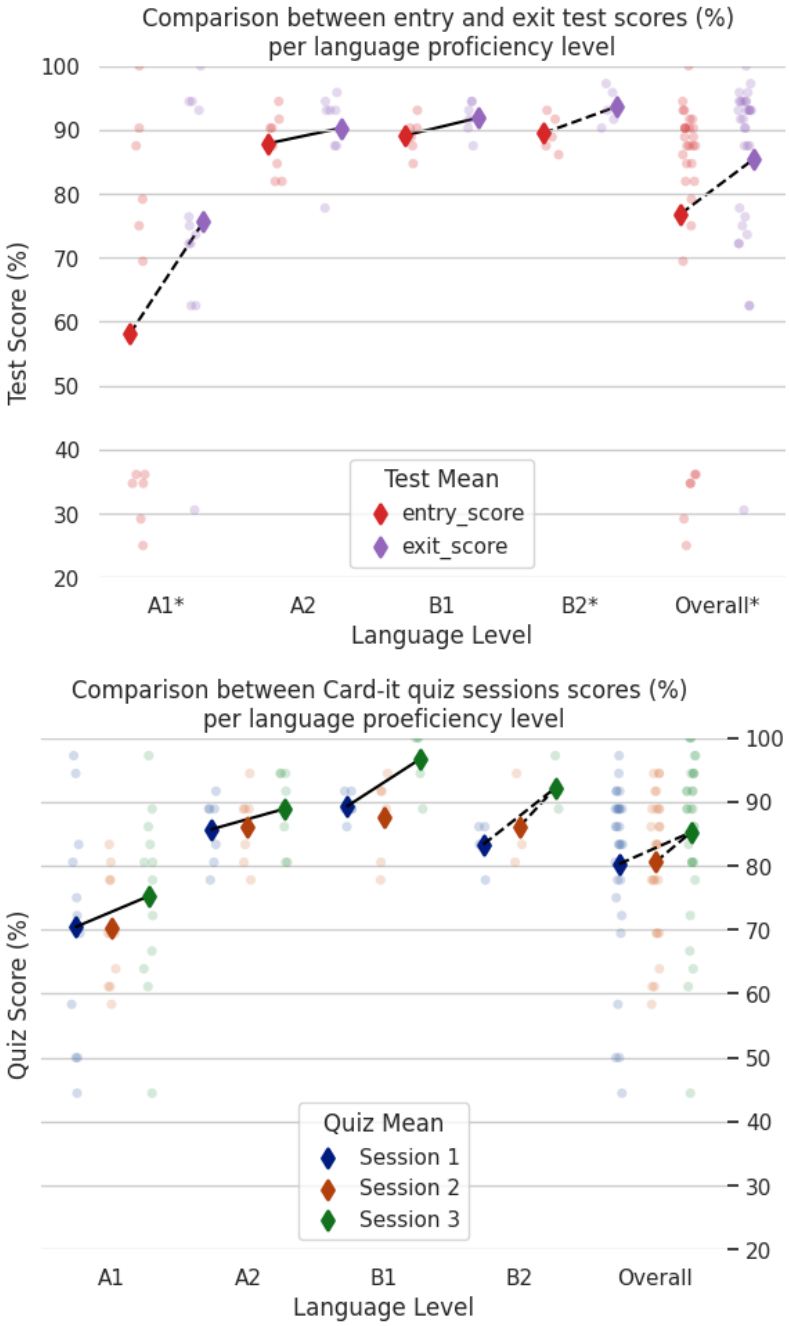


Figure 5.6: $\text{entry}_{\text{test}}$ and $\text{exit}_{\text{test}}$ score distribution per language proficiency level, which shows participants of all language levels improved their test scores after the Card-it study. A similar trend can be observed in the distribution of Card-it quiz scores throughout the three sessions.

Table 5.3: Card-it usage frequency and test score improvements.

Usage	Participants	% Improvement $entry_{test} \times exit_{test}$
Participants who used Card-it at least once...		
... between each and all sessions	9	53%
... between Sessions (1 and 2), or (2 and 3)	5	29%
Participants who did not use Card-it between sessions	17	4%

between Sessions 1 and 2, or 2 and 3, improved by 29%. In contrast, those who did not use Card-it between any sessions improved only by 4%. However, there was not a statistically significant difference between the relative improvements between sessions and tests.

CARD-IT HELPED PARTICIPANTS IMPROVE THE SAME, INDEPENDENTLY OF PREVIOUS KNOWLEDGE. There was similar improvement between sessions for the set of verbs to which participants were exposed (*exposed* improved by 72% between sessions 1 and 2, 78% between sessions 2 and 3, and 85% between sessions 1 and 3) in both the knowledge test and Card-it quizzes versus the set of verbs missing (*missing* improved by 72% between sessions 1 and 2, 75% between sessions 2 and 3, and 83% between sessions 1 and 3) but tested in the Card-it quizzes. The Paired Wilcoxon signed-rank test showed both sets significantly improved between sessions 1 and 3 ($p < .05$).

5.4.2.4 *Card-it Usability*

After each study session, participants answered usability questions about their experience using Card-it; questions were asked in positive and negative framings, and the results were consistent. All questions use a 5-point Likert agreement scale ("strongly agree" – "strongly disagree"). Regarding Card-it's overall usability; whether or not it is easy and intuitive to use. Participants were also asked for their opinions on specific Card-it functions, in particular, its quiz format, feedback information, and hint function. The final usability survey, after study session 3, included questions to assess if participants would recommend Card-it to other learners and how they rate Card-it compared to other language learning applications they may have used to learn Italian.

Figure 5.7 illustrates that the participants rated Card-it as an overall easy application to use, and its features were rated as helpful for

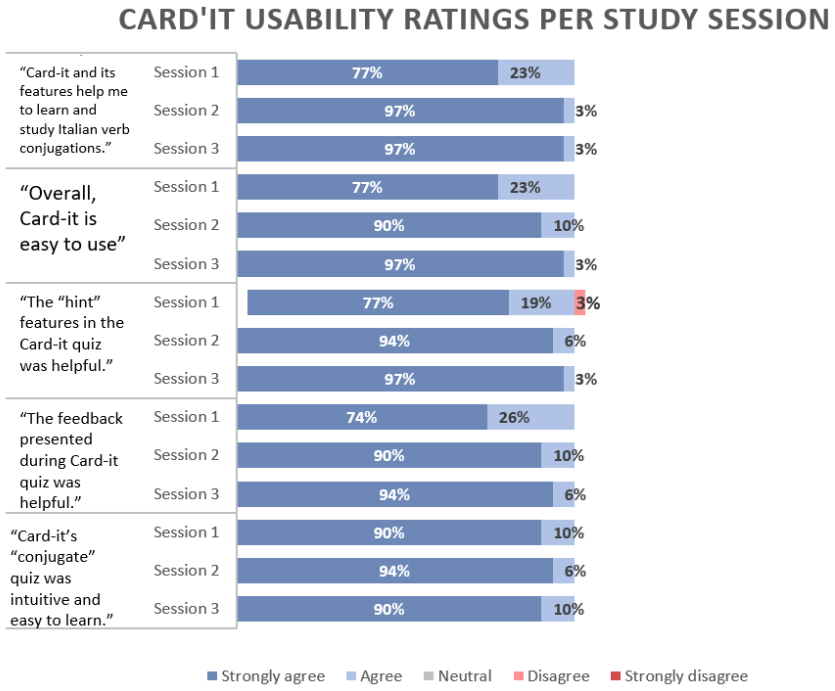


Figure 5.7: Results for the items targeting both overall and specific Card-it features and user experience from the usability surveys after each study session. In all items, participants rated Card-it very positively, and one can observe an increase in the category "strongly agree" over the course of the study.

learning and studying Italian verb conjugations. While all answers were either "strongly agree" or "agree," there was an increase in the choice "strongly agree" throughout the multiple study sessions. Figure 5.7 also shows ratings for specific features of Card-it. The participants evaluated the quiz format "Conjugate" as intuitive and easy to use, and its feedback and hint function as helpful.

In the final usability survey after study session 3, the main interest was in the participants' overall rating of Card-it concerning possible future use. The vast majority stated that they would recommend Card-it for beginner learners of Italian ("strongly agree" = 97%; "agree" = 3%), with a decrease when asked about intermediate learners ("strongly agree" = 77%; "agree" = 23%) and advanced learners ("strongly agree" = 65%, "agree" = 23%; "neutral" = 10%; "disagree" = 3%). When asked, all participants could picture themselves continuing to use Card-it beyond the scope of the study. Some added comments on how they would incorporate Card-it in their study routine: *"I would use Card-it to supplement class lessons and do revision time and*

time again" (P20) or that they would be *"making use of the verb analyzer, the quiz feature and others"* (P31).

Lastly, participants were asked to rate several statements that compared Card-it to other tools or apps they used or know of, to study Italian. In all cases, Card-it was perceived as a better option, including *"Card-it is a better option for supplementing a course or class lessons"* ("Strongly Agree" = 77%, "Agree" = 23%), and *"Card-it is more appropriate for the study of verb conjugations"* ("Strongly Agree" = 90%, "Agree" = 10%). Appendix D contains all items.

5.4.2.5 *Discussion of Learner Evaluation*

Across the three sessions, learners improved their verb conjugation scores significantly. Our main goal for the learner evaluation was to observe an overall improvement in the participants' performance on verb conjugation over the span of their participation in our learner study, i.e. an improvement that can be attributed to their use of Card-it for three weeks. As the results for the knowledge test in the entry and exit quiz show, a statistically significant improvement ($p < .001$) can be observed, confirming our initial hypothesis. When looking at the improvement between the entry and exit quizzes with respect to language proficiency levels, one can see that A1 and B2 improve significantly ($p < .05$). While this result is less surprising for A1 learners, one may not expect such a strong improvement for B2 learners. On the contrary, B2 learners were already expected to know the quizzed tense and verb forms and to achieve high scores throughout the entry and exit quiz, as well as all three quiz sessions. When looking at participants who self-declared as B2 and their significant improvement between the $entry_{test}$ and $exit_{test}$; it may be attributed to either them overestimating their level of proficiency or to the small sample size ($N = 5$).

Regarding the amount of time participants spent using Card-it in addition to the quiz sessions, participants who used Card-it the most also showed larger test score improvements between $entry_{test}$ and $exit_{test}$ scores (53%, 29%, and 4%; for those who used it between all sessions, between only any two sessions, and those who never used it in between any sessions, respectively). Despite those improvement differences not being statistically significant, it may indicate that Card-it might be helpful when used regularly; a hypothesis is that

this can be evaluated in future work along with a larger pool of participants.

In summary, Card-it is easy to use and helpful to learners. The participants regard Card-it as a fitting tool for autonomous and classroom learning. Throughout the study majority participants added comments such as *"I have never seen a similar tool to study conjugations that was that easy to use"* (P38), *"I am a self-taught learner at times and Card-it is a great app for learning and understanding tenses"* (P28), and *"it is a good option for self-taught learners"* (P21). When allowed to leave comments or questions about their experience using Card-it in general, some participants mentioned that *"Card-it is educative"* (P25), *"[an] awesome application for studying verb conjugations"* (P40), and *"[Card-it] has a good UI suitable for studying"* (P36). Concerning its design, one participant explicitly wrote, *"I like the card format of displaying information and the flipping of the cards"* (P33).

While majority of the comments were positive, a around 10% chose to not leave any comments, and across the three sessions some comments were composed by constructive feedback or description of an usability issue. For example, *"I expected the verb analyzer to show more information about the verb analyzed"* (P20), *"I think without the explanation videos it would have been challenging for me to understand, what to do. But the short videos were enough so that I had no problems then"* (P17), and *"I confess to having a great learning experience. It would be nice if Card-it taught the whole Italian grammar"* (P27).

5.5 OVERALL DISCUSSION, LIMITATIONS, AND FUTURE WORK

Card-it successfully integrates computational linguistics resources in an interactive web application for learning and training Italian verb conjugations. While the presented evaluations and its current implementation include Italian only, the technical pipeline demonstrates that other languages can be integrated into the application, given the availability of their equivalent FSM. While Card-it is technically built on a linguistic tool, the application interface and feature design were focused on providing a simple and friendly experience for learners to achieve their learning goals. The evaluations showed that Card-it is indeed simple and easy to use and that learners found it to be an appropriate method to study Italian verb conjugations as a complement to other studying activities (i.e., classes) or for those who are

autonomous learners. The learner evaluation also indicated that Card-it's design and flexibility allow learners to plan, control, and evaluate their learning goals. The same positive results were observed when learners were asked about it and stated that they could see themselves using Card-it in the future. While there was no baseline or screen based on previously used applications, in comparison to other language learning applications used by participants in the past, Card-it was still favoured by the participants for conjugation training.

While our iterative design practices and evaluations have led to positive results, Card-it can be further improved to address some of the current limitations mentioned by experts and learners. This includes sentence examples, translations, and the pronunciation of verbs. Upon reflecting on these valuable comments, there is a list of proposed improvements that can inform future CALL research and future work directions for Card-it.

Another limitation of this project is the learner evaluation of Card-it, which was conducted with a relatively small sample of learners without the control or rigid assessment of participants' language proficiency levels, limiting the generalization of findings. Larger-scale studies with diverse learner profiles would provide more robust evidence of its effectiveness and usability.

Possible other features to implement in Card-it include a gamification mode that would leverage the social aspects of language learning, fill-in-the-blank exercises to practice conjugations on a sentence/contextual level, and a spaced-repetition model to recommend specific cards or decks for further study based on learners' performance. Additionally, integrating another Italian linguistic tool, specifically an implemented syntactic grammar, to include exercises at a sentence and syntactic level will likely benefit the use cases for both teachers and learners.

Future work aims to conduct a longitudinal evaluation study in real-life Italian classrooms in collaboration with teachers to tailor Card-it to learner needs further and expand its usability in a classroom setting. Given Card-it's modular architecture, where the Finite State Lexicon provides the database of verb conjugations used, the application can easily be adapted to another language by including an FSM of said language.

5.6 SUMMARY

This chapter presents Card-it, a web-based digital flashcard application that leverages an FSM to enable the semi-automatic creation of study materials for learning Italian conjugation. It discussed the existing CALL application for learning or studying Italian and described Card-it's iterative design process and features.

The learner evaluation showed that Card-it supports autonomous learners of Italian as a foreign language in studying Italian verb forms. The expert feedback showed promising results for Card-it as a supplementary tool to aid teachers and their students in a classroom setting. The future work includes implementing gamified features, spaced-repetition algorithms for more personalized learning, and extra contextual linguistic information such as example sentences.

6

LANGEYE

A mobile application to promote autonomous vocabulary learning through situated- and micro-learning opportunities.

In SLA, one of the known challenges is to create contextual learning opportunities in non-native environments. For example, accessing or visiting English-speaking environments requires considerable resources to provide English learners with immersive, practical or meaningful ESL experiences [53]. This project focuses on contextual vocabulary learning using the learner's real-world objects. While accessing L2 environments might be challenging, LangEye uses the learners' own real-world objects via their smartphone cameras to curate L2 vocabulary and micro-learning activities.

LangEye is aimed at self-motivated autonomous learners providing casual but meaningful study sessions; it leverages the practicability and portability of mobile devices as a means to contextual learning with real-world objects. In other words, LangEye promotes contextual vocabulary learning via micro-learning as a mobile application. Furthermore, LangEye demonstrates a practical mobile-based variation of the data-driven framework *"WordSense"* for situated learning proposed in 2017 by Vazquez et al. [137]. Vazquez et al.'s work focused on mixed reality platforms to overlay vocabulary information onto the learner's real-world [137]. While mixed-reality platforms are more immersive, mobile devices are more accessible and affordable to users.

One of the issues with current MALL applications is the lack of modularity and flexibility for study sessions' order and length according to the learner's current interests (physical surroundings) and available time and energy to study. Thus, one design concern was providing short learning interactions that can be extended according to the learner's availability and willingness to continue the engagement. Furthermore, all study sessions or activities are learner-driven, either generated based on the learner's 'memories' (saved object pictures) or on the objects currently surrounding them in the real environment. While, in practice, the application relies on and is limited

by the accuracy of computer vision and object labelling algorithms, this learner-centred interaction gives learners more control over their learning curriculum.

6.1 "REAL WORLD" OBJECTS AND MALL

Recalling Section 3.4, some MALL applications integrate the real world in language learning via location-based activities or lessons [60, 129]. Others provide 'anytime' learning opportunities [8, 21, 36, 39, 131]. Similarly, LangEye's goal is to create 'anytime' learning opportunities that leverage the ubiquity of smartphones. Although LangEye's 'Scan&Find' set of activities was not evaluated, this mode enables learners to dynamically generate short activities based on the objects available in their surroundings. Yet, the 'memories' allow for long-term curriculum building based on images captured on the learner's smartphone.

Related to LangEye's take on creating exploratory activities out of the real world or household objects is the multi-platform (both desktop and mobile) game Influent. Influent[1] is a commercial serious game version of Howland et al. [64]'s SanjigenJiten CALL system in a 3D virtual game environment for visual-spatial learners. Influent enables learners to explore and interact with trivial objects in a 3D-modelled household; the objects can be saved as vocabulary learned. After acquiring some vocabulary items, a gamified mode (mini-games) is activated where the learner needs to search and select given objects within the virtual environment — this evaluates vocabulary recall and recognition.

'Read My World' is a mobile application developed by one of Microsoft's internship teams. This project aimed at increasing literacy in adult English learners [150]. This app focuses on vocabulary building, spelling, and pronunciation powered by several APIs that can recognize objects' names or detect and augment text based on photos taken with the user's smartphone. Words searched for are stored to build a personal dictionary; the app also creates a few word games from the user's saved vocabulary.

Similarly to 'Read My World', LangEye uses the learner's mobile camera to interact with real-world objects. We use data APIs (such as Google Vision API, OpenAI's ChatGPT API, etc) to detect objects

1 https://playinfluent.com/

and display their name, definition, and sample usage sentences. Like Influent and 'Read My World,' LangEye saves pictures taken or explored objects as vocabulary learned through its 'memories' feature. While the ability to save vocabulary for study for personalized learning is also explored by others (i.e., Arakawa et al. [8] and Tsourounis and Demmans Epp [131]), LangEye's interaction relies on images as the main input from the learner and the other learning information such as descriptions and sample sentences are fetched from other data-driven resources. The creation of 'memories' using text entries, such as the name of an object, is not possible in LangEye's current version. On the other hand, LangEye provides learners with activities (mini-games) that exercise their recognition and recall both from the learners 'memories' or onsite by scanning the real environment around learners.

6.2 LANGEYE — DESIGN AND FEATURE EXPLORATION

LangEye's learner interaction design used the smartphone camera as the main input modality to build vocabulary. The app creates 'anytime' learning opportunities by allowing the learner to interact with objects around them:

- take pictures of objects which they can interact with in-situ via editing or exploring the picture ('memory')
- take pictures of objects now, but choose to edit or explore the picture ('memory') later
- scan (take a three to 10 seconds video) the environment and start exploring the objects around them with dynamically generated activities (see Section 6.2.6)
- start a 'Review Memories' session with a desired length — 3 to N words for up to 3 phases per session (see Section 6.2.4).

In this section, I discuss LangEye's features and design considerations.

6.2.1 *Languages Supported*

LangEye currently supports the following source or target languages: English, French, Spanish, and Portuguese. The *source language* is the

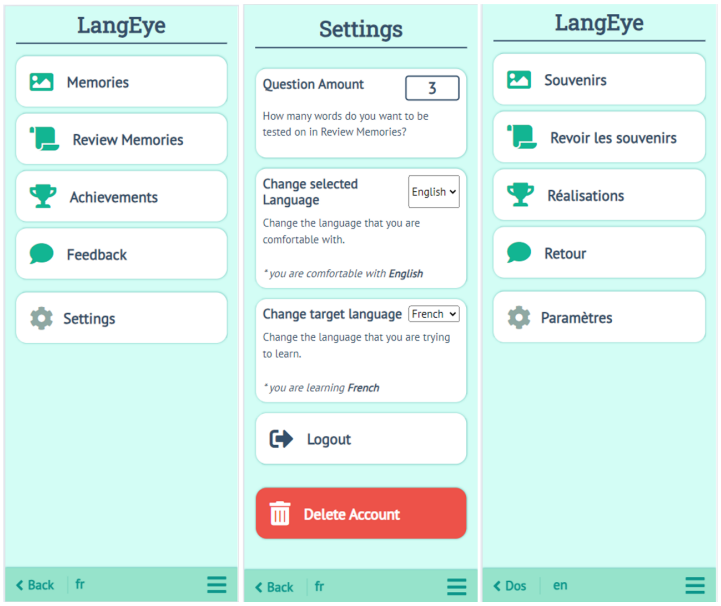


Figure 6.1: The left panel shows LangEye's UI with English as the source language. The middle panel shows LangEye's 'Settings' menu, where the source and target languages can be changed, and it determines the number of words selected for the 'Review Memory' session. The right panel is the UI when French is the source language and English is TL.

language the learner is familiar with or learning from, and the *target language* is the language to be learned. The UI elements, such as the menu and activity instructions, can be set to any of the listed languages as the *source language*. Likewise, the names of objects and learning content (i.e., definitions and example sentences) can be displayed in any listed language as the *target language*. Learners can define their source and target language in the 'Settings' menu option as shown in Figure 6.1.

For example, to demonstrate how the data-driven framework can support different languages, assume we set LangEye's source language to English and the target language to French. All API calls and resources are English-based due to its abundance of data and computational resources. When choosing a different French as TL, LangEye displays the UI text, such as menu names and activity instructions in English. Yet, the language learning content, such as sample sentences and vocabulary words, is displayed in French (supported by Machine Translation (MT)) with the option to 'translate' on demand to English (source language). If the source language chosen is also different from English, LangEye uses MT to display the UI elements in the chosen language, too.

In this way, LangEye can support a variety of high-resource languages, where MT works well. Although LangEye currently offers 5 languages, its architecture is designed to support adding as many high-resource languages as needed.

A learner can also create usage (history) data related to any pair of languages. The history data saved from the learners' interactions with the 'Review Memory' activity, such as accuracy for a 'memory' in a review phase, is associated with that language pair configuration. However, the 'memories' themselves are shared across any language pair but are respectively translated to the appropriate paired languages. Although, learners could only learn French as their TL from English to simplify the user study.

6.2.2 *Capturing the Learner's Environment*

LangEye's main input modality for new vocabulary is via the learner's smartphone camera. To capture the real objects around the learner, LangEye offers two different modes: **Picture** and **Scan&Find**.

PICTURE MODE. The picture mode is the most casual study session out of the two. Using this mode, the learner can aim their device's camera at an object they wish to interact with using the TL and snap a picture of it. In response, the app returns a list of five likely labels (names) for the object sorted by probability (most likely object name as the top result); the learner can explore the definition and a sample sentence for each of the labels provided, choosing to pick a new top label based on that information. Alternatively, they can agree with the results presented initially. On the same screen, learners also have the option to listen to the pronunciation of the object name. At this point, they can choose to take another picture to explore more objects, or they can end their study session. The app saves all the objects, pictures and names in the learner's 'memories' for further interaction.

SCAN&FIND MODE. The Scan&Find requires the learner to scan their real-world surroundings for a few seconds to initiate the study session. In the background, the app detects the objects in the learner's scene, which are used for generating and presenting an activity randomly chosen from five activity types designed to help learners rec-

ognize and recall the objects around them using the TL. All the objects detected in the scene can also be saved to the learner's 'memories' if they choose to do so.

6.2.3 *Memories*

The memories feature enables the ability to save and recall acquired vocabulary items. In this case, the vocabulary items are associated with the pictures taken by the learners themselves. Therefore, the learned words are tied to a familiar object picture, which is more effective when compared to unfamiliar or no pictures for vocabulary learning [68, 72, 117]. As shown in Figure 6.2, learners can scroll through their gallery of memories and open a specific picture to visualize information on that object. Memories (images) can be deleted along with their associated quiz data and labels using the *'delete image'* button.

EDITING LABELS. To mitigate possible CV and MT errors, learners have the option to add their own label to the picture by pressing the *'add label'* button — see Figure 6.3; this prompts the learner to add a text label they may find most appropriate to their picture if that is not included within the original options given by the app. When entering their own label, learners type the text label in their source language, the app presents them with the TL translation — this means learners do not need to know the TL name. The app presents the label(s) added by the learner with their information (definition and sample sentence).

The text-to-speech function is available for any text label as a pronunciation reference for the word, but it is limited to the accuracy of the API. Learners who judge a label as inappropriate to the memory can long press it to delete it (Figure 6.4). Lastly, they can start a 'memories review' quiz based on their memories from LangEye's main menu to support self-evaluation of their vocabulary (memories).

6.2.4 *Reviewing Memories*

This feature is a classic quiz-style review of the collected memories (vocabulary words). The *'memory review'* has three phases or types of quiz questions:

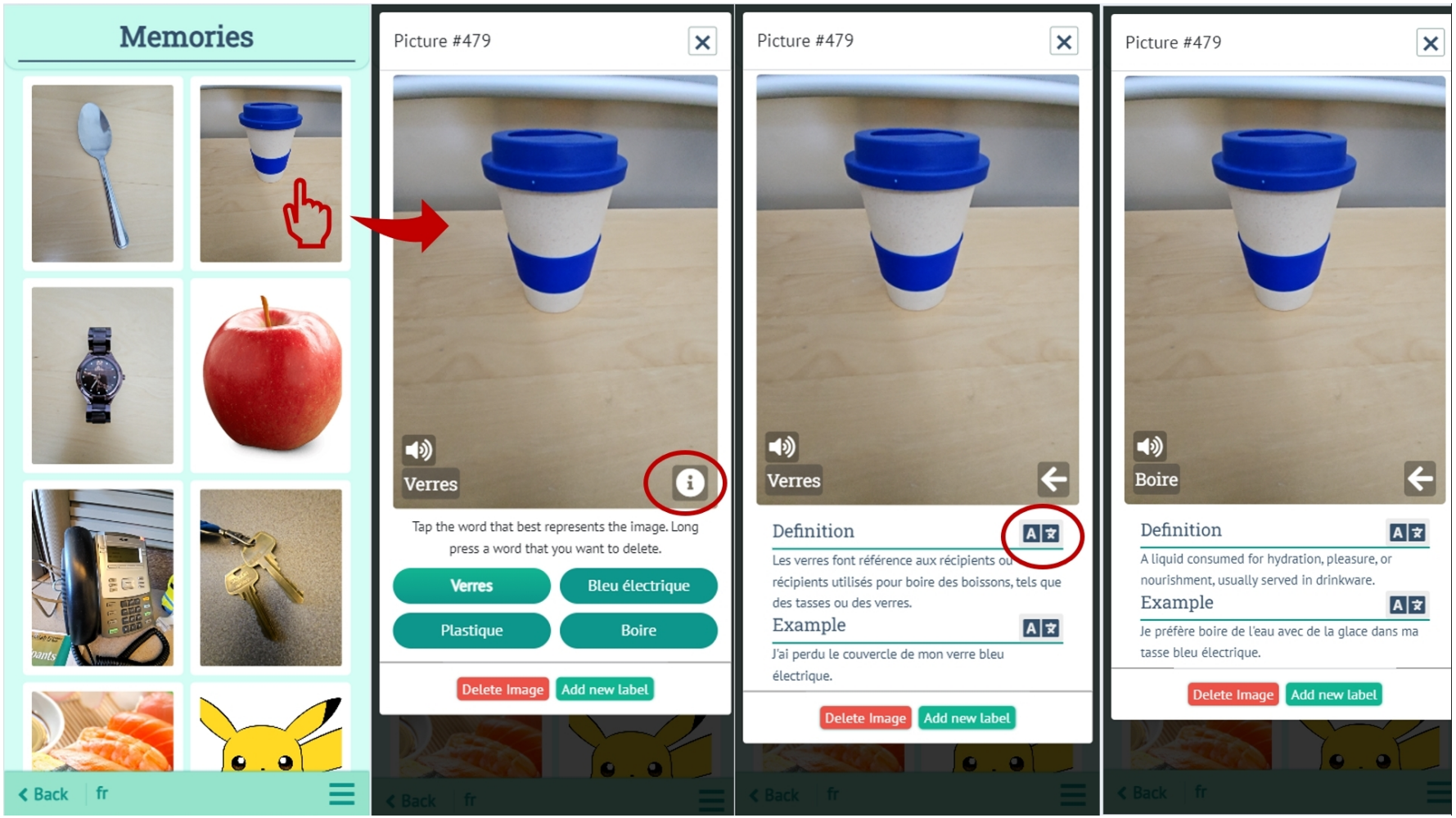


Figure 6.2: Screenshot of the 'Memories' feature. On the first screen from the left, we can see the roll of memories. The second screen shows the overview of individual memory with its respective text labels (object names resulted from the Computer Vision (CV) API). Upon clicking on the information icon on the memory, LangEye shows the definition and an example sentence in the TL, as shown on the third screen. The fourth screen shows the word definition translated to English using the *'translate'* icon on the example.

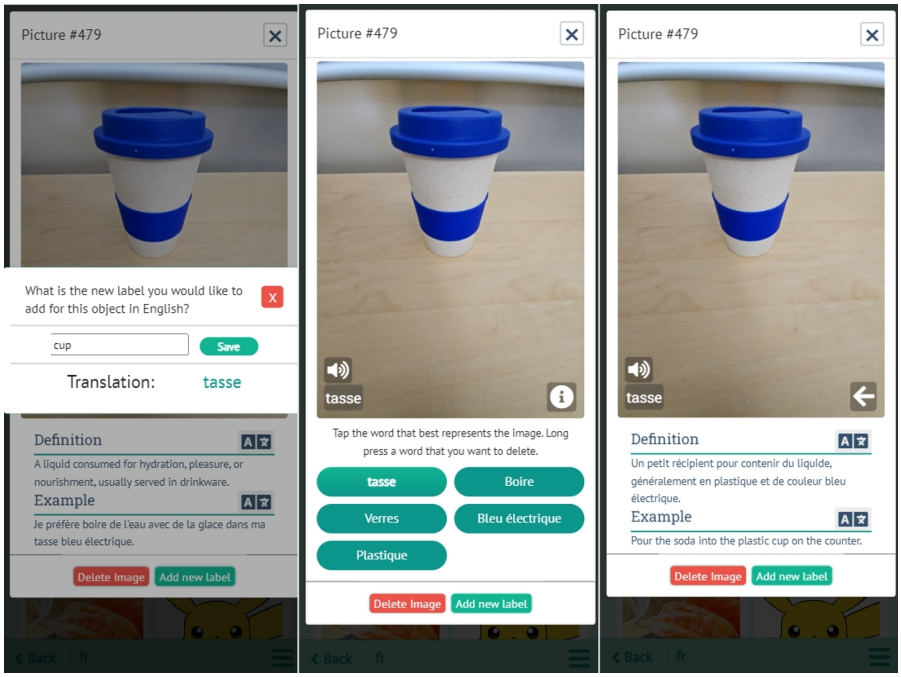


Figure 6.3: When the learner decides to add a new label, or name, to the object in the memory, they can press the *'add new label'* button. As shown in the first panel, this allows them to insert the new label by typing it in their source language. The app displays the translation in the TL. By saving the new label, it will be displayed as an additional option in the memory (second panel). When selecting their customized labels, the information displayed is equivalent to the original (automatic) labels where the sample sentence and given definitions are also machine translated — the third panel.

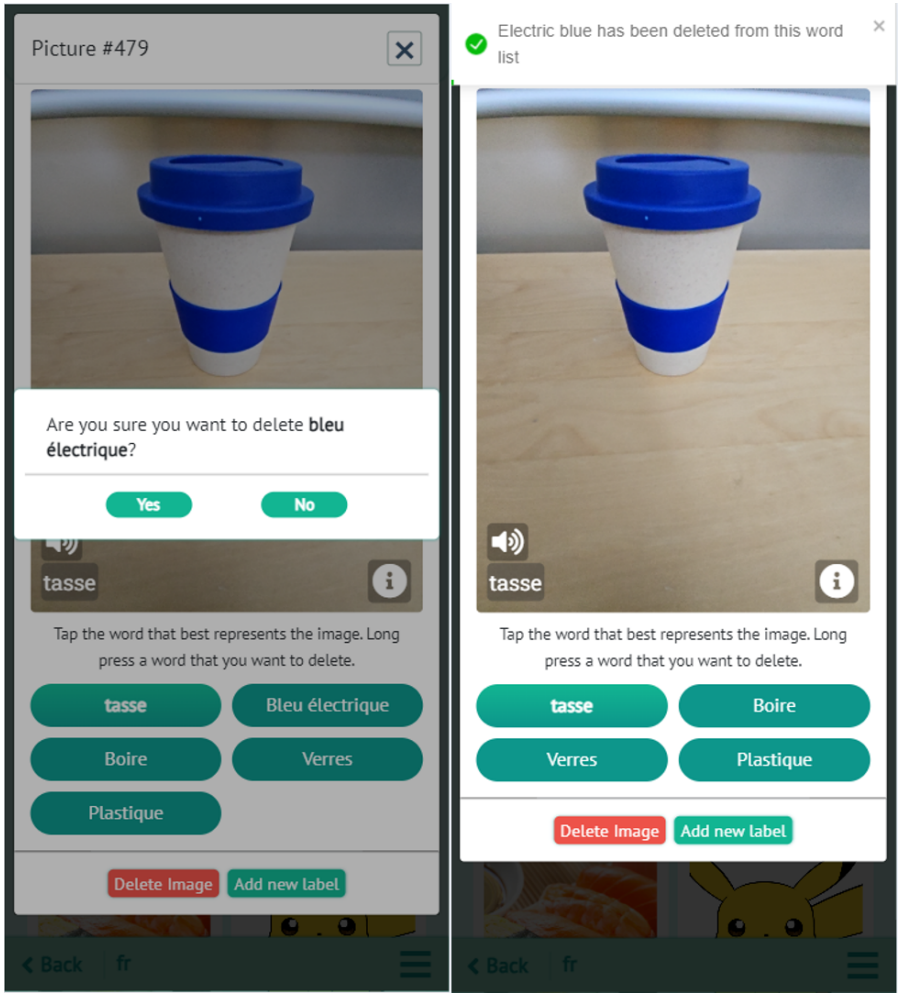


Figure 6.4: If the learner decides that any of the labels (either custom or original) is not suitable to the memory, they can long press it to delete it from the memory.

- **Phase 1:** object name recalling from memories; auto-generated multiple choices using pictures from the learner's memories.
- **Phase 2:** word collocation within a sample sentence; fill in the blank by choosing the correct picture (object name) from auto-generated multiple choices using pictures from the learner's memories.
- **Phase 3:** same activity as Phase 2, but without multiple choices. In this phase, the learner must spell (type) the correct object name themselves.

The system chooses N memories (words/objects) to review during a Memory Review session. For each phase, learners are tested on those same N words. The phases are progressive; for example, to at-

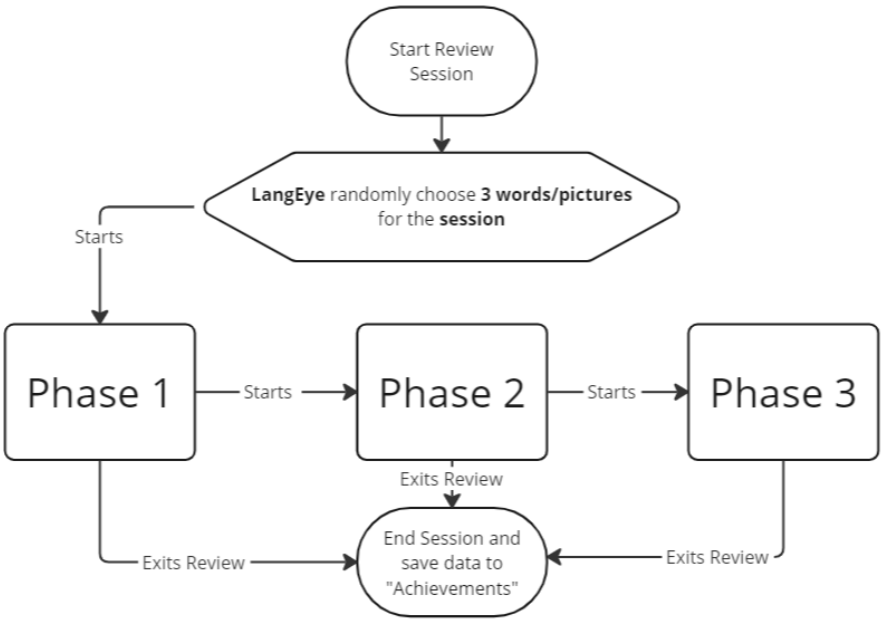


Figure 6.5: Flowchart shows the different choices for learners after starting a review session for three (N = 3) memories. The session can end in between phases, and the higher phases need to be accessed in an ordered manner.

tempt Phase 3, the learner must have completed Phases 1 and 2 in order. Learners can choose to end the review session between phases. The complete flowchart is shown in Figure 6.5. This progressive ordered phase was designed to increase the complexity of tasks and reduce system help at each level.

This style of activity was inspired by the card party game called Monikers[2], where players take turns trying to guess the name or concept represented in cards. In the first round, clue givers can say anything they want except for the name itself. For the second round, clue givers can only say one word. In the final round, clue givers can not say anything; they can only use gestures and charades. Similarly to the Memory Review, the participants work on the same set of words in all rounds. This eliminates words/objects or cards already answered out of the answer options, which, over time, can help with vocabulary memorization and acquisition. In the same way that Monikers' names are learnt within the context presented in the card, the words/objects are contextualized in full sentences presented during Phases 2 and 3.

2 https://boardgamebarrage.com/wiki/Monikers

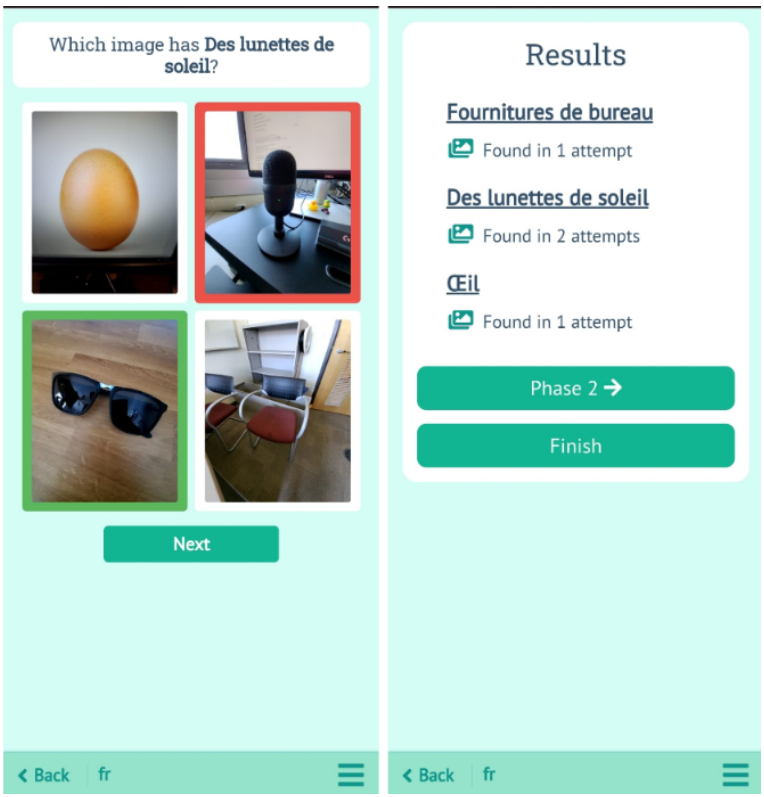


Figure 6.6: Screenshot of a 'Memory review' Phase 1 quiz created from the learner's memories.

In the Memory Review, during **Phase 1**, learners are prompted with "Which picture contains a <object>?" and are given four pictures as answer choices from their gallery — including the picture with the object in question. Their task is to choose the correct object (picture) as shown in Figure 6.6. Every time the learner chooses a picture, LangEye indicates whether their pick is correct by showing a red (incorrect) or green (correct) border. This way, learners get 4 chances or guesses to match the object name and their picture correctly. This guarantees the learner knows the three words/objects being reviewed in the session when proceeding to the next phase.

If they proceed to **Phase 2**, learners will again be tested on the same three words/objects. In this phase, learners are prompted with a set of "Fill in the blank" questions where they are expected to use the same vocabulary but to fill the gap on given example sentences as shown in Figure 6.7. Similarly to Phase 1, they answer this activity by picking from four pictures from their memories. When the learner picks a picture, the equivalent word is shown in the sentence gap, allowing them to reflect on their choice before submitting it as their

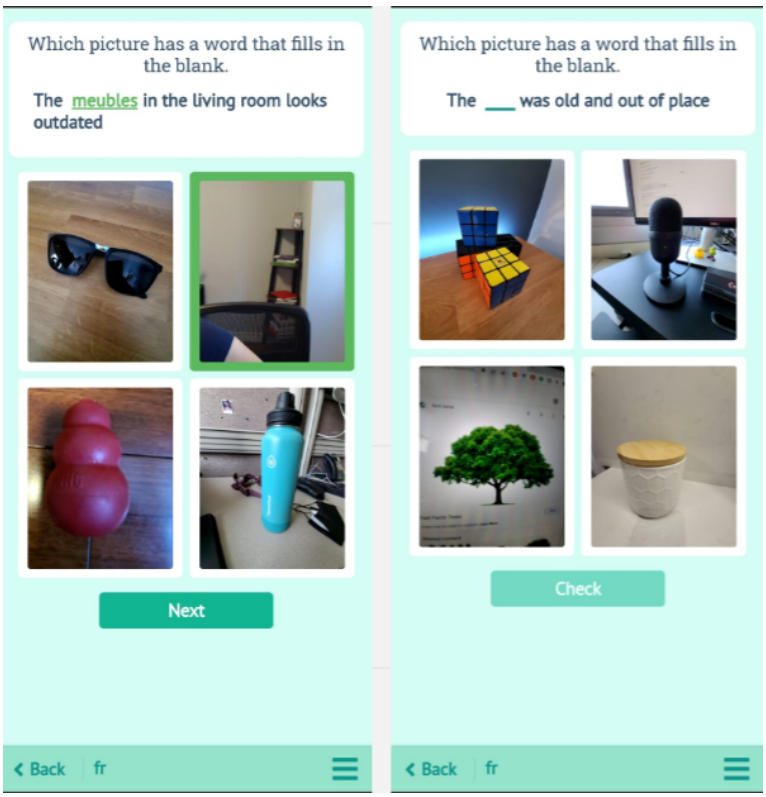


Figure 6.7: Screenshots of two examples of a 'Memory review' Phase 2 quiz created from the learner's memories. Specifically, this is a fill-in-the-blank question. The right panel shows a correct choice sample, and the left panel shows an unattempted question sample — if the learner does not attempt, they are not allowed to proceed.

answer. This activity takes more than recalling the word/object; it also requires the learner to understand how to place or collocate the word within the full sentence. If the incorrect word/object is chosen, LangEye displays the correct choice before the learner can move to the next task or activity. The sentence, however, is displayed in the source language; this choice was made to support beginner learners who might have little understanding of the TL grammar. This is a design trade-off that can affect more advanced learners who wish to complete the activity in the TL instead. The Phase 3 aims to mitigate this affordance.

The activity in **Phase 3** is also a "Fill in the blank" question. However, during Phase 3, the learner is not presented with multiple choices (see Figure 6.8; they must recall the three words/objects studied in the session and type their names to answer the question. While this phase has no visual clue (object picture), a 'Clue' button can reveal the answer's definition on demand. Even though learners are supposed to

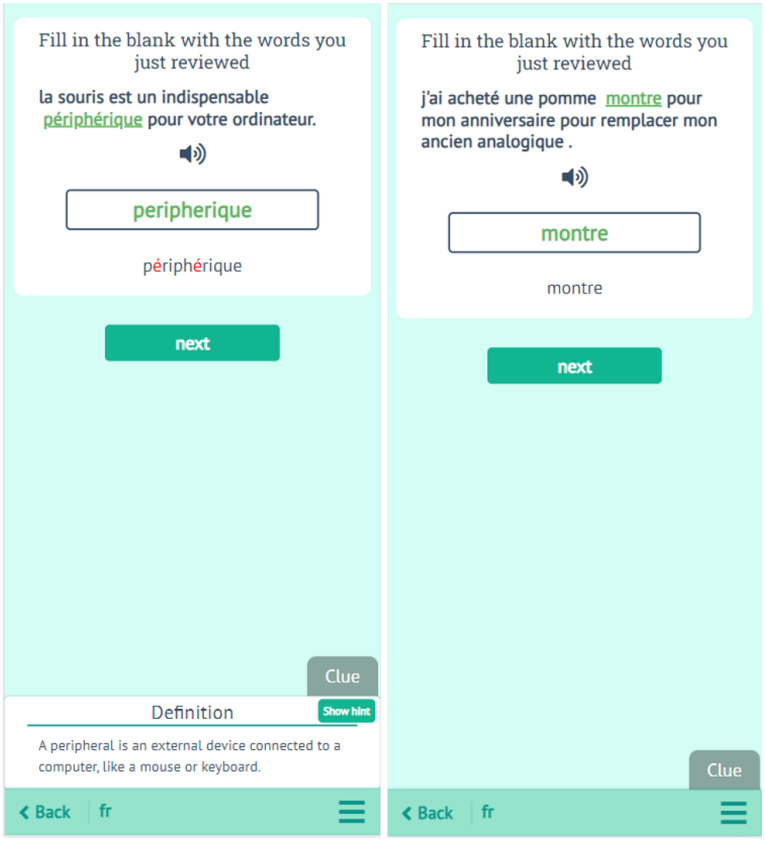


Figure 6.8: Screenshot of a 'Memory review' Phase 3 quiz created from the learner's memories. Specifically, this is a fill-in-the-blank question.

spell the object's name, LangEye considers typos as correct answers. Yet, the correct spelling will be highlighted to provide corrective feedback to the learner. In contrast to Phase 2, this phase presents the sentence in the TL; this activity was designed to target more advanced learners who are prepared to read and understand the TL's grammar. In summary, the goal is that over time, a beginner learner can become more effective at Phase 3 as they become more proficient at the TL, and that LangEye can help bridge the gap between beginner and more advanced learners.

Table 6.1 shows a comparison of the different activity task types, answer support, corrective feedback, cognitive task, and language support for each of the phases. Phase 1 is the simplest type of activity and most appropriate for beginner learners. Phase 2 adds cognitive complexity by requiring syntactic vocabulary knowledge and a complete sentence; however, it provides the familiar picture support as multiple choice answer. Phase 3 requires the most cognitive load from learners, where they must spell (type) their answers in a given

sentence; to balance the difficulty, learners can attempt Phase 3 using the optional 'Clue', which displays the word's definition. This last Phase also provides flexibility when checking for the word's spelling to provide positive corrective feedback to learners.

Table 6.1: Features and language support comparison between the three Memory Review phases. From Phase 1 to 3, the visual support of images is reduced, and the usage of the target language is increased, resulting in the later phases being more challenging. As illustrated in Figure 6.5, learners can decide to save and exit their review session in between phases.

		Phase 1	Phase 2	Phase 3
Task type	Identify picture	✓	—	—
	Fill in the blank	—	✓	✓
Answer support	Picture	✓	✓	—
	Multiple choice	✓	✓	—
	Type/Spell	—	—	✓
	Clue	—	—	Word definition
Answer and corrective feedback	Instantaneous check	✓	—	—
	Submit and check	—	✓	✓
	Corrective feedback	✓	✓	✓
Cognitive task	Word recall	Picture	Picture	Spell
	Word collocation	—	✓	✓
Language *Source: Source Language* *TL: Target Language*	UI elements	Source	Source	Source
	Vocabulary word	TL	TL	TL
	Sample sentences	—	Source	TL
	Word definition	—	—	Source

6.2.5 *Achievements*

The 'Achievements' feature (see Figure 6.9) on LangEye collects the Memory Review history and accuracy in each attempted phase. For phase one, it shows the average number of attempts or guesses taken. For phases two and three, it displays the average of the accuracy ($\frac{\#\text{timesCorrect}}{\#\text{timesTaken}}$). It allows learners to track their progress for each word/picture. Achievements allow learners to sort the words by accuracy, best to worst, or vice-versa, as shown in Figure 6.9's first two panels; meanwhile, highlighted in the bottom of the first panel, the initials "fr", indicate that these achievements are for a learner of French (TL).

When the learner clicks on a word from the achievements, more detailed information is shown, including the associated picture and the information for each phase per attempt, along with its date. On attempts that the learner did not proceed to all phases, the accuracy is shown as "N/A". Figure 6.9 shows an example where the last two panels display the detailed information on the French word the learner has done the worst so far: *"Ciseaux"* (scissors).

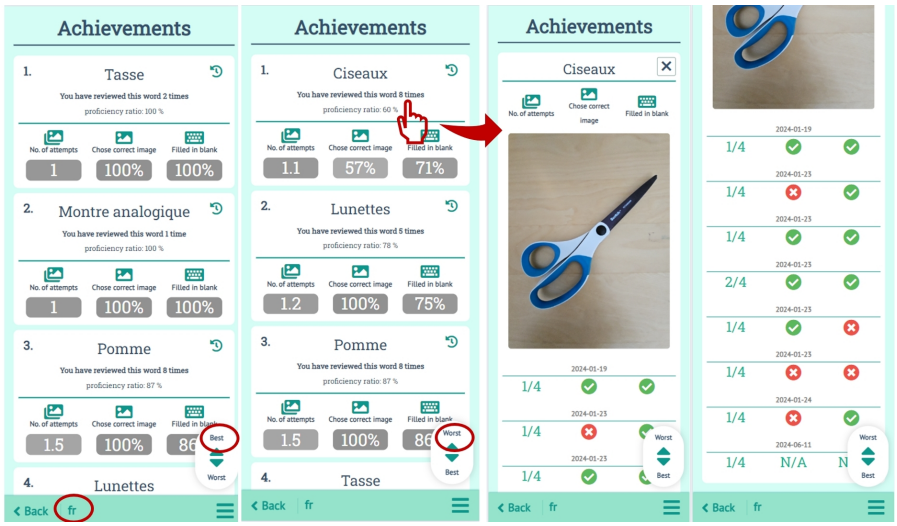


Figure 6.9: Screenshot of the Achievements feature where learners can track their progress for each vocabulary tested on a Memory Review session. Clicking on a word lets learners view more detailed information on individual attempts. The achievements shown in the picture are for a learner of French (TL) from English (source language).

6.2.6 *Scan&Find Activities*

The Scan&Find mode, originally named Eye Mode, is a mode for exploration of the learners' real environment. In this set of activities, I examined the interactive design of learning activities that use the smartphone's camera as the primary input and lens for the learner's language learning experience. The ideas implemented here are experimental and serve as a prototype, aiming to create a more immersive experience. As AR and XR devices become more affordable and multimodal generative AI technologies continue to evolve, the practicality of the ideas discussed in this subsection becomes increasingly feasible. Thus, the user evaluation presented in this chapter does not include this mode due to technical difficulties and the robustness of

the state-of-the-art technology at the implementation time (Summer 2022).

Figure 6.10 shows the series of screens to use the Scan&Find mode. The steps of activating this mode are performed once per study (login) session. That means learners can use the same scan to perform multiple activity rounds for the same or different activity types. Alternatively, the learner may scan their environment again to replace the previous scan of the session. The default scan length is 10 seconds, but the learner can redefine it to a number between 3 and 10 seconds on the' Settings' menu. The option to re-scan is particularly helpful to mitigate any issues with object detection or when the learners decide to move on to a different (physical) setting.

OBJECT EXPLORATION. In this activity (Figure 6.11), learners are presented with a list of objects detected in their environment; the learners will choose one of those words in the list which they wish to find and take a picture of. This activity exercises the learner's word recognition and reinforces the connection between the TL word and objects in their familiar surroundings. Learners can opt to look for objects they know the name of or for objects they are curious to learn more about. When finding the object, the learner can save its snapshot into their memories for further study and memorization. Lastly, they can keep looking for objects from the list indeterminably.

OBJECT IMPOSTER. In this activity (Figure 6.12), the learners are presented with a list of objects detected in their environment, in addition to one *imposter* object which was not detected but is contextually related to the list; learners are tasked to choosing the imposter object in the list. Upon finding the imposter, the learner can proceed to a different activity. If chosen to return to this same activity, the learner might be presented with the same or a different imposter option. In the future, this activity can be improved by providing an option to look up information on the imposter object, including its definition, sample usage, and a stock image (from a dataset or AI generated). This would allow them to transform the imposter into a new memory.

OBJECT SEARCH. Similarly to the *object exploration* activity, in this activity (Figure 6.13), the learners are tasked with locating and taking

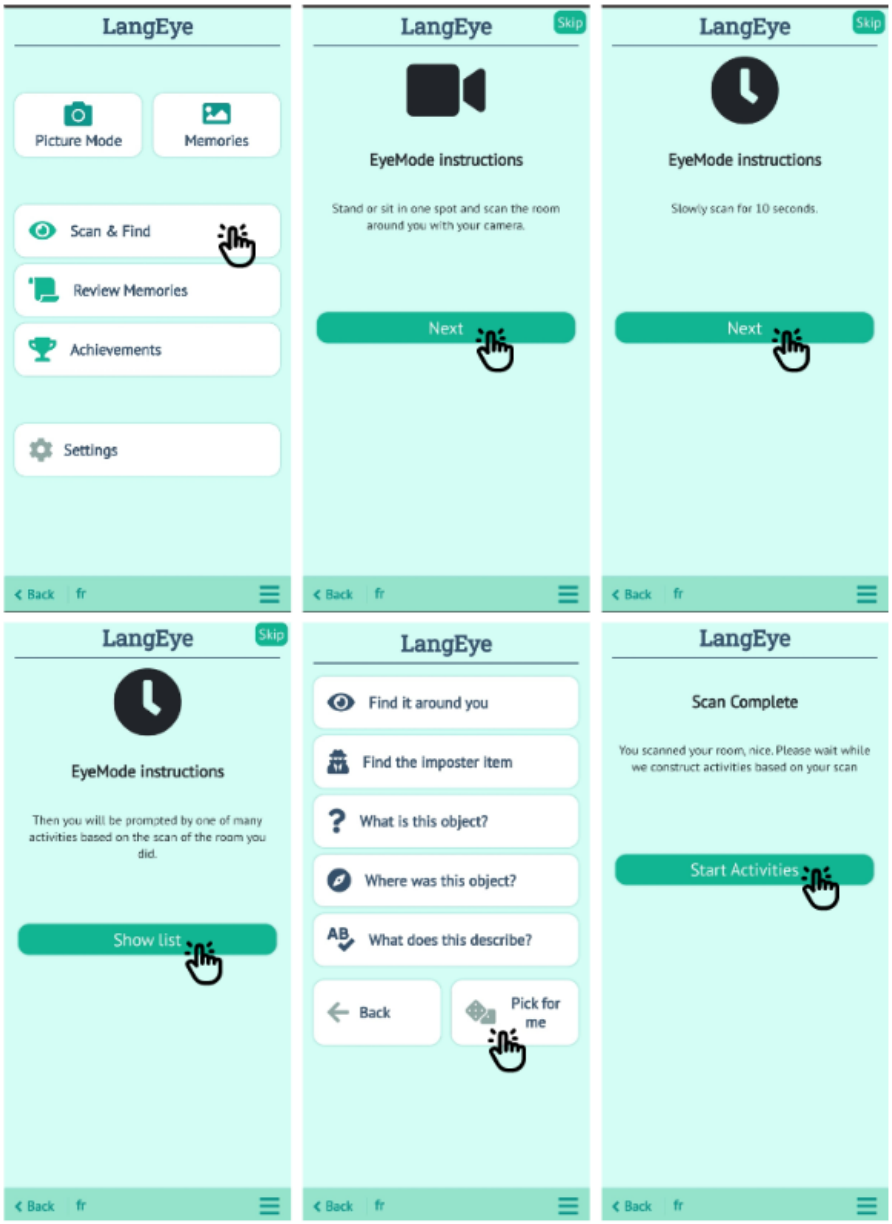


Figure 6.10: From left to right, and top to bottom; this sequence of screenshots shows the main menu screen of LangEye and the sequences of instructions when accessing the Scan&Find mode. The bottom middle screen lists the possible activities for the eye (Scan&Find) mode. The same environment scan may be used to perform multiple rounds of activities (either the same or different ones).

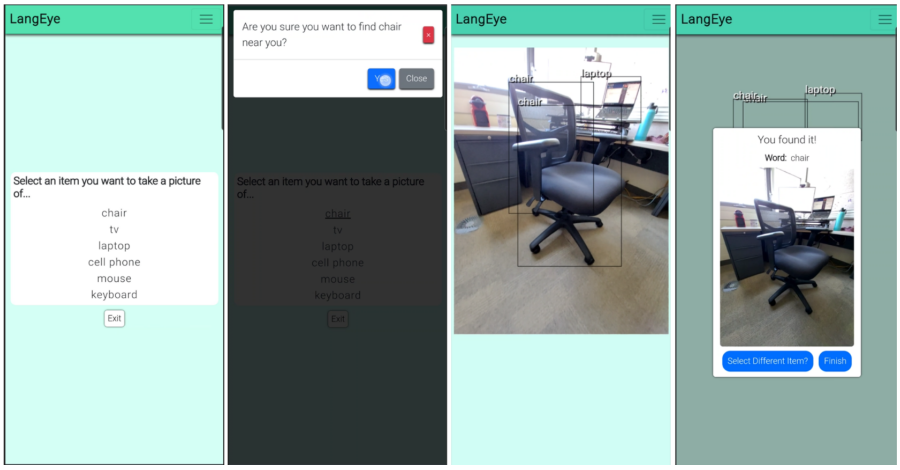


Figure 6.11: Screenshot of the sequences of the activity: *object exploration*. The app presents a list of objects detected in the learner's environment. The learner can select which object they would like to try finding with their camera and proceed to aim the camera at the object. The app detects the desired object and automatically presents it as found. The learner can choose to save this object in their memories.

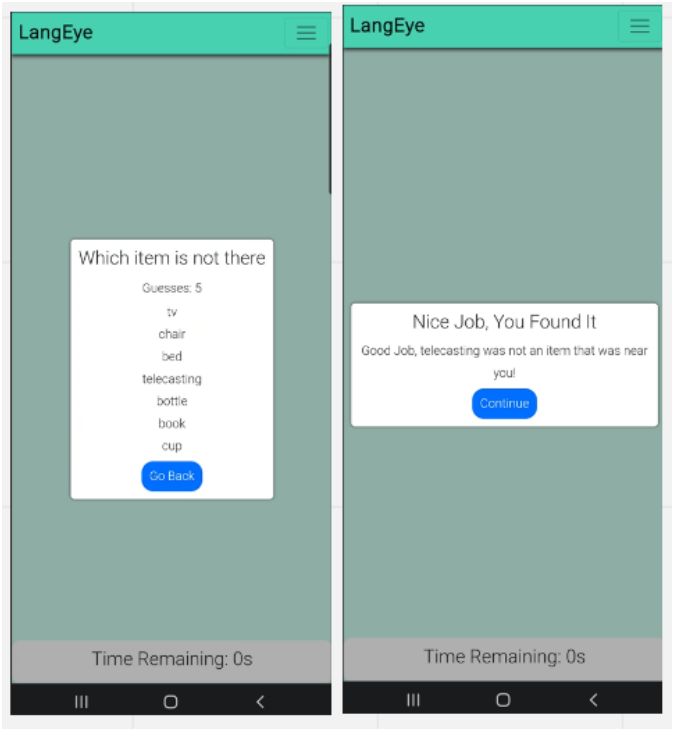


Figure 6.12: Screenshot of the sequences of the activity: *object imposter*. In this activity, the app presents a list of objects found in the learner's environment and an *imposter* object that is not included in the original list. The learner's task is to select the name of the imposter object.

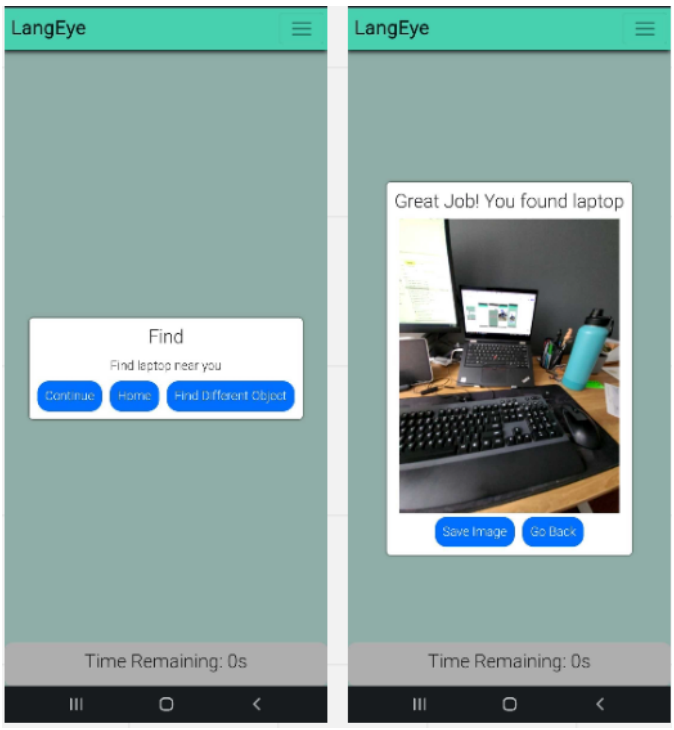


Figure 6.13: Screenshot of the sequences of the activity: *object search*. The app prompts the learner to find a specific object from their environment. Upon detecting the object with their camera, the learner may save it to their memories.

a picture of one specific previously detected object in the scene. The main difference is that in *object search*, the app automatically assigns the object to be found, and the search occurs once, while in the object exploration, the learner selects the object to be found themselves, and they may choose to keep on finding several objects. After finding the prompted object, learners will have the option to re-start the activity (or start a different one), which may prompt them to find different objects in the new search. Upon finding the object, the learner may save it in their memories.

OBJECT RECOGNITION. In this activity (Figure 6.14), using a picture of an object from the scene, the app asks learners to type that object's name. The learner is given multiple tries to guess the correct word entered. This activity should use a cropping algorithm for future iterations to better center and indicate the specific object for recognition. At the prototype's current state, the presented snapshot can display multiple objects, which may lead to confusion and frustration from the learners. On top of that, the app could give the

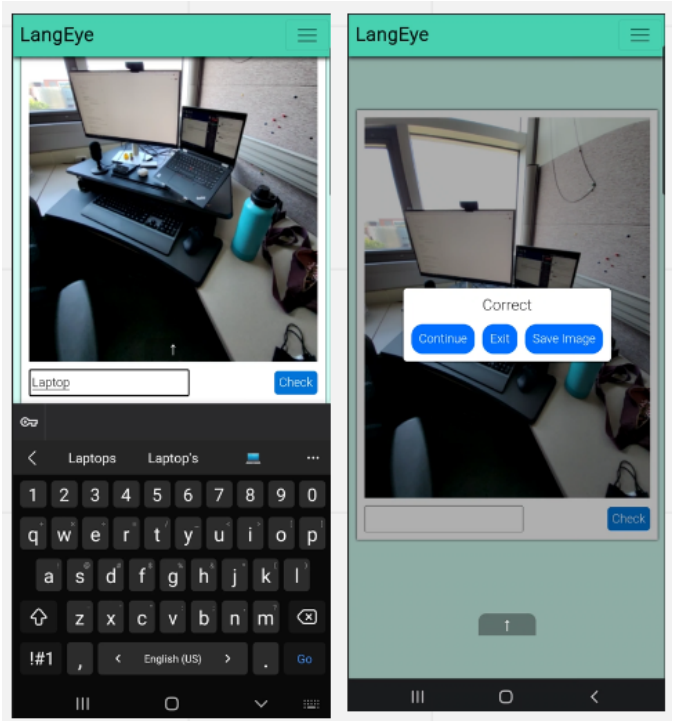


Figure 6.14: Screenshot of the sequences of the activity: *object recognition*. The app presents the learner with a snapshot of the environment scan and asks them to type in the name of the object displayed. Upon correctly entering the object's name, the learner may save it to their memories.

learner multiple options to match the object's name, making it a faster activity and decreasing possible challenges related to orthography and language-specific keyboard types for learners, particularly novice learners. Alternatively, the app can display possible character blocks for the learner to click (as opposed to typing) and spell the object's name; this would reduce the possibilities of characters or syllables to pick from, which would likely lead to a better User eXperience (UX), and help mitigate for situations where multiple objects can be spotted on the snapshot presented.

OBJECT ACTION. In this activity (Figure 6.15), the learner is presented with an action verb semantically related to one of the objects in the scene; their task is to locate that object and take a picture of it. For example, 'Find an object you can *"drink"* from' in their environment, where a *"bottle"* is present. The learner would be expected to associate the action described by the verb to one of the objects in the scene and take a picture of it. This can help them link verbs to their nouns, contributing to linguistic awareness and language pro-

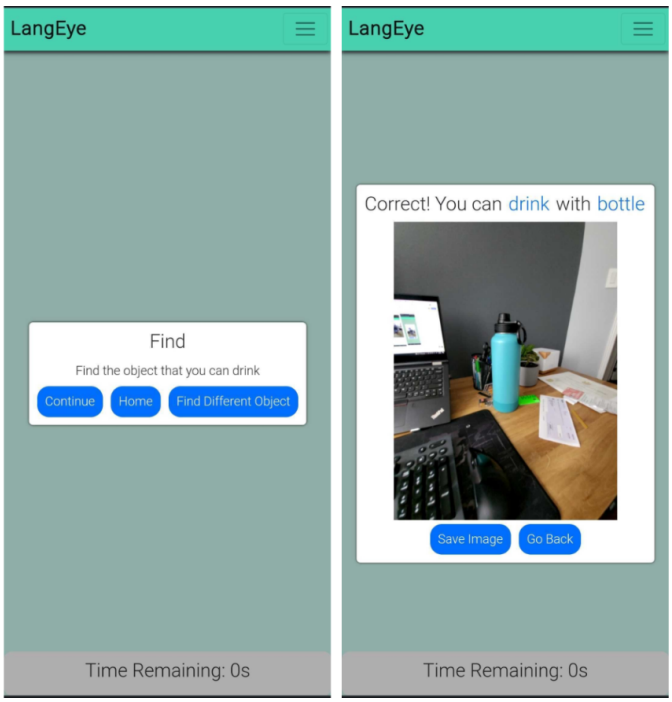


Figure 6.15: Screenshot of the sequences of the activity: *object action*. The app prompts learners to find a specific object related to an action verb from the detected objects. The learner will locate the respective object in their environment and take a picture of it. The picture may be saved in the learner's memories.

duction. The learner can choose to find a different object (using its respective action verb) if they are uncomfortable answering the first prompt. Upon finding the object, the learner may save it to their memories. As future work, generative AI models can be used to produce better action verb options or even short sentences to lead learners to deduce the desired object.

6.2.7 *Treasure Hunt Activity*

Like the Scan&Find mode, the treasure hunt activity was experimental and not tested. Furthermore, it was not fully implemented due to technical challenges at the time of implementation. This section goes over the design behind it.

This activity (Figure 6.16) begins with an object picture from either the picture mode or from the learner's saved 'memories'. From the initial picture (O_1), the learner is prompted to find a second object (O_2) that is "contextually" similar to O_1. Using Word2vec model, a NLP technique for obtaining vector representations of words' seman-

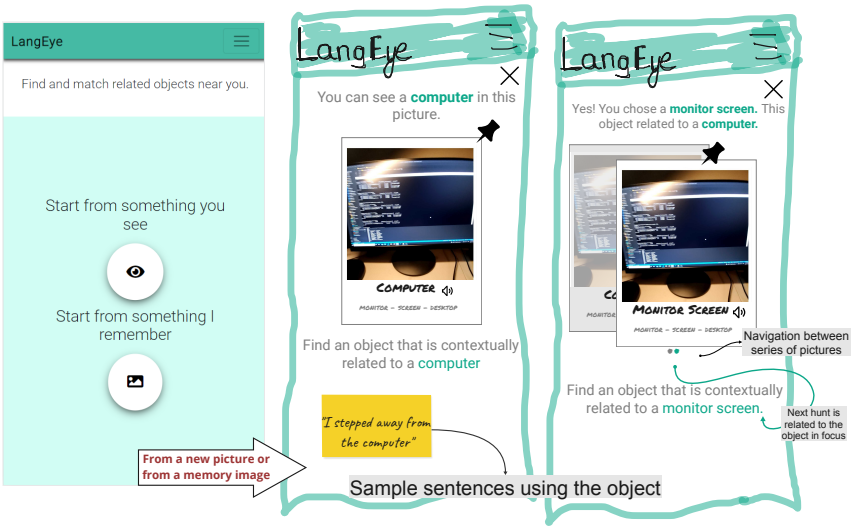


Figure 6.16: Sketch of the sequences of the activity: *Treasure Hunt*. From either a memory or taking a new picture, the learner can start a new treasure hunt from the object detected, which entails finding a contextually similar object next. The pictures from the hunt would be automatically saved to the learner's memories.

tic relationships, the app calculates the semantic similarity between both objects (names). Based on the semantic difference, the app allows learners to keep trying with different objects if incorrect (semantic difference is larger than a threshold), or it congratulates the learner for fulfilling the quest. If successful, the learner can either exit the activity or continue by finding any third object (O_3) similar to O_2. Thus, the activity can be as short as finding one object, O_2 and as long as finding N objects (O_N).

This activity was designed to allow learners to leverage their L1 knowledge of the real environment around them — the contextual relationship between objects, to acquire the TL vocabulary. At the time of development, generative AI models such as ChatGPT were not as widely available nor as well developed. For future improvements, I would suggest using a LLM to help determine context relationships and coherence between objects, considering the learner's physical environment. Furthermore, personalized profiles and models can be applied to improve the UX further.

6.3 IMPLEMENTATION

LangEye is a mobile web application implemented using NodeJS [51] and ReactJS [125] for the front and back end, respectively. As a web application, the learner does not need to install anything to use the app, and their data is stored remotely in our servers under their created account. To use LangEye, the learner must allow it to access the device's camera, as it is a core piece of our interaction design.

This chapter explores design choices for a learner-driven immersive MALL application for contextual vocabulary acquisition. Therefore, most of the technology behind the key interactive components of LangEye are third-party APIs. Nonetheless, the presented work is limited by those resources' state-of-the-art object detection and recognition. See a list of the APIs used in the implementation of LangEye:

- Google Vision API for object detection and recognition [57]
- OpenAI API GPT-3 for word definitions and sample sentences [1]
- Google translate API [58]
- Google text-to-speech API [56]

6.3.1 *Creating & Retrieving Memories*

The creation of new memories can originate from two main user interactions (1) the Picture Mode and (2) uploading an image from their device's gallery. Either way, once a new image is added, the pipeline of creating a new memory is the same. Figure 6.17 shows the main steps in that pipeline, including annotating objects within a picture using image-object detection tools — Google Vision Object Detection and storing the image along with its metadata (i.e., object names (labels), translated labels and their definitions). Thus, the memory creation can be split into two parts, image classification and image contextualization.

IMAGE CLASSIFICATION. Classifying images starts with the user adding a new image. The image is sent to the Google Vision Object Detection API, which classifies all the objects within the image. LangEye stores all the labels (detected objects) in an array of fixed lengths of four, sorted by the most confident objects the API has detected.

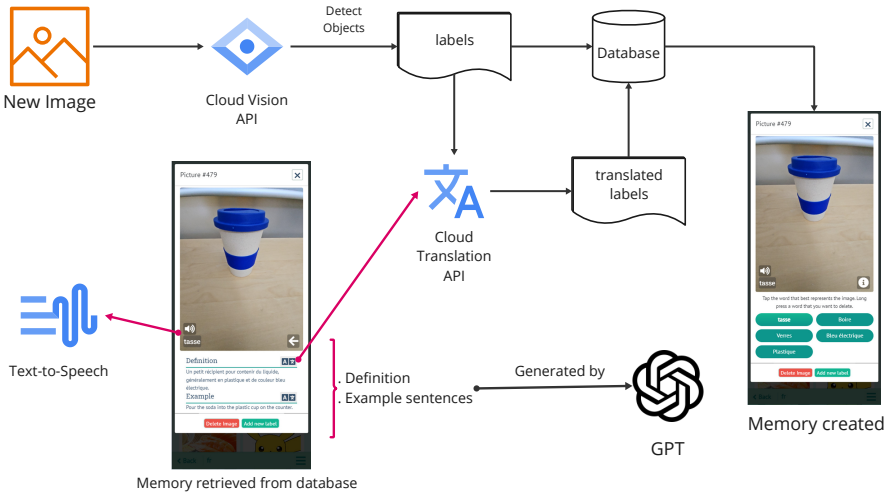


Figure 6.17: Diagram highlighting the integration of Google's and OpenAI's usage in LangEye's memories. From a new image, either taken in the app or upload from the gallery, LangEye uses Google's Vision API to extract objects and their names (labels). Each label is supplemented using Google Translation API and OpenAI's GPT to store the learner's current TL translated label, definition, and example sentence in English. Definitions and sentences are translated on demand when the memory is retrieved by the learner. Translated labels are added to the database as the learner switches to a different TL.

IMAGE CONTEXTUALIZATION. Once the image has been successfully analyzed and its objects classified, and the image and labels are saved in the database, LangEye stats the image contextualization process. It concurrently translates the labels into the learner's current source or TL, if not English; and it generates an example sentence and definition for each label using OpenAI's chat completion API. While the translated labels are saved, the definitions and example sentences are translated on demand, only the English counterparts are stored. This reduces the number of calls to the OpenAI API which reduces the monetary cost, but also the computing cost of redundant calls.

Thus, when a memory is loaded on the app, the metadata is retrieved from the database. If a different language is being used at that moment, a new set of translated labels is added to the database. Otherwise, the translation API is used as needed.

6.3.2 *Memory Review*

The general workflow of the Memory Review works as described in Section 6.2.4. This sections aims to clarify on the N words selection for the review session.

WORD SELECTION. For word selection, LangEye follows a pseudo-random approach following the categories: (1) words that have not yet been tested; (2) words with low accuracy; and (3) words with high accuracy. For $N = 3$ the selection algorithm randomly, ideally, selects one word from each category. If it fails to find a word in any category, it will give preference to pooling a word from a different one following the enumerated order for categories (1), (2), and (3). If the N is higher than three, the same principle applies. The selection algorithm will continue pooling the words until N is reached while aiming to balance the three categories. As illustrated in Figure 6.5, all the N words selected are used in each of phases. The goal is to balance reviewing new words (no baseline of accuracy), words the learner is not proficient with (low accuracy), and learners not forgetting about known words (high accuracy).

API INTEGRATION WITHIN REVIEW PHASES. Phase 1 does not require integration with any external services as it pools multiple choices from the learner's memories. However, for Phases 2 and 3, LangEye uses the OpenAI API to fetch a sentence, where the corresponding word is replaced by a blank, to be filled by the learner as the solution to the respective phase and activity.

6.4 USER EVALUATION METHODOLOGY

To evaluate LangEye's potential as a language learning tool and the app's usability, I designed a 1-week longitudinal user study with learners of French as a second language. This study was reviewed by the University of Ontario Institute of Technology (UOIT)'s REB (file #17447).

The study aims to answer the following research questions:

- **[R1]:** Are learner-curated pictures to learn vocabulary better for motivation than stock images in LangEye?

- **[R2]:** What is the potential of LangEye as a language learning tool for acquiring vocabulary in a foreign language?

The study randomly assigns participants to control and camera groups to answer these questions. The **control group** is assigned a version of the app that does not use the smartphone camera but rather contains a set of pre-defined vocabulary and pictures stored. They could use the app's Memory, Memory Review, and Achievements. The **camera group** was assigned the full app version — including the Picture mode that allows them to take and upload images. Neither group had access to the Scan&Find mode. Both groups experienced LangEye with English as the source language and French as the TL to facilitate the recruitment process and the analysis of the results. The study had two sessions:

- **Session 1:** in-lab app training session, language learning background, and a post-session survey with learners to investigate their first impression regarding app usability.
- **Session 2:** online exit semi-structured interview to investigate their opinion after using the app for at least five days.

The sessions were scheduled five to ten days apart. Participants were asked to use the app at least once daily for five days between the two sessions. After completing their daily tasks, they were instructed to submit a short usability and feedback survey. In the days between sessions, they received daily reminder emails.

RECRUITMENT. Both versions deployed were modified to only support learners of French (TL) from English (source language). Thus, I recruited 20 learners of French as a foreign language. Participants were split into two groups, **control** and **camera**, of 10 people. Participants were compensated with $10 CAD for the first session and $20 CAD for the second session, with a maximum total of $30 CAD for both sessions. Participants were recruited using posters hung on high-traffic areas of the campus.

DATA COLLECTION AND STUDY STRUCTURE. Each study session was 50 minutes long. We collected the participants' survey responses using Google Forms, the voice and video recordings for each session, and the participants' app usage data, including the photos taken be-

tween sessions — for the camera group —, which were also collected and used for analysis.

6.4.1 *Control Group Memory Curation*

One of the main challenges in comparing the two group of participants is to curate the control group's memories. My approach to this problem was to run both session of the study with the camera group first. This allowed me to use the collection of memories created by that camera group as the control group's memories. On average, the camera group collected 11 pictures ($Median = 9; min = 5; max = 19$). The aggregation of duplicates of the camera group memories resulted in a total of 24 objects (memories) listed below:

apple	floor	lemon	speaker
bag	flooring	paper	spoon
cup	food	peripheral	stapler
dishware	fork	scissors	tableware
drinkware	glasses	serveware	watch
eyewear	hand	slipper	wood

To create uniformity in the images, avoid bias toward the quality of images taken by the camera group, and preserve the participants' privacy, OpenAI's Dall-E 3 was used to create the images for each of these memories. The prompt included the object name and instructions for the illustration to be "realistic" to mimic a photo taken of the object — sample shown in Figure 6.18. The pairs of all AI-generated images and labels can be found in Appendix E. This approach was used to present a similar curation of vocabulary while not adding words that might not have been added using a smartphone camera.

6.4.2 *Study Session 1: Introducing LangEye*

For both groups, Session 1 was held on campus in a controlled lab environment for human studies.

ROOM SETUP. The room was set up with a collection of office and other household objects that participants could explore using the LangEye application. Therefore, we pre-selected eight objects that

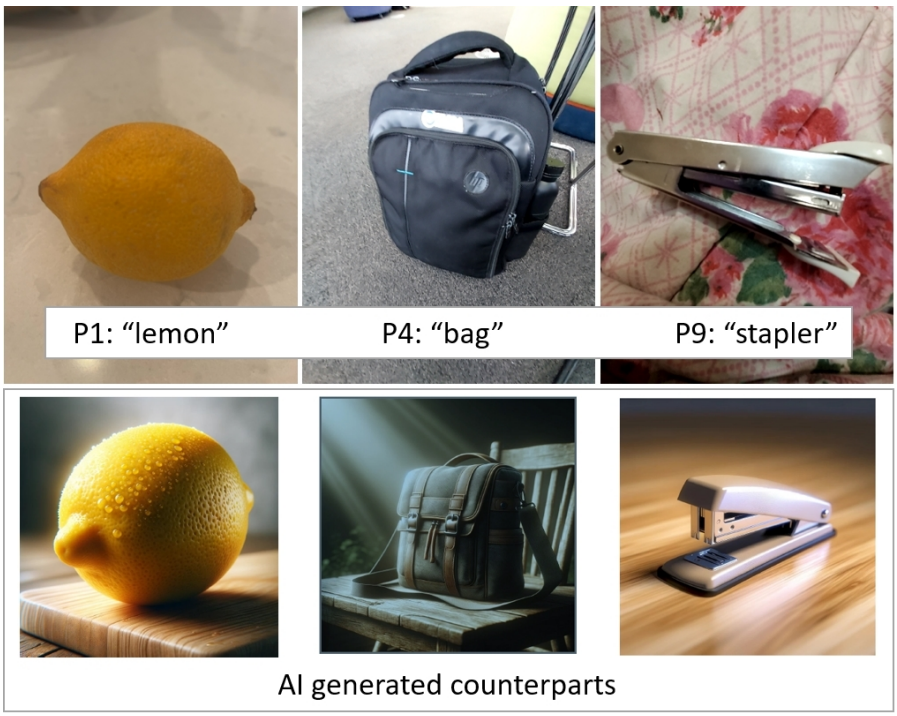


Figure 6.18: Sample images taken by participants and their AI-generated counterparts used in the study for the control group.

learners would likely find in their households (see Figure 6.19). This setup was used for the camera group. For the control group, the room setup was the same, but no objects were offered. The control group interacted with pre-curated memories.

PRE-SESSION BACKGROUND QUESTIONNAIRE. Before participants were presented the app and started working on their training tasks, they were asked to fill a short background survey including questions on their French learning and language learning apps.

TRAINING TASKS. Participants received visual and verbal information on the general goal of the app, followed by a demonstration of its main features. All participants were given a chance to ask the investigator questions at any moment; they also had a printed document with a tutorial on using each feature step-by-step with pictures. Participants were asked to explore all the main features at least once before proceeding to the post-session questionnaire.

Both groups of participants received training; the main difference is that the camera group participants had the camera mode for taking pictures, whereas the control group did not. Thus, the control group

Figure 6.19: The top picture shows the room set up with the video camera. The bottom picture shows the eight objects available for the participants' training. The camera group was invited to explore (take pictures and create memories) as many objects as they wished.

started by exploring and editing memories, while the camera group started by taking pictures to curate their memories.

POST-SESSION QUESTIONNAIRE. This survey evaluated the usability and practicality of LangEye as a language learning tool. Learners were asked about their perceived app usefulness.

6.4.3 *Between Sessions*

At the end of Session 1, participants were instructed to use the app between sessions 1 and 2 of the study; they were asked to use LangEye at least once daily for 5 days. Depending on availability, their second session occurred five to ten days after their first session. During this period, participants received daily reminders to use the app and complete a short feedback form, which included error reporting and usability questions.

6.4.4 *Study Session 2: Exit Semi-Structured Interview*

Participants were asked similar usability questions from the post-session questionnaire in this session. In addition, they answered questions about their experience using the app between sessions, including their perceived learning of French vocabulary, the accuracy of the object recognition and labelling features and the usage of their self- and pre-curated memories to build vocabulary for the camera and control group, respectively.

The interviews were recorded using the built-in feature of Google Meets which included video, audio, and a transcript from the meeting. During the interview, I used the equivalent form from the Study Session 1 questionnaire to support note-taking and structure of participants' answers. After each session, I reviewed the transcript and the video to verify its data integrity as well as accuracy of the automatic transcript. Both the structure form notes and the transcript were used for the qualitative analysis of responses.

6.5 STUDY RESULTS

This section presents the results from the user study, including the pre-session, post-session, exit interview, and software usage data.

QUALITATIVE DATA ANALYSIS. The questionnaire and the semi-structured interview open-ended questions were coded into categories following commonly mentioned themes in the participants' answers. The original supporting quotes were linked to these categories. For some questions, I used a systematic coding process with questions such as "What is your most and least favourite feature?", where the codes were the names of the existing features. An open coding including a thematic analysis approach was taken when participants were describing their overall experience using the app, those there were aggregated and presented with respective supporting quotes during discussions.

6.5.1 *Pre-session background questionnaire*

PARTICIPANTS BACKGROUND. Participants' ages were collected in broad standard age categories. They split across two groups: 18–24 (N = 15) and 25–30 (N = 5) years old. All participants declared to be fluent in English (N = 20), while a minority of participants declared English as their first language (N = 6). While most participants (N = 15) self-declared as beginner (A1 and A2) learners in both groups, 9 out of the 15 were in the camera group. Thus, the control group had the most advanced (B1 and B2) learners, 4 out of 5. This also corresponds with the proportion of participants in the control group who had been learning French for longer than 1 year. See Table 6.2 and Table 6.3 for the data summary.

Participants also declared a variety of languages that they are fluent in, other than English: Tamil (N = 5), Hindi (N = 4), Urdu (N = 3), Gujarati (N = 2), Cantonese (N = 2), Mandarin, Igbo, Spanish, Telugu, and Punjabi; 4 participants did not declare another language. Some other languages were mentioned when asked about any other languages they are not fluent in but are trying to learn, in addition to French: Arabic (N = 2), Hindi, Spanish, Italian, Dutch, Japanese, German, Norwegian, Cantonese; 11 participants were not studying any other language.

Table 6.2: Summary of participants' background information per study group: camera and control.

Attribute	Camera	Control	Total
Age			
18–24 years old	6	9	15
25–30 years old	4	1	5
L1			
Other	8	6	14
English	2	4	6
French level			
A1	5	5	10
A2	4	1	5
B1	–	4	4
B2	1	–	1

Table 6.3: Table shows participants' main method for learning French and the duration of their studies. The study was run in Canada, a bilingual country (English and French are official languages). Thus, French immersion schools are commonly available in Canadian education. Courses for French ("Course at school") as a foreign language are also common in Anglophone schools. And other "French course" or classes are easily accessible in language institutes. Participants who indicated "None" were never enrolled in a course or followed a specific method.

French Study		Camera	Control	Total
Duration	**Method**			
1 week or less	None	3	1	4
less than 6 months	French immersion	–	1	1
	Online course or resource	1	–	1
	None	1	–	1
1 year+	Course at school	–	1	1
	Online course or resource	1	1	2
5 years+	Course at school	1	5	6
	French course	1	–	1
10 years+	French course	1	–	1
	French immersion	1	1	2

TECHNOLOGY FOR LANGUAGE LEARNING. When asked about technologies, including websites and applications they have used to study French, most participants mentioned Duolingo (N = 12); other mentions were Udemy, Coursera, Renshuu, word reference, Learn French, and Memrise. On the other hand, 8 participants *"do not currently use any apps"*, where 5 belong to the control group. Only 1 participant from the camera group declared to use apps for French learning *"daily"*. The other 7 participants said to use apps *"Once a week or less"*, where 3 and 4 are from the camera and control group, respectively. The other 4 participants declared to use apps *"3 times a week or less"* (1 — camera) and *"6 times a week or less"* (2 and 1 for camera and control, respectively).

Equally, in both groups, *"mobile"* was chosen as their preferred platform for language learning (N = 14). 5 participants chose *"desktop"* (3 and 2 from the camera and control groups, respectively) and 1 from the control group chose *"either"* platform.

LANGUAGE LEARNING GOALS. The most popular top motivating factor for participants to study French was *"Career-related"* (N = 11 — 7 and 4 for camera and control, respectively), followed by *"Leisure-"* (N = 4) and *"Travel-"* (N = 3) related. Lastly, *"Academic-"* (N = 1) and *"Family-"* (N = 1) related, for camera and control groups, respectively.

Furthermore, participants also declared their personal importance of mastering each of the four competencies (speaking, listening, reading, and writing) in French using a 5-point Likert Scale. Their answers are tabulated in Table 6.4.

6.5.2 *Post-session 1 questionnaire & Session 2 Semi-structure Interview*

USABILITY. Participants were asked to rate their experience using LangEye during Session 1 based on their first impressions and during the exit interview (Session 2) after have used the app for five days or more. Figure 6.20 shows these ratings across sessions for each LangEye feature they were exposed to. Session 1 scores were overall higher than its respective Session 2 for all groups. The Memories and Achievements features dropped the most for the control and camera group, respectively.

Table 6.4: Participants level of importance for them to master each of the competence skills in French. Writing seems to be the least important skills out of the four skills.

Skill	Mean	Median	MIN (1)	MAX (5)
Speaking				
Camera	4.1	4	2	5
Control	4.3	4.5	3	5
Listening				
Camera	4.2	4	2	5
Control	4.4	4	4	5
Reading				
Camera	4.2	4.5	2	5
Control	4.1	4	3	5
Writing				
Camera	3.9	4	2	5
Control	3.3	3	2	5

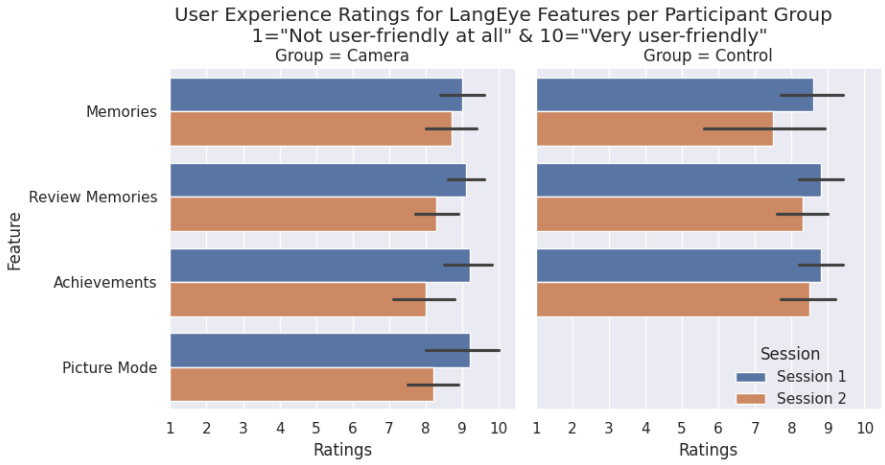


Figure 6.20: Using a 10-point user experience scale where (1) "Not user-friendly at all" and (10) "Very user-friendly" participants provided ratings on each of four LangEye features: *Memories*, *Review Memories*, *Achievements*, and *Picture Mode* — for camera group only. The chart on the left represents the camera groups and the control group is shown on the right. Overall, Session 1 had higher ratings than Session 2, and the camera group had higher ratings than the control group.

COMFORT AND CONTROL: POST-USAGE PERCEPTIONS. Using a 5-Point Likert scale, participants from both groups were comfortable using LangEye for reviewing vocabulary learnt and for learning new vocabulary (Table 6.5). While the camera group displayed similar comfort level for both activities, the control group showed lower levels for learning new vocabulary. This is likely due to the pre-curated and limited set of vocabulary they were given. Meanwhile the camera group, can continuously expend their vocabulary using the Picture mode. Beyond that, during their session 1, 4 out of 5 participants explicitly asked about options for expanding the list of pre-curated memories (vocabulary). In terms of self-efficacy or control, both groups reported similar ratings.

MOTIVATION LEVEL: PRE- AND POST-USAGE PERCEPTIONS. Using a 5-Point Likert scale, participants self-reported their overall motivation levels for learning French in both sessions (Table 6.5). While the motivation level for the camera group remained similar for both sessions, the control group reported lower motivation during Session 2 in comparison to Session 1. In addition, the control group reported overall lower motivation levels in comparison to the camera group.

MOST AND LEAST FAVOURITE LANGEYE FEATURES. When asked about their most and least favourite LangEye features, the most liked ones by the camera group were the Picture Mode (4), Review Memories (3), and Memories (3); P9 (A2) commented *"[Memories can be] used to retrieve words even if not around anything picture worth."* P8 (A2) commented that the Picture mode allows them to learn in real time *"Picture mode learning in real time, make feel "learning with stuff around me" like a calculator."*

The most liked ones by the control group were Review Memories (8), Memories (1), and Achievements (1); where one person specifically mentioned the Review Memories' Phase 1, adding it was *"easy and straightforward to use"* (P20 — A1). P15 (B1) commented on the Memories *"well organized, liked [the option to] add or remove words, also dictionary aspect to translate, which is good."*

For the camera group's least favourite LangEye feature they indicated Review Mode's Phase 3 (3), Phase 2 (2), Picture Mode (2), Achievements (2), and P7 (A1) insisted they did not have a least favourite feature — None (1). Several participants indicated that the

Table 6.5: Tabulation of the self-reported levels of (1) comfort using LangEye for reviewing and learning vocabulary post-study, (2) control over learning post-study, and (3) motivation for learning French per session.

Comfort level reviewing vocabulary		**Mean**	**Median**	**MIN (1)**	**MAX (5)**
Camera		4.3	4	3	5
Control		4.4	4.5	3	5
Comfort level learning new vocabulary					
Camera		4.2	4	4	5
Control		4	4	3	5
Perceived level of control over learning					
Camera		4.3	4.5	3	5
Control		4.3	4	3	5
Perceived level of motivation for learning French per Session					
Camera	Session 1	4.3	4	3	5
	Session 2	4.3	4.5	3	5
Control	Session 1	4.1	4	3	5
	Session 2	3.8	4	2	5

sentences displayed in Phases 2 and 3 were at times confusing, for example P1 said *"second stage sentences don't make sense, makes it more confusing"*, and P8 (A2) added *"more confusing than helpful"*. For Phase 3, participants felt overwhelmed with the task where they were presented a full French sentence and suggested *"the 3 phases is good. . . [for Phase 3] just translate a particular word as needed; like Duolingo allows to check the particular word in the sentence in French."* (P7 — A1).

For the control group's least favourite LangEye feature, the mentioned ones were Achievements (4), Memories (3), Review Memories' Phase 2 (1) and Phase 3 (1), and None (1). Participants indicated Achievements as their least favourite and justified this choice for not having used it as much as the other features; therefore, they found it *"less interactive"* (P19), and they wished it to have more options for reviewing their past attempts; for example *"split per phase section"* (P11). For the Memories, P20 explained that *"Word definitions when adding label don't make much sense. . . not straightforward; too abstract."*. In addition to the similar comments regarding sentence confusion for phases 2 and 3, P15 added *"phase 2 adds a bit of confusion — image that has wood, objects that are on top of wood, which was hard [to choose the correct option]."*

6.5.2.1 *LangEye versus other language learning apps.*

Participants were asked to consider their experience with LangEye and compare it to other language learning apps, in a general sense, they have used or might know of. They were followed-up with their specific opinion about LangEye and other vocabulary-specific learning apps or resources.

LEARNING VOCABULARY WITH PICTURES. Participants agreed that using picture along with vocabulary words is positive for learning; LangEye differs itself from regular vocabulary or translation of a word for having a visual aid in addition to the definition and sample usage.

"[...] image is so much better; I don't know what the word was but with the picture it makes more sense, and easier for me. I'm not sure about what other people think. It's easier to memorize and associate the vocab with object." P15 (B1) from the control group.

"better feel [than dictionaries] on memorization and create associations to images..." P20 (A1) from the control group.

"Pictures help a lot as it serves as a different input." P14 (A1) from the control group.

LEARNING VOCABULARY WITH THEIR SELF-CURATED MEMORIES. For the camera group, the images used were taken by their smartphone camera using the Picture Mode. Their responses were mostly focused on the helpfulness of using their self-curated images. Overall, participants like the sense of control and personalization that curating their memories gave them.

"The picture factor; [LangEye] gives vocabulary based on pictures I take, and adds a sense of personalization into it." P8 (A2) from the camera group.

"The vocabulary resonates with our daily living, my own things, real life, closely resonating with my own environment not like the commercial images...[to] learn, [to] register [vocabulary] with my own pictures...[to] remember it easier and helps with a foreign language to have familiar objects." P3 (A1) from the camera group.

"With this apps you choose to learn words relevant to you, not just any [word]. Duolingo teaches basic, but you can't learn something you use ev-

eryday you need to progress through it…[LangEye] is a lot more engaging, more personal. [In a] dictionary is hard to find the word and you can't go back and review, this [in this app] you can review. You are likely to remember it…your own picture is easier to remember [the vocabulary]." P1 (B2) from the camera group.

"*[Pictures of] stuff around you makes it easier to remember.*" P6 (A2) from the camera group.

"*In this app [LangEye] you are more in control of the content…at the same time more limited to your own content…other apps [there are] more options and less in control.*" then suggested "*a mix of pre-populated and also add your own stuff would be great…*" P9 (A2) from the camera group.

"*picture mode stands out, as most apps have their own taylor made, own agenda [i.e. curriculum], in this app [LangEye] we have more control in learning pattern and pace by taking pictures of daily life, more flexible than other apps…When look up words on dictionary, when I take a picture it [LangEye] automatic tells the name in French, which is way more efficient and convenient…other positive difference is the picture is something I always see and use in my everyday life so it's easy to associate and recall to the words I learned in the app.*" P5 (A1) from the camera group.

MULTIPLE CONCEPTS ASSOCIATED TO THE SAME IMAGE. Another distinctive trait of the Memories in LangEye is the usage of CV to identify and label objects in the image, this causes images to be associated with a set of labels (concepts). This characteristic can be helpful with contextual learning of vocabulary. However, it can also confuse learners when multiple memories contain the same word/-concept as reflected in the feedback for the Memories Review's Phase 2 and Phase 3. While learners can edit labels, novice or beginner learners can get overwhelmed with the decision making.

"*Other apps might have pictures of only one word or object…[in LangEye] the pictures have more context and words, good for intermediate learners…other people might not enjoy if there too many pictures with the same words.*" P14 (A1) from the control group.

"*other apps show exact words; here is good, you, you get options.*" P2 from the camera group.

"*I like it [LangEye]…it uses image for learning, the image it seems to teach…like…different concepts.*" P20 (A1) from the control group.

"this [LangEye] reminds me of Real Kana[3] [app for learning Japanese]... you can choose what font to use [for Japanese writing], customizing ... similar effect of the images, associate cutlery and fork and know they can associate multiple terms to the same image." P16 (B1) from the control group.

6.5.3 *Learner Comments and Feedback*

LANGUAGE-SPECIFIC CONSIDERATIONS. Several participants indicated issues with the Review Memories' Phase 2 due to the translation of the sentence into English. In French, articles and nouns are gendered; when translating the sentences for the fill-in-the-blank activity the gender information is lost in English (where nouns have no gender). This caused confusion when choosing the correct answer for some participants, particularly if their were not beginners — A2 to B2 learners. Therefore, some suggested that the gender of nouns should be indicated (P1, P8, P9, P16).

APP REMINDERS AND GAMIFICATION. While some participants explicitly appreciate the lack of in-app reminders or incentives for daily or frequent engagement (i.e., P12); more participants expressed to miss having incentives and daily reminders, particularly for reviewing vocabulary (i.e., P1, P2, P16).

"nice to not have constant reminders; no in app notifications. I like the complete control of when to use, instead of pressure to tune in'.' P12 (A1) from the control group.

"Notification to study in the app should help for review.". Suggestion by P1 (B2) from the camera group.

CONTENT CUSTOMIZATION. Including both groups, all participants appreciated and made mention of having images in the memories; which they felt to be a positive aid for learning and memorizing vocabulary. All participants from the camera group made mention of the positive impact of having personalized images attached to their vocabulary.

'This app adds personal attachment to the picture. The more connection to something the easier to remember [it]... like the picture I took of a calculator

3 https://www.realkana.com/. Accessed in July 2024.

I used in the past 10 years has more connections [than a stock photo]... with the app, do [study] anywhere in the phone rather than the dictionary... here can focus on the words that are important to me. Some thing more customized to them.". P8 (A2) from the camera group.

AI GENERATED CONTENT AND ROBUSTNESS. Participants from both groups indicated inconsistencies with object detection, which caused uninteresting labels to be attached to memories. For example, if all objects are on top of a wooden table, those objects' memories will likely include the keyword (label) "wood". Another known issue with object detection for learner-taken pictures is the framing of the object itself; a clean background is not always available when the picture is taken. Clutter in the background and/or having multiple object in frame can cause mislabelling or unintended labelling for memories. While the editing of labels mitigates this issue, not all learners may realize the mislabelling due to unfamiliarity with the TL. Another option to reduce mislabelling is to allow learners to crop their images or create bounding boxes on their target objects.

"*Sometimes, the picture, the object on camera... it doesn't get the object in the forefront but the background; small bug would be a perfect app.*" P10 (A1) from the camera group.

In addition, participants raised opportunities to improve the AI-given definitions, which they observed to be too technical at times. And, sentences used in the Review Memories or as example, at times, may not be (1) contextually correct or (2) not contain the vocabulary word, but a variation of it. For example, the sentences given for "wood" (noun) contain "wooden" (adjective) instead; although contextually it can have the same meaning, the word studied is a noun and the produced one is an adjective. Similarly, this unintentional behaviour from the LLM could bring a learning opportunity to users, but at the same time, it will likely confuse them.

"*In a 'Memory,' at times, the definition or sample sentence didn't make sense.*" P19 (B1) from the control group.

REVIEW MEMORIES DESIGN. Beginner learners (A1), in particular, preferred the Review Memories' Phase 1 and Phase 2, and concerns about the Phase 3. Due to their limited knowledge of the language — some of them declared that this was their first experience trying to

learn French, the Phase 3 was overwhelming or frustrating to them. At the same time, when asked if they had considered finishing their review session after completing Phase 2 and skipping Phase 3 altogether, they declared to have forgotten they could skip it or that they wanted to be able to finish the session (all three phases). In contrast, more advanced learners found issues with Phase 2 where they were given the English sentence both due to the gendered nouns misrepresentation, but also because they wanted the sentences to be in French. Overall, participants enjoyed the multiple phase system in the review session.

"I'd recommend the app with better features for beginners; phase 3 is a large sentence, beginners cannot understand it, its overwhelming…until second phase is good." P2 (A1) from the camera group.

"Phase 3 question could have an English translation before and after the activity." P17 (A1) from the control group.

"For review memories, the second one, the sentence in English is not helpful because of the gender of the words. But I like the three phases." P9 (A2) from the camera group.

"Second stage sentences don't make sense, makes it more confusing…first and third stages were fine." P1 (B2) from the camera group.

"The review, I had to pick this one as my favorite…I like how it has three phases, by matching it contextually (word and sentences; word and picture, word and English plus the listening the sentence pronunciation is really good)." P11 (B1) from the control group.

AESTHETICS IMPROVEMENTS. Many participants (9 out of 20) suggested to make the app more colourful, or to have multiple colour schemes to choose from. They also mentioned that they like seeing more animations when interacting with the interface and sounds when actions are performed. For example, there should be a bell for correct and incorrect answer checking. Some buttons (i.e., the translate) should be larger to make tapping easier.

6.5.4 *Between Sessions*

In between sessions, I had two main data sources: the daily feedback for submissions, and the app usage data and logs.

6.5.4.1 *Daily Feedback Forms*

DAILY FEEDBACK RATINGS. Participants were asked to submit a daily feedback form after using the app in between sessions. While the number of daily submissions (Figure 6.21) remained somewhat stable or the control group (from 7 to 5), the camera group had the most decline (from 7 to 2). When aggregating both groups, at Day 1 there were 14 submissions, which was reduced to 7 at Day 5. The charts in Figures 6.22 and 6.23 show the participants' ratings (5-Point Likert scale for agreement) per day. The difference in the volume of submissions makes it difficult to compare across groups, but overall, the ratings lean toward "Strongly agree" to "Neutral" throughout the study days. Here are the statement items participants were asked to rate:

- "Overall, this app is easy to use."
- "I'm having fun using this app."
- "I have learned new French words using this app."
- "I feel more in control of my French vocabulary learning progress and content since using this app."

When analyzing the control group ratings for the item "I have learned new French words using this app," it shows a different trend from other items having more negative ratings. This might be related to a couple factors: (1) the control group had more individuals with higher proficiency levels (N = 4 for B1–B2) than the camera group (N = 1 for B1–B2), and (2) the control group could not add new words. While the first factor likely influences this item the most, both are related. The second factor can also have influenced the ratings for "I feel more in control of my French vocabulary learning progress and content since using this app.". When compared to the camera group which shifted to positive ("Strongly agree" and "Agree") ratings, the control group remained with a split between "Strongly agree", "Agree", and "Neutral".

TECHNICAL ISSUES OR DIFFICULTIES. 26% of submissions declared some technical difficulties or issues with the app. Some were related to lag or delay loading labels or information sourced on-demand from APIs while using mobile networks (i.e., during a bus

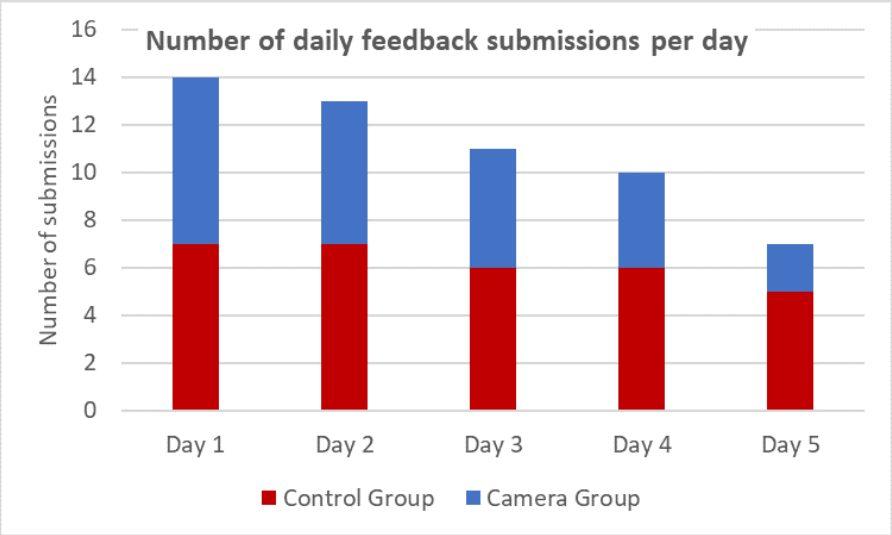


Figure 6.21: Starting with a total of 14 submission in the first day of study, it declined to 7 for day 5. While the submissions remained somewhat stable or the control group (from 7 to 5), the camera group had the most decline (from 7 to 2). The charts in Figures 6.22 and 6.23 show the participants ratings per day.

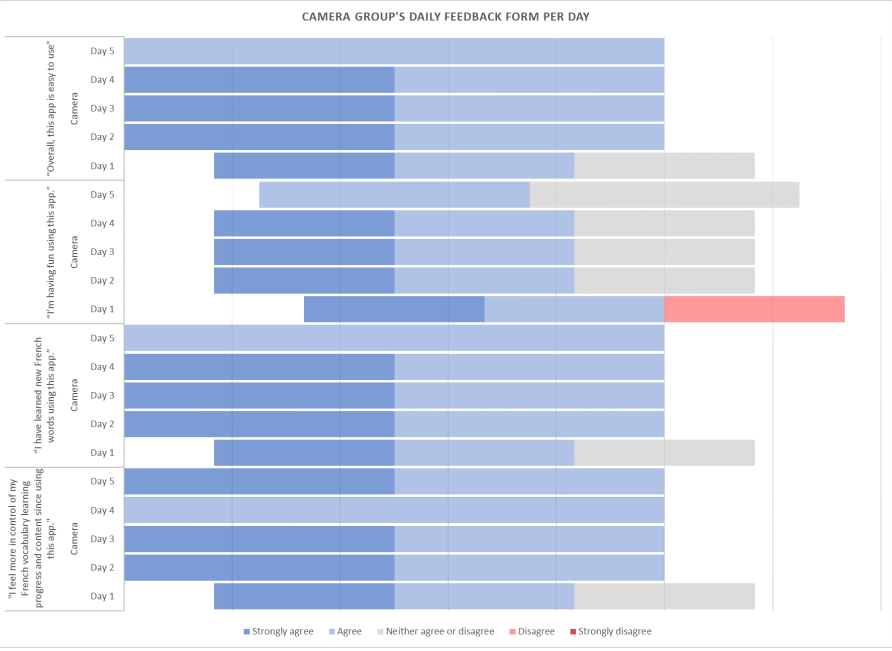


Figure 6.22: Camera group participants' ratings for the daily feedback form per day. The Figure 6.21 shows the number of responses per day. While at Day 1 there were 7 submissions, that number declines along the days. This chart shows the distribution of the respondents answers to the 5-Point Likert scale agreement statement items at each day, from bottom (Day 1) to top (Day 5) at each item. Although there is a shift to "Strongly agree"/"Neutral" as days pass the number of responses are reduced.

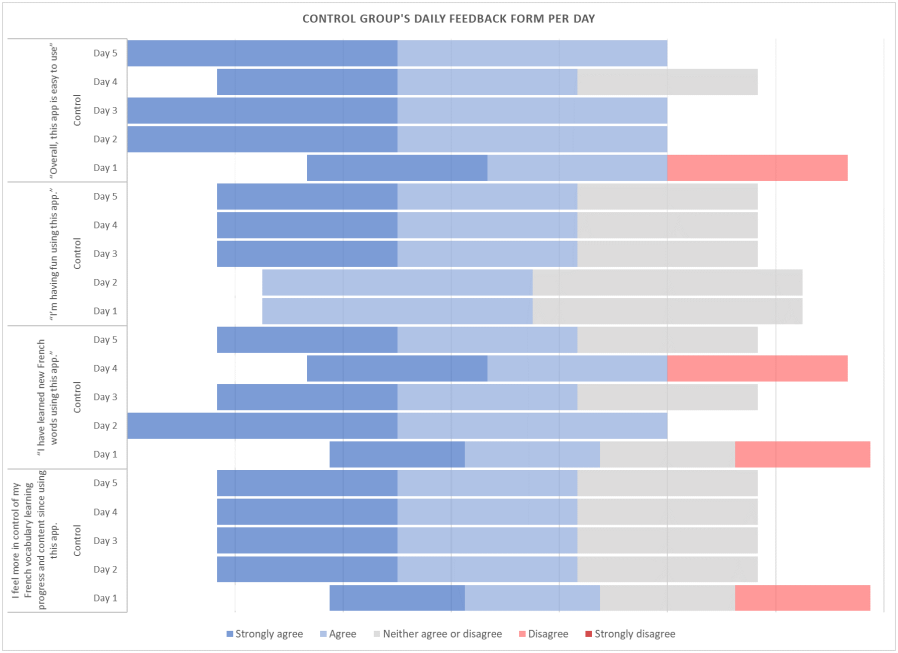


Figure 6.23: Control group participants' ratings for the daily feedback form per day. The Figure 6.21 shows the number of responses per day. While at days 1 and 2 there were 7 submissions, that number declines to 5 at Day 5; which is higher than the camera group's Day ($N = 2$). This chart shows the distribution of the respondents answers to the 5-Point Likert scale agreement statement items at each day, from bottom (Day 1) to top (Day 5) at each item. Although there is a shift to "Strongly agree"/"Neutral" as days pass the number of responses are reduced.

ride). However, the majority of the issues reported were related to the quality of the sentences or incorrect labelling of objects, as also reported during the interview reported in this section.

6.5.4.2 *LangEye Usage: Software Data*

Unsurprisingly, among the core LangEye features, the Picture Mode was the most used by the camera group; followed by Review Memories, Memories, and Achievements. Without the Picture More, the control group follow the same pattern. The camera group also had 58% of the total (both groups) actions performed. At the same time, the camera (48%) and control (52%) group had a similar volume of actions for other features (not core). Table 6.6 contains the number of actions and proportion across features and participant groups.

Table 6.6: Number of actions logged per participant group and the proportion of actions for each feature. Features are split in "core" and "other".

Actions	**Group**		
Core Features	**Camera%**	**Control%**	**Both%**
Picture Mode	37%		21%
Review Memories	28%	45%	35%
Memories	24%	34%	28%
Achievements	11%	21%	16%
Num. Actions	**262**	**187**	**449**
Other Features			
Text-to-Speech	61%	63%	62%
Settings	5%	3%	4%
Feedback	12%	11%	11%
Login	22%	23%	23%
Num. Actions	**327**	**351**	**678**

The "Text-to-Speech" feature was the most used among the other features of LangEye. This feature was accessible from the Memories and from within the Memories Review. So, this feature is technically a lower-level function, which explains the higher number of actions.

6.6 DISCUSSION

LangEye is a full application that explores the implementation of generative AI and other Machine Learning (ML) technologies to enhance language learning. While I focused on vocabulary learning using the smartphone camera as the main input, the results from the presented user study help guide future work related to generative AI and education, as well as mixed-reality applications. Many of the technical challenges faced in LangEye are limitations of the state-of-the-art of the technologies used. However, since implemented many more have studied and presented better prompting techniques that can help mitigate some of the issues found. For example, using a few-shot prompting technique to help providing simple example sentences. In this section, I discuss the takeaways from the user study, and I present the answer to my research questions for the user study and some discussion to guide future work.

6.6.1 *Takeaways*

BALANCE BETWEEN PRE- AND SELF-CURATED CONTENT. Self-curation leads to more control and flexibility, but the trade-off is the work (cost) involved in the creation of the content. Thus, learner curation increases learner autonomy, but it should be a tool in addition to pre-curated content. For most learners, self-curation is a means to expand and supplement a base curriculum. For the current state of LangEye, this study design choice helped control the content of learning with minimum "teacher" intervention for curriculum creation. Thus, this method can be followed to create a curriculum in any language, not limited to French or to the researcher's language proficiency. While this decision imposes limitations to the results, it also allows the research to focus on understanding the learners' perspective of being the sole creator of their language learning app. Future work can feature pre-curated content along with the flexibility provided by LangEye and focus on other aspects of learner interaction, such as their willingness to curate content and to deviate from the pre-curated curriculum.

THE AFFORDANCES OF AI-GENERATED CONTENT FOR LANGUAGE LEARNING. The language used by AI agents can sometimes be

overly complex or technical, which may pose challenges for learners. To address this issue, prompt engineering techniques can be employed. For instance, using the instruction "make the explanation suitable for a language learner at A1 level" can help simplify the content. Additionally, the ambiguity of images and contextual concepts can confuse learners. One solution is to improve quiz design by ensuring clearer options. For example, in a multiple-choice quiz, instead of listing "Fruit" alongside various fruits as options when only one answer is expected, it might be better to use this format in a different type of activity, such as "choose all appropriate options," where all images of fruits would be considered correct. Furthermore, inconsistencies in translation and word sense can also be problematic, but adopting better prompt engineering practices may help minimize these issues.

LANGUAGE-SPECIFIC NEEDS. Although LangEye aimed to provide a flexible architecture that does not rely on language expertise to generate a learning curriculum, this presents some trade-offs with accuracy and considerations for language-specific traits. While LangEye can support any languages given their respective translation models, the study showed issues with French gendered articles and Phase 3 translations. Similar traits can be found in Romance languages (i.e., Portuguese) and other languages such as German. Therefore, the current structure provides a base framework for supporting several languages, some degree of language expertise is need to make the appropriate adaptions.

SELF-CURATED CONTENT FOSTERS PERSONALIZATION AND PROVIDES FLEXIBILITY. The study showed that learners overall approve of using picture to learn vocabulary, however, the camera group indicated some positive effects of using and learning with their personal objects. Furthermore, they indicated that it allowed them to create attachment to the vocabulary, which they thought to be a more memorable experience.

REVIEW MEMORIES STRUCTURE. At its present form, the Review Memories require improvements in the integration of generative AI, but its designed structure and set of activities were overall perceived positively and helpful by learners.

6.6.2 *Research Questions*

Based on the user study results, I can discuss the answer to the following research questions:

- **[R1]:** Are learner-curated pictures to learn vocabulary better for motivation than stock images in LangEye?
- **[R2]:** What is the potential of LangEye as a language learning tool for acquiring vocabulary in a foreign language?

[R1] LEARNER-CURATED MEMORIES. The results indicate that learner-curated pictures (vocabulary) lead to **higher self-perception of control over learning and provide more flexibility**. The camera group also reported **higher motivation levels compared to the control group**. Furthermore, it also fostered a sense of customization and affinity with the content, which was perceived positively by the camera group participants. Thus, the answer to [R1] is positive; learner-curated pictures are better for motivation when learning vocabulary. Furthermore, the study indicates LangEye as a tool to promote autonomy on language learners when considering the following:

1. self-reported confidence levels in the study;
2. personalizing increases motivation (i.e., Leong et al. [88]); and
3. self-efficiency (control over learning) promotes autonomy (Littlewood [92] and Ushioda [135]).

[R2] POTENTIAL OF LANGEYE FOR VOCABULARY ACQUISITION. The study did not deploy a specific vocabulary knowledge test, neither pre- or post-study to measure learning objectively. However, the study indicated that LangEye promotes autonomy for learners, which indicates potential for language learning. Although there many steps to improve the user experience as well as the consistency with the AI generated content, learners were receptive of LangEye and generally optimistic about the future adoption or recommendation of it for vocabulary learning.

6.6.3 *Limitations & Future Work*

The study only included learners of French, even though the framework enables support for other languages. Future work should include a more diverse set of learners and languages. Another limitation is the participant population, where many participants self-reported as beginner (A1) learners, were truly novices — they started learning French as they joined the study. Ideally, I would want to control and screen for people who have been studying French (or another TL) for longer, even if at beginner level. Because vocabulary learning is only on aspect of learning a language, and LangEye was designed for supporting vocabulary acquisition, only.

Another limitation is that both groups of participants did not work on the study concurrently. Given that the participants are primarily undergraduate or college students, they may be working at the study at busier times of their school term. Thus, the difference in time frame can yield differences in their engagement with the app between sessions. This might be related to the difference and decline of answers to their daily feedback forms while the app usage does not differ much between groups.

Based on the feedback received from learners, improvement in the app includes:

- Improve word selection for Review Memories, make it more consistent, some participant reported erroneous behaviour of word selection.
- Add explicit learner feedback for the memory where it can be reviewed more frequently.
- Notification system within the learner's mobile phone for the daily reminders; it would make it more organic than via emails.
- Improve the sentence and definition generation.
- Add language-specific traits; i.e., indicating noun's gender in English translations.
- Introduce a reward system for editing and correcting labels within memories. This can redesign the label editing as a learning tool.

In the future, I want to control the number of days or amount of usage more systematically. Similarly, despite reminders, many participants did not fill out their daily feedback forms which would allow me to better track their sentiment throughout the study period.

6.7 SUMMARY

This chapter presented the design, feature exploration, implementation, and user evaluation of LangEye, a mobile web application designed to enhance contextual vocabulary acquisition through immersive MALL. LangEye explores the implementation of innovative activities such as object recognition, object action, and treasure hunts, enabling learners to engage with their environment to learn new vocabulary. LangEye is a data-driven application, integrating APIs like Google Vision for object detection, OpenAI GPT-3 for generating word definitions, Google Translate for translations, and Google Text-to-Speech for pronunciation.

I presented the development process, focusing on memory creation and review functionalities, which leverage generative AI APIs to generate definitions, sentences, and contextual usage examples. The user evaluation methodology includes a one-week longitudinal study with French language learners, comparing the experiences of a control group using pre-defined images and a camera group utilizing the full app's capabilities.

The study's results demonstrate LangEye's effectiveness in motivating learners and facilitating vocabulary acquisition through learner-curated content. Feedback from participants highlights the app's usability and potential to significantly improve language learning outcomes through engaging and interactive experiences. Finally, the work presented in this chapter demonstrates the potential of combining data-driven methods with learner-centered design to promote autonomy and enhance the language learning process.

7

CONCLUSION

From the lessons I learned, I invite you to share my vision and dream my dream. I hope for a future where language learning is organic: accessible, driven by data, technology, and personal interest and experiences. We shall not hide precious computational resources behind a technical wall. Systems should be user-centred and data-driven.

This dissertation described the design, implementation, and evaluation of three data-driven applications for language learners aiming to promote autonomy. The takeaways from these projects highlight the need for flexible, personalized, and data-driven approaches to language learning. My work contributes to the academic understanding but also provides practical solutions that can be implemented in educational settings to enhance language acquisition through learner-centred apps.

The H-Matrix project introduced a novel hierarchical matrix visualization technique for error analysis in language learning. By allowing detailed error analysis and hierarchical organization, H-Matrix helps educators and learners understand common mistakes and focus on areas needing improvement. This project emphasizes the importance of data-driven insights in tailoring educational content to individual learner needs.

Card-it leverages a FSM analyzer to facilitate the creation of digital flashcards for learning Italian verb conjugation. The Card-it application supports both studying and testing using flashcards, offering quizzes and immediate corrective feedback. Another feature of Card-it is its classroom management capabilities, which allow teachers to create classrooms, share collections, and monitor student progress through detailed statistical overviews. This feature helps teachers tailor their instruction to better meet student needs and track their progress over time.

The LangEye project explores the integration of the learners environment into their language learning journey. Although some of the features (i.e., Scan&Find and Treasure Hunt) were not included in the user evaluation; they present relevant ideas that can be explored in

future work, or even in the mixed-reality context. That said, LangEye still leverages concepts present in augmented reality to help learners capture real-world objects and associate them with language memories. By integrating environmental interactions, LangEye makes vocabulary learning contextual and relevant for learners.

Based on these projects, language learning applications should be flexible and personal, to empower learners using learner-centred design. Data-driven learning should aim to increase accessibility of data and linguistic resources. Generative AI is a powerful tool that support learner-curated and personalized content, but it also presents trade-offs, such as content appropriateness, accuracy, and TL-specific traits. Thus, generative AI along with emergent technologies such as mixed-reality platforms present potential to expand the ideas discussed in this dissertation.

7.1 SUMMARY OF KEY LEARNINGS AND THEORETICAL IMPLICATIONS

The findings of this dissertation underscore the importance of learner-centred design principles, which are rooted in theories of autonomous language learning and self-regulated behaviour. Across all three projects, learner autonomy was promoted by granting users control over their content, pacing, and learning strategies. These projects showed that fostering motivation and engagement relies on a balance of structure and freedom, where learners can interact with tailored content while maintaining the flexibility to pursue their unique learning paths toward language learning autonomy.

Furthermore, my research highlighted that personalized and adaptive technology-driven approaches can effectively support language learners at different stages of their autonomy continuum. Integrating data-driven insights with user-centred design methods offers powerful means to enhance both engagement and learning outcomes.

7.2 METHODOLOGICAL REFLECTIONS AND CHALLENGES

Several methodological insights emerged from this work. The combination of user interviews, expert evaluations, and iterative design processes provided rich qualitative and quantitative data to inform the development and evaluation of each application. However, challenges

such as scalability, generalizability of results across diverse learner populations, and biases in data collection were encountered. For example, LangEye's reliance on real-world object recognition raised concerns about data accuracy, while Card-it's flashcard generation was constrained by linguistic rules.

7.3 SUMMARY OF LIMITATIONS AND CHALLENGES

While the applications presented meaningful advances in learner autonomy, several limitations exist. The main limitations include: (1) data dependence, (2) bias, (3) scope, and (4) timing.

DATA DEPENDENCE. Tools like LangEye rely heavily on accurate data generation and recognition algorithms. Any inaccuracies or limitations on that process can potentially affect the quality of the learning content generated.

BIAS. Biases in data collection, such as object selection in LangEye or the demographics of participants in all the studies, may influence the applicability of findings. Efforts were made to mitigate these biases, but their impact must be acknowledged.

SCOPE. The focus on specific aspects of language learning, such as morphology and vocabulary, means broader competencies like conversation and writing were not fully addressed. Addressing these limitations will require expanding data sources, refining AI algorithms, and exploring broader linguistic competencies in future work.

TIMING. The research presented i nthis dissertation was conducted over an extended period (2017 to 2024), during which significant advancements occurred in the educational technology and AI fields. Rapid changes, such as the emergence of large language models and augmented reality-based tools, posed challenges in keeping the research cutting-edge by the time of its completion. Future work should be more adaptive to these fast-paced developments, incorporating newer technologies as they emerge to maintain relevance and innovation.

By acknowledging these limitations, this research lays a foundation for future exploration and refinement of technology-driven language

learning tools that better support learner autonomy in a rapidly evolving digital landscape.

7.4 LESSONS LEARNED AND FUTURE WORK

Looking forward, several directions are promising for furthering this research. Advances in AI, augmented reality, and adaptive learning could allow for even more personalized and immersive experiences. Specifically, H-Matrix can provide visual support to LLM agents natural language explanations and output regarding common transfer effects to teachers. Card-it could leverage LLMs to incorporate sample sentences of the verb forms in cards or in new quiz activities. Lastly, LangEye could leverage augmented reality to create richer interactions with real-world vocabulary contexts. Future projects could explore integrating conversational practice, writing support, and more robust user feedback mechanisms to create comprehensive learning ecosystems.

Reflecting on what would be done differently, a more diverse range of participants and longer-term studies would be prioritized to capture more generalizable data. Moreover, the integration of ethical considerations surrounding AI-driven personalization and data privacy is crucial as these technologies evolve. Particularly, future or extensions of these projects should consider the representations of uncertainty and innacuracies of the generative AI outputs.

7.5 SUMMARY OF CONTRIBUTIONS

This work demonstrates that by combining data-driven insights with learner-centred design, technology can empower language learners to take control of their educational journey. The tools developed offer a glimpse into a future where learners are empowered by technology, and language learning is deeply personalized, adaptive, and contextually relevant. This approach holds the potential to transform not only how languages are learned but how learners perceive and engage with new languages throughout their lives.

In summary, this dissertation presents the following contributions:

- H-Matrix — An application to visualize and analyze linguistic features in a large learner corpus to identify transfer effects (Chapter 4).
- Card-it — A digital flashcard application to enable teachers and learners to teach and study Italian verb morphology (Chapter 5).
- LangEye — A mobile application to promote L2 immersion and situated vocabulary acquisition (Chapter 6).

Part III

APPENDIX

A

SEMI-STRUCTURED INTERVIEW QUESTIONNAIRE

REB Application: #14426

Target audience: Non-native English speakers recruited on campus

Introduction:

- Hi! Thank you for volunteering in the study. My name is Mariana and I will be conducting this interview.
- This study has been approved by the UOIT Research Ethics Board.
- The overall goal of this study is to better understand the problems faced by non-native English speakers and to develop an application to support the improvement of English skills and language confidence. For this reason, we are performing this interview in order to get to know more about non-native English speakers background, language proficiency, possible language barriers and English learning tools used.

Consent Form:

- Here are some key points from the consent form:
- You are required to sign this consent form before we start and your compensation will be given to you at the end of this session.
- You should know that the interview session will be audio recorded and I will also be taking notes for future data analysis.
- All questions in this interview are optional and you may decline to answer any question without any loss of compensation.
- It's important to point out that nothing collected in this interview can ever interfere with any personal or academic record.
- You should also be aware that the collected data when anonymized could be made public for research purposes, for example in a research publication. Audio recordings will not be released in any form, however, de-identified quotes may be included in publications.

Finally, here is the consent form <hand form>. Please read it carefully and sign it. If you have any doubts please feel free to ask.

Study Agenda:

- Thank you! Now let me explain to you some details about how this session is going to proceed:

First, I'm going to ask you some background questions, followed by some questions on English learning. Then, I'm going to ask about tools for learning / practicing English. I am now going to turn on the recording.

<turn recording on>

Interview:

A) Background Questions

1) Gender: F / M / Other
2) Education level:
3) Home country:
4) First language:
5) Any other languages spoken:
6) How long has been living in an English-speaking country?

B) English Learning

1) How often English is used in your home country?

If needed, suggested scale:

- Rarely
- Not usual
- Frequent
- Very frequent

2) What English level do you consider yourself to be?

If needed, suggested scale:

- Low-Intermediate
- Intermediate
- Upper-Intermediate
- Advanced

3) How often English is used in your daily life?

If needed, suggested frequency of usage:

Almost exclusively English

Only at academic/work environment

4) Generally, which language you feel more comfortable using? Does it change according to the situation?

Ask to elaborate. Ex.: mother language is better for talking about personal matters, English better for explaining academic/technical matters.

5) Was there any topic/set of rules of English that have been particularly hard for you when learning?

If needed give some examples: plural, past participle conjugation, passive voice, etc.

6) Do you feel like there is one or many specific English mistakes you keep on committing even though you technically you know they are wrong? *If yes,* Please tell me about them.

If needed give some examples: plural, past participle conjugation, passive voice, etc.

7) Did you ever feel like your first language has ever affected you when learning English?

Ask to elaborate if any effect is mentioned.

8) How old were you when you started learning English?

9) How old were you when you reached your current English level?

10) How did you primarily learn English?

If course at home country, any specific licensed methodology?
If ESL course abroad, where?
If self-thought, any specific technique you can mention?

11) Do you take part in any English learning activities on an ongoing basis?

If clarification is needed, mention tutoring, lessons, groups like Conversation Cafe.

C) English Learning App

1) Have you used/ are you using any software/app for learning English?
If a prompt is needed, provide popular examples: Duolingo, Rosetta Stone.
   - ▢ Yes. Which?
   - ▢ No. Why not? (Skip to 4)

2) Do you feel the app you use does everything you need or you would want any extra features? Which functions do you feel are missing?

3) Do you feel the app you use has some feature(s) you particularly really like?

4) Have you used/ are you using software/app to help with writing in English?
   If a prompt is needed, provide popular examples: WordQ, Grammarly.
   - ▢ Yes. Which?
   - ▢ No. Why not?

5) Do you feel the app you use does everything you need or you would want any extra features? Which functions do you feel are missing?

6) When you look for an app to help you with English learning, what are you mainly looking for?

7) If they say getting better in the long-term, then ask if they feel it is working. And how they notice that.

If needed say: short-term as in getting better grades for your essays, long-term as in getting better in English in general in order to not need an app in the future.

8) Do you have any tips that can help other non-native English speakers to improve their language skills?

9) Would you be interested to be contacted in case we perform follow up studies related to today's interview? If so, please provide your contact information. *(Hand paper to provide this info – separate from study notes!)*

Thank you:
Thank you very much for participating! That will be all for me today. Is there anything you would want to add related to the topic? Do you have any further questions for me?

<Turn recording off>

Thank you again. Your compensation is here.

If you have any other questions or comments, or if you want to withdraw your data from the study, feel free to contact me at any time. If you have any concerns about this study you may contact my supervisor or the UOIT Research Ethics Board as detailed on your copy of the consent form.

B

CARD-IT EXPERT EVALUATION MATERIALS

1. Letter sent to participants after they signed up, containing information on study structure and study material access.

Dear teacher,

Thank you for your interest in our research!

As stated in your letter of information, you must complete three steps to finish the study. We recommend that the steps are followed in order :

1. Watch Card-it's video demonstration (less than 8 minutes long)
    - This video will present the application's main features so you can test it yourself. To make it more accessible, we are providing the video with English and German narration; both videos have the same content, and you may watch one or the other. You don't need to watch both.

        English narration: [anonymized for submission]

        German narration: [anonymized for submission]

2. Test the Card-it application yourself (flexible – recommended 10 to 20 minutes)
    - At this step, you may log into Card-it with the provided login information and spend as much time as you wish testing the application. You will be given access to both a "teacher" and a "student" account if you wish to compare both types of accounts.
    - App URL: [anonymized for submission]
    - "Teacher" type account credentials

        Username: [anonymized for submission]
        Password: [anonymized for submission]

    - "Student" type account credentials

        Username: [anonymized for submission]
        Password: [anonymized for submission]

    - We are attaching a list of recommended tasks you might perform using Card-it. You are not expected to use this list of tasks, they are optional suggestions only.
3. Tell us about your experience using Card-it by filling out a survey (approx. 30 minutes)
    - Survey: [anonymized for submission]

To be considered for this study, **please complete the steps above before October 14th, 2022**.

After completing all the steps, you will receive a thank-you message with your compensation on the email provided in your survey response within 48 hours. If you do not receive it, please contact the researcher [anonymized for submission] by forwarding a copy of your survey response.

Please contact us if you have any questions or require an extension to complete the study steps. You also may refer to the "letter of information" attached to your invitation email for more information on the project.

Thank you again for your interest in our research.

Sincerely,
[anonymized for submission]

2. List of recommended tasks for exploring Card-it

Possible Usage Ideas:

These are just possible ideas to let you explore the functions of Card-it. You may look around and explore the tool yourself just as well. Feel free to follow or deviate from these as much as you would like.

Teacher Account:

- Log into the teacher test account and have a look at the pre-created flashcard collections.
- Go to "Classroom" and click on Italian I.
- Do one of the available quizzes
- After the quiz, you can have a look at the learning statistics: Remember to choose the right deck and quiz type to see your result. Clicking on the score gives you a more detailed view.
- Create a new classroom of your choice (e.g. My Italian Class). Right down the code students need to join your class.
- Go back to home and create a new collection of your choice (e.g. Irregular Verbs)
- Add a deck of your choice to that collection (e.g. finire & andare)
- Add this collection to your classroom (My Italian Class)

- Go to verb search, enter verbs of your choice (e.g. finite & andare) and select values for tense, mood, etc. as you like (e.g. imperfetto, remote, futuro semplice, futuro anteriore, congiuntivo, etc.)
- Select all verbs and add them to your previously created deck “finire & andare” in the collection (Irregular Verbs)

Student Account

- log in, and explore the existing collections and decks
- Go to “Classroom” and join the Class you just created in the teacher account using the code.
- You may explore the deck and do some quizzes. (Note: if you log into your teacher account again, you will see the quiz results under Classroom >> Choose Class >> Learning Statistics.

3. Survey Questions

Demographics

In this section, we want to know a bit more about you and your experiences as a teacher.

How long have you been teaching language? (Multiple Choice Question)

- ☐ Less than 1 year
- ☐ More than 2 but less than 5 years
- ☐ More than 5 but less than 10 years
- ☐ More than 10 years
- ☐ Prefer not to say

General Usability

We want to understand how easy or difficult our application is to use.

Please answer this section from a teacher's perspective

(If you feel uncomfortable answering any of the required questions, just enter "NA" in the text answers or choose "Prefer not to say".)

Please comment on which functions or features of Card-it do you find most useful. (Free text)

Please comment on which functions or features of Card-it do you find least useful. Feel free to add any ideas or suggestions for improving it. (Free text)

On a scale of 1 to 5, how intuitive or easy to use do you find the app in general? (Multiple Choice Question)

- ☐ 1 - very difficult
- ☐ 2 - difficult
- ☐ 3 - neither difficult or easy
- ☐ 4 - easy
- ☐ 5 very easy

☐ Prefer not to say

Please use this space if you wish to elaborate on your rating above. (Free text)

Was there any feature you'd wish Card-it would have? (Free text)

Quiz mode

In this section, we want to know more about your expert opinion using Card-it's quizzes.

Please answer this section from a teacher's perspective.

(If you feel uncomfortable answering any of the required questions, just enter "NA" in the text answers or choose "Prefer not to say".)

On a scale of 1 to 5, how easy would you rate the tool's testing of the learning progress in the quiz mode? (Multiple Choice Question)

- ☐ 1 - very difficult
- ☐ 2 - difficult
- ☐ 3 - neither difficult or easy
- ☐ 4 - easy
- ☐ 5 very easy
- ☐ Prefer not to say

Please explain your rating in the previous question. (Free text)

On a scale of 1 to 5, how appropriate do you find the "conjugate" quiz mode? (Multiple Choice Question)

- ☐ 1 - very inappropriate
- ☐ 2 - inappropriate
- ☐ 3 - neither inappropriate or appropriate
- ☐ 4 - appropriate
- ☐ 5 very appropriate
- ☐ Prefer not to say

Please use this space if you wish to elaborate on our rating above. (Free text)

On a scale of 1 to 5, how appropriate do you find the "identify tense" quiz mode? (Multiple Choice Question)

- ☐ 1 - very inappropriate
- ☐ 2 - inappropriate
- ☐ 3 - neither inappropriate or appropriate
- ☐ 4 - appropriate
- ☐ 5 very appropriate
- ☐ Prefer not to say

Please use this space if you wish to elaborate on our rating above. (Free text)

In addition to the quiz modes presented, can you think of a different quiz type that we missed? Feel free to suggest any. (Free text)

Classroom Management

In this section, we want to know more about your expert opinion using Card-it's classroom management features.

Please answer this section from a teacher's perspective. (If you feel uncomfortable answering any of the required questions, just enter "NA" in the text answers or choose "Prefer not to say".

On a scale of 1 to 5, how would you rate the tool's usability with respect to classroom management? (Multiple Choice)

- ☐ 1 - very bad
- ☐ 2 - bad
- ☐ 3 - neither bad or good
- ☐ 4 - good
- ☐ 5 very good
- ☐ Prefer not to say

Please use this space if you wish to elaborate on our rating above. (Free text)

On a scale of 1 to 5, how easy did you find it to create a class? (Multiple Choice)

- ☐ 1 - very difficult
- ☐ 2 - difficult
- ☐ 3 - neither difficult or easy
- ☐ 4 - easy
- ☐ 5 very easy
- ☐ Prefer not to say

Please use this space if you wish to elaborate on our rating above. (Free text)

On a scale of 1 to 5, how easy did you find it to add decks and cards to a class? (Multiple Choice)

- ☐ 1 - very difficult
- ☐ 2 - difficult

- ☐ 3 - neither difficult or easy
- ☐ 4 - easy
- ☐ 5 very easy
- ☐ Prefer not to say

Please use this space if you wish to elaborate on our rating above. (Free text)

In addition to the class management features presented, can you think of different features that we missed? Feel free to suggest any. (Free text)

Adopting and integrating Card-it into your classroom

In this section, we want to know more about the potential integration of Card-it into real-live classroom activities.

Please answer this section from a teacher's perspective.

If you feel uncomfortable answering any of the required questions, just enter "NA" in the text answers or choose "Prefer not to say".

Imagine you created a classroom in the app for your students. Can you tell us about if and how you would see yourself using the app in a regular semester? (Free text)

Imagine you created a classroom in the app for your students. Are there functions that you see yourself using most? (Free text)

In case you're skeptical about using the app with your students, what would the app need to be useful to your teaching? (Free text)

Card-it usability for students/learners

In this section, we would like you to imagine yourself in the role of a language learner in your class, or an independent learner using Card-it on their own, and try to answer the following questions from that perspective.

Please answer this section from a learner's perspective.

If you feel uncomfortable answering any of the required questions, just enter "NA" in the text answers or choose "Prefer not to say".

What is your impression of the tool's usability when it comes to learning/acquiring Italian verb conjugation? (learner perspective) (Free text)

Which functions do you find most useful? (learner perspective) (Free text)

Do you miss anything? (learner perspective) (Free text)

Final remarks

In this section, you have the option to leave us any general comments, questions, or feedback regarding your experience using Card-it either from a teacher's or student's perspective.

Do you have any comments/ideas/suggestions in general? Please use this paragraph for any general comments or questions you'd like to leave us. (Free text)

Do you consent to be contacted for future studies related to this project using the email contact provided in this form? (Free text)

C

COLLECTION OF VERBS USED IN CARD-IT'S LEARNER STUDY

Table 1. List of all 72 quiz prompts in the knowledge quiz in $entry_{test}$ and $exit_{test}$. Each prompt consisted of a multiple choice question, presenting them with three answer options.

Prompt	**Verb**	**Answer Options**		
Futuro Semplice Indicativo (lui/lei/Lei)	trovare	ebbero trovato	troveresti	troverà
Presente Indicativo (tu)	dare	do	dessi	dai
Presente Condizionale (io)	dare	dai	darei	avranno dato
Futuro Semplice Indicativo (lui/lei/Lei)	mettere	metterà	metterebbe	metterei
Presente Condizionale (tu)	erudire	erudisti	erudiscano	erudiresti
Presente Indicativo (lui/lei/Lei)	mettere	mettesse	metto	mette
Presente Condizionale (io)	suggere	suggerei	suggiate	suggiamo

Futuro Semplice Indicativo (lui/lei/Lei)	erudire	erudente	erudirà	erudisco
Presente Condizionale (tu)	trovare	avranno trovato	trovasti	troveresti
Futuro Semplice Indicativo (tu)	suggere	suggevo	suggerai	suggerebbero
Presente Indicativo (tu)	suggere	suggi	suggessero	suggente
Presente Condizionale (lui/lei/Lei)	riputare	riputerò	riputaste	riputerebbe
Futuro Semplice Indicativo (lui/lei/Lei)	riputare	riputi	riputa	riputerà
Presente Indicativo (tu)	dovere	devi	avessero dovuto	dovemmo
Presente Indicativo (tu)	trovare	trovando	trovi	avessi trovato
Presente Indicativo (lui/lei/Lei)	dovere	deve	hanno dovuto	dobbiamo

Presente Condizionale (tu)	dare	daresti	dessero	daremo
Presente Condizionale (io)	dire	direi	dicesti	diciamo
Presente Condizionale (io)	riputare	riputeremmo	riputassimo	riputerei
Futuro Semplice Indicativo (tu)	dovere	dovevamo	dovrai	aveste dovuto
Presente Condizionale (io)	trovare	troverebbe	troverei	troverete
Futuro Semplice Indicativo (lui/lei/Lei)	dare	abbia dato	darà	abbiano dato
Futuro Semplice Indicativo (io)	suggere	sugga	suggevamo	suggerò
Presente Indicativo (lui/lei/Lei)	erudire	erudirono	erudisce	erudirete
Presente Indicativo (io)	dire	dico	diresti	abbia detto

Futuro Semplice Indicativo (tu)	dire	dirai	direbbero	dissero
Presente Condizionale (lui/lei/Lei)	erudire	erudissi	erudirebbe	erudisce
Presente Condizionale (lui/lei/Lei)	dare	darebbe	diedero	abbiate dato
Futuro Semplice Indicativo (tu)	erudire	erudisci	erudissimo	erudirai
Futuro Semplice Indicativo (io)	dare	davi	avrai dato	darò
Presente Condizionale (lui/lei/Lei)	suggere	sugga	suggerebbe	suggiamo
Presente Indicativo (io)	mettere	metto	mettano	metti
Futuro Semplice Indicativo (io)	trovare	troverebbero	troverò	trovare
Presente Condizionale (lui/lei/Lei)	dire	direbbe	avesti detto	dice

Presente Condizionale (lui/lei/Lei)	trovare	hai trovato	trovarono	troverebbe
Futuro Semplice Indicativo (io)	dovere	dovessero	dovrò	dovetti
Presente Indicativo (lui/lei/Lei)	trovare	avresti trovato	trova	abbiamo trovato
Presente Indicativo (io)	riputare	riputasti	riputerò	riputo
Presente Indicativo (io)	suggere	suggessero	suggo	suggessi
Futuro Semplice Indicativo (tu)	riputare	riputerai	riputaste	riputassero
Futuro Semplice Indicativo (lui/lei/Lei)	dire	dite	dicevano	dirà
Presente Indicativo (tu)	erudire	erudisci	erudivano	erudirebbe
Futuro Semplice Indicativo (tu)	dare	darai	avendo dato	dai

Presente Condizionale (io)	dovere	devi	avete dovuto	dovrei
Futuro Semplice Indicativo (tu)	trovare	troverai	trovi	avremo trovato
Presente Indicativo (lui/lei/Lei)	dire	dicessero	dice	direbbe
Presente Condizionale (tu)	mettere	mettereste	metteresti	mettemmo
Presente Indicativo (io)	dovere	devo	avranno dovuto	avete dovuto
Futuro Semplice Indicativo (io)	erudire	erudirò	erudisse	erudivi
Presente Condizionale (lui/lei/Lei)	mettere	metterebbe	metterete	metté
Presente Condizionale (tu)	dire	diciamo	dica	diresti
Presente Indicativo (tu)	dire	dici	diresti	dissi

Futuro Semplice Indicativo (io)	riputare	riputi	riputerò	riputassimo
Futuro Semplice Indicativo (io)	dire	dicevano	hanno detto	dirò
Futuro Semplice Indicativo (lui/lei/Lei)	suggere	suggerebbero	suggevamo	suggerà
Presente Indicativo (io)	trovare	trovò	aveste trovato	trovo
Presente Condizionale (tu)	suggere	suggeresti	suggemmo	suggi
Presente Condizionale (lui/lei/Lei)	dovere	hai dovuto	dovrete	dovrebbe
Presente Condizionale (tu)	dovere	dobbiamo	dovresti	dovrai
Presente Indicativo (lui/lei/Lei)	dare	dava	avreste dato	dà
Presente Indicativo (tu)	mettere	metti	hai metto	mettevano

Futuro Semplice Indicativo (lui/lei/Lei)	dovere	dovrà	avevo dovuto	deva
Presente Indicativo (tu)	riputare	riputi	hai riputato	riputeresti
Presente Indicativo (io)	erudire	erudisco	erudivi	erudirono
Presente Condizionale (io)	mettere	metterà	metterei	mettere
Futuro Semplice Indicativo (tu)	mettere	mettessi	metteva	metterai
Presente Condizionale (io)	erudire	erudirei	erudisco	erudiamo
Presente Indicativo (io)	dare	darò	diede	do
Presente Indicativo (lui/lei/Lei)	suggere	suggessero	suggevo	sugge
Presente Indicativo (lui/lei/Lei)	riputare	riputante	riputa	riputava

Presente Condizionale (tu)	riputare	riputeresti	riputino	riputerò
Futuro Semplice Indicativo (io)	mettere	metterò	mettevo	messo

D

SUPPLEMENTARY GRAPHS FROM CARD-IT'S SESSION 3 USABILITY SURVEY

Figure 1. Results for the question on recommending Card-it for advanced, intermediate, and beginner learners of Italian. Overall participants agreed that Card-it would be suitable for the different proficiency levels.

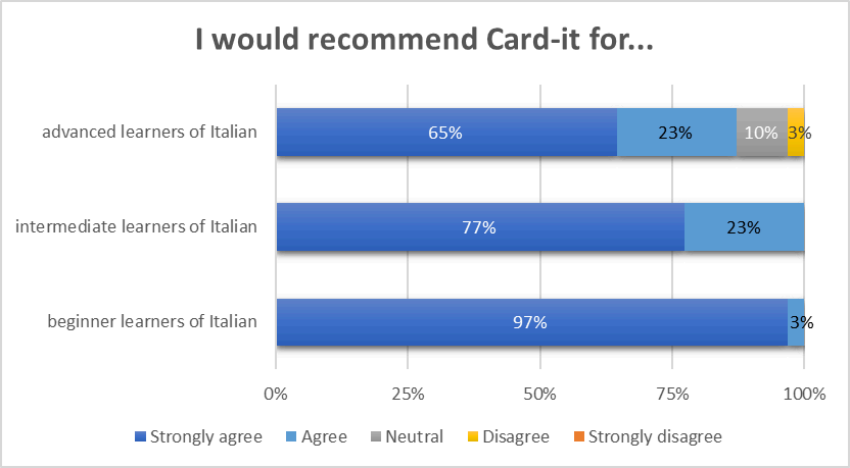


Figure 2. Results for the participants ratings regarding Card-it's flexibility for studying verb conjugation and its ability to enable them to plan, control, and evaluate their learning of Italian conjugation.

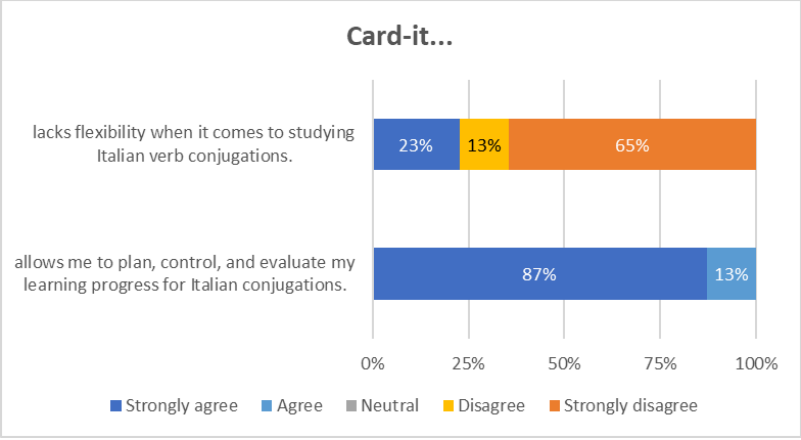
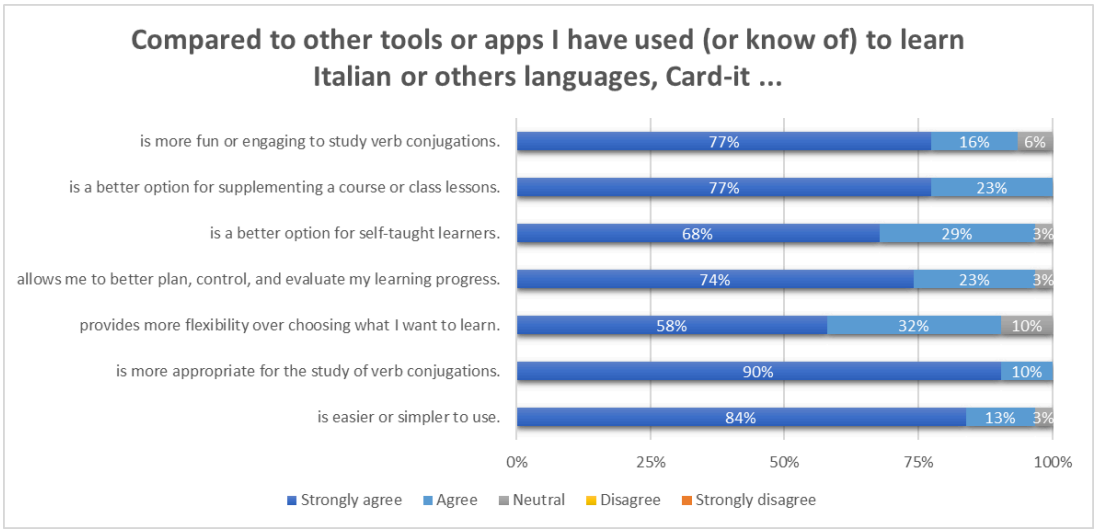


Figure 3. Results for the question where participants were asked to directly compare Card-it to other language learning tools or apps they have used before. Card-it was rated positively on all the different aspects according to their opinions. Particularly, Card-it was perceived as a more appropriate app for studying verb conjugations (90% “strongly agree” and 10% “agree”).



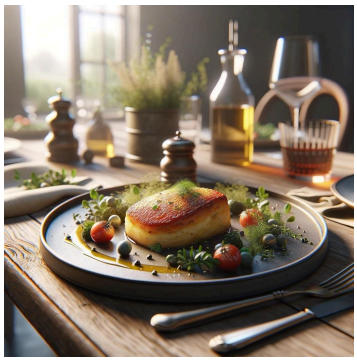
E

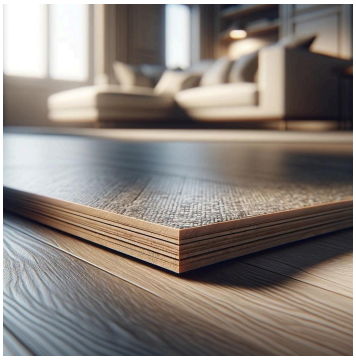
LIST OF LANGEYE'S AI GENERATED IMAGES

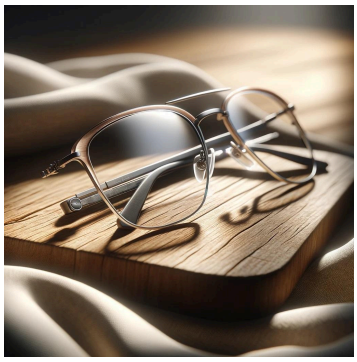
You can find the list of AI generated images for each respective memory label upload to LangEye's Memories for the control group participants.

Image	Item	Labels
	Wood	{Wood,"Natural material",Trunk,"Natural landscape"}
	Tableware	{Tableware,Dishware,Drinkware, "Kitchen utensil"}
	Glasses	{Glasses,"Vision care","Eye glass accessory",Eyewear}

	Drinkware	{Drinkware,Tableware,Water,Liquid}
	Dishware	{Dishware,Tableware,Drinkware,Serveware}
	Spoon	{spoon,Tableware,Cutlery,Wood,"Kitchen utensil"}
	Fork	{fork,Tableware,Dishware,Cutlery,Wood}

	Food	{Food,Tableware,Table,Dishware}
	Peripheral	{Peripheral,"Input device","Personal computer","Computer keyboard"}
	Cup	{Cup,Tableware,Drinkware,Dishware}
	Watch	{Watch,"Analog watch",Silver,Clock}

	Bag	{Bag,"Luggage and bags","Flash photography","Camera accessory"}
	Apple	{apple,Food,Plant,Fruit,Mcintosh}
	Serveware	{Serveware,Tableware,Drinkware,Dishware,Cup}
	Floor	{floor,Rectangle,Wood,Textile,Flooring}

	Eyewear	{Eyewear,Glasses,"Vision care",Furniture,Table}
	Slipper	{Slipper,Property,Furniture,Comfort,Wood}
	Paper	{Paper,"Automotive design",Lamp,"Tints and shades",Ceiling}
	Lemon	{Lemon,Food,Rangpur,Fruit,"Sweet lemon"}

	Stapler	{Stapler,"Output device",Wood,"Office equipment","Communication Device"}
	Speaker	{Speaker,"Camera lens","Camera accessory",Window,"Cameras & optics"}
	Scisors	{Scissors,Wood,"Office supplies","Office instrument",Eyewear}

BIBLIOGRAPHY

[1] OpenAI. (2023). ChatGPT (Mar 14 version) [Large language model]. https://chat.openai.com/chat.

[2] Ahmad Fayez Mutlaq Al-Zu'be. "The difference between the learner-centred approach and the teacher-centred approach in teaching English as a foreign language." In: *Educational research international* 2.2 (2013), pp. 24–31.

[3] Ruthi Aladjem and Bibiana Jou. "The linguistic landscape as a learning space for contextual language learning." In: *Journal of Learning Spaces* 5.2 (2016).

[4] Abdulhameed A Alhuwaydi. "A Review on vocabulary learning-designed MALL applications in the EFL context." In: *Theory and Practice in Language Studies* 12.10 (2022), pp. 2191–2200.

[5] Antonie Alm. "Language Learning with Netflix: extending out-of-class L2 viewing." In: *2021 International Conference on Advanced Learning Technologies (ICALT)*. IEEE. 2021, pp. 260–262.

[6] Tanja Angelovska. "Cross-linguistic awareness of adult L3 learners of English: A focus on metalinguistic reflections and proficiency." In: *Language awareness* 27.1-2 (2018), pp. 136–152.

[7] Lene Antonsen, Ryan Johnson, Trond Trosterud, and Heli Uibo. "Generating modular grammar exercises with finite-state transducers." In: *Proceedings of the second workshop on NLP for computer-assisted language learning at NODALIDA*. Vol. 17. 2013, pp. 27–38.

[8] Riku Arakawa, Hiromu Yakura, and Sosuke Kobayashi. "VocabEncounter: NMT-powered vocabulary learning by presenting computer-generated usages of foreign words into users' daily lives." In: *Proceedings of the 2022 CHI conference on human factors in computing systems*. 2022, pp. 1–21.

[9] Kathleen Bardovi-Harlig. "The relationship of form and meaning: A cross-sectional study of tense and aspect in the interlanguage of learners of English as a second language." In: *Applied Psycholinguistics* 13.3 (1992), pp. 253–278.

[10] Kenneth R. Beesley and Lauri Karttunen. *Finite State Morphology*. Stanford: CSLI Publications, 2003. ISBN: 9781575866840. URL: http://www.fsmbook.com.

[11] Jacopo de Bernardinis, Carlo Castagnari, and Giorgio Forcina. "ICALL-IT : an Intelligent Computer-Assisted Language Learning Platform for the Italian language ICALL-IT : an Intelligent Computer-Assisted Language Learning." In: March 2016 (2015). DOI: 10.13140/RG.2.1.4343.2083.

[12] Yves Bestgen, Sylviane Granger, and Jennifer Thewissen. "Error patterns and automatic L1 identification." In: *Approaching language transfer through text classification* (2012), pp. 127–153.

[13] Renaud Blanch, Rémy Dautriche, and Gilles Bisson. "Dendrogramix: A hybrid tree-matrix visualization technique to support interactive exploration of dendrograms." In: *2015 IEEE Pacific Visualization Symp. (PacificVis)*. IEEE. 2015, pp. 31–38.

[14] Daniel Blanchard, Joel Tetreault, Derrick Higgins, Aoife Cahill, and Martin Chodorow. "TOEFL11: A corpus of non-native English." In: *ETS Research Report Series* 2013.2 (2013).

[15] Françoise Blin. "CALL and the development of learner autonomy: an activity theoretical study." PhD thesis. The Open University, 2005.

[16] Cristina Bosco, Vincenzo Lombardo, Leonardo Lesmo, Daniela Vassallo, et al. "Building a Treebank for Italian: a Data-driven Annotation Schema." In: *Proceedings of SECOND INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION (LREC 2000)*. ELDA. 2000, pp. 99–105.

[17] Joan Bresnan. *Lexical-Functional Syntax*. Oxford, UK: Blackwell Publishers, 2001.

[18] Alisa Burova. "Industrial xReality (XR): Design, Application Scenarios, and Adoption Strategy in the Context of Industrial Maintenance." In: (2024).

[19] M.D. Bush and R.M. Terry. *Technology-enhanced Language Learning*. ACTFL foreign language education series. National Textbook Company. ISBN: 9780844293967.

[20] David R Byrd and Branden Lansing. "Electronic flashcards inside the classroom: Practical and effective." In: *Journal of Language Teaching and Learning* 6.2 (2016), pp. 1–13.

[21] Jason Byrne. "Anytime Autonomous English MALL App Engagement." In: *International Journal of Emerging Technologies in Learning (iJET)* 14.18 (2019), pp. 145–163.

[22] Arthur Caetano, Alyssa Lawson, Yimeng Liu, and Misha Sra. "ARLang: An outdoor augmented reality application for portuguese vocabulary learning." In: *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 2023, pp. 1224–1235.

[23] Alexander Campolo and Kate Crawford. "Enchanted determinism: Power without responsibility in artificial intelligence." In: *Engaging Science, Technology, and Society* (2020).

[24] Nathan T Carr, Kyle Crocco, Janet L Eyring, and Juan Carlos Gallego. "Perceived benefits of technology enhanced language learning in beginning language classes." In: *IALLT Journal of Language Learning Technologies* 41.1 (2011), pp. 1–32.

[25] Francine Chambers. "What do we mean by fluency?" In: *System* 25.4 (1997), pp. 535–544.

[26] Wai Meng Chan and Seo Won Chi. "Popular media as a motivational factor for foreign language learning: The example of the Korean wave." In: *Media in foreign language teaching and learning* (2011), pp. 151–188.

[27] George Chinnery. "You've got some GALL: Google-assisted language learning." In: *Language Learning & Technology* 12.1 (2008), pp. 3–11.

[28] Collins-Dictionary. *Autonomy*. In: *collinsdictionary.com dictionary*. URL: https://www.collinsdictionary.com/dictionary/english/autonomy (visited on 07/08/2023).

[29] Collins-Dictionary. *Independence*. In: *collinsdictionary.com dictionary*. URL: https://www.collinsdictionary.com/dictionary/english/independence (visited on 07/08/2023).

[30] Mark A Conroy. "Internet tools for language learning: University students taking control of their writing." In: *Australasian Journal of educational technology* 26.6 (2010).

[31] Piero Cosi, Rodolfo Delmonte, S Biscetti, Ron A Cole, Brian Pellom, and S van Vuren. "Italian literacy tutor-tools and technologies for individuals with cognitive disabilities." In: *InSTIL/ICALL Symposium 2004*. Citeseer. 2004.

[32] Peter Crosthwaite and Vít Baisa. "A user-friendly corpus tool for disciplinary data-driven learning: Introducing CorpusMate." In: *International Journal of Corpus Linguistics* (2024).

[33] David Crystal. *English as a global language*. Cambridge university press, 2012.

[34] Marlies De Zeeuw, Rob Schreuder, and Ludo Verhoeven. "Processing of Regular and Irregular Past-Tense Verb Forms in First and Second Language Reading Acquisition." In: *Language Learning* 63.4 (2013), pp. 740–765. ISSN: 00238333. DOI: 10.1111/lang.12023.

[35] Hélène Deacon and John Kirby. "Morphological awareness: Just "more phonological"? The roles of morphological and phonological awareness in reading development." In: *Applied Psycholinguistics* 25 (Apr. 2004), pp. 223–238. DOI: 10.1017/S0142716404001110.

[36] David Dearman and Khai Truong. "Evaluating the implicit acquisition of second language vocabulary using a live wallpaper." In: *Proceedings of the SIGCHI conference on human factors in computing systems*. 2012, pp. 1391–1400.

[37] Bruno Di Biase and Camilla Bettoni. "The development of Italian as a second language." In: *Grammatical development in second languages: Exploring the boundaries of Processability Theory* (2015), pp. 117–148.

[38] Alaa Ali Dinar et al. "Influence of the Semantic Aspects of Mother Tongue on Learning English as a Foreign Language." PhD thesis. Sudan University of Science and Technology, 2016.

[39] Tilman Dingler, Dominik Weber, Martin Pielot, Jennifer Cooper, Chung-Cheng Chang, and Niels Henze. "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*. 2017, pp. 1–12.

[40] Jihua Dong, Yanan Zhao, and Louisa Buckingham. "Charting the landscape of data-driven learning using a bibliometric analysis." In: *ReCALL* 35.3 (2023), pp. 339–355.

[41] Zoltán Dörnyei and Ema Ushioda. "Teaching and researching motivation." In: (2021).

[42] Ted Dunning. "Accurate methods for the statistics of surprise and coincidence." In: *Computational Linguistics* 19.1 (1993), pp. 61–74.

[43] Sarah Elaine Eaton. *Global Trends in Language Learning in the 21st Century*. ERIC, 2010.

[44] Darren Edge, Elly Searle, Kevin Chiu, Jing Zhao, and James A Landay. "MicroMandarin: mobile language learning in context." In: *Proceedings of the SIGCHI conference on human factors in computing systems*. 2011, pp. 3169–3178.

[45] Rod Ellis. *Understanding second language acquisition*. Vol. 31. Oxford university press, 1989.

[46] Rod Ellis. *The study of second language acquisition*. Oxford University, 1994.

[47] Felix Engelmann, Sonia Granlund, Joanna Kolak, Marta Szreder, Ben Ambridge, Julian Pine, Anna Theakston, and Elena Lieven. "How the input shapes the acquisition of verb morphology: Elicited production and computational modelling in two highly inflected languages." In: *Cognitive psychology* 110 (2019), pp. 30–69.

[48] Gülşen Eryiğit, Fatih Bektaş, Ubey Ali, and Bihter Dereli. "Gamification of complex morphology learning: the case of Turkish." In: *Computer Assisted Language Learning* 36.8 (2023), pp. 1421–1449.

[49] Nicolas F Fernandez, Gregory W Gundersen, Adeeb Rahman, Mark L Grimes, Klarisa Rikova, Peter Hornbeck, and Avi Ma'ayan. "Clustergrammer, a web-based heatmap visualization and analysis tool for high-dimensional biological data." In: *Scientific data* 4 (2017), p. 170151.

[50] Isabeau Fievez, Maribel Montero Perez, Frederik Cornillie, and Piet Desmet. "Promoting incidental vocabulary learning through watching a French Netflix series with glossed captions." In: *Computer Assisted Language Learning* 36.1-2 (2023), pp. 26–51.

[51] Open JS Foundation. *ReactJS*. Accessed in June 2024. URL: https://nodejs.org/en.

[52] Shaun Gallagher et al. "The narrative alternative to theory of mind." In: *Radical enactivism: Intentionality, phenomenology, and narrative* (2006), pp. 223–229.

[53] Nicola Galloway and Rachael Ruegg. "The provision of student support on English Medium Instruction programmes in Japan and China." In: *Journal of English for Academic Purposes* 45 (2020), p. 100846.

[54] Gerhard Gassler, Theo Hug, and Christian Glahn. "Integrated Micro Learning–An outline of the basic method and first results." In: *Interactive computer aided learning* 4 (2004), pp. 1–7.

[55] Malcolm Gladwell. *Outliers: The story of success*. Little, Brown, 2008.

[56] Google. *Google Text-To-Speech API*. Accessed in June 2024. URL: https://cloud.google.com/text-to-speech/.

[57] Google. *Google Vision API*. Accessed in June 2024. URL: https://cloud.google.com/vision.

[58] Google. *Google Vision API*. Accessed in June 2024. URL: https://cloud.google.com/translate/.

[59] Sylviane Granger, Estelle Dagneaux, Fanny Meunier, and Magali Paquot. *International corpus of learner English*. 2009.

[60] Ari Hautasaari, Takeo Hamada, Kuntaro Ishiyama, and Shogo Fukushima. "Vocabura: A method for supporting second language vocabulary learning while walking." In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3.4 (2019), pp. 1–23.

[61] Trude Heift. "Corrective feedback and learner uptake in CALL." In: *ReCALL* 16.2 (2004), pp. 416–431. ISSN: 14740109. DOI: 10.1017/S0958344004001120.

[62] Catherine Regina Heil, Jason S Wu, Joey J Lee, and Torben Schmidt. "A review of mobile language learning applications: trends, challenges, and opportunities." In: *The EuroCALL Review* 24.2 (2016), pp. 32–50.

[63] Shanna Li Ching Hollingworth and Wesley Willett. "FluencyAR: Augmented Reality Language Immersion." In: *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 2023, pp. 1–3.

[64] Robert Howland, Sachi Urano, and Junichi Hoshino. "SanjigenJiten: computer assisted language learning system within a 3d game environment." In: *International Conference on Advances in Computer Entertainment Technology*. Springer. 2012, pp. 262–273.

[65] Rick H Hoyle and Michelle R Sherrill. "Future orientation in the self-system: Possible selves, self-regulation, and behavior." In: *Journal of personality* 74.6 (2006), pp. 1673–1696.

[66] Chia Hsu, Yu Chen, Yu-Jung Liu, Yu-Cheng Chang, and Min-Jui Lee. "Spelland: Situated Language Learning with a Mixed-Reality Spelling Game through Everyday Objects." In: *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 2023, pp. 1–6.

[67] Pingying Hu and Jiaxiu Zhang. "A pathway to learner autonomy: a self-determination theory perspective." In: *Asia Pacific Education Review* 18 (2017), pp. 147–157.

[68] Wu-Yuin Hwang, Holly SL Chen, Rustam Shadiev, Ray Yueh-Min Huang, and Chia-Yu Chen. "Improving English as a foreign language writing in elementary schools using mobile devices in familiar situational contexts." In: *Computer assisted language learning* 27.5 (2014), pp. 359–378.

[69] George Iber. "English language learning in the Malaysian school setting: Where can we find 10,000 hours? A theoretical perspective." In: *Advances in Language and Literary Studies* 7.4 (2016), pp. 46–50.

[70] Carl James. *Errors in language learning and use: Exploring error analysis*. Routledge, 2013.

[71] Scott Jarvis and Magali Paquot. *Native language identification*. 2015.

[72] Sook-Hi Kang. "The effects of a context-embedded approach to second-language vocabulary learning." In: *System* 23.1 (1995), pp. 43–55.

[73] Regina Kaplan-Rakowski and Alice Gruber. "Low-Immersion versus High-Immersion Virtual Reality: Definitions, Classification, and Examples with a Foreign Language Focus." In: *12th International Conference Innovation in Language Learning*. 2019, p. 552.

[74] Laksnoria Karyuatry. "Grammarly as a Tool to Improve Students' Writing Quality: Free Online-Proofreader across the Boundaries." In: *JSSH (Jurnal Sains Sosial dan Humaniora)* 2.1 (2018), pp. 83–89.

[75] Hasan Kaya and Gülşen Eryiğit. "Using finite state transducers for helping foreign language learning." In: *Proceedings of the 2nd Workshop on Natural Language Processing Techniques for Educational Applications*. 2015, pp. 94–98.

[76] Daniel A Keim and Daniela Oelke. "Literature fingerprinting: A new method for visual literary analysis." In: *Proc. of the IEEE Symp. on Visual Analytics Science and Technology (VAST)*. 2007, pp. 115–122.

[77] Eileen Kintsch, Donna Caccamise, Marita Franzke, Nina Johnson, and Scott Dooley. "Summary Street®: Computer-guided summary writing." In: *Handbook of latent semantic analysis*. Psychology Press, 2007, pp. 275–290.

[78] Blanka Klimova. "Evaluating impact of mobile applications on EFL university learners' vocabulary learning–A review study." In: *Procedia Computer Science* 184 (2021), pp. 859–864.

[79] Ekaterina Kochmar. *Error detection in content word combinations*. Tech. rep. University of Cambridge, Computer Laboratory, 2016.

[80] Ekaterina Kochmar and Ekaterina Shutova. "Modelling semantic acquisition in second language learning." In: *Proc.of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*. 2017, pp. 293–302.

[81] K Kranthi. "Technology enhanced language learning (TELL)." In: *International Journal of Business and Management Invention* 6.2 (2017), pp. 30–33.

[82] Eric Krokos, Catherine Plaisant, and Amitabh Varshney. "Virtual memory palaces: immersion aids recall." In: *Virtual reality* 23.1 (2019), pp. 1–15.

[83] Raffaele Lampugnani. *The Rosetta Stone Language Library Italiano I*. 1998.

[84] Diane Larsen-Freeman. "Not so fast: A discussion of L2 morpheme processing and acquisition." In: *Language Learning* 60.1 (2010), pp. 221–230.

[85] Diane Larsen-Freeman and Michael H Long. *An introduction to second language acquisition research*. Routledge, 2014.

[86] Jean Lave. *Situated learning: Legitimate peripheral participation*. Cambridge university press, 1991.

[87] Inhye Lee. "Effects of Contact with Korean popular culture on KFL learners' motivation." In: *The Korean Language in America* 22.1 (2018), pp. 25–45.

[88] Joanne Leong, Pat Pataranutaporn, Valdemar Danry, Florian Perteneder, Yaoli Mao, and Pattie Maes. "Putting Things into Context: Generative AI-Enabled Context Personalization for Vocabulary Learning Improves Learning Motivation." In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. CHI '24. Honolulu, HI, USA: Association for Computing Machinery, 2024. ISBN: 9798400703300. DOI: 10.1145/3613904.3642393. URL: https://doi.org/10.1145/3613904.3642393.

[89] Kelvin Leong, Anna Sung, David Au, and Claire Blanchard. "A review of the trend of microlearning." In: *Journal of Work-Applied Management* 13.1 (2020), pp. 88–102.

[90] Fan Li, Si Fan, and Yanjun Wang. "Mobile-assisted language learning in Chinese higher education context: A systematic review from the perspective of the situated learning theory." In: *Education and Information Technologies* 27.7 (2022), pp. 9665–9688.

[91] David Little. "Language learner autonomy: rethinking language teaching." In: *Language Teaching* 55.1 (2022), pp. 64–73.

[92] William Littlewood. ""Autonomy": An anatomy and a framework." In: *System* 24.4 (1996), pp. 427–435.

[93] Hui-ju Liu. "Learner autonomy: The role of motivation in foreign language learning." In: *Journal of Language Teaching and Research* 6.6 (2015), p. 1165.

[94] Matthew Wycliffe Lucas and Androula Yiakoumetti. "Cross-linguistic awareness-raising practices can enhance written performance in EFL classes in Japanese universities." In: *International journal of bilingual education and bilingualism* 22.8 (2019), pp. 958–973.

[95] Amel Lusta, Özcan Demirel, and Behbood Mohammadzadeh. "Language corpus and data driven learning (DDL) in language classrooms: A systematic review." In: *Heliyon* (2023).

[96] E. Macaro. *Target Language, Collaborative Learning and Autonomy*. Modern languages in practice. Multilingual Matters, 1997. ISBN: 9781853593680. URL: https://books.google.ca/books?id=2bnmj2yqe5AC.

[97] Thomas Mayer, Christian Rohrdantz, Miriam Butt, Frans Plank, and Daniel Keim. "Visualizing vowel harmony." In: *Linguistic issues in language technology* 4.2 (2010), pp. 1–33.

[98] Garold L Murray. "Exploring learners' CALL experiences: A reflection on method." In: *Computer Assisted Language Learning* 12.3 (1999), pp. 179–195.

[99] William Nagy, Virginia Berninger, and Robert Abbott. "Contributions of Morphology Beyond Phonology to Literacy Outcomes of Upper Elementary and Middle-School Students." In: *Journal of Educational Psychology* 98.134-147 (2006).

[100] Carolina Nobre, Nils Gehlenborg, Hilary Coon, and Alexander Lex. "Lineage: Visualizing multivariate clinical data in genealogy graphs." In: *IEEE Trans. Vis. Comput. Graphics* 25.3 (2018), pp. 1543–1558.

[101] Carolina Nobre, Marc Streit, and Alexander Lex. "Juniper: A Tree+ Table Approach to Multivariate Graph Visualization." In: *IEEE Trans. Vis. Comput. Graphics* 25.1 (2019), pp. 544–554.

[102] Terence Odlin. *Language transfer: Cross-linguistic influence in language learning*. Cambridge University Press, 1989.

[103] Terence Odlin. "Crosslinguistic influence in second language acquisition." In: *The Encyclopedia of applied linguistics* (2013).

[104] Maurizio Oliva. *Italian for Everyone*. 2000.

[105] Allan Paivio, James M Clark, et al. "Dual coding theory and education." In: *Draft chapter presented at the conference on Pathways to Literacy Achievement for High Poverty Children at The University of Michigan School of Education*. Citeseer. 2006.

[106] Kevin Papin and Regina Kaplan-Rakowski. "An exploratory analysis of the impact of learners' first language on vocabulary recall using immersive technologies." In: *CALL for widening participation: Short papers from EUROCALL* (2020), pp. 266–271.

[107] Zhenhui Peng, Xingbo Wang, Qiushi Han, Junkai Zhu, Xiaojuan Ma, and Huamin Qu. "Storyfier: Exploring Vocabulary Learning Support with Text Generation Models." In: *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 2023, pp. 1–16.

[108] Elena Pizzuto and Maria Cristina Caselli. "The Acquisition of Italian Verb Morphology in a Cross-Linguistic Perspective." In: *Other Children, Other Languages. Issues in the theory of Language Acquisition*. Ed. by Yonata Levy. New York: Psychology Press, 1994, pp. 137–188. DOI: doi.org/10.4324/9780203773208.

[109] Eva Puimège, Maribel Montero Perez, and Elke Peters. "Promoting L2 acquisition of multiword units through textually enhanced audiovisual input: An eye-tracking study." In: *Second Language Research* 39.2 (2023), pp. 471–492.

[110] Fawaz Ali Ahmed Qasem. "Exploring Spelling Errors with Relation to the Phonological Syllable Structure in the Writings of Saudi ESL Learners." In: *International Journal of English Linguistics* 10.6 (2020), pp. 152–161.

[111] Fawaz Qasem. "Crosslinguistic influence of the first language: Interlingual errors in the writing of esl saudi learners." In: *Macrolinguistics* 8 (2020), pp. 105–20.

[112] Judith Rispens, Catherine McBride-Chang, and Pieter Reitsma. "Morphological awareness and early and advanced word recognition and spelling in Dutch." In: *Reading and Writing* 21 (Aug. 2008), pp. 587–607. DOI: 10.1007/s11145-007-9077-7.

[113] Christian Rohrdantz, Michael Hund, Thomas Mayer, Bernhard Wälchli, and Daniel A Keim. "The World's Languages Explorer: Visual Analysis of Language Features in Genealogical and Areal Contexts." In: *Computer Graphics Forum*. Vol. 31. 3pt1. Wiley Online Library. 2012, pp. 935–944.

[114] Richard M Ryan and Edward L Deci. "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being." In: *American psychologist* 55.1 (2000), p. 68.

[115] Richard M Ryan and Edward L Deci. "Self-regulation and the problem of human autonomy: Does psychology need choice,

self-determination, and will?" In: *Journal of personality* 74.6 (2006), pp. 1557–1586.

[116] Dominik Sacha, Yuki Asano, Christian Rohrdantz, Felix Hamborg, Daniel Keim, Bettina Braun, and Miriam Butt. "Self organizing maps for the visual analysis of pitch contours." In: *20th Nordic Conf.of Computational Linguistics*. 2015, pp. 181–189.

[117] Azizova Fotimakhon Saidbakhramovna, Rahmatova Nargiza Valijonvna, and Kurbanbayeva Dilnoza Sharofidinovna. "The method of educating vocabulary in a foreign language or target language." In: *Linguistics and Culture Review* 5.S1 (2021), pp. 1649–1658.

[118] Faramarz Samifanni. "The Fluency Way: A Functional Method for Oral Communication." In: *English Language Teaching* 13.3 (2020), pp. 100–114.

[119] Helmut Schmid, Marco Baroni, Erika Zanchetta, and Achim Stein. "Il sistema 'tree-tagger arricchito'–The enriched TreeTagger system." In: *IA Contributi Scientifici* 4.2 (2007), pp. 22–23.

[120] Francesco Scordato. *Italian Encounters*. 2001.

[121] Sauro Shannon and Carol Chapelle. *The handbook of technology and second language teaching and learning*. 2017.

[122] Mariana Shimabukuro, Jessica Zipf, Shawn Yama, and Christopher Collins. "Evaluating Classroom Potential for Card-it: Digital Flashcards for Studying and Learning Italian Morphology." In: *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA) (Short Paper + Demo)*. To appear. 2023.

[123] Frances Shiobara. "The 10,000 Hour Rule and What it Means for Language Teaching." In: ().

[124] Roumyana Slabakova. "Features or parameters: which one makes second language acquisition easier, and more interesting to study?" In: *Second Language Research* 25 (Apr. 2009), pp. 313–324. DOI: 10.1177/0267658308100291.

[125] Meta Open Source. *ReactJS*. Accessed in June 2024. URL: https://react.dev/.

[126] David J Steinhart. *Summary Street: An intelligent tutoring system for improving student writing through the use of Latent Semantic Analysis*. University of Colorado at Boulder, 2001.

[127] Marcus Taft. "Interactive-activation as a Framework for Understanding Morphological Processing." In: *Language and Cognitive Processes - LANG COGNITIVE PROCESS* 9 (Aug. 1994), pp. 271–294. DOI: 10.1080/01690969408402120.

[128] Fabio Tamburini, Matias Melandri, et al. "AnIta: a powerful morphological analyser for Italian." In: *LREC*. 2012, pp. 941–947.

[129] Nguyen Tran, Shogo Kajimura, and Yu Shibuya. "Location- and Physical-Activity-Based Application for Japanese Vocabulary Acquisition for Non-Japanese Speakers." In: *Multimodal Technologies and Interaction* 7.3 (2023), p. 29.

[130] Reut Tsarfaty, Djamé Seddah, Yoav Goldberg, Sandra Kübler, Yannick Versley, Marie Candito, Jennifer Foster, Ines Rehbein, and Lamia Tounsi. "Statistical parsing of morphologically rich languages (spmrl) what, how and whither." In: *Proceedings of the NAACL HLT 2010 First Workshop on Statistical Parsing of Morphologically-Rich Languages*. 2010, pp. 1–12.

[131] Stephen Tsourounis and C Demmans Epp. "Learning dashboards and gamification in MALL: Design guidelines in practice." In: *The international handbook of mobile-assisted language learning* (2016), pp. 370–398.

[132] Shotaro Ueno and Osamu Takeuchi. "Effective corpus use in second language learning: A meta-analytic approach." In: *Applied Corpus Linguistics* 3.3 (2023), p. 100076.

[133] Ema Ushioda. "The role of motivation." In: *Learner Autonomy* ().

[134] Ema Ushioda. "Motivation as a socially mediated process." In: *Learner autonomy in the foreign language classroom: Teacher, learner, curriculum and assessment* (2003), pp. 90–102.

[135] Ema Ushioda. "Why autonomy? Insights from motivation theory and research." In: *Innovation in language learning and teaching* 5.2 (2011), pp. 221–232.

[136] Aaditya Vaze, Alexis Morris, and Ian Clarke. "CuriosityXR: Context-aware Education Experiences with Mixed Reality and Conversation AI." In: *2024 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR)*. IEEE. 2024, pp. 41–49.

[137] Christian David Vazquez, Afika Ayanda Nyati, Alexander Luh, Megan Fu, Takako Aikawa, and Pattie Maes. "Serendipitous language learning in mixed reality." In: *Proceedings of the 2017 CHI conference extended abstracts on human factors in computing systems*. 2017, pp. 2172–2179.

[138] Christian Vázquez, Lei Xia, Takako Aikawa, and Pattie 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.

[139] Mirvan Xhemaili. "Literature review on the role of mother tongue in learning and teaching English for specific purpose." In: *Journal of Education and Practice* 4.18 (2013), pp. 39–43.

[140] Choongil Yoon. "Concordancers and dictionaries as problem-solving tools for ESL academic writing." In: *Language Learning & Technology* 20.1 (2016), pp. 209–229.

[141] Aiqing Yu and Guy Trainin. "A meta-analysis examining technology-assisted L2 vocabulary learning." In: *ReCALL* 34.2 (2022), pp. 235–252.

[142] H Gülru Yüksel, H Güldem Mercanoğlu, and M Betül Yılmaz. "Digital flashcards vs. wordlists for learning technical vocabulary." In: *Computer Assisted Language Learning* 35.8 (2020), pp. 2001–2017.

[143] Ameena Zafar. "Error analysis: A tool to improve English skills of undergraduate students." In: *Procedia-Social and Behavioral Sciences* 217 (2016), pp. 697–705.

[144] Nurkhamimi Zainuddin. "Technology Enhanced Language Learning Research Trends and Practices: A Systematic Review (2020-2022)." In: *Electronic Journal of e-Learning* 21.2 (2023), pp. 69–79.

[145] Eros Zanchetta and Marco Baroni. "Morph-it! a free corpus-based morphological resource for the italian language." In: *Corpus Linguistics 2005* 1.1 (2005).

[146] Eros Zanchetta, Marco Baroni, et al. "Morph-it! A free corpus-based morphological resource for the Italian language." In: *Proceedings of corpus linguistics conference series 2005 (ISSN 1747-9398)*. Vol. 1. University of Birmingham. 2005, pp. 1–12.

[147] Yalun Zhou and Michael Wei. "Strategies in technology-enhanced language learning." In: *Studies in Second Language Learning and Teaching* 8.2 (2018), pp. 471–495.

[148] Jessica Zipf, Mariana Shimabukuro, and Christopher Collins. "Card-IT: a Dynamic FSM-based Flashcard Generator for Learning Italian Verb Morphology." Talk in European Association of Computer Assisted Language Learning - EUROCALL 2021, online. 2021. URL: https://whova.com/web/euroc1_202108/Agenda.

[149] LanguageTooler GmbH. *LanguageTool API - Spell and Grammar Checker*. Accessed in June 2019. URL: https://github.com/languagetool-org/languagetool.

[150] Microsoft. *Read My World: Learn vocabulary from the world around you with new Microsoft Garage project*. Accessed in June 2022. URL: https://www.microsoft.com/en-us/garage/profiles/read-my-world/.

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

https://bitbucket.org/amiede/classicthesis/

Final Version as of December 11, 2024 (E-Scholar).