From Screens to Words: Understanding the Effects of Multimodality in Digital Second Language Vocabulary Learning

Paula Janjić

100341326

School of Psychology, University of East Anglia

January 2025

A thesis submitted for the degree of

Doctor of Philosophy

Under the supervision of

Prof. Kenny R. Coventry and prof. John P. Spencer

Author's Declaration

This PhD was funded through a studentship provided by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Actions grant agreement No 857897. All the studies were approved by University of East Anglia School of Psychology Ethics Committee. The work from this PhD was presented at international and interdisciplinary conferences, published and submitted for publication to journals listed below. The studies submitted for publication are presented in the thesis in manuscript format.

Conferences:

Janjić, P., Coventry, K. R. (October, 2024). From screens to words: understanding the role of modality in digital second language vocabulary learning [Paper presentation]. *International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA)*. Zagreb, Croatia.

Janjić, P., Coventry, K. R. (October, 2023) Exploring the use of digital language learning resources in the UK: Insights from second language teachers [Paper presentation]. *Advances in the Study of Language Development and Literacy Learning in the Digital Age*. Porto, Portugal.

Janjić, P., Spencer, J.P., Coventry, K. R. (July, 2022). Second language vocabulary learning: When do pictures speak louder than words? [Poster presentation]. *44th Annual Meeting of the Cognitive Science Society.* Toronto, Canada (hybrid).

Janjić, P., Coventry, K. R. (October, 2021). The influence of learning methods and word properties on second language vocabulary learning outcomes. [Talk] *2nd International Workshop on Psycholinguistics, Neurolinguistics and Clinical Linguistics.* Novi Sad, Serbia (online).

Janjić, P., Coventry, K. R. (September, 2021). Do pictures speak louder than words when learning a second language…or does it depend on word class? [Talk] *Understanding Language and Literacy Development in the Digital Age*. Sevilla, Spain.

Manuscripts published:

Janjić, P., Arikan, G., Gudde, H. B., Murphy, J. J. C., Sivaram, L., & Coventry, K. R. (2024). The role of pointing gestures and eye gaze in second language vocabulary learning. *Discourse Processes*, 1–19. https://doi.org/10.1080/0163853X.2024.2343625

Janjić, P., & Coventry, K. R. (2024). From screens to words: Understanding the role of modality in digital second language vocabulary learning. In D. G. Sampson, D. Ifenthaler, & P. Isaías (Eds.), *Proceedings of the Conference on e-Learning and Digital Applications*. IADIS Press.

Janjić, P. & Coventry, K. R. (under review). Digital language learning resources: Analysis of software features and usage patterns in UK schools. *Frontiers in Education*, 10, Article 1532802. https://doi.org/10.3389/feduc.2025.1532802

Manuscripts submitted for publication:

Janjić, P., Spencer, J.P. & Coventry, K.R. (under review). Second language vocabulary learning: The interplay between learning modality, testing modality and word class. *Cognition*.

Acknowledgements

Throughout these years, I often felt deep gratitude for this opportunity, but my greatest appreciation has always been reserved for the people who shaped it.

First, I would like to thank my supervisor, prof. Kenny Coventry. His continuous support, expertise and guidance made every challenge solvable and every deadline achievable. Above all, his genuine care and understanding were invaluable in creating a supportive environment. I am also grateful to my secondary supervisor prof. John Spencer, for his valuable insights and fresh perspective. I extend my appreciation to student interns Laksha, Joe, Lucy and Piotr for their interest and involvement.

In Kenny's lab, I was fortunate to be surrounded by most encouraging people from the very beginning – Gozdem and Harmen. Through video calls during covid, Gozdem's support reached me even before I reached Norwich, and continued to follow me all the way through. And as if that weren't enough, she introduced me to the brilliant Agatha, Alice, Greta and Laura - most amazing 'support bubble' I could have wished for. From everyday lunches, chats and jokes to shared interests and meaningful conversations, our time together quickly formed into lasting friendships. Same goes for Harmen - having someone interested in debugging R code and celebrating little triumphs along the way was priceless. Your constructive approach and calm attitude helped me many times and on many levels.

The whole time, I was surrounded by an incredibly supportive community of PhD students and researchers at UEA. Walks and coffees with Pavla or Bea always managed to lift me up, just as did book clubs, journal clubs, bouldering and gardening. Same goes for each of our chats over coffee breaks where we shared our challenges and wins along the way.

This support was extended beyond UEA through the e-LADDA international training network. Meeting with these bright, funny and friendly people, both online and in person, was always a source of inspiration and motivation - I truly enjoyed working with you. This also goes to all the supervisors within the project who patiently guided us.

I would also like to acknowledge my early mentors - prof. Marijan Palmović, prof. Miroslav Vrankić, prof. Sanja Simleša and prof. Adrià Rofes. By sharing your knowledge and enthusiasm, leading exciting projects, and offering support and encouragement, you provided the foundation for this rewarding path.

I couldn’t imagine navigating the way through academia through all these years and countries without Kristina Burum. From studies at ERF, Erasmus internships and all the way to MSCA and our PhDs – we’ve taken all the steps together, filling our inbox with memes and voice messages along the way. Although we didn’t end up in the same place - you knew it all and understood some parts better than anyone.

One never-ending source of support have always been my family and friends. With our neighbours in Norwich we shared so much more than the address. With Shams in particular - it was priceless to have a friend like you to share the experiences of both motherhood and academia. To all my friends in Croatia, I am thankful for all the love I received through your video calls, messages, visits and updates. You made it possible to feel present in two places at the same time.

To my parents, grandparents and sister – thank you for always being my safe place and a source of unreserved support and care. Knowing that I can always rely on you encouraged me to embrace new opportunities and face new challenges.

To my children – you may not fully remember these years, but each of you added extra magic and joy to this journey. Watching you grow along the way made every step more memorable and more meaningful. Your hugs, smiles and silly faces would always make all the worries seem small. You were, without a doubt, the best well-being support that I could have asked for.

And finally, to Matija. Although this thesis is about vocabulary, I still can’t find the right words to describe your role in everything. You shared every trouble and celebrated every achievement as if they were your own. You made it possible, you made it fun, you cheered me along the way. This accomplishment is as much yours as it is mine.

There are countless other wonderful people whose kindness and support helped me during these years. Before starting the PhD, I was warned that it could feel lonely and isolating at times. However, thanks to each and every one of you, this journey will always feel like a most wonderful collaboration.

Abstract

Learning a second language (L2) is a multimodal process encompassing linguistic, social, and sensory-motor experiences. While technological advancements enable multimodal input in learning settings, its effectiveness in L2 vocabulary learning varies. This thesis explores the complex dynamics of second language vocabulary learning by examining distributional and embodied accounts of language processing, and investigating social aspects of digital language learning. The first set of experiments extends research on L2 learning modality by examining previously underexplored variables of congruency between learning and testing modalities, word class, and translational ambiguity. Results reveal that different word classes benefit from different modalities, with implications for distributional, embodied, and hybrid accounts of language processing. The second series of experiments investigates how social cues (eye-gaze and deictic gestures), critical in first language acquisition, affect L2 vocabulary learning in digital contexts. Through manipulating distractor items and agent-target distance, results suggest that the effectiveness of these social cues does not completely translate to screen-based L2 vocabulary learning environments. The final study shifts to practical application, developing a framework for evaluating language learning software based on research-driven practices and theoretical foundations. This framework was implemented with UK language teachers to assess learning software and investigate classroom technology implementation. The thesis concludes by integrating these findings to address theoretical and practical implications of multimodality in second language vocabulary learning.

Table of Contents

List of Figures

List of Tables

From Screens to Words:

Understanding the Effects of Multimodality in Digital Second Language Vocabulary Learning

"We shape our tools and thereafter they shape us.

These extensions of our senses begin to interact with our senses."

Culkin, 1967

on the nature of educational media

Chapter 1 | General Introduction

Communication is inherently multimodal, extending far beyond spoken words to encompass gestures, facial expressions, body language and spatial relations (Goldin-Meadow, 2014; Kendon, 2004; Levinson & Holler, 2014; McNeill, 2012). This integrated interplay of verbal and non-verbal elements shapes how we learn and process language, from early development through adulthood (Kendon, 2014; Levinson & Holler, 2014; Vigliocco et al., 2014). Although multimodal approaches are now widely recognised, language research has historically been dominated by approaches focusing on language as an abstract symbolic system independent of other cognitive and sensorimotor processes (Chomsky, 1965; Fodor, 1983). A significant shift occurred within the last few decades when language began to be seen as fundamentally embodied, multimodal and social, emphasizing the integration of multiple sensory, motor and cognitive systems (Barsalou, 1999; Barsalou, 2010; Lakoff & Johnson, 1980; Meteyard et al., 2012; Paivio, 1971; Zwaan & Kaschak, 2006).

While these notions are well-accepted in studies of language processing and first language acquisition, their influence on research involving second language learning is more recent and therefore still evolving (Atkinson, 2010; Douglas Fir, 2016; Li & Jeong, 2020; Mitchell et al., 2019). In practice, this leads to foreign language vocabulary learning still frequently relying on out-of-context associations between L1 and L2 words (Li & Jeong, 2020), contrasting with the natural language learning process which is situated, embodied and multimodal – a disconnect this thesis aims to address.

Confronting the task of word learning in a classroom context or as an adult (Hartshorne et al., 2018) becomes even more challenging when coupled with conventional, frequently decontextualized learning methods. This can quickly start to cause frustration and stagnation in the learning process. Despite the increased emphasis across governments on recognising the importance of language learning (American Academy of Arts & Sciences et al.,, 2020; European Commission, 2018), as well as technological advances and digitalization of teaching, language learning outcomes still fall short of expectations (Brecht, 2015; Collen, 2020; Costa & Albergaria-Almeida, 2015). These persistent issues highlight the importance of finding strategies to master L2 vocabulary – methods that align with how we naturally process and retain information. With this in mind, the main focus of this thesis is to comprehensively investigate the multimodal context of L2 vocabulary learning, one of the foundations of second language learning.

The need to develop a deeper understanding of multimodal learning is particularly urgent as new technologies continue to rapidly evolve. Just as Culkin (1967.) recognized the reciprocal relationship between the design of educational tools and learning almost sixty years ago, this interaction between technology and our sensory processing remains central today, particularly in the area of second language vocabulary learning. Currently, research findings in this domain are frequently inconsistent and fragmented, failing to provide a comprehensive answer on how to best support initial vocabulary learning (Boers et al., 2017; Li & Jeong, 2020). This further weakens the already fragile link between scientific research and technological application. Technological innovations have developed rapidly and in isolation from developments in second language teaching and learning theories and practices (Li & Lan, 2022), resulting in a variety of software solutions of questionable quality (Kolak et al., 2021; Shing & Yuan, 2016). Furthermore, the lack of insight into the practical application of both the research findings and technological solutions within the classroom setting also creates a disconnect between research and evidence-based practice. Therefore, the second broader goal of this work is to address the need for convergence across theory, applications and teaching by investigating different perspectives (Figure 1).

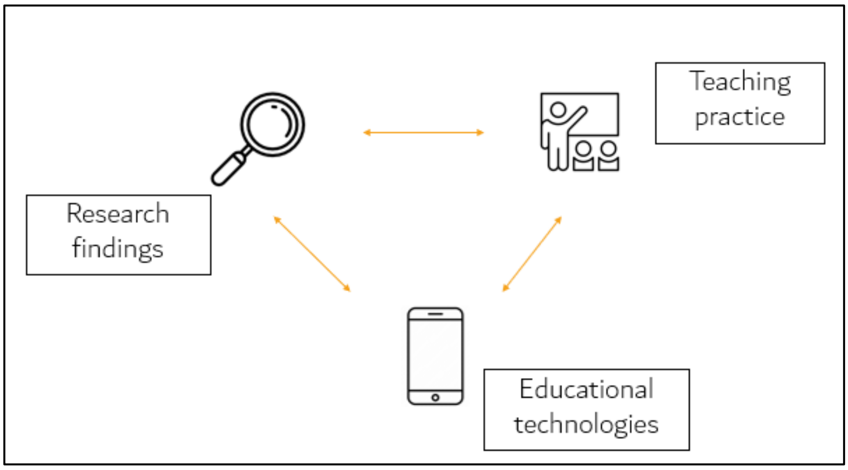

Figure 1. Figure illustrating the 'ideal' relationship between research findings, teaching practice and educational technologies Each component influences and is influenced by the others, creating an iterative process that forms a foundation for comprehensive answers to research questions while simultaneously enhancing educational practice and the effectiveness of technological tools

This aligns with ongoing scientific efforts to integrate cognitive science with educational practices and technology (Douglas Fir Group, 2016; Li & Lan, 2022). Currently, there is a missed opportunity to create an iterative process where research would inform practice, and practical experience would in turn inform further research, as well as technological design. As Li and Lan (2022) note in their review of second language learning in multimedia contexts – "*theories and practices for the learning and teaching of second languages have lagged behind the pace of (...) technological innovations*" (p.361). Therefore, the changing landscapes of second language learning call for collaborative and transdisciplinary efforts to enable better understanding of language learning processes, as well as real-life application of obtained knowledge.

Due to the complexity of second language vocabulary learning process, there are many factors that require thorough investigation. This work will narrow in on the initial vocabulary learning, as a foundation for further development of L2 (Nation, 2001). More specifically, the focus will be on multimodality, an aspect of second language vocabulary learning that is closely connected to the emerging technologies, as well as the teaching practice. While emphasizing cognitive perspectives, a combination of experimental research, software evaluation and real-world application will be employed to enable a comprehensive investigation of multimodal approaches in language learning.

The remainder of this chapter establishes a theoretical foundation and rationale for the subsequent studies. It begins by examining language through a multimodal lens, followed by an exploration of second language learning theories and their development in the context of multimodality. Finally, the chapter focuses specifically on L2 vocabulary learning, reviewing the current gaps in the field and introducing the motivation for the following work.

1.1. The multimodal nature of language

Over the past century, the understanding of cognition has undergone a radical transformation. It moved away from focusing on purely internal mental processes towards a more comprehensive framework that acknowledges the interconnected nature of mind, body and environment (Barsalou et al., 2008; Glenberg et al., 2013; Meteyard et al., 2012).

This paradigm shift was reflected across different disciplines, including psychology and linguistics. In classic studies, Paivio's (1971) Dual Coding Theory demonstrated that both verbal and non-verbal codes are included in processing of concrete words, paving the way for further studies on the embodied and multimodal nature of language. As Ellis (2019) points out, since our experiences are multimodal, contextualized, dynamic and situated, so are our memory representations, which reflect those experiences (Barsalou 2008; Varella et al., 1991). In this view, the quality and depth of cognitive representations are inherently tied to the richness of multimodal experiences (Barsalou, 2008; Ellis, 2019; Paivio, 1990).

Within this evolving landscape, embodiment approaches view cognitive processes, including language, as not abstract and disembodied, but grounded in the body's sensory and motor systems, as well as interactions with the physical and social environment (Shapiro, 2014; Wilson, 2002). However, the understanding of embodiment has changed from early notions of 'strong' embodiment that focuses on modality-specific, lower-level experiential representations, to 'weak' or 'graded' conception. This contemporary view sees embodiment as inherently multimodal and layered (Dove, 2022), allowing for a combination of embodied and abstract representations within cognitive processes. Following this, the situated nature of language enables the use of both linguistic and multimodal resources to accomplish social actions and convey meaning effectively (Murgiano et al., 2021; Hawkins, 2019).

However, with a growing amount of research on multimodality and embodiment, different interpretations emerged over time and across different fields. In multimedia learning and human-computer interaction research, multimodality is often viewed in a more restricted sense, focusing on integration of audio-visual information (see Montero Perez, (2020) for an overview). On the other hand, a broader definition of multimodality in language acknowledges the importance of the non-arbitrary dimension of language that consists of different sensory-motor experiences that also include the social dimension as a crucial part of multimodality of language (Murgiano et al., 2021; Perniss, 2018). This view is also adopted within this thesis which investigates the roles of images, text, deictic cues and the interpersonal space, although adjusted to the context of screen-based learning. This aligns with the new insights into both language learning and language processing that saw language processes as fundamentally shaped by the interplay between sensory-motor systems and the environment (Atkinson, 2010; Ellis, 2019).

The views on language as multimodal and embodied were supported by both neuroimaging and behavioural studies. For example, an fMRI study by Hauk et al. (2004) showed how reading action-related words (“lick”, “pick”, “kick”) activated somatotopically corresponding motor areas related to the tongue, hand and foot, suggesting that meaning is directly grounded in bodily experience rather than existing as pure abstract symbols (Barsalou, 2008; Fischer & Zwaan, 2008; Pulvermüller, 2013). In a similar vein, Rueschemeyer et al. (2010) found greater motor system activity when reading words referring to easily manipulable objects (e.g. “spoon”, “hammer”) than for less manipulable objects (e.g. “building”). In line with these findings, behavioural studies showed that motor action, as well as sensory input can interfere with or facilitate sentence processing (and vice versa), depending on whether there is a match or mismatch between the actions and the sentence content. These effects have been confirmed across different studies both within the motor (Kaschak & Borregine, 2008; Meteyard et al., 2007; Zwaan & Taylor, 2006), and the perceptual domain (Meteyard et al., 2008; Zwaan & Kaschak, 2008; Zwaan et al., 2002), demonstrating how language is embodied and rooted in our experience of the world.

This is additionally supported by findings from language development research, showing that children heavily rely both on embodied experiences (Smith & Gasser, 2005) and on social cues from their environment (Tomasello, 2003; Yu & Ballard, 2007). The acquisition of first language is inherently multimodal, with children learning through a rich combination of sensory, motor and social experiences (Yu & Smith, 2012). From a very young age, children demonstrate the ability to track patterns in speech signals (Saffran et al., 1996), while simultaneously integrating sensory-motor information from their environment (Yu & Ballard, 2007). This is further enhanced by developing joint attention abilities and starting to utilize social cues such as gaze and pointing gestures in language learning (Baldwin & Moses, 2001; Tomasello, 2003). This synchronized emergence of speech and gesture in infancy additionally highlights the link between action and language (Iverson & Goldin-Meadow, 2005).

Taken together, these findings speak in favour of the interactive, embodied and multimodal nature of language, which leads to significant implications for both theory and practice across different disciplines. The growing evidence challenged existing approaches and methodologies and introduced novel perspectives, especially within cognitive and learning sciences research. However, as Perniss (2018) argues, progress in the field of language research relies on redefining the concept of language to adequately capture its nature as a complex system that operates across multiple modalities. One of the fields that also needed restructuring in line with the emerging findings was the field of second language learning.

1.2. Second language learning through the lens of multimodality

Since their emergence in cognitive research, multimodal perspectives have become well-established in the cognitive sciences. However, their influence on second language research is only beginning to take shape (Atkinson, 2010; Ellis, 2019; Li & Jeong, 2020). The gradual expansion beyond purely cognitive and structural views reflects an emerging recognition that second language learning is not solely a product of internal cognitive mechanisms. In the following section a range of theoretical perspectives are explored to help indicate how multimodality has become central to understanding L2 learning, laying the groundwork for investigating its role in digital L2 learning.

1.2.1. The emergence of multimodality in theories of L2 learning

1.2.1.1. Foundational and early cognitive approaches to L2 learning

The field of second language acquisition research appeared as a distinct field of enquiry in the late 1960s (Myles, 2010). Still, being a highly interdisciplinary field, its progress largely followed the developments in the fields of linguistics, cognitive psychology and educational sciences, with findings from language processing theories significantly influencing the field of second language learning. At that time, the field was dominated by behaviorist theories, where language learning was conceptualized primarily in terms of habit formation through repeated exposure, memorisation and reinforcement (Lado, 1964; Skinner, 1957). Evidence for these early behavioural models was largely drawn from animal learning paradigms and observations of rote memorization, though their applicability to the complexity of human language acquisition was later challenged (see Mitchell et al., 2019 for a historical overview). From a modern perspective, these models are seen as overly simplistic, lacking the cognitive depth to explain creative language use and the internal processes of learners. For instance, the reliance on simple stimulus-response associations would likely not stand up to current experimental designs that probe more advanced cognitive mechanisms.

With the advent of Universal Grammar (Chomsky, 1959; Corder, 1967), some researchers argued that L2 acquisition tapped into the same innate and specialized language faculty that guided L1 acquisition (Flynn, 1983). Although these perspectives broadened the view of how learners internalize language structures, they often maintained a focus on mental representations and rule-based processing, leaving the embodied, social and environmental dimensions underexplored. These theories struggled to account for the significant role of individual differences, input quality and interaction, aspects that newer theories increasingly emphasize (Ellis, 2008).

In parallel, cognitive and information-processing models began to highlight how learners build language proficiency by transforming explicit and declarative knowledge into more implicit, procedural knowledge through repeated practice and attention to linguistic patterns (Bialystok, 1978; Krashen, 1981; McLaughlin, 1987). These approaches drew on psychology and neurolinguistics to explain how learners store, retrieve and automate L2 skills. While pivotal in illuminating the psychological foundations of acquisition, and often supported by early reaction times studies or analyses of learning curves (as reviewed in DeKeyser, 2007), purely cognitive models still often failed to capture all the aspects of real-world language use which involves social interactions, contextual factors and multimodal cues. From the current perspectives, they have limited ecological validity, which more recent embodied and sociocultural theories aim to address (Li & Jeong, 2020).

1.2.1.2. The shift towards current multimodal perspectives of L2 learning

A major shift in second language research emerged with interactionist and sociocultural approaches that initially applied Vygotsky's sociocultural framework (1978) to L2 learning. In line with this, such apporaches stressed the importance of social context in shaping how languages are learned and used (Atkinson, 2002; Lantolf, 1995; Swain, 2005). From this perspective, language is seen as a tool in which social cues become mediators of learning, shifting the focus to the interactive and dynamic nature of language.

A further step was taken by more recent connectionist, emergentist and usage-based theories (Ellis, 2019; MacWhinney, 2012; Rebuschat & Williams, 2012) that aim to explain how linguistic knowledge emerges from repeated, and importantly, situated, encounters. Learners are thought to pick up on distributional and frequency patterns in the input, but crucially, such input is seen as context-bound and multimodal (Ellis, 2008). These theories find strong support in computational modeling and corpus linguistics (Ellis, 2019). While these approaches do well at explaining pattern acquisition, they are still developing ways to fully incorporate the nuances of social interaction and individual learner agency.

Research shows that visual, auditory and social cues, as well as actions and emotional states play a key role in how learners notice, process and retain linguistic information (Douglas Fir, 2016). This embodied perspective is also seen in the work of Li and Jeong (2020), who propose the Social L2 learning (SL2) Hypothesis. They argue that engaging learners' perceptual, sensorimotor and emotional capacities within social interactions improves language learning outcomes by reinforcing more direct mapping of new language forms to their referents. This active and embodied social participation is what they posit enables a more direct, L1-like grounding of L2 words to their meanings.

Central to the more recent theories is the increasing emphasis on second language learning as a multimodal, embodied and socially situated process (Atkinson, 2010; Douglas Fir, 2016; Ellis, 2019), that is also dynamic and shaped by the context. This represents the current theoretical framing for the work reported in this thesis. These properties become additionally relevant in the era of digitalisation where different modalities converge, thus bringing naturalistic, L1-like language learning contexts closer to the L2 learners in the classroom environment, as will be discussed in the chapter 4. As the digital contexts rapidly change, it is crucial to investigate what effect multimodal learning environments have on the nature of second language processing and learning. Thus, as The Douglas Fir Group (2016) notes, one of the aims within the field of second language learning research is to characterize both linguistic and non-linguistic forces that influence the process and the outcomes of second language learning.

1.2.2. The multimodal link between L1 and L2 acquisition

As Ellis (2019) notes, while first language acquisition is relatively well-researched, the field of second language learning has received less attention. Although both processes share similarities, there are also marked differences. The degree to which these processes align continues to spark discussion.

One line of research tends to emphasize the fundamental differences between first and second language acquisition, underlining that adults learning L2 rely on distinct cognitive mechanisms from those utilized by infants and children acquiring their native language (Bley-Vroman, 2009; DeKeyser, 2000). At the same time, other perspectives favour a fundamental continuity hypothesis, which suggests that mechanisms for language learning operate continuously across the lifespan, with differences stemming from different contexts and conditions in which learning unfolds (Ellis, 2019; Krashen, 1981).

Proponents of fundamental differences between L1 and L2 learning suggest that they result from different mechanisms (Bley-Vroman, 2009). They support the 'Critical Period Hypothesis' (Lenneberg, 1967), suggesting a biologically determined effect of age on language learning. On the other hand, continuity-based perspectives argue that many of the same underlying mechanisms that guide L1 acquisition, such as statistical learning, sensitivity to input frequency and interactive and meaningful usage also apply to L2, but the significant outcome differences result from different environments in which language learning takes place (Ellis & Wulff, 2015; Krashen, 1994; Li & Jeong, 2020; MacWhinney, 2012). While children are able to use the rich contexts, sensorimotor experiences and relevant social interactions, adults tipically lack all of this in the classroom context where their learning usually takes place. Approaches such as the 'Unified Competition Model' (MacWhinney, 2012) and Krashen's 'Input Hypothesis' (1994) propose that adults' capacity to acquire languages does not diminish with age; rather they often lack appropriate input in form of opportunities for perception, action and social use that are readily available to children in their first language learning. Still, they recognize that the already well-established L1 representations in adults affect how L2 input is processed and acquired (Ellis, 2007). This particularly affects low-salience features and forms that do not directly map to already entrenched L1 categories, in which the L1 structures persist against changes during later L2 learning (Ellis, 2007; Hernandez et al., 2005; Li & MacWhinney, 2013).

In L1 acquisition, the multiple viewpoints on the relative importance of different cues for language development are combined within the 'Emergentist Coalition Theory' that acknowledges that lexical acquisition is dependent on multiple types of input, but that children will differently value these cues at different stages of their development (Golinkoff & Hirsh-Pasek, 2006; Hollich et al, 2000). Early on, they might assign more value to the perceptual cues, but as they learn to better understand the social cues, they begin to increasingly rely on them in the process of acquiring vocabulary. However, less is known about how different types of cues support adult L2 vocabulary learning.

Hartshorne et al. (2018) recently showed how the critical period for second language acquisition extends to far later than previously thought, almost to adulthood (17.4 years old). This study, employing a large dataset, offers a rigorous examination of age effects, challenging earlier, more rigid interpretatons of the 'Critical Period Hypothesis'. Increasing evidence also suggests that adult language learners can indeed approach L1-like representations when they engage in social and context-rich practices that mirror aspects of child L1 acquisition (Li & Jeong, 2020). Li and Jeong (2020) demonstrate that L2 words encoded through socially interactive conditions activate brain areas similar to L1 processing. Their use of neuroimaging techniques provides direct evidence for the neural plasticity uderlying L2 learning in enriched contexts, offering a stronger empirical basis than earlier behavioral observations alone. They propose a Social L2 learning (SL2) hypothesis that suggests that appropriate learning contexts that are interactive, multimodal and driven by action and perception can enhance L2 acquisition even in adulthood. Through sensorimotor and social input, learners can bypass the L2-L1 links and instead forge direct links between words and their embodied representations (MacWhinney, 2012). With the technological advances present today the more immersive input becomes possible even within typical classroom environments (Li & Jeong, 2020). When used in the right way, this can have positive implications for adult L2 learning.

While there is continuing debate over the extent to which L1 and L2 learning processes align, a growing body of research supports a fundamental continuity hypothesis: adult L2 learners retain the ability to acquire language much like children, provided they have meaningful multimodal input, social interaction and a rich, contextualized environment (Douglas Fir, 2016; Li & Jeong, 2020). Such evidence has direct implications for how we approach second language teaching, as it suggests that harnessing L1-like conditions of learning can significantly enhance L2 development, even for late learners. This evolving understanding of the relationship between L1 and L2 acquisition suggests the need for more integrated approaches of language teaching that have to find the optimal strategies to support L2 learning from the very beginning. Since words form the building blocks of language, in the following sections we turn the focus to the first step on the path to learning a second language – vocabulary learning.

1.3. L2 vocabulary learning – where are we now?

An essential, and complex, first step in second language learning is acquiring vocabulary (Kim et al., 2018; Nation, 2001). Learning new L2 vocabulary requires both creating new representations of word forms and their meanings and establishing new connections between them (Eddington & Tokowicz, 2013; Kroll & Stewart, 1994; Pavlenko, 2009). Two processes necessary for this to take place are word encoding and word retrieval (Morett, 2019). Word encoding refers to the initial stage of processing and storing of new vocabulary. Effective encoding helps create strong memory traces, allowing learners to build connections between new words and their existing knowledge. Retrieval is the ability to access and produce stored words when needed, which often requires repeated exposure and use in various contexts. Encoding is closely related to retrieval, as we will demonstrate in the second chapter, investigating the nature of this interplay. Both processes are essential for long-term retention, which is why enhancing them through targeted strategies can significantly improve second language vocabulary acquisition (Baxter et al., 2022).

A crucial prerequisite to improving L2 vocabulary learning is understanding of the cognitive processes involved. This includes how native and foreign languages are represented and how they interact with each other during language acquisition and use. Models of bilingualism differ in how they conceptualize the formation of lexical-semantic systems, resulting in different implications for improving effective learning. This chapter will provide an overview of influential theories, models and research that address these issues, with a special section that relates it to the multimodality aspect.

1.3.1. Models of initial L2 vocabulary learning

There are multiple ways in which models of bilingualism conceptualize lexical-semantic systems and their forming. One of the early and influential works on this topic was produced by Potter et al. (1984) in a study testing different hypotheses of second language word learning process. The ‘Word Association Hypothesis’ proposes that during second language learning the L2 words are connected to corresponding native language (L1) words. This link is used to reach underlying concepts needed to understand and produce words. In contrast, the ‘Concept Mediation Hypothesis’ suggests that words from both L2 and L1 operate independently, with direct links to the conceptual system (Figure 2). A third hypothesis introduced an intermediate model, which is the combination of the two initial hypotheses, suggesting that this process is dynamic. According to the ‘Intermediate Hypothesis’, early-stage bilinguals rely on their native language to process L2, but as proficiency increases, L2 becomes independent and directly related to concepts. Although study outcomes supported the Concept Mediation Hypothesis, other research suggests that it might have been due the proficiency levels in the beginner learner group, which was already quite advanced; a similar study by Chen and Leung (1989) tested the same hypotheses and found support for the Intermediate hypothesis, leaving the question open. These early experimental designs, while foundational, often used relatively small sample sizes and specific learner populations (e.g.

often university students), which might limit the generalizability of their findings by today's standards. The conflicting results also highlight the complexity of the issue and the need for more nuanced models.

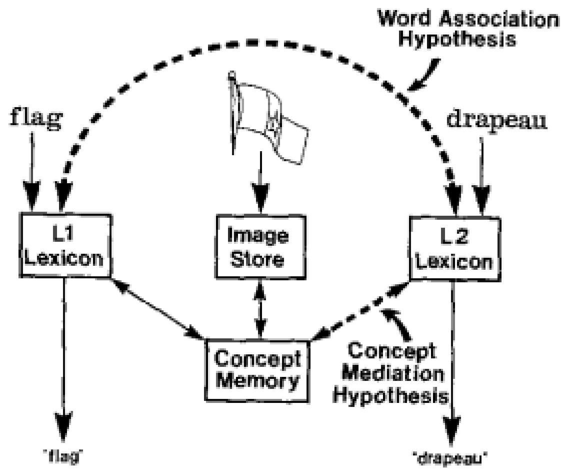

Figure 2. Illustration of the Word Association and Concept Mediation Hypotheses (Potter et al., 1984)

Further research continued to build on these hypotheses. One of the influential models that represents this process in second language learning that integrated these hypotheses is the 'Revised Hierarchical Model' (RHM) by Kroll and Stewart (1994) presented in Figure 3. Its value is in that it offered a structured explanation of how L1 and L2 words are organized and accessed by proposing a clear distinction between lexical and conceptual representations. The RHM offered hypotheses about how L1 and L2 processing differs, as well as describing the developmental trajectory of L2 vocabulary learning. It explains how the initial learning of new words in L2 is mediated by the words in L1, which is used to reach the underlying concepts. As the learner increases their proficiency, they becomes able to access the concepts directly from L2, without the need for mediation from L1 words, which makes the retrieval and understanding faster and more fluent (Kroll et al., 2010). Although the intial version of the

RHM was not explicitly multimodal, it served as a foundation to examine how new forms of input – such as images – might facilitate forming of direct concept links.

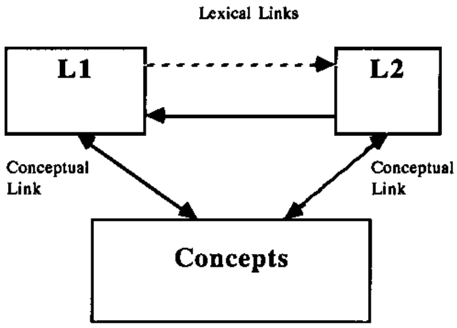

Figure 3. The Revised Hierarchical Model (Kroll & Stewart, 1994)

Over the years, the RHM was further developed and refined to adapt to new insights about multimodal L2 vocabulary learning. This evolution demonstrates the model's adaptability and the field's growing recognition of multimodality, moving it from a purely lexical model to once incorporating richer sensory inputs. In the 'Revised Hierarchical Model with Image', the influence of images on second language learning was included in the model, thus encompassing multimodal dimensions (Yoshii, 2006; Figure 4). Using an image would provide a direct conceptual link during the encoding stage of L2 vocabulary learning, which serves as an additional means of accessing the underlying concept during the word retrieval (Chen & Leung, 1989). Similar effects have been found with the use of iconic gestures (Allen, 1995; Macedonia, 2019), which are also thought to enhance the initial L2 word learning compared to the traditional word-to-word methods, by forming a direct path between the new word and the underlying concept. This idea will be expanded and tested in the first study presented in the following chapter.

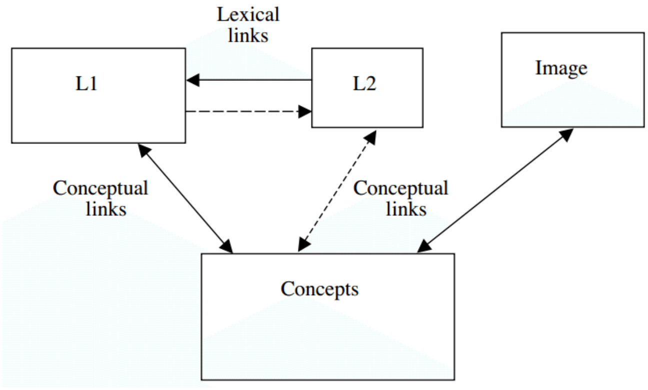

Figure 4. The Revised Hierarchical Model with Image (Yoshii, 2006)

A step further in developing the RHM was modifying it to account for varying cross-linguistic relations between words and their underlying concepts. Since the L1 translations do not always directly correspond to the L2 word meaning, Kroll and Tokowicz (2001) adapted the RHM to account for translation-ambiguous words. However, they did not include all the ways in which the translational ambiguity can occur. Therefore, the 'Revised Hierarchical Model of Translational Ambiguity' (RHM-TA; Figure 5) was developed (Eddington & Tokowicz, 2013). Within this model, panel *a* shows form ambiguous translations when a single word in the source language (L1) has multiple translations in the target language (L2). Panel *b* depicts meaning ambiguous translations, where a word in L1 translates into multiple L2 words and concepts. The direction of the ambiguity can also be opposite (from L2 to L1), but it is not depicted here (see Bracken et al., 2016).

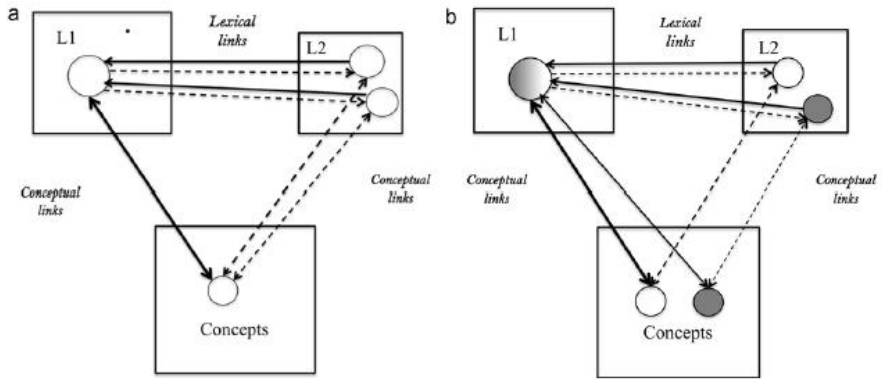

Figure 5. The Revised Hierarchical Model of Translational Ambiguity (RHM-TA; Eddington & Tokowicz, 2013). The RHM-TA incorporates translation-ambiguous words in the L1 to L2 direction derived from different sources. Panel A depicts synonym translation-ambiguous words and panel B shows meaning translation-ambiguous words. To examplify panel A, English word *shy* (L1) could in German be translated to both *schüchtern* and *scheu* (L2), which both share the same meaning (concept). In panel B, the L1 word *odd* can have two meanings - *uneven number* and *strange*, corresponding to two different concepts. However, in German as L2, there is no single word that encompasses both meanings, but there are two words, creating a different type of translation ambiguity.

However, Jarvis and Pavlenko (2007) note that most of the earlier L2 vocabulary learning models do not address the cross-linguistic differences that occur in the way the concepts are structured. The relationship between concepts across languages is rarely straightforward. Words and categories in one language often cover narrower or broader ranges of meanings in the other language, leading to complex overlapping relationships. (Malt et al, 2003). The different manner in which different languages carve up the semantic space, provides a 'filter' through which concepts are both viewed and encoded (Malt et al., 2015; Slobin, 1996). The complexity in cross-linguistic conceptualization also intersects with the idea of multimodality, where different types of links between L1 and L2 concepts might lead to distinct mechanisms of learning, thus benefitting from different modality effects.

The central issues in the theories of L2 learning so far had been the ways of mapping forms to meaning, but not the nature of the representation itself, which was assumed to be common. However, in second language learning it is not always the case that the new word has to be mapped onto an existing concept, but sometimes the existing concept has to be restructured and recreated to fit the new language. This has implications for the effectiveness of different L2 vocabulary learning methods and modalities at input, as will be shown in chapter 2. Therefore, in their 'Modified Hierarchical Model' Pavlenko (2009; see Figure 6) divides the underlying area of concepts into L1 specific categories, L2 specific categories and shared categories to account for the translational ambiguity on a deeper level. Pavlenko's model represents a significant current development, offering a more nuanced view of conceptual restructuring, supported by evidence from bilinguals' complex cross-linugistc mappings (Pavlenko, 2009; Jarvis & Pavlenko, 2007).

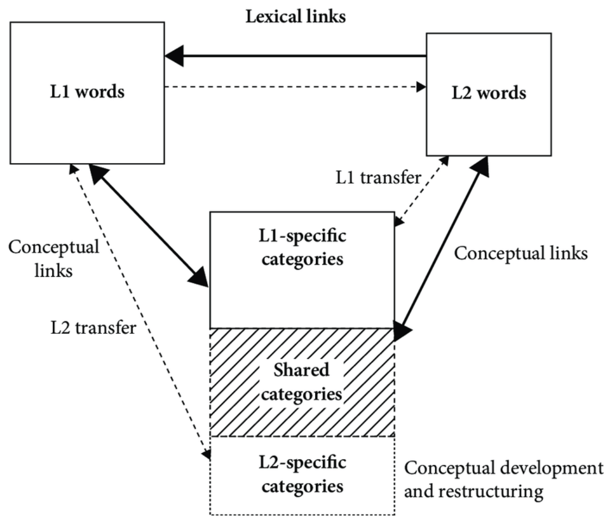

Figure 6. The Modified Hierarchical Model (Pavlenko, 2009). The model incorporates specific conceptual categories in relation to L1 and L2. Additionally, it includes different types of transfer and links between conceptual and lexical elements of the model.

On the other hand, the ‘Bilingual Interactive Activation’ (BIA) and ‘Bilingual Interactive Activation Plus’ models (BIA+) offer a different perspective on bilingual word learning than RHM (Dijkstra & Van Heuven, 2002; see Figure 7). The BIA and BIA+ are computational frameworks built on a monolingual Interactive Activation model of English word recognition (McClelland & Rumelhart, 1981). Both models assume that lexical access is non-selective with regard to language, meaning that words from both languages become activated during word processing. While RHM is a hierarchical model with separate lexical levels for L1 and L2, BIA+ is a network model that posits an integrated lexicon for both languages. This means that the former implies a greater reliance on L1 translations for understanding, while the latter suggests parallel activation of words from both languages. Its non-selective lexical access framework emphasizes how words from both languages are simultaneously activated during recognition. The BIA+ model incorporates orthographic, phonological, and semantic representations, providing a detailed explanation of how bilinguals process words and resolve competition between L1 and L2 during word recognition. However, it does not directly address how additional modalities might be integrated into bilingual lexical access. The structure of the BIA+ model is composed of four main levels: feature, letter, word, and language levels. The feature level processes the basic visual elements, such as lines and curves; the letter level processes the letters; the word level contains representations of words from both languages; and the language level contains representations of the languages a bilingual knows.

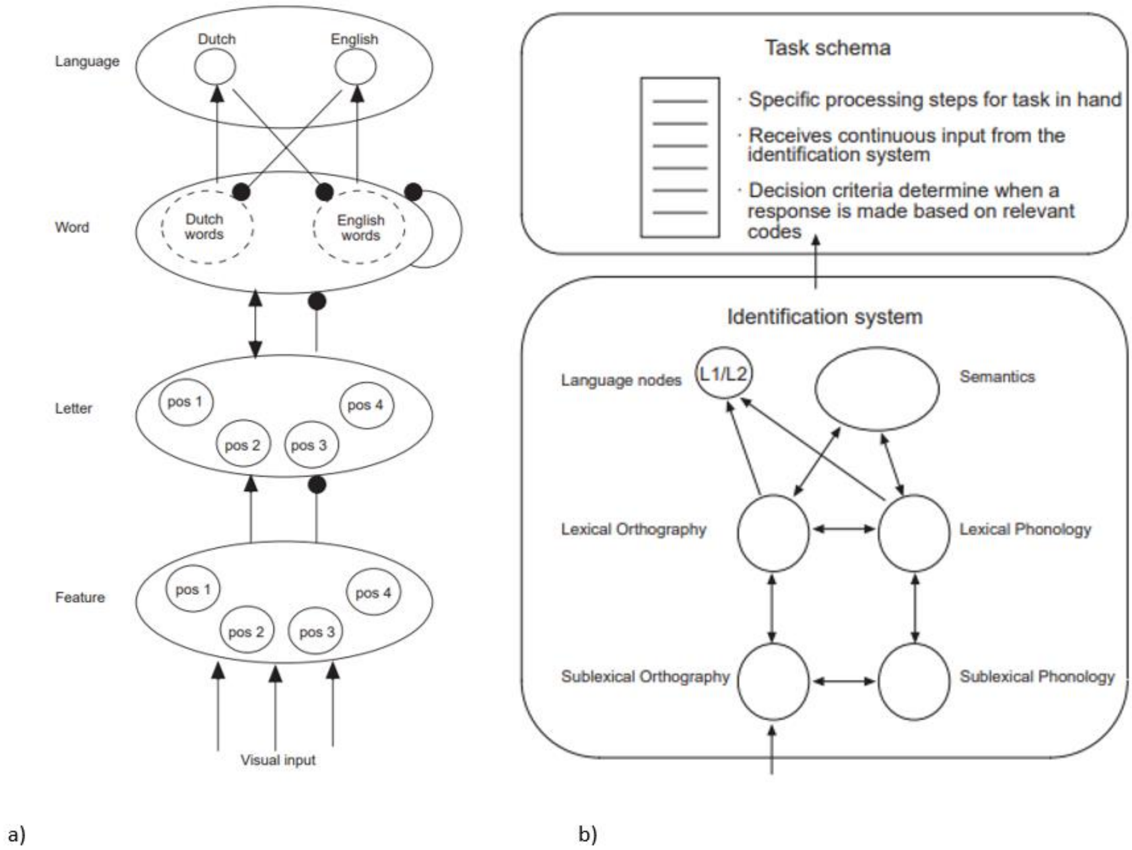

Figure 7. Schematic representation of the BIA (a) and BIA+ (b) models (Dijkstra & Van Heuven, 2002)

Both BIA+ and RHM models offer valuable insights into bilingual language processing with their own strengths and limitations. The RHM model is better equipped to explain developmental changes and asymmetries in language processing in bilingual individuals. Its primary focus is on language production and the gradual strengthening of connections between words and concepts. In contrast, the BIA+ model is more effective at explaining word recognition in more proficient bilinguals.

To address these differences, the 'Multilink model' was created as a unified computational model that integrates key aspects of the BIA+ and RHM models (Dijkstra et al., 2019). While the BIA+ model primarily focuses on language recognition and the RHM model on language production, the Multilink model seeks to integrate both processes. This integration is achieved through a localist-connectionist network that simulates the activation of various word representations (orthographic, phonological, and semantic). It considers the role of word-level variables such as word frequency, word length, orthographic similarity and phonological neighbourhood. This very recent model signifies a current trend towards theoretical integration, aiming to reconcile strengths of previously competing frameworks. As a newer model, its empirical validation is ongoing, but it represents a promising direction for future research (Dijkstra, 2019). However, it still does not model for some of the important aspects of second language learning, such as translational ambiguity and multimodality. The first iteration includes meaning representations that are either completely shared or completely separate, with the ambiguous multiple mappings planned for future stages. Likewise, the multimodal effect of pictures in the process is identified as one of the important aspects to address. Although originally focused on textual and phonological cues, the model could be expanded to consider how multimodal input (such as images, gestures, and sounds) influences L2 learning.

The nature of the Multilink model is in line with the reasoning by Li and Xu (2023), who argue for the need to broaden the current models by combining theoretical frameworks. Integrating different disciplines and methodologies to create pluralist models can improve the understanding of bilingual language learning. While pluralism in a competitive sense can emphasize exclusivity, where different theories are evaluated against each other, compatible pluralism supports the coexistence of diverse, mutually consistent models, each addressing different facets or dimensions of a phenomenon, acknowledging the complexity of the processes of interest.

To put it in words of Kroll (2010): "*No model should be left behind*". Each of these models represent different factors that are acknowledged as crucial in language learning and that affect encoding processes in different ways. The ways in which encoding can take place are multiple, each with specific consequences for the vocabulary learning outcomes. However, many important aspects of word learning such as the input and output modality, cross-linguistical differences and word properties have yet to be sufficiently explored and integrated into the current models. Therefore, these are the areas to be addressed to provide a more complete understanding of language as a complex integrated system that engages multiple sensory modalities.

1.3.2. Theories of multimodality applied to second language vocabulary learning

Recent approaches to second language learning increasingly view this process as embodied and holistic. These approaches move away from traditional methods that rely on word-to-word translations, focusing instead on creating deeper, concept-based links that try to mirror the natural acquisition of first language in early childhood. This section provides an overview of prominent embodied theories that drive current experimental research in the area of L2 vocabulary learning.

1.3.2.1. Foundational theories informing multimodal L2 vocabulary learning

The importance of multimodality during the encoding phase has been recognized by many prominent language learning theories and models that can be adapted to second language learning. One of the frequently used theories in this context is the ‘Dual Code Theory’ (Paivio & Csapo, 1973), which claims that accompanying verbal information with visual information (pictures) leads to the forming of a mental image that consists of two codes – a verbal and a visual. Since these codes are processed and stored in the different domains of the mind, they make the word more easily retrievable and more durable in memory. While foundational and influential, Dual Code theory originated from general memory research (Paivio, 1971, 1990). Its application to L2 learning has been fruitful, but a critical perspective might note that the inclusion of linguistic or L2-specific factors such as word properties or complexities of cross-linguistic conceptual differences should still be improved (Jarvis & Pavlenko, 2009).

1.3.2.2. Current embodied and multimodal perspectives in L2 vocabulary learning

Building on this understanding of multiple processing channels, theories of embodied cognition (Barsalou, 2008) expand the understanding of lexical representations by suggesting that they are linked to perceptual features across all senses. These approaches posit that learners' mental representations of concepts, objects, and actions are deeply shaped by their bodily experiences (e.g., using hands, feet) and sensory input (e.g., auditory, visual). This perspective contrasts with traditional theories arguing that cognition operates in distinct, separate modules, with language functioning independently from sensory perception and physical actions. The emphasis on physical action as an important aspect of learning inspired studies to investigate its role in L2 learning. This led to a number of recent studies on the use of observing and producing iconic gestures as a tool to enhance the learning of L2 vocabulary (Macedonia et al., 2019; Morett, 2019; Repetto et al., 2017; Rosenthal-von der Putten & Bergmann, 2020).

Drawing on similar ideas, Mayer et al. (2015) propose 'The Multisensory Learning' theory, grounded in cognitive neuroscience. They suggest that enriched sensory experiences harnesses additional brain regions and thereby strengthens the encoding process. The premise is that when more sensory channels are engaged, a broader neural network underpins the newly learned words, making them easier to recall later.

All the models described share a common notion that multimodality offers the potential to establish the connections between L2 words to concepts sooner than previously thought. The idea behind it is that by modifying the nature of vocabulary training, initial encoding and later retrieval processes can become more efficient. Experimental studies have followed up on these ideas and applied them to L2 learning settings to test the effects of multimodality (Comesaña et al., 2012; Macedonia, 2019; Morett et al., 2012). However, the evidence accumulated so far does not always support the same method. This highlights a crucial aspect of current research, which is the need for rigorous experimental designs that can disentangle the effects of different modalities, while controlling for cofounding variables. Thus, the central question of the way in which L2 words should be presented and learned to promote initial encoding and facilitate later retrieval remains open.

1.3.3. Narrowing the lens: Modalities at encoding from different perspectives

The idea that the learning method influences bilingual memory organisation and processing can be traced back to several decades ago (Ervin & Osgood, 1954 as cited by Lotto & De Groot, 1998). These early observations, while insightful, often relied on less controlled methodologiest, less rigid analysis procedures and smaller participant samples compared to current standards (Durrant et al., 2022). Consequently, while they highlighted important variables, the precise mechanism and generalizability of their findings were often underspecified.

Still, the choice of teaching method has consistently been shown to play an important role in second language vocabulary acquisition. For example, Laufer and Rozovski-Roitblat (2015) examined how the number of exposures during L2 word learning compares to the influence of learning method in shaping learning outcomes evaluated by different task types. They exposed learners to different reading-based methods of vocabulary learning and found that the number of exposures alone did not significantly impact acquisition. Rather, the method of instruction proved to be the determining factor in vocabulary acquisition. Additionally, the effectiveness dependent on the task requirements. Although the focus of this study was not to examine multimodality, it underscores how choosing the right teaching method can make learning more efficient by reducing the number of exposures needed to learn a word. Still, theoretical perspectives on the optimal teaching methods differ. One of the points of difference is the role of different modalities present at the encoding stages.

Distributional accounts emphasize the importance of linguistic context and statistical patterns of words co-occurring in language input (Griffiths et al., 2007; Louwerse, 2011), suggesting that exposure to language alone can be sufficient for effective word learning. In contrast, embodied theories highlight the role of sensory and motor experiences in grounding word meanings (Barsalou, 2008; Pulvermüller, 2013), proposing that multimodal methods, such as incorporating images or gestures, enhances learning by enriching mental representations. Given the critical role of teaching methods in vocabulary acquisition, two main approaches have been highlighted as emerging from these perspectives: Verbal methods, which focus primarily on the linguistic and distributional aspects of language and Multimodal methods, which integrate perception and action systems. The following sections will explore these two approaches, as well as introduce more recent hybrid and social perspectives.

1.3.3.1. Foundational frameworks: Distributional and Multimodal approaches to L2 vocabulary learning

1.3.3.1.1. Verbal modality at input and distributional accounts of L2 vocabulary learning

The traditional approach used as a first step in language learning was to associate L2 words with their L1 translations, utilizing the already existing connections between L1 words and meaning. Word-to-word associations facilitate the formation of initial form-meaning links essential for early stages of language acquisition where contextual understanding is limited, and learners benefit from straightforward and direct associations (Altarriba & Knickerbocker, 2011; Dronjic, 2019; Lotto & De Groot, 1998; Schmitt, 2008). This approach enables learners to quickly establish lexical connections, allowing cognitive resources to be freed up for further processing of language forms (Schmitt, 2008).

Despite this approach being deemed traditional and often considered less successful compared to multimodal approaches, many studies comparing pictures and text as input modalities still show the advantage of traditional verbal text-to-text methods in initial L2 word learning (Altarriba & Knickerbocker, 2011; Martín-Luengo et al., 2023; Morett et al., 2012; Repetto et al., 2017; Zhang & Zhang, 2024; Lotto & de Groot, 1998). For example, Martín-Luengo et al. (2023) explored the effectiveness of using pictures versus translations to aid learning, as well as measuring learners' confidence levels. Although learners reported feeling more confident when using word-picture pairs, their performance was actually better when learning through L2-L1 word pairs. Similarly, Morett et al. (2012) discovered that, for adult learners, neither images nor gestures provided any significant advantage in second language vocabulary acquisition compared to word-word learning methods.

However, a methodological appraisal of these studies reveals that there are elements that may limit the generalizability of conclusions. For example, there is a lack of control over psycholinguistic properties of the stimuli (Altarriba & Knickerbocker, 2011; Zhang & Zhang, 2024), which are known to significantly influence learning and processing speed (Brysbeart et al, 2014; Kuperman et al., 2012). When such properties are not controlled, the observed results might stem from inherent differences in item difficulty. While some studies controlled certain aspects of word properties (Martin-Luengo, 2023; Repetto et al., 2017), showing improved practices, the consistency of such controls across all research remains variable.

Furthermore, the statistical analyses employed in many of these studies warrant consideration. Several studies (Altarriba & Knickerbocker, 2011; Martin-Luengo et al., 2023; Morret et al., 2012; Repetto et al., 2017, Lotto & de Groot, 1998) relied on traditional Analysis of Variance (ANOVA), which van identify overall effects, but often requires data aggregation that can obscure item-level and participant-level variability. Critically, some of these analyses did not include by-participant (F1) and by-item (F2) analyses, a practice long recommended to ensure that findings generalize across both participants and stimuli (Clark, 1973; Raaijmakers et al., 1999). The absence of robust by-item and by-participant analyses, particularly when words are not matched on psycholinguistic properties, makes it difficult to rule out specific effects driving the results. Out of the highlighted studies, only Zhang & Zhang (2024) employed a mixed effects model, which can account for random effects of participants and items, without requiring data aggregation (Locker et al., 2007). The recent move towards generalized linear mixed models is getting more prominent and is one way of advancing the field, allowing for a more nuanced and robust analysis compared to other methods (Jaeger, 2008).

Also, the text benefit findings might be limited to certain conditions only. While L1-L2 pairings are efficient in creating initial connections in some words, other words require additional information in form of linguistic context (Coventry & Guijarro-Fuentes, 2012). Theoretical perspectives, such as distributional accounts, propose that understanding word meaning requires more than just associating it with a direct L1 translation (Ellis, 2008). Instead, the meaning of a word is shaped by its co-occurrence with other words within the language (Zhao, 2020). According to distributional accounts, the frequency with which words appear alongside other words influences learners' understanding of their meanings. This approach suggests that even basic verbal learning involves recognizing recurring patterns and relationships between words, which can help learners generalize their knowledge to new vocabulary. This has been proven particularly effective when learning word types recognized as more challenging for L2 learning, such as prepositions (Morett, 2019; Zhao et al., 2020) or words that are marked by cross-linguistic differences in meanings (Jarvis & Pavlenko, 2007; Talmy, 2000). This type of learning is also associated with large language models (LLMs) which achieve high levels of linguistic competence by following patterns, regularities and rules (despite still lacking functional competence; see Mahowald et al., 2024). These insights highlight the importance of linguistic context in learning new words, in line with the distributional accounts. This can be particularly relevant for certain types of words and cross-linguistical relations where there is more ambiguity present between the translations and meanings.

Taken together, there is evidence for the benefit of text as a modality in initial second language vocabulary learning. Still, this should be further investigated, as there are certain drawbacks to methodological and analytic choices in the current works, which limits the comprehensiveness of findings. This will be further explored in the second chapter.

1.3.3.1.2. Multimodal input and embodied accounts of L2 vocabulary learning

Multimodal approaches challenged the traditional, purely verbal methods of learning by integrating multimodal inputs like pictures, gestures, and videos, which current technological advances make increasingly accessible. As was already mentioned in the previous sections, embodiment approaches suggest that learning is optimized when it taps into the learner’s physical interactions with the world, enhancing word learning by connecting vocabulary to sensory-motor experiences (Craik & Lockhart, 1972; Paivio & Csapo, 1973).

However, the implementation of multimodal methods requires careful consideration, as each modality has its own unique processing requirements and places distinct demands on the learning mechanisms involved. While some studies show a clear advantage for multimodal approaches like using pictures during vocabulary learning, the results across studies are not always consistent. The existing literature shows conflicting findings regarding their individual effects on L2 vocabulary learning outcomes (Wolf et al., 2019). Due to the scope of this work, this section will focus on pictures as a multimodal input and review current literature.

A growing body of research shows how using pictures as a modality during second language vocabulary learning improves word encoding and retention as compared to the method of verbal learning through translation pairs (Bates & Son; 2020; Boddaert et al., 2021; Carpenter & Geller, 2020; Morett, 2019; Tonzar et al, 2009). This was confirmed in a study by Comesaña et al., (2009) in which they proved that learning with pictures had a direct impact on the type of lexical-semantic connections that were forming during L2 vocabulary learning, as predicted by RHM. In this experiment, children aged 11-12 years learned L2 words with traditional verbal method or with pictures and were later tested using the semantic interference task – an experimental paradigm that measures how meaning-based relationships between words affect language processing. In this task participants had to choose correct L2 word translations with some of the pairs including a distractor based on semantic similarity. Assuming that picture-based learning would help form links between L2 and the semantic system, the distractor based on semantic similarity would cause children to make errors more frequently than if the connection was only formed on the phonological L2-L1 basis. The results indeed showed that semantic interference had an effect only on children that learned new words with pictures. These outcomes led to the conclusion that between these two learning methods applied to children, only learning through associating with pictures established direct connections to the underlying concepts. This is in line with Yoshii's adapted Revised Hierarchical Model (2006) depicting how introducing pictures as input at learning creates direct connection to the underlying concepts, as opposed to reaching them only via L1 word translations (Figure 4).

However, a closer examination of the methodologies employed in these studies again reveals similar issues related to simuli control and statistical analysis, which necessitates a cautious interpretation of the reported effects. For instance, some studies supporting this effect, such as Bates & Son (2020) and Morett (2019) did not rigorously control for key psycholinguistic properties of the target words. Moreover, the statistical approaches adopted vary in their robustness. Bates & Son (2020) employed t-tests and Carpenter & Geller (2020) and Tonzar et al. (2009) used ANOVAs without reporting by-item and by-participant analyses, meaning that they did not account for the item-level or participant-level variance that raises concerns about generalizability of findings.

Additionally, the picture superiority effect in L2 learning is not universally observed. Bisson (2015) highlights that the evidence for the picture superiority effect L2 learning is less consistent compared to findings from general memory research. For example, Lotto and De Groot (1998) found no clear advantage for pictures in explicit second language word learning. In fact, their study showed that learners performed better when they were presented with L2-L1 word pairs, rather than L2 words paired with pictures. Since then, different studies have had the same outcome, as shown in detail in the previous section.

Additionally, a potential drawback of multimodality in word learning is that the use of pictures can divert learners' attention away from the written form of target words. Barcroft (2015) found that, while focusing on word meanings represented through visual cues enhances retention of meaning, it often comes at the expense of retaining the word’s form, such as its spelling or pronunciation. Pictures tend to guide learners toward understanding the concept or referent but do not emphasize the orthographic structure of the word itself. This division of attention, supported by research that investigated the effect of cognitive load (Boers et al., 2009; Martín-Luengo et al., 2023; Zhang & Zhang, 2024), suggests that learners may prioritize processing the visual stimulus over the word's form, ultimately hindering full retention of the vocabulary item.

The conflicting findings presented so far suggest that, while Paivio and Csapo's memory research (1973) supports the picture superiority effect, its applicability to L2 learning may be more nuanced and dependent on specific learning conditions. The shift from general memory principles to L2 learning applications requires consideration of factors such as learner proficiency, nature of L2 items, specific cognitive demands of learning and testing tasks, as well as cross-linguistical ambiguity. For example, the focus on concrete nouns is a significant limitation in the existing literature, as it potentially oversimplifies the debate by not accounting for word types where pictorial representation is less straightforward (e.g. abstract concepts, prepositions). This frequently occuring methodological choice restricts the generalizability of findings claiming superiority for either verbal or picture-based methods. Moving forward, the challenge lies in obtaining a thorough understanding of all the factors at play and finding tailored strategies that consider the interplay between the relevant factors. The experiments presented in chapter 2 will address this gap.

1.3.3.2. Contemporary developments: Social and Hybrid perspectives to L2 vocabulary learning

1.3.3.2.1. Social cues and social accounts of L2 vocabulary learning

Similar to the grounding effects from sensory-motor input such as pictures and gestures, a type of enrichment and grounding is thought to take place when learning and processing linguistic stimuli in a social context (Çetinçelik et al., 2021; Kreysa et al., 2018; Li & Lan, 2020). Language learning is inherently a social process, shaped by both cognitive mechanisms and social interactions. Therefore, it is important to recognize the powerful role that social cues such as gaze, gestures and other nonverbal behaviours play in guiding learners' attention, reinforcing meaning and supporting vocabulary acquisition (Çetinçelik et al., 2021).

Nonverbal social cues include multiple actions, such as facial expressions, posture, haptics, vocal cues, eye gaze and different types of gestures (Carmichael & Mizrahi, 2023). While iconic gestures have been studied as a helpful tool to enhance second language vocabulary learning (Macedonia, 2019; Repetto et al., 2017; Morett, 2019, Rosenthal-von der Putten & Bergmann, 2020), less is known about the influence of other types of gestures or other nonverbal social cues in this domain. Still, there are indications that other types of social cues might also contribute to initial second language learning. An overview will be introduced in this section, with more detailed information following in the third chapter.

During development, understanding social cues such as pointing gestures and eye gaze is tightly coupled to the emergence of joint attention, which sets the stage for early language acquisition (Brooks & Meltzoff, 2008; Goldin-Meadow, 2007). Numerous studies have shown that gaze and pointing in infancy are correlated with later language and vocabulary growth (e.g. Brooks & Meltzoff, 2008; Çetinçelik et al., 2021; Kirk et al., 2022; Law et al., 2012; Tenenbaum et al., 2015). By facilitating joint attention, these social cues help learners link words to the correct referents, understanding what the speaker is referring to. For L2 learners, joint attention is particularly important because it offers a non-verbal means to disambiguate the meaning of new words and attending to form-meaning relationships (Douglas Fir Group, 2016). Social cues, such as gaze direction, gestures, and changes in tone or pitch, help make these relationships more transparent, allowing learners to better understand and remember new words. Studies found that bilingual children, compared to monolingual children, were more attuned to the speaker's eye gaze when learning new words, helping enhance their learning outcomes (Brojde et al., 2012). Brojde et al. (2012) found that in unambiguous condition both groups used both cues, but when presented with conflicting information during word learning (eye gaze versus object properties), bilingual children prioritized eye gaze, as opposed to monolingual children. Although findings from children are compelling, a critical question for L2 adult research is whether this heightened attunement persists and translates also to adult L2 learning, and what specific types of social cues are most effective for adult learners.

Social cues also have an important place within classroom settings, where they enhance communication and support more effective learning (She & Fisher, 2002). A language teacher may use pointing gestures or gaze-shift to direct the learner's attention to an object or concept or iconic gestures to illustrate the meaning of a word or use (Valenzeno et al., 2003). Such actions provide scaffolding, helping learners to notice important linguistic forms, and enhance their ability to make connections between spoken words and their meanings (Matsumoto & Dobs, 2017). The use of non-verbal cues such as eye contact and gesturing can also promote immediacy effects, thus reducing the perceived distance between communicators and increase the learner's engagement with the material, motivation, as well as recall of information (Allen et al., 2006; Witt et al., 2004). Both Allen et al. (2006), and Witt et al. (2004) did meta-analyses on the relationship between teachers' verbal and non-verbal immediacy and learning, showing that it correlates not only with perceived learning and engagement, but also with enhanced learning outcomes.

Despite their demonstrated benefits in L1 and general classroom setting, there is a lack of studies investigating social cues in L2 vocabulary learning context. The potential benefits of their implementation have been explored mostly with social robots and digital agents in L2 word learning contexts (Belpaeme et al., 2018; Bergmann & Macedonia, 2013; Demir-Lira et al., 2020; Tsuji et al, 2021; Vogt et al., 2019). However, findings have been inconsistent, suggesting that there is more research needed to understand the role of social cues in contexts other than in-person social interactions.

In an eye-tracking study that looked at the effects of speaker's gaze on first language comprehension and acquisition, gaze helped direct attention to the target object more rapidly, but did not lead to improved memory performance (Kreysa et al., 2018). This might suggest that a speaker's gaze might be of limited importance for slower processes such as consolidating word-object relations in memory or might even come with a cost of dividing attention while learning (Kreysa et al., 2018), which might also be reflected in L2 vocabulary learning.

Promising findings relating to the role of social cues in L1 acquisition, processing and general classroom settings, could potentially be also applied in the L2 learning context, as was reported before for other findings from memory research that apply both in L1 and L2 (see Montero Perez, 2020). However, at the moment the application of social cues in L2 contexts remains insufficiently understood. This represents a missed opportunity, as L2 learners, particularly in digital environments, could potentially benefit from the attentional guidance and conceptual grounding that human social interactions naturally provide. As digital learning tools and virtual agents become increasingly common, understanding how cues like gaze, gestures, and nonverbal behaviours can enhance or hinder vocabulary acquisition is critical. The concern is that the gap between the rapidly growing numbers of digital learning tools and robust research in this area is increasing, potentially leading to suboptimal or even counterproductive learning experiences (Hirsh-Pasek et al., 2015). The current understanding of how specific cues like gaze and gesture contribute to L2 learning outcomes is fragmented, thereby limiting our ability to design truly effective digital learning tools. Further research could unlock more effective ways to integrate these cues into L2 learning, ensuring that learners benefit from the same attentional and cognitive advantages that social interactions provide. To do so, systematic investigation is needed into how learners perceive and interpret screen-based social cues. By increasing our understanding, L2 teaching methods can become optimized and more aligned with natural language learning processes.

1.3.3.2.2. Hybrid accounts – integrating distributional, embodied and social perspectives

The inconclusive results of the debate between verbal and multimodal methods in L2 vocabulary learning point to a possibility of a more complex underlying interaction between different factors. As research into language processing evolves, current approaches that integrate both distributional and embodied perspectives have gained prominence. Hybrid accounts propose that these two approaches do not necessarily need to be viewed as conflicting, but complementary, as both sensorimotor and linguistic networks are proposed to be engaged in language processing. Furthermore, they include social and pragmatic aspects of language (Borghi et al., 2019; Andrews et al., 2009; Andrews et al., 2014). As Borghi et al. (2019) argued, language is not just a tool for encoding and decoding meanings—it is also a form of social action.

In the context of lexical networks, hybrid approaches recognize that word meaning is shaped by its physical and perceptual associations, as well as the statistical patterns in language use. Crucially, they recognise the role of specific word-level properties and how they differently engage these networks (Borghi et al., 2019). The integration of these two perspectives has led to the development of theories such as Words as Social Tools (WAT) (Borghi & Binkofski, 2014) and Language is an Embodies Neuroenhancement and Scaffold (LENS) (Dove et al., 2020), which propose that both embodied experiences and distributional linguistic data contribute to how words are processed and understood. These specific theories offer valuable frameworks by making differential predictions based on word type. For instance, concrete nouns, which are closely tied to sensory-motor experiences, tend to activate embodied representations more strongly, while abstract words, such as abstract nouns rely more on linguistic processing and the statistical relationships between words. A critical point for L2 research is how these differential weightings of embodied versus distributional information might influence the learning of new L2 vocabulary.

Still, for some types of words the acquisition patterns between L1 and L2 has rarely been considered together (Coventry & Guijarro-Fuentes, 2012). Learners may find it easier to acquire vocabulary related to objects, places, or actions they can physically interact with or visualize. On the other hand, abstract terms and grammatical words (e.g., conjunctions, auxiliary verbs) might require more focus on distributional learning, where the meaning is inferred from how the word is used in relation to other words. For example, learners might struggle with the meaning of a certain preposition unless they can understand its usage across multiple contexts, where its meaning is derived not from any direct sensory experience but from its linguistic relationships.

The strength of hybrid accounts lies in their theoretical capacity to accommodate a wider range of empirical findings and acknowledge the multifaceted nature of language. While this approach could reconcile different frameworks by offering a more nuanced view of language processing, it has yet to be transferred into the domain of L2 vocabulary research. A critical challenge lies in their operalization in designing studies that can effectively tease apart and measure the contributions of distributional, embodied and social factors. Currently, most of the studies focus only on concrete nouns as the learning target. At the same time, through emphasizing the role of different word types and their interactions with cognitive processes, hybrid models encourage a deeper consideration of how language is processed and learned. Aside from integrating theories, hybrid accounts also draw attention to the challenge of identifying important factors that affect language processing and learning. Understanding these factors could result in finding ways to optimize current learning methods and enhance initial second language vocabulary learning, as will be shown in the first experimental study.

1.4. Precis of the Thesis

The review above has identified gaps present in understanding the complex and multimodal nature of L2 vocabulary learning, providing a motivation for the work presented in the remainder of this thesis, which approaches the topic of L2 vocabulary learning from multiple complementary perspectives. First, two series of experimental studies provide an insight into previously overlooked factors in L2 vocabulary learning (chapters 2 and 3). The first two experiments (chapter 2) investigated the effects of different modalities at learning. Prior results have been inconclusive, possibly due to previously unconsidered variables. The studies in chapter 2 presented this research question from broader frameworks of distributional versus embodied approaches, considering crucial factors and their interactions; testing methods and properties of the learning content. The careful consideration of factors resulted in deeper understanding of the L2 vocabulary learning process. The outcomes of the first studies indeed highlighted the importance of the testing modality and word class interactions, as well as confirming the differences present between ambiguous and unambiguous words. The broader goal of this study is to move the spotlight from investigating input modalities in isolation. Rather, it draws the attention to the complex dynamics between encoding and retrieval modalities in initial L2 learning, as well as the crucial effects of different word types that drive successful L2 word learning.

Chapter 3 presents a series of experiments examining the effects of social cues within a screen-based vocabulary learning setting. Although the importance of such cues has been well-established for L1 learning, less is known about their importance for L2 learning. Particularly understudied is how these effects translate into the digital L2 vocabulary learning setting. To answer this, a set of experiments examined the role of deictic cues (eye gaze, pointing gestures) in an online L2 vocabulary learning task by systematically manipulating their availability and congruency, as well as the position of the named objects relative to the interlocutor.

Chapter 4 shifts away from the experimental setting and presents findings on the practical aspect of L2 word learning through survey methodology. This study collects insights from second language teachers on their use of digital language learning tools in their classrooms via a questionnaire study. In addition, it introduces a theory-based evaluation framework for assessing second language learning software, incorporating both linguistic and educational perspectives to measure software quality and its alignment with research-based practices. The results identify the areas in which current language software aligns with educational and linguistic principles, but also highlights the areas for improvement.

Finally, the discussion (chapter 5) integrates findings from both experimental and practical studies and contextualises them from a transdisciplinary perspective. In doing so, it will circle back to address the fundamental questions raised about both theoretical and applied second language vocabulary learning.

The empirical studies presented are formatted as manuscripts that are either published or submitted and under review (detailed in Author Declaration) and will be presented in the following chapters. Each chapter will begin with an introduction to the relevant research context and connect it to the broader theme of this thesis.

Chapter 2 | The Interplay between Learning Modality, Testing Modality and Word Class in Second Language Vocabulary Learning

2.1. Introduction to Chapter 2

In Chapter 1 we established that the modality of input plays an important role in the language learning outcomes. While technological advances have created new possibilities for immersive, embodied and interactive learning, the abundance of design options without guidance can result in a product that is overwhelming for the users. Therefore, grounding design decisions in empirical research is essential to balance multimodal elements with learner's processing cabilities, thereby optimizing the development of educational technologies (Meyer & Moreno, 2003). However, our review in the previous chapter revealed inconsistent findings when comparing different modalities of learning. Proposed explanation for inconsistencies range from methodological choices to unconsidered factors that might play a moderating role, as debated in hindisight in some of the previous works (Altarriba & Knickerbocker, 2011; Boers et al., 2017; Comesaña et al., 2012).

For instance, due to the associative nature of memory, the processes of learning and testing are inherently linked, with the context used for encoding stimuli directly impacting their later retrieval (Li & Jeong, 2020; Hald et al., 2016). However, in vocabulary learning research testing methodologies vary greatly, each with distinct cognitive demands and implications for measuring learning outcomes. The type of the testing task (Kroll et al., 2010; Zhang & Zhang, 2024), its (in)congruence with the learning modality (Chen & Leung, 1989; Lotto and De Groot, 1998), as well as the time of testing (Carpenter & Olson, 2012; Tonzar et al., 2009) can differently affect the testing results. In line with this, Wolf et al. (2019) argue how studies investigating modality effects may have produced conflicting results due to the lack of one or more crucial controls regarding the task modality. This motivated the present studies to systematically investigate the influence of the testing task modality as a potential determinant of the learning modality effect in the process of second language vocabulary learning.

Another important factor influencing the learning process is the nature of learning content. Words differ on a range of dimensions that influence the way we process and represent their meaning (Hauk & Tschentscher, 2013; Pulvermüller et al., 2010) and how challenging they are to learn (Borghi, 2019). So far, the research in this domain has began to examine the differences between concrete and abstract words (Mayer et al., 2015; Repetto et al., 2017), cognates and non-cognates (Comesaña et al., 2012; Poarch et al., 2015; Tonzar et al., 2009) and emotionally valenced words (Ferré et al., 2009). The results confirmed that these differences can indeed affect L2 vocabulary learning outcomes. Still, most of the studies examining different learning modalities fail to account for these effects, as they are predominantly done on concrete nouns only, making it difficult to generalize results (as discussed in Butler, 2019; Poarch et al., 2015). Knowing how not all words are learned equally well and how not all words are processed in the same way (Romero Lauro, 2013), it is possible that they would also benefit from different learning strategies (Butler, 2019; Degani & Tokowicz, 2010; Gleitman et al., 2005). This further motivated the research questions in this set of studies to include the less considered effects of word type.

In the self-contained manuscript paper that follows, elements influencing the success of L2 word learning will be analyzed. Well-known factors such as the type of input will be re-evaluated in a new context, showing that the currently predominant way of studying it in isolation leads to inconclusive results. In contrast, the importance of frequently overlooked factors such as word type and testing modality will be highlighted, examining how they influence learning outcomes.

2.2. Second Language Vocabulary Learning: The Interplay between Learning Modality, Testing Modality and Word Class

This chapter has been submitted to Cognition journal as:

Janjić, P., Spencer, J.P., Coventry, K.R. (under review). Second language vocabulary learning: The interplay between learning modality, testing modality and word class. *Cognition*

The numbering will be adjusted to fit in with the thesis format.

2.2.1. Abstract

Theoretical accounts of second language (L2) vocabulary learning vary as a function of the degree to which words are grounded in non-linguistic representations or are determined distributionally. However, to date empirical research testing the roles of sensorimotor and linguistic processes in L2 vocabulary have failed to produce consistent results. The most frequent focus has been on the influence of learning modalities on L2 vocabulary learning performance, but studies usually fail to consider possible moderating variables, such as word properties or the congruence between modalities at learning and testing. The present experiments address these gaps by systematically examining how learning modality, testing modality, word class, and translational ambiguity interact to influence L2 vocabulary learning. In two experiments (N = 320 monolingual adults), the effects of picture-based versus text-based learning and testing modalities on learning of different types of words were investigated. Results showed that learning modality effects were not universal, but rather depended on the interaction between learning and testing modality. Furthermore, results revealed a significant and consistent impact of both testing modality and word class on learning outcomes.

Specifically, text-based testing was particularly advantageous for verbs and prepositions, indicating that different word classes engage distinct cognitive processes and grounding mechanisms. These findings integrate both embodied and distributional perspectives by acknowledging the roles of both sensorimotor and linguistic networks, thus supporting hybrid language processing models. Moreover, we present a model of L2 vocabulary learning that recognises the importance of aligning testing modality with word class. Such a model may play a role in the development of refined instructional strategies and digital learning tools that more effectively harnesses both visual and textual modalities to enhance L2 vocabulary learning.

2.2.2. Introduction

Mastering a second language (L2) is a complex process that begins with vocabulary learning. It provides a foundation on which to build more advanced language skills, and is a significant predictor of later listening and reading comprehension tasks (Grabe, 2009; Li & Zhang, 2019; Stahr, 2009). However, understanding of the mechanisms underlying vocabulary learning in a second language learning remains limited (Jeong et al., 2021; Kim et al., 2018; Schmitt & González-Fernández, 2019). There has been much focus on the modality of learning as a crucial factor in determining success of vocabulary learning (Carpenter & Geller, 2020 Comesaña et al., 2009; Morett et al., 2012). Approaches vary form those advocating the importance of presenting words together with images or actions, consistent with 'embodied' approaches to language learning (Allen, 1995; García-Gámez et al., 2018; Macedonia et al., 2011; Mayer et al., 2015; Morett, 2019; Repetto et al., 2017; Rosenthal-von der Putten & Bergmann, 2020; Tellier, 2008), versus accounts which maintain that initial vocabulary learning relies on mapping new L2 words onto existing L1 words (Cook, 2010; Dronjic, 2019; Jahangard, 2022; Kroll & Linck, 2007; Kroll & Stewart, 1994; Laufer & Girsai, 2008; Nation, 2003; Schmitt, 2008).

As we review below, empirical results regarding the relative effectiveness of grounding words in non-linguistic contexts versus presenting words in only a linguistic context has yielded inconclusive results. We argue that such inconsistencies are a result of two key factors that have hitherto not been considered; testing modality and word class. In the studies reported in this paper, we fully cross learning modality and testing modality across different word classes (nouns, verbs, prepositions) to examine how these factors *in combination* determine the effectiveness of new vocabulary learning in a second language. Prior to presenting the studies, we first review current findings and theories as a stepping stone to presenting a new framework placing different classes of words and retrieval as key drivers of the learning modality effect in L2 vocabulary learning.

2.2.2.1. Modality of input in L2 vocabulary learning models

Theoretical perspectives on vocabulary learning range from 'distributional' accounts (e.g. Griffiths et al., 2007; Louwerse, 2011) that place primary emphasis on the importance of linguistic context and statistical patterns of words co-occurring with each other within language input, to 'embodied' approaches that emphasise the importance of sensory and motor experiences in grounding word meanings embodied (e.g. Barsalou, 2008; Pulvermüller, 2013). Embodied perspectives lead to hypotheses that multimodal strategies, like the use of pictures or gestures, may improve encoding by enriching mental representations (Comesaña et al., 2009; Repetto et al., 2017). On the other hand, distributional approaches would suggest emphasis on linguistic exposure and context as a more effective means of language learning.

In line with theories of first language learning, research on L2 vocabulary learning has recently shifted focus to consider multimodality as key factor. Embodied theories have attracted attention in second language learning research, with studies testing the effects of combining sensory-motor experiences with linguistic input (Comesaña et al., 2009; Morett et al., 2012; Repetto et al., 2017; Rosenthal-von der Putten & Bergmann, 2020; Wang & Lee, 2021). The idea of enhancing word learning by linking language to sensory-motor experiences is supported by several embodiment theories. For instance, 'Dual Code Theory' (Paivio & Csapo, 1973) posits that combining verbal and visual information creates dual mental codes, thus enhancing encoding and retrieval. Meyer's 'multimodal processing model' (2007) shows that presenting content in parallel modalities creates integrated verbal and pictorial models. Similarly, the Connectivity model (Klimesch, 1987) suggests that enriching words with sensory experiences, enlarges and connects the network of concepts, facilitating faster processing.

In contrast to embodied approaches, L2 vocabulary instruction has until recently been dominated by methods that emphasize L1-L2 connections and linguistic input. These traditional approaches that rely on the use of first language for initial L2 vocabulary learning (Schmitt, 2008; Dronjic, 2019; Kroll & Stewart, 1994). Creating lexical connections through L1 words is thought to help in establishing the initial form-meaning link and in storing novel L2 items (Barcroft, 2002; Kroll & Curley, 1988; Laufer & Girsai, 2008), bootstraping learning by freeing up cognitive resources and allowing focus on the form (Schmitt, 2008). Only once the initial link is created, can the learner benefit from receiving more contextualized types of word knowledge, which would mean that the choice of the optimal method would not be multimodal by default, but depend on multiple different factors.

Various models of bilingual processing have been proposed to conceptualize the formation of lexical-semantic systems and the connections between first language (L1) and second language (L2) lexicons. Potter and colleagues (1984) were among the first to empirically test how L2 words are represented and accessed in relation to L1 words, proposing two influential hypotheses. The 'Word Association Hypothesis' posits that L2 words are primarily linked to their L1 translations and learners rely heavily on their native language to access meaning. For example, an L2 learner encountering the Spanish word *mesa* might initially access the L1 word *table* before accessing the underlying concept of a table. On the other hand, the 'Concept Mediation Hypothesis' suggests that L2 words are directly linked to non-linguistic concepts. For instance, a proficient Spanish learner might directly link *mesa* to the concept of a table without mentally translating it to *table* first.

Kroll and Stewart's (1994) 'Revised Hierarchical Model' can also be seen as an integration of both Word Association and Concept Mediation, offering a framework for different stages of L2 learning. According to this model, during early L2 learning, learners rely heavily on their L1 lexicon, using L1 words as a bridge to concepts. As proficiency in the L2 increases, learners establish direct links to concepts, facilitating faster and more fluent retrieval. This model has inspired a number of studies investigating methods for strengthening conceptual links between L2 words and their meanings through different teaching strategies, including lexical versus semantic approaches (Altarriba & Knickerbocker, 2011; Boddaert, et al., 2021; Comesaña et al., 2009).

The traditional word learning where L2 words are paired with their translations is thought to result in weak representations that are not directly linked to the word meaning, but use an alternate route through L1 lexicon to reach the underlying concepts (Li & Jeong, 2020). On the other hand, 'embodied learning' approaches try to achieve semantic encoding, where the connections from second language words are formed directly to their relating concepts (Kroll & Stewart, 1994), mimicking natural first language acquisition and potentially forming stronger and more durable connections (Mayer et al., 2015). As we next review, empirical studies testing between these approaches have produced inconclusive results.

2.2.2.2. Overview of empirical findings

To explore studies that have compared the two methods of L2 vocabulary learning described above, we ran a search through two main databases (Scopus and Web of Science) in June 2024. The search terms included 'second language vocabulary learning modality', 'L2 vocabulary learning modality', 'second language picture(/image) vocabulary learning', 'second language text(/word) vocabulary learning' and similar variations. The search results were screened for peer-reviewed studies that looked into the effectiveness of picture versus text as a learning modality in L2 vocabulary context. We only included studies in experimental settings, as classroom context introduces additional variability. Also, only explicit and not incidental learning paradigms were included.

As can be seen in Table 1, 19 experimental studies were found that fit the selection criteria. Before unpacking the findings, it is clear that the results across studies are inconsistent.

Of the 19 studies, six report significant picture benefit with effect sizes ranging from small[1] ($\eta p^2 = .08$) to considerable ($\eta p^2 = .409$) (Bates & Son, 2020; Boddaert et al., 2021; Carpenter & Geller, 2020; Comesaña et al., 2009; Morett, 2019; Webber, 1978). On the other hand, three studies report a significant text effect with some studies showing smaller effect sizes (ηp^2=.03) and others indicating large effect sizes, up to ηp^2=.44 (Comesaña et al., 2012; Lotto & De Groot, 1998; Martín-Luengo et al., 2023;). Additional three papers report no learning modality effect found (Bergmann, 2020; Miyakoda et al., 2011; Repetto et al., 2017; Rosenthal-von der Putten &). Finally, seven studies show different effects within the same study, depending on the nature of the task or word properties, indicating potential mediating effects of different factors (Altarriba & Knickerbocker, 2011; Carpenter & Olson, 2012; Emirmustafaoğlu & Gökmen, 2015; Farley et al., 2012; Morett et al., 2012; Tonzar et al., 2009; Zhang & Zhang, 2024).

Table 1. Table presenting an overview of research on the effect of picture-modality versus text-modality on second language vocabulary learning outcomes. Only studies examining explicit learning methods have been included.

Study	**Participants**	**Learning modalities**	**Testing cue modalities**	**Learning x Testing modality crossed**	**Word type**	**Additional effects tested**	**Findings**
Webber, 1978	Children M_{age} = 9,5 yo N = 42	Picture and text	Picture and text (L1-L2)	Yes	Concrete nouns	Cognate status	Picture modality advantage (η_p^2 = .318)
Lotto & De Groot, 1998	Adults M_{age} = not reported (students) N = 64	Picture and text	Picture and text (L1-L2)	Yes	Concrete nouns	Cognate status; Frequency of words	Text modality advantage (η_p^2 = .033); Effect of congruency on RT (congruent advantage; η_p^2 = .133)
Comesaña, Perea, Pineiro & Fraga, 2009 (Experiment 2)	Children M_{age} = 10-11 yo N = 48	Picture and text	Text (L2-L1 pairs recognition: response times)	No	Concrete nouns	Semantic interference task	Picture modality advantage in facilitating conceptual processing (η^2 = .24; η^2 = .25)
Tonzar, Lotto & Job, 2009	Children M_{age} = 9-13 yo N = 229	Picture and text	Picture and text (congruent only; L1-L2)	No	Nouns	Cognate status	No learning modality effect in immediate testing; Picture modality advantage in delayed

[1] Labels for effect sizes as recommended in Lakens (2013).

							testing (η_p^2 = .08; η_p^2 = .10)
Altarriba & Knickerbocker, 2011	Adults M_{age} = 19.3 N = 84	Picture and text	Text (L2-L1 pairs recognition)	No	Concrete nouns	Picture type (black & white or colored); semantic priming task	No modality effect (Exp1)
	Adults M_{age} = 20.07 N = 15						Text modality advantage (Exp2) (d = 1.14; d = 1.36) modality effect η^2 = .22
Miyakoda, Kaneko & Ishikawa, 2011	Adults M_{age} = not reported (students) N_{exp1} = 64 N_{exp1} = 40	Picture and text	Not specified (L2-L1)	No	Different word classes (not controlled)		No learning modality effect
Carpenter & Olson, 2012 (Experiment 2)	Adults M_{age} = not reported (students) N = 24	Picture and text	Picture and text (congruent only; L1-L2)	No	Concrete nouns	Free recall of L1 words tested; Testing time	Inconclusive findings – no modality effect at Test 1; picture advantage at Tests 2 & 3 (η_p^2= .409; η_p^2 = .256)
Comesaña, Soares, Sanchez-Casas & Lima, 2012	Children M_{age} = not reported (students) N = 48	Picture and text	Text (L2-L1 pairs recognition: reaction times)	No	Different word classes (not controlled)	Cognate status	Text modality advantage for reaction times (η^2 = .53) and scores (η^2 = .22)
Farley, Ramonda & Liu, 2012	Adults M_{age} =10.87 N = 87	Picture and text	Text (L2-L1)	No	Concrete nouns; abstract nouns	Concreteness; Testing time	Picture modality advantage for abstract words (d = .26 (immediate); d = .28 (delayed)); No modality effect for concrete words
Morett, Gibbs & MacWhinney, 2012	Adults M_{age} =20.25 N = 26	Picture and text	Text (L2-L1) (Exp1)	No	Different word classes (not controlled)	Active vs passive learning mode; gestures; Testing time	Text modality advantage (η_p^2 = .26; η_p^2 = .44)
	Adults M_{age} = 20.64 N = 26		Audio (Exp2)				No learning modality effect
Emirmustafaoğlu & Gökmen, 2015	Children M_{age} = ~12 N = 75	Picture and text	Picture and text (congruent only; L1-L2)	No	Concrete nouns	Testing time	Picture modality advantage (d = .54 (immediate); d = 1.03 (delayed))
	Children M_{age} = ~12 N = 60		Picture and text (L1-L2)	Yes			No learning modality effect
Repetto, Pedroli & Macedonia, 2017	Adults M_{age} = 30.45 N = 20	Picture and text (+ gestures)	Text (L1-L2 production; L2-L1 production and recognition; free recall)	No	Concrete nouns; abstract nouns	Concreteness	No learning modality effect
Morett, 2019	Adults M_{age} = 18.70 N = 28	Picture and text (+ gestures)	Text (L1-L2); free recall; depiction	No	Concrete nouns; concrete verbs	Testing mode	Picture modality advantage (OR (text) = .89) or N/A)
Bates & Son, 2020	Adults M_{age} = 18-19 yo N = 45	Picture and text	Text (L1-L2)	No	Different word classes (not controlled)	Proficiency level	Picture modality advantage (d = .69)
Carpenter & Geller, 2020	Adults M_{age} = not reported	Picture and text	Picture and text (congruent	No	Concrete nouns	Confidence judgements	Picture modality advantage (η_p^2 = .22)

	(students) N = 25		only; L1-L2)				
Rosenthal-von der Putten & Bergmann, 2020	Adults M_{age} = 24.2 N = 26	Picture and text (+ gestures)	No cue (free recall (L1, L2 and L1-L2))	No (issue avoided by not using any cues)	Nouns; adjectives; verbs	Word type	No difference between text and picture modalities in L1-L2 free recall; overall effect of enrichment present (gestures);
Boddaert, Casalis & Mahe, 2021	Children M_{age} = 8.83 N = 40	Picture and text	Picture	No	Concrete nouns	Relatedness between words and distractors	Picture modality advantage (d = .23)
Martín-Luengo, Hu, Cadavid & Luna, 2023	Adults M_{age} = 21.88 N = 113	Picture and text	Text and picture (congruent only; pairs recognition)	No	Concrete nouns	Confidence judgements; different scripts	Text modality advantage (η_p^2 = .035)
Zhang & Zhang, 2024	Children M_{age} = 11-12 yo N = 43	Picture and text	Text (L2-L1 production and recognition)	No	Not specified	Content related input vs paralinguistic related input	No modality effect on recognition; picture advantage on production

The effectiveness of different L2 vocabulary learning methods is typically tested by comparing the word learning condition consisting of translations between L1 to L2 to a multimodal experimental condition. One of the early research efforts exploring the role of multimodality in second language vocabulary learning was conducted by Webber (1978). Participants were presented either with foreign language words paired with first language translations or with images picturing referent objects or concepts. The results supported the idea that using non-verbal stimuli, such as images, could enhance the learning and recall of foreign language vocabulary by linking L2 words directly to their concepts, paving the way for later studies examining the impact of multimodal approaches and their benefits compared to the traditional learning methods. Similar studies, comparing pictures and text continued to emerge (Table 1), and confirm the picture advantage and embodiment approach (Bates & Son, 2020; Boddaert et al., 2021; Carpenter & Geller, 2020; Carpenter & Olson, 2012; Farley et al., 2012; Morett, 2019; Tonzar et al., 2009).

In contrast, a number of studies show that approaches relying on L2-L1 connections can produce comparable (Miyakoda et al., 2011; Repetto et al., 2017) or even superior results compared to multimodal L2 vocabulary learning (Lotto & De Groot, 1998; Martín-Luengo et al., 2023). A recent study by Martín-Luengo et al. (2023) examined the role of pictures versus translations as a learning aid, while also rating learners' confidence in their learning performance. It was found that learners were more confident about word-picture pairs, but the actual performance was better in the word-word condition ($\eta p^2 = .035$). Similarly, a study by Morett et al. (2012) found that neither images nor gestures facilitated second language vocabulary learning in adult learners compared to the word-word learning input.

We argue that the inconsistent results between studies may be due to i) a general failure to consider modality at test, and ii) different word classes examined. Only three studies considered and fully crossed modalities at learning with modalities at testing, to account for the effect of (in)congruency (Emirmustafaoğlu & Gökmen, 2015; Lotto & De Groot, 1998; Webber, 1978). Five further studies used two modalities at testing, but did not fully cross them (Carpenter et al., 2009; Carpenter & Geller, 2020; Emirmustafaoğlu & Gökmen, 2015; Martín-Luengo et al., 2023; Tonzar et al., 2009). Instead, they tested only with congruent modalities (text cues in the text learning condition and picture cues in the picture learning condition). Moreover, only one of these studies included different word classes as a variable (Rosenthal-von der Putten & Bergmann, 2020). Other studies almost exclusively used nouns as target words (N=13), of which the majority used only concrete nouns (N=9). The rest of the studies either did not report the word class (N=3) or used different word classes, but without controlling for them (N=2).

2.2.2.3. Testing Modality as a Determinant of L2 Vocabulary Learning

It is well established in memory research that the nature of the testing task, as well as the relationship between learning and testing is fundamental to learning performance (Godden & Baddeley, 1975; Grant et al., 1998; Morris et al., 1977; Tulving & Thomson, 1973). Tulving & Thomson (1973) proposed the *encoding specificity principle*, which posits that the difference between the cues and processes at encoding and at retrieval affects learning outcomes, with recall improving when the retrieval conditions mirror the learning conditions. Their experiments included studying words as target items that were paired with different words. At retrieval, they tested the efficiency of different words as retrieval cues and concluded that "*the effectiveness of particular cue depends on how the to-be-retrieved item was encoded at input*" (p.371). This idea was extended into the influence of environmental contexts, showing that when they match between retrieval and encoding, the recall improves (Godden & Baddeley, 1975; Grant et al., 1998). This has been complemented by the *transfer-appropriate processing* framework suggested by Morris et al. (1977) that proposes that the type of mental processing used at retrieval and encoding influences the recall. In their study, participants processed words in two different ways during encoding (semantic processing vs rhyme processing). The retrieval outcomes depended on which type of processing the test required, with congruency between encoding and retrieval leading to better outcomes.

Although the effects of congruency during learning are well known in memory models, research on L2 vocabulary typically does not account for this, as reviewed in the previous section. Despite investigating the effects of picture versus text modalities at learning, the modality of cues at testing is predominantly textual (Table 1), which might bias results and obscure the true impact of multimodal learning, as was discussed in hindsight in some of the studies listed (Altarriba & Knickerbocker, 2011; Comesaña et al., 2012). For example, a study by Altarriba & Knickerbocker (2011) tested the effects of different modalities at encoding during L2 word learning by observing the reaction times in the priming paradigm. Contrary to their predictions, the results were in favour of the text modality compared to picture modality, which they explained with the effect of the transfer-appropriate processing (Morris et al., 1977). In other words, since the modality at testing (text) was the same as one of the modalities used at encoding, it led to an advantage in that particular condition.

The few studies of L2 vocabulary learning that do consider modality at retrieval are dispersed over different modalities of interest, such as gestures (Morett, 2019) or different learning methods, such as vocabulary learning through reading (Chun and Plass, 1996), listening (Jones, 2004) or videos (Jelani & Boers, 2018). Chun and Plass (1996) explored the impact of multimedia annotations on L2 vocabulary learning through reading, using both production and recognition tests that incorporated both pictorial and written items. While annotations as supplementary learning aids typically provide a definition or explanation of the word either in the L2 or in the native language (L1) of the readers, the multimedia environment allows for annotations to take form of both text, pictures and videos. The study observed a significant improvement in students' performance when the testing format matched the modality of the information presentation, consistent with both the encoding specificity principle (Tulving & Thomson, 1973) and the transfer-appropriate processing framework (Morris et al., 1977).

The three picture-vs-text L2 vocabulary learning studies that did fully cross learning and testing modalities in their designs show mixed findings (Emirmustafaoğlu & Gökmen, 2015; Lotto & De Groot, 1998; Webber, 1978). Webber (1978) presented both stimuli modalities at testing to control for the confound, and reported an advantage of picture modality at learning ($\eta p^2 = .318$). However, this study did not analyse the effect of congruency, including it only as a counterbalancing feature. On the other hand, Lotto and De Groot (1998) tested both the learning modality and the congruency effect. The study included adult participants previously unfamiliar with the target language. During the learning phase, each participant was presented with a screen showing a sequence of words paired either with their translation in the native language, or with a picture representing its meaning. In the testing phase, half of the participants in each of the two groups was tested in the modality congruent to the modality at learning and the other half in the incongruent condition, resulting in four groups total. Despite controlling for the congruency effect in the similar way as Webber (1978), they found opposite results - learning outcomes were found to be overall better with text modality at learning ($\eta p^2 = .033$). Regarding the congruency, in both learning modalities the congruent condition yielded better results, showing that congruency between modalities at learning and testing facilitates word learning. However, in their study involving 12-year old intermediate learners, Emirmustafaoğlu and Gökmen (2015) did not find the congruency effect. Instead, a universal benefit from picture modality at testing was observed.

Regardless of the evidence for the congruency effect and the effect of the testing modality, most of the multimodality studies have tested learning outcomes through text modality only (Table 1). This creates a confound in the research design and potentially obscures the effects of the learning modality. The few studies that use two modalities at testing do not fully cross learning and testing modalities (Carpenter & Geller, 2020; Carpenter & Olson, 2012; Martín-Luengo et al., 2023; Tonzar et al., 2009), and thus fail to account for the inherently different effects that testing tasks can have on retrieval. A significant difference between text and pictures as a modality at retrieval has been found in L1 word naming times (Fraisse, 1960; Potter & Faulconer, 1975), showing the importance of defining the influence of task type on the testing outcomes and conclusions drawn from them.

Taken together, these arguments suggest that, when testing effects of the learning modality, it is crucial to consider testing modality.

2.2.2.4. Word Properties as a Determinant of L2 Vocabulary Learning

Out of the reviewed studies considered above (Table 1), only Rosenthal-von der Putten and Bergmann (2020) included different word classes (nouns, verbs, adjectives) as a variable. Their study focused on the digital learning environment and investigated how different modalities presented by a virtual agent influence L2 vocabulary learning outcomes. In their work, they compared three modalities – two “enrichment conditions” (gestures, pictures) and one “no enrichment” (control) condition, consisting of L2 word and its L1 translation, presented alongside the virtual agent. This study also sidestepped the (in)congruence between the learning modality and the modality of cues presented at testing by testing with a free recall method in which learners are asked to retrieve and produce previously learned vocabulary items without any cues. The participants (N=26) were asked to recall as many L1 words, L2 words and L1-L2 word pairs as possible as a measure of their learning outcomes. Although the main focus of this study was to investigate and compare gesture and picture conditions, we focused on the picture vs text (control condition) comparison, relevant to our own study. The L1-L2 recall found a gesture advantage compared to both text and pictures, but no difference between picture and text conditions. Crucially, the overall study found that different word types benefited from different modalities at learning. Nouns benefitted more from gesture modality, verbs from picture modality, whereas neither of the enrichment modalities brought an advantage for adjectives. However, since the focus of the study was on investigating enrichment, there is variability related to the presentation of stimuli between picture condition and text (control) condition. While the text condition presents only text, the picture condition presents both text and picture. The different number of modalities present across conditions introduces additional constructs of attention, processing speed and cognitive load that may mask the effect of the modality on its own (Boers et al., 2017). So while these results suggest that word class is an important factor in understanding the effect of modalities at learning, still more research with different designs are required to fully understand their relationship.

In L1 learning, there is an established literature indicating that different word types are grounded and processed in different ways, relying to varying degrees on distributional versus embodied representations (Borghi & Binkofski, 2014; Bultena et al., 2013; Markostamou & Coventry, 2022; Pulvermüller et al, 2010; Vigliocco et al., 2011; Zhang et al., 2013). For example, while 'concrete' nouns can be effectively represented visually (where an image can successfully capture the most important features), prepositions require additional linguistic information in order to determine how their appropriate use (Coventry et al., 1994; Coventry & Garrod, 2004). Some explanations attribute the difference between the word classes to the varying levels of concreteness (for competing evidence, see Martin & Tokowicz, 2020), which refers to how directly a word's meaning can be perceived through the senses (Brysbaert et al., 2014). Although there are exceptions, nouns overall tend to be more concrete and verbs and prepositions tend to be more abstract (Borghi & Binkofski, 2014). Since the abstract words rely more on linguistic experiences, their representation is more affected by the linguistic variations across languages than that of concrete words that are more constrained by the environment and grounded in sensorimotor information (Borghi et al., 2019; Gentner & Boroditsky, 2001). For instance, in English, whether an object is labelled a *plate* versus a *dish* affects whether the appropriate preposition to use is *on* or *in*, which cannot be represented through a single image (Coventry et al., 1994). Instead, distributional and contextual information are required to understand the correct usage (Figure 8).

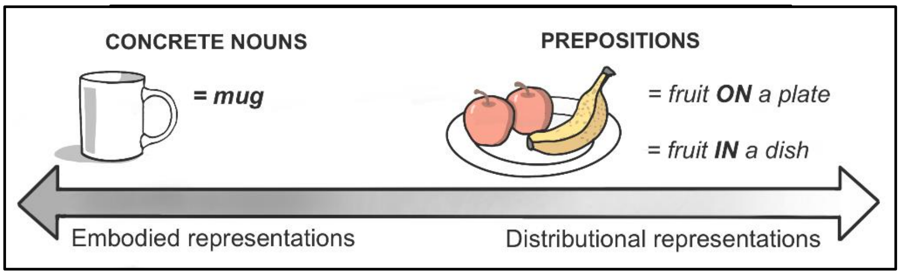

Figure 8. Figure showing the continuum between words that can be represented via sensorimotor information, such as concrete nouns, versus words that depend on linguistic context, such as prepositions.

Following this, linguistic context would have a stronger influence on the conceptualization of relations that are expressed by verbs and prepositions than on conceptualization of objects expressed by nouns.

This has implications for L2 learning as well. A study by Mestres-Missé et al. (2010) used fMRI to compare brain activation patterns during initial noun and verb learning in a second language, showing different patterns of activation and different learning outcomes between the two word classes. While in an MRI scanner, adult participants were introduced to pseudowords simulating nouns versus verbs embedded within L1 sentences. Subsequently, participants performed a word recognition task. The results showed that learning nouns was associated with greater activation in the visual object processing area (left fusiform gyrus), whereas verbs displayed larger activation in left middle posterior temporal gyrus and left posterior inferior frontal gyrus, which might be due to higher cognitive demands associated with relational processing of verbs. This indicated that even at the initial stages of second language vocabulary learning there are distinct neural patterns differentiating word classes. Also, the behavioural results showed that nouns were learned better and were associated with faster reaction times, in line with previous research (Gillette et al., 1999). The explanation for both the noun advantage and different neural activation was attributed to the different conceptualization of meaning in nouns versus verbs. The authors argue that while nouns are more frequently related to individual referents, meaning of verbs emerges from relations with other elements of the sentence, similarly as argued for prepositions (Coventry et al., 1994; Coventry & Garrod, 2004).

In line with these findings, behavioural studies have repeatedly shown that nouns are learned earlier in first language acquisition and are easier to learn in second language learning as well (e.g., Gentner, 1982; Gillette et al., 1999; Ellis & Beaton, 1993; Waxman et al., 2013). Considering the variability in the way different word types are processed and represented, it is possible that they would also benefit from different learning strategies. For example, Farley and colleagues (2012) compared how learning modality affects concrete versus abstract words. The results showed that picture modality benefited abstract, but not concrete words. Moreover, Rosenthal-von der Putten and Bergmann (2020) found that the effect of the learning modality interacted with the word class variable. This was observed despite the words in this study being comparable with respect to concreteness, imageability and familiarity. These results show that it is likely that, rather than having one optimal learning method, the best learning strategy depends on the type of linguistic material learned.

Building on this idea, the influence of word class is closely tied to another factor affecting learning outcomes in L2 word learning - cross-linguistic ambiguity. Cross-linguistic ambiguity occurs when different linguistic, cultural and social contexts within which the words are acquired result in some translation equivalents sharing only a part of their conceptual representational elements (De Groot, 1993). Frequently words and categories in one language refer to a subset or a superset of classes labeled in the other language; leading to cross-cutting and overlapping between words (Malt et al., 2003). This has been documented in many domains and across different word types (Bowerman & Choi, 2001; Malt et al., 2015), resulting in formal L2-L1 'translation equivalents' rarely sharing all aspects of meaning (de Groot, 1993).

Research done on prepositions showed that the variability between languages in how they describe space makes it particularly challenging for L2 learners to learn the patterns and the conditions under which a particular preposition is used (Coventry & Garrod, 2004). Similarly, studies found that the meanings of verbs are more ambiguous across languages than that of nouns (Prior et al., 2007; Smith et al., 2012). Furthermore, a study by Majid, Jordan, and Dunn (2015) showed how abstract words exhibit more variation in meaning across languages than concrete words. In their study involving 20 Germanic languages, participants named items in four domains: colour, body parts, containers, and spatial relations. While the meanings of colours, body parts, and containers were relatively consistent across languages, spatial relations showed the greatest variation. These results are in line with the proposal that the conceptualization of objects (concrete nouns) is more constrained by the environment and less dependent on language, which also makes it less variable across languages (Gentner & Boroditsky, 2001). On the other hand, language has a stronger influence on how relations expressed by verbs and prepositions are conceptualized, therefore leading to more significant cross linguistic differences (Coventry et al., 2011).

This holds true even in etimologically closely related languages, like Dutch and English. For example, in English certain relations between objects would be described using a preposition "on" (e.g. pencil on a desk, picture on a wall, ring on a finger), whereas Dutch language uses three different prepositions, based on the type of support between the objects (e.g. horizonal versus vertical support). Therefore, it is not always enough to only map a new L2 word onto an existing concept, but the existing concept has to sometimes be restructured and recreated during the encoding phase (Pavlenko, 2009). This creates difficulties in mapping L2 words onto the L1 forms or meanings since the conceptual space is differently divided and there isn't a one-to-one correspondence (Pavlenko, 2009).

It is well established that cross-linguistically ambiguous words are more difficult to learn in second language than unambigous words (Degani et al., 2016; Degani & Tokowicz, 2010; Tokowicz & Kroll, 2007). In the study by Degani & Tokowicz (2010), native English speakers learned pairs of Dutch and English words. The words were divided into unambiguous (one Dutch word corresponds to one English word) and ambiguous (two Dutch words correspond to one English word) word types. The ambiguous words were further divided into words with multiple forms (two Dutch translations for one meaning of the English word) and words with multiple meanings (each Dutch translation corresponds to a different meaning of the ambiguous English word). The results showed that participants translated ambiguous words more slowly and with less accuracy, and they had difficulty recognizing the correct translations.

Jarvis and Pavlenko (2007) note that most L2 vocabulary learning models do not address these cross-linguistic differences that occur in the way the concepts are structured. The central issues in the theories of L2 learning have for long been about the ways of mapping forms to meaning, but not the nature of the meaning representation itself, which was assumed to be common. However, the new word can not always be simply mapped onto an existing concept, but sometimes the existing concept has to be restructured and recreated to fit the new language (Degani & Tokowicz, 2010). This difference results in two distinct encoding processes that potentially benefit from different learning strategies (Barcroft & Sunderman, 2008), which should be investigated in the context of initial L2 vocabulary learning.

Overall, the majority of L2 vocabulary learning studies so far have been conducted on concrete nouns (see Table 1). Two studies from Table 1 expanded their focus to abstract nouns (Farley et al., 2012; Repetto et al., 2017) and one study included a balanced number of concrete nouns and verbs (Morett, 2019), but did not analyse the differences between them. Another six studies do not explicitly report on word class (Bates & Son, 2020; Comesaña et al., 2012; Miyakoda et al., 2011; Morett et al., 2012; Zhang & Zhang, 2024). However, from the reported word lists it can be observed that a mix of word classes have been used, but not systematically controlled. This results in difficulties in generalizing the present findings (Butler, 2019). To obtain a more complete understanding of L2 vocabulary learning, it is crucial to recognize the distinct properties that different types of words bring into the learning process.

2.2.2.5. The present studies

Given that word learning outcomes are dependent on a complex interaction between multiple factors (Tonzar, 2009), the present experiments set out to investigate the effects of modality of learning, modality of testing and word class on learning of ambiguous and non-ambiguous L2 vocabulary. In the two experiments presented below, participants learned new vocabulary (nouns, verbs and prepositions in a pseudo-language) by watching them appear paired either with text modality or with picture modality.

In accordance with theories and models, such as the Paivio's Dual Coding Theory (Paivio and Csapo, 1973) and the Revised Hierarchical Model (Kroll and Stewart, 1994), it was expected that pictures would bring a general advantage to the vocabulary learning outcomes. However, given that this advantage was not always consistent in previous studies (Bisson, 2015), we hypothesised that it would vary as a function of word class, testing modality and word ambiguity. Specifically, word class was expected to modulate the outcomes, especially considering that nouns as a word class tend to be easier to represent as a picture, whereas prepositions are more dependent on the learner tracking the pattern of word co-occurrences (Coventry & Guijarro-Fuentes, 2008). Finally, testing in a modality congruent with the learning modality was expected to result in better learning outcomes, in line with the encoding specificity principle (Tulving & Thomson, 1973).

2.2.3. Experiment 1

The first experiment investigated the effects of learning modality on L2 vocabulary acquisition. Specifically, we examined how pairing pseudowords with either picture or text stimuli influenced recall and recognition across various word classes (nouns, verbs, and prepositions), word ambiguities and testing modalities, fully crossing learning and testing modalities.

2.2.3.1. Methodology

2.2.3.1.1. Participants

Participants were monolingual English-speaking adults aged 18 to 30, who had either normal vision or corrected-to-normal vision, and no known learning disabilities or hearing issues. Recruitment took place via online recruitment platforms. Participants enlisted via the university's internal recruitment system (Sona Systems, www.sona-systems.com) and received course credits as per the university's policy, whereas individuals joining through the Prolific platform received financial compensation in line with Prolific's guidelines. The target sample size (n=160) was determined in accordance with recommendations in Brysbaert and Stevens (2018), as well as by using G*Power software to calculate the sample size for power of 80%, alpha= 0.05 and conservative effect size of d=.25.

2.2.3.1.2. Design

Participants learned pseudowords in an online experimental setting (Gorilla Experiment Builder; www.gorilla.sc). They were first randomly divided into either picture modality or text modality groups for learning. Each of these two groups were further randomly divided into either picture or text modality for testing, resulting in four between-participant groups (Figure 9).

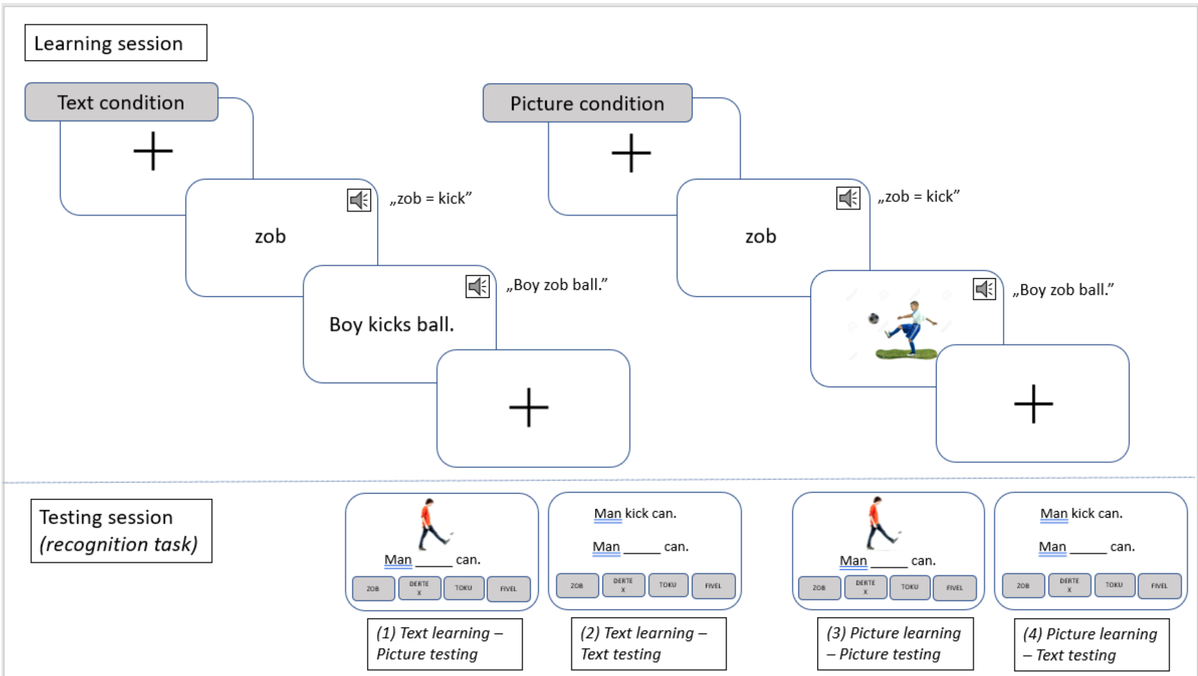

Figure 9. Figure showing the sequence of trials comparing text and picture learning conditions. The final row indicates four between-participant groups that result from crossing picture and text modalities at learning and testing.

Two variables relating to the properties of the words were manipulated within participants. The word class variable had three levels (nouns, verbs and prepositions) and the ambiguity variable had two levels (translation ambiguous and translation unambiguous).

The total design was 2x2x3x2 mixed factorial design (learning modality x testing modality x word class x ambiguity).

2.2.3.1.3. Experimental stimuli

Word stimuli. Twenty words from each of the three word class categories (nouns, verbs, prepositions) were carefully selected. Specifically, the nouns selected were concrete nouns, the verbs were restricted to action verbs and prepositions chosen were spatial prepositions. Half of the words from each group had an unambiguous, one-to-one relationship between the English word and the pseudoword. The main condition for selecting the other half of the words was that they are found to be cross-linguistically ambiguous between English and another language (e.g. the concept of the English preposition *on* is differently represented in Dutch, where there are three different words for *on* (*aan, op* and *om*, depending on the type of support provided to the object – vertical or horizontal), meaning that learning those words for the English speaker would be different because of the need to establish a new distinction, as oppose to learning a word *apple,* which translates unambiguously between the two languages). Another condition was that the distinction could be represented within a short sentence and/or an image. Finally, the words were placed in the context of five different sentences, which was particularly important for the ambiguous condition because they enabled a clear representation of the distinction between the two meanings.

Pseudoword stimuli. Pseudowords were taken from the English Lexicon project database (Balota et al., 2007) and then split into two groups and balanced on their length, number of orthographic neighbours, sum of the bigram count and mean reaction time by using

LexOPS R package (Taylor et al., 2020). This was done to ensure that ambiguous and unambiguous pseudowords are equal in their psycholinguistic properties.

The length of the pseudowords was different between the word class categories to reflect the typical phonological characteristics of a specific word class, which was shown to affect language processing (Farmer et al., 2006; Monaghan et al., 2010) and is particularly important in artificial language learning studies (Monaghan et al., 2011). Studies have shown that closed class words such as prepositions had fewer syllables than open class words such as nouns and verbs (Kelly, 1992; Morgan et al., 1996). Likewise, nouns are found to be longer than verbs (Kelly, 1992; Morgan et al., 1996). To reflect this, the average length of prepositions was 3.7 letters, with a range between 2 and 4 letters. The length of verbs was an average of 4.6 letters, with a range of 4 to 6 letters and the average length of nouns was 5.6 letters, with a range of 4 to 7 letters per word. The whole list of words and pseudowords can be found in the Appendix A.

Audio and picture stimuli. To create the auditory stimuli, Google Voice software was used. Using the text-to-speech technology, sentences and words were created using a voice of a native British English speaker. Images representing relevant sentences were found using Google's search engine filtering images under Creative Commons Licence and platforms for sharing copyright-free images (Freepik, Pexels, Unsplash).

2.2.3.1.4. Procedure

The experiment took place over the course of 8 days. The first day involved only a learning session, whereas the following day included both a learning and a testing session. The third session took place after a week (day 8) and included a delayed testing session, since in the previous literature the modality difference did not always have an effect on immediate learning outcomes, but would sometimes enhance the encoding process for the longer term learning outcomes. The third session also included verbal and visuospatial tests (Mill Hill

Vocabulary Test, MHVT; Raven & Court, 1998, The Matrix Reasoning test, MR; Wechsler, 2009) to control for individual differences.

Initially, participants were briefed about taking part in an experiment about learning words in a new language. They received information regarding the experimental setup, including its duration and structure, which encompassed both learning and testing phases over the course of 8 days. At the beginning of the experiment, participants were presented with a list of the pseudowords to be learned and asked to read through and retype them. The goal of this familiarization task was to facilitate learning outcomes during the following learning session. Before starting the learning part of the experiment, participants engaged in practice trials which demonstrated what the learning session will look like. The items in the practice trials were not included in the learning sessions and they consisted of one example from each word class category.

During the learning sessions, participants saw text showing the target pseudoword with a simultaneous auditory presentation of the pseudoword read out loud and followed by its translation to English (as in previous studies; Macedonia et al., 2011; Morett, 2014). After the initial display, the word would appear embedded in context. The context would be a short sentence of 3-4 words that would contain the pseudoword. The sentence would be read out loud, as in the first display, by an artificially generated voice. The audio of sentence would appear either paired with a text or with a picture reflecting the content of the sentence (see Figure 9).

The learning phase of the experiment took approximately 40 minutes, with each item presented for a total of 6.5 seconds (2500ms for the word and the translation and 4000ms for the word within a sentence), with two 2-minute breaks. Each word appeared five times, and each time embedded within a different sentence. The order in which the words would appear was randomised. At the end of the first learning phase, the participants were instructed to come back the next day at the same time (after 24 hours have passed).

The learning phase for the second day was the same as in the first day. Additionally, the second day introduced a testing session. The testing phase took place immediately after the learning phase and comprised two tasks - one targeting production of newly learned words and one assessing their recognition. Before the beginning of each task, participants received instructions and one trial task. In the production task participants were shown the context (picture or text, depending on the group) and below a sentence that was missing a target word. They were asked to name the pseudoword by typing their answer into the bracket below the testing stimuli. In the recognition task, instead of writing down the response, participants chose between four possible options (see Figure 9). The incorrect options were the other pseudowords that were included in the learning session.

The outcome measure was accuracy score (binary: correct/incorrect). To account for potential (minor) spelling mistakes in the production tasks, answers which contained a spelling error of one Levenshtein distance from the correct answer —meaning they had one letter added, substituted, or omitted—were also considered correct.

2.2.3.2. Analysis

Data in both experiments were analysed using generalized linear mixed-effects models with the logit link function and binomial distribution. Model specification was hypothesis driven to reflect research questions and experimental conditions to avoid multiple testing and overfitting (Barr et al., 2013; Meteyard & Davies, 2020). Learning modality, testing modality, word class and word ambiguity and their interactions were included as fixed effects and words and participants intercepts were included as random effect factors to account for the repeated observations of the individuals and items. The four-way interaction was not included due to the convergence issues. Contrast coding method used was deviation coding to facilitate the interpretation of potential main effects. P-values were estimated using the significance criterion of $\alpha < 0.05$. Separate analysis was conducted for each testing task and session.

2.2.3.3. Results & Discussion

Overall performance across the testing times and testing tasks can be seen in Table 2. Both in the immediate and in the delayed testing task participants demonstrated positive learning outcomes. In the recognition task they scored well above the chance level (0.25) and in the production task they gave correct answers to approximately one third of the tasks.

Table 2. Proportion of accurate responses on each of the testing tasks. Standard deviations are shown in the brackets.

	Immediate	Delayed
Recognition	0.59 (0.49)	0.58 (0.49)
Production	0.35 (0.48)	0.29 (0.45)

Table 3. Table showing the overview of significant results across different testing tasks and testing times in Experiment 1.

Significant effects	Recognition task		Production task	
	Immediate	Delayed	Immediate	Delayed
Learning modality	n.s.	n.s.	n.s.	n.s.
Ambiguity	$\chi2$ =60.39, p<.001	$\chi2$ =57.99, p<.001	$\chi2$ =25.82, p<.001,	$\chi2$ =15.71, p<.001
Word class	$\chi2$ =14.2, p<.001	$\chi2$ =13.40, p=.001	$\chi2$ =34.87, p<.001	$\chi2$ =43.69, p<.001
Testing modality	$\chi2$ =5.88, p=.015	n.s.	n.s.	n.s.
Learning modality*Ambiguity	$\chi2$ =7.98, p=.005	$\chi2$ =9.47, p=.002	$\chi2$ =28.12, p<.001,	$\chi2$ =28.5, p<.001
Learning modality*Word class	n.s.	n.s.	$\chi2$ =13.63, p=.001,	$\chi2$ =11.33, p<.001
Ambiguity*Testing modality	n.s.	$\chi2$ =5.05, p=.025	n.s.	$\chi2$ =10.81, p=.001
Wordclass*Testing modality	$\chi2$ =36.09, p<.001	$\chi2$ =46.52, p<.001	n.s.	$\chi2$ =14.9, p<.001
Learningmod.*Ambiguity*Test. mod.	$\chi2$ =26.9, p<.001	$\chi2$ =51.48, p<.001	$\chi2$ =12.83, p<.001	$\chi2$ =3.98, p=.046

Word class*Ambiguity*Testing mod.	n.s.	X^2=11.99, p=.002	n.s.	n.s.

2.2.3.3.1. Recognition: Immediate

The model summary of the mixed-effects regression can be seen in Appendix A, with the main significant results presented in

Table **3**. Analysis of the learning data showed no significant main effect of learning modality. However, significant main effects of ambiguity, χ^2 =60.39, p<.001, word class, χ^2 =14.2, p<.001, and testing modality, χ^2 =5.88, p=.015, were found. Non-ambiguous words (M=0.716, SD=0.451) were learned more successfully than ambiguous words (M=0.455, SD=0.498) and nouns (M=0.665, SD=0.472) were learned better than prepositions (M=0.553, SD=0.497) and verbs (M=0.538, SD=0.499), consistent with the results of previous studies (Degani & Tokowicz, 2010; Van Hell & de Groot, 1998). Overall, testing modality showed an advantage of testing with text (M=0.613, SD=0.487) over testing with pictures (M=0.558, SD=0.497). However, all three predictors were also involved in higher order interactions, indicating that it is the interplay between these variables that is important.

The relationship between learning modality and ambiguity showed that the learning modality did not have an effect when testing within the same category. Learning ambiguous words with text or with picture led to effectively same results, and the same was found for unambiguous words. (The difference between learning modalities was only present when comparing modalities across different levels of ambiguity, and therefore isn't informative).

On the other hand, different testing modalities had a significant effect within the same category of word class (Figure 10). A closer inspection of the interaction between word class and testing modality showed that the testing outcomes were better when both verbs and prepositions were tested with text (M=0.58, SD=0.494; M=0.585, SD=0.493) than with picture modality (M=0.497, SD=0.5; M=0.52, SD=0.5).

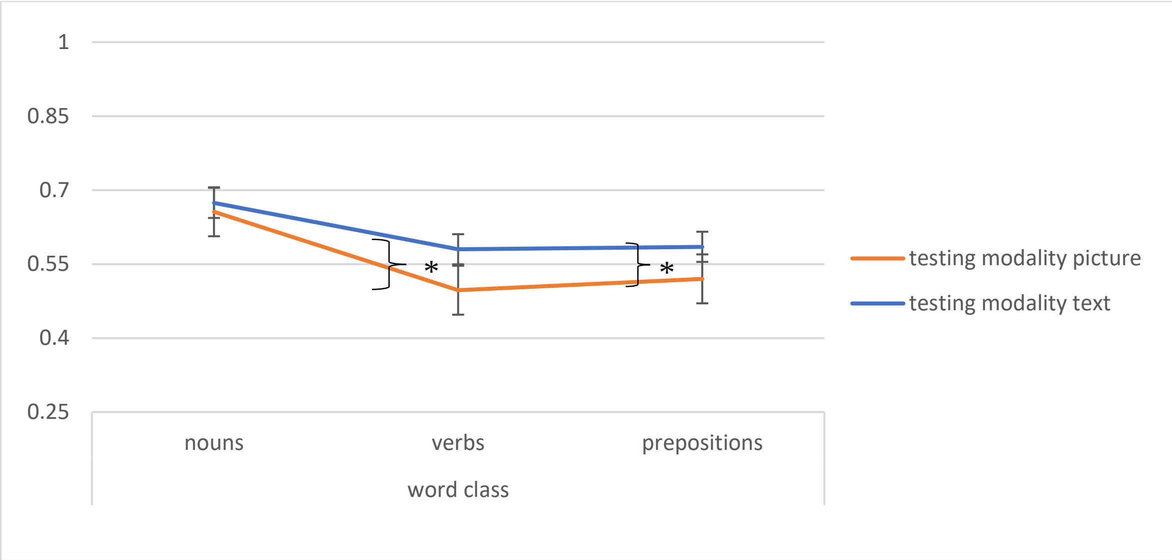

Figure 10. Plot showing the interaction between testing modality and word class in the immediate recognition task. The baseline on the graph is set at 0.25 to indicate chance level performance (25%) on the recognition task.

Finally, a significant three-way interaction was observed between learning modality, testing modality and ambiguity variables (χ^2 =26.9, p<.001) (Figure 11). A post-hoc comparison and estimated marginal means were used to establish the nature of the interaction. To begin with, significant differences were present only in the unambiguous group of words, as can be seen in the plot. The analysis showed that the group that learned in the picture modality and was tested with text (picture-text (M=0.796, SD=.044)) performed significantly better than both of the groups in congruent modalities (picture-picture, M=0.688, SD=.057; text-text, M=0.711, SD=.055). This interaction shows the crucial effect of the testing task on the learning outcome. When both groups were tested with the text-task, the difference was in favour of the picture-learning modality. However, if the groups were both tested with the picture-task, there was no significant difference, suggesting again that the difference between the learning modality depends on the testing modality. However, the influence is not driven by the congruency between the learning and testing modality as expected. Instead, the best performance was observed in the incongruent picture-learning/text-testing group.

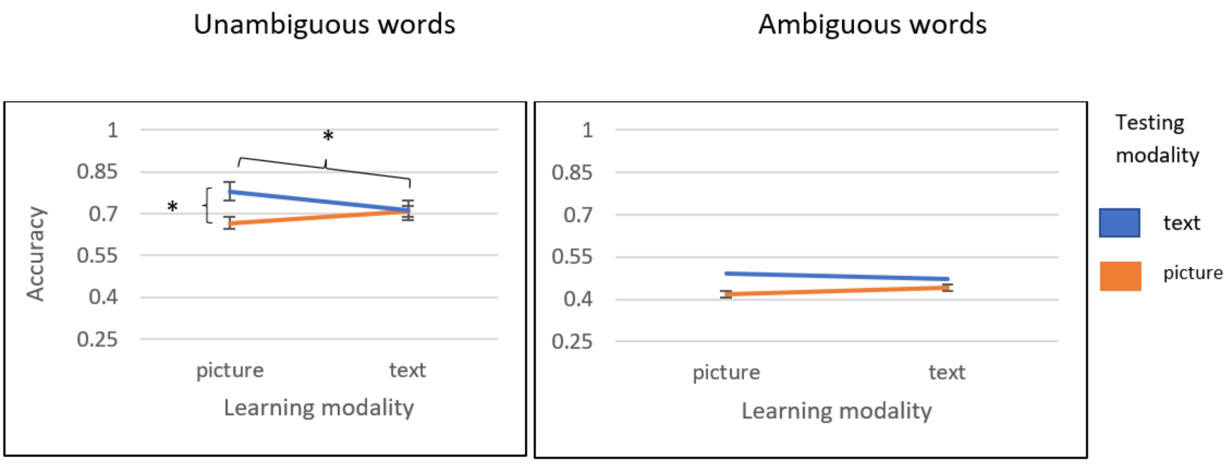

Figure 11. Plot showing the interaction between learning modality, testing modality and ambiguity in the immediate recognition task. The baseline on the graph is set at 0.25 to indicate chance level performance (25%) on the recognition task.

2.2.3.3.2. Recognition: delayed

The analysis of the delayed recognition task was conducted with the same model (resulting in the summary output shown in the

Table **3** and the Appendix A). As in the immediate recognition task, there was no main effect of the Learning modality, but there were significant main effects of ambiguity and word class in the same direction as before. Unambiguous words (M=0.715, SD=0.498) were learned better than ambiguous words (M=0.454, SD=0.498) and nouns (M=0.663, SD=0.473) were learned better than prepositions (M=0.555, SD=0.497) and verbs (M=0.535, SD=0.499). There was no main effect of testing modality.

In addition to the same two-way interaction between ambiguity and learning modality identified in the first analysis, a two-way interaction between ambiguity and the Testing modality was found to be significant. This time, the text modality at testing (M=0.739; SD=0.439) showed an advantage compared to the picture modality (M=0.69, SD=0.463), but only in the unambiguous condition.

These interactions were involved in a higher-level interaction, involving learning modality, testing modality and ambiguity (X^2=51.480; p<.001), with the same pattern present as for immediate recall. The differences were found only in the unambiguous words where again the effect of the picture learning modality was modified by the modality of the testing task. This resulted in the picture-learning group that was tested with text (M=0.779, SD=0.415) scoring higher (β=0.483, p=<.001) than the picture-learning group that was tested with pictures (M=0.665, SD=0.472). The difference was also found between the picture-text and text-text groups (M=0.715, SD=0.452), where again the picture-text condition resulted in higher scores (β=1.733, p=0.005).

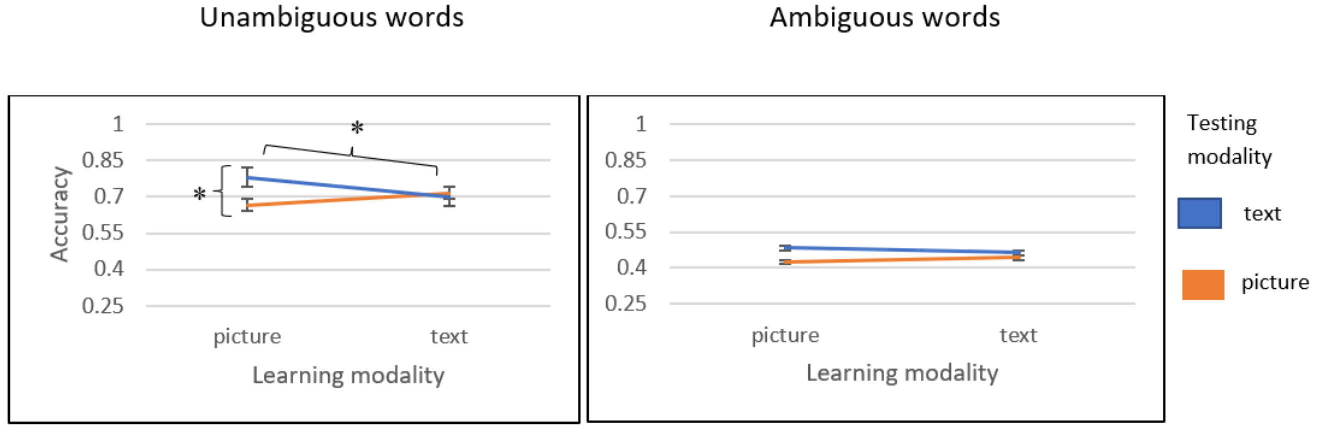

Figure 12. Plot showing the significant three-way interaction between learning modality, testing modality and ambiguity in the delayed recognition task. The baseline on the graph is set at 0.25 to indicate chance level performance (25%) on the recognition task.

In line with the results from the immediate recognition task, the interaction between word class and testing modality was significant. The results again demonstrated a positive effect of the text modality at testing on verbs and prepositions. However, this time they were further modified by ambiguity in a significant three-way interaction between ambiguity, word class and testing modality, χ^2=11.999, p=.002. The significant interaction of interest appeared only in the unambiguous word group (see Figure 13), between the preposition tested with text (M=0.694, SD=0.461) and prepositions tested with pictures (M=0.626, SD=0.484). The scores for unambiguous preposition learning were higher when tested with text modality than with the picture modality (β=0.784, p=0.042). The same was found for unambiguous verbs (β=0.784, p=0.012) where the verbs tested with text (M=0.694, SD=0.461) showed an advantage over the verbs tested with pictures (M=0.603, SD=0.489).

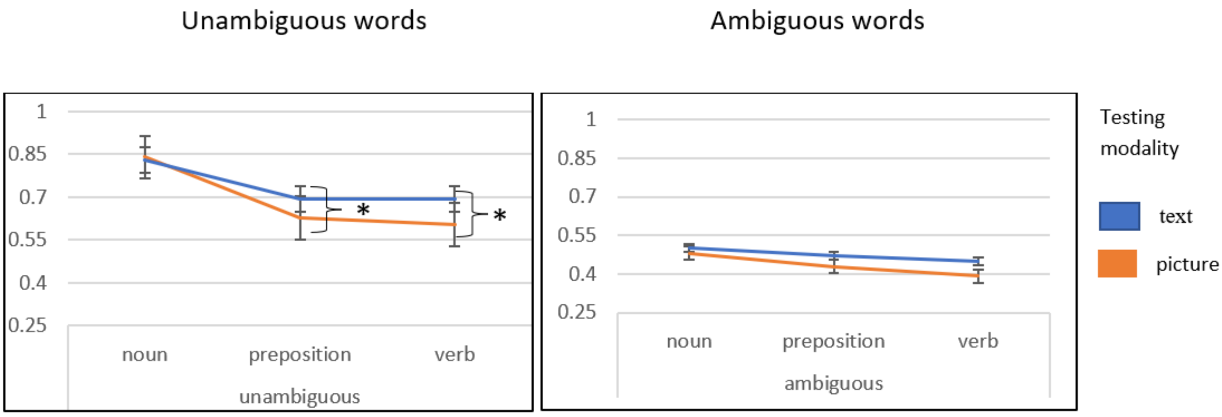

Figure 13. Plot showing the significant three-way interaction between testing modality, word class and ambiguity in the delayed recognition task. The baseline on the graph is set at 0.25 to indicate chance level performance (25%) on the recognition task.

2.2.3.3.3. Production: immediate

The results of the model with the outcomes of the immediate production task are shown in the

Table **3**. As expected, overall accuracy results are lower than in the recognition task (Table 2), but the pattern of the results remained the same. Significant main effects of word class and ambiguity were found, with the same patterns of nouns having a higher mean score (M=0.464; SD=0.499) than verbs (M=0.323; SD=0.448) and prepositions (M=0.323; SD=0.468), as well as unambiguous words (M=0.421; SD=0.494) having a higher mean score than ambiguous words (M=0.289; SD=0.454). No significant main effect of learning or testing modality was found.

Two-way interactions between testing modality and ambiguity, and testing modality and word class that were found in the recognition tasks, were not found in this analysis. However, there was a significant interaction between learning modality and word class (χ^2=13.63, p=.001), but it was driven by the already established differences between nouns and other two word classes. Similarly, the significant learning modality-ambiguity interaction was centred around the difference between ambiguous and unambiguous words.

However, the same three-way interaction between learning modality, testing modality and ambiguity again reached significance, χ^2=12.827; p<.001. The picture-text group (M=0.489, SD=0.500) had a significantly higher score (β=0.483, p=<.001) than the text-picture group (M=0.414, SD=0.493) when learning unambiguous words. The results are shown in the plots in Figure 14, which follows a similar pattern as in the two recognition tasks (see Figure 11 and Figure 12).

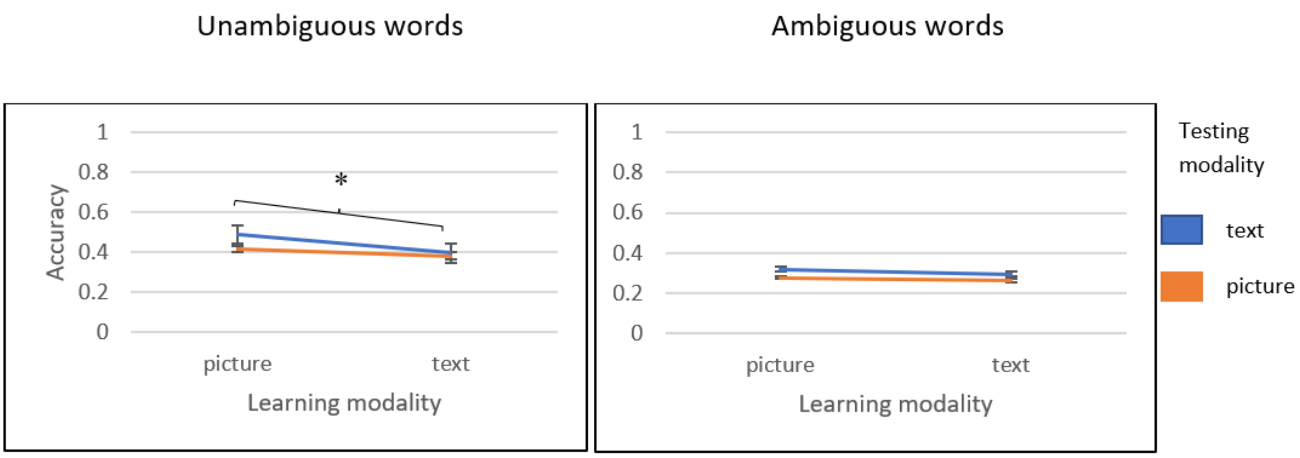

Figure 14. Plot showing the significant three-way interaction between learning modality, testing modality and ambiguity in the immediate production task

2.2.3.3.4. Production: delayed

The final analysis examining the results of the delayed production task largely again followed the same pattern. There was again no main effect of the learning modality, as opposed to the main effects of word class and ambiguity which were again present.

Again, the effects in the two-way interactions between learning modality and ambiguity, and learning modality and word class were driven by the differences between different word classes and different levels of ambiguity.

Similarly as before, a three-way interaction was found between learning modality, testing modality and ambiguity. However, this time the interaction was only marginally significant, χ^2=3.975, p=0.046) and the post-hoc analysis did not reveal any significant differences of interest, other than the expected differences between the unambiguous and ambiguous word learning outcomes (Appendix A).

2.2.3.3.5. The analysis of individual differences

To control for individual differences, verbal and visuospatial tests (Mill Hill Vocabulary Test, MHVT; Raven & Court, 1998, The Matrix Reasoning test, MR; Wechsler, 2009) were analysed. Due to the complexity of the models, this data was not fitted as fixed factors, but analysed separately to test for group differences using a one-way ANOVA. Groups did not differ significantly on verbal scores, $F(3, 156) = .898$, $p = .444$, indicating comparable baseline characteristics between groups. Analyses of the visuospatial abilities revealed a significant difference between groups, $F(3, 156) = 2.923$, $p = .036$. To investigate whether this difference affected previous results, this measure was included as a covariate in the main analysis. The analysis showed that across all four analyses all of the primary results remained unchanged (coefficients varied by <0.1%), indicating that pre-existing group differences in visuospatial ability did not confound the experimental findings. The detailed descriptives values and analyses outputs can be found in the Appendix A.

2.2.3.3.6. Summary of results for Experiment 1

This experiment tested the second language vocabulary learning outcomes, while considering underexplored variables of word class, testing modality and ambiguity to enable a more thorough examination of the learning modality effect.

Across all four analyses consistent main effects of word class and ambiguity emerged. The pattern of results showed that nouns were easier to learn than verbs and prepositions, and unambiguous words were easier to learn than ambiguous words. Also consistently, there was no main effect of modality. Instead, the modality effect was always expressed through interaction with other variables, as will be summarised here.

A consistent pattern was found between testing modality and word class in both recognition tasks, confirming the prediction that different word classes benefit from different modalities. The analysis showed that the modality of testing had a significant impact on the learning scores of prepositions and verbs, but not of nouns. In both cases, testing with text showed an advantage compared to testing with pictures. This corresponds to the approaches that posit that different word classes are grounded and processed in different ways (Bultena et al., 2013; Markostamou & Coventry, 2022; Zhang et al., 2013). Current results speak in favour and further extend them to a second language learning context. The pattern of these results is in line with the work explaining how concrete words such as nouns are more visually grounded, whereas more abstract word classes, such as verbs and prepositions are more linguistically grounded (Bird et al., 2000; Borghi et al, 2019; Gleitman et al, 2005).

A significant three-way interaction between learning modality, testing modality and ambiguity was found in three out of four analyses conducted. It showed that the condition where the picture modality is presented during learning and text modality during testing leads to significantly better results with unambiguous words. The same pattern was not observed for the ambiguous words. Although confirming the prediction regarding the importance of the testing modality, the direction of this effect is unexpected. Having a combination of pictures at learning and text at testing lead to best outcomes goes against the encoding specificity principle (Tulving & Thomson, 1973), which would predict that congruency between learning and testing would bring an advantage. However, when comparing pictures and text modalities at learning while using only text modality to test (as in the majority of previous studies (see Table 1)), picture modality emerges as superior, as seen in immediate and delayed recognition tasks. This is in line with previous research which tested only with text modality (Bates & Son, 2020; Boddaert et al., 2021; Carpenter & Geller, 2020;). Still, including the picture modality at testing in the present experiment showed that the advantage of picture as a learning modality disappears. These findings suggest that the background processes affecting L2 vocabulary learning outcomes are more nuanced.

The findings so far consistently underline the importance of the testing modality and word class in understanding second language vocabulary learning processes, often overlooked in past research. However, the unexpected direction of the congruency effect within the learning modality-testing modality interaction motivates the next experiment to better understand the influence of testing modality.

2.2.4. Experiment 2

One of the characteristics of the first experiment was that the examples of retrieval cues during testing (textual and visual contexts used along the target words) were not the same as during learning. The purpose of this was to avoid participants merely pairing words with a particular image without processing its meaning, especially important given that ambiguous words were a part of the testing paradigm. Still, we expected to find the effect of the modality congruence, as we did not find previous research that would suggest the modality congruence effect to be separable from the cue novelty between learning and testing. However, the advantageous effect of modality congruence was not found. One possible reason is that the modality congruence advantage observed in previous research was driven by the cues being identical, but not necessarily because of the same modality. To tease these effects apart, the second experiment included cue novelty as a variable by including both identical and different cues at testing. The second goal was to test if the findings from the first experiment replicate.

At the same time, the ambiguity variable was removed because all the effects in the first study were present only in the unambiguous condition. The reason for that might be much greater processing demands in encoding that requires creating new concepts (Barcroft & Sunderman, 2008), which resulted in overall lower learning scores and possibly obscured the expected effect. Therefore, only the words with a direct (one-to-one) relationship between the L1 and L2 concepts were retained. Other aspects of the study remained the same to ensure comparability between studies.

2.2.4.1. Materials and methods

2.2.4.1.1. Participants

The study recruited 162 participants under the same constraints as the previous study and from the same recruiting platforms.

2.2.4.1.2. Design

The design was the same as Experiment 1, with learning modality, testing modality and word class as variables. However, in this experiment ambiguity variable was replaced by cue novelty as the fourth variable (consisting of two levels, “new” and “old”). The total design was 2x2x3x2 mixed factorial design (learning modality x testing modality x word class x cue novelty).

2.2.4.1.3. Experimental stimuli

The stimuli remained the same for the learning condition, with the different arrangement between the learning and testing stimuli to account for the new manipulation described above. For half of the participants the testing stimuli remained as they were in the first study (different examples than in the learning stimuli) and in the other half the testing stimuli were the same as during learning session. Due to removing the ambiguity condition, the number of words to be learned was smaller than in the previous experiment, as only unambiguous words were kept. This resulted in a number of 10 per word class, for a total of 30 words.

2.2.4.1.4. Procedure

Due to the smaller number of words to be learned, the sessions in this experiment were shorter, taking around 30 minutes each day. Other than that, the procedure remained the same.

2.2.4.2. Results & Discussion

Overall performance across the testing times and testing tasks is shown in Table 4. Both in the immediate and in the delayed testing task participants demonstrated positive learning outcomes. The scores are higher compared to the previous experiment where ambiguous words were involved.

Table 4. Proportion of accurate responses on each of the testing tasks. Standard deviations are shown in the brackets.

	Immediate	Delayed
Recognition	0.81 (0.39)	0.75 (0.46)
Production	0.52 (0.5)	0.30 (0.46)

As in the first experiment, the main (significant) results are presented in a tabular form for ease of understanding and overview (Table 5).

Full analyses outputs (with both significant and non-significant results) can be found in the Appendix A. Consistent with the previous experiment, there was no overall main effect of the learning modality for any of the individual analyses. Likewise, the main effect of word class was always present, indicating that nouns are learned better than verbs and prepositions, in line with previous research.

Table 5. Table showing the overview of significant results across different testing tasks and testing times in Experiment 2

Significant effects	Recognition task		Production task	
	Immediate	Delayed	Immediate	Delayed
Learning modality	n.s.	n.s.	n.s.	n.s.
Word class	$\chi2$ =53.33, p<.001	$\chi2$ =42.84, p<.001	$\chi2$ =13.05, p<.001	$\chi2$ =22.23, p<.001
Learning modality*Testing modality	$\chi2$ =4.34, p=.037	$\chi2$ =7.05, p=.008	n.s.	n.s.
Learning modality*Word class	n.s.	n.s.	$\chi2$ =6.38, p=.041	n.s.
Word class*Testing modality	$\chi2$ =14.09, p<.001	n.s.	$\chi2$ =7.45, p=.023	$\chi2$ =8.51, p=.014
Word class*Cue novelty*Testing modality	$\chi2$ =9.43, p=.009	n.s.	n.s.	$\chi2$ =6.21, p=.045

2.2.4.2.1. Recognition: Immediate

In the analysis of the immediate recognition scores several interactions were found to be significant (see Appendix A). A significant two-way interaction was detected between learning and testing modality. However, the post-hoc analysis revealed only marginally significant differences (p=0.062), which will not be reported.

There was also a significant two-way interaction between testing modality and word class, with a further higher order three-way interaction with cue novelty, χ^2=9.426, df=2, p=.009. The analysis revealed that the interaction between testing modality and word class was present only in the new cue condition, but not in the old cue condition (see Figure 15). When the cue was novel, the scores for prepositions were significantly higher (β=0.411, p = .011) when tested with text (M = 0.774, SD = 0.419) than with pictures (M = 0.662, SD = 0.473).

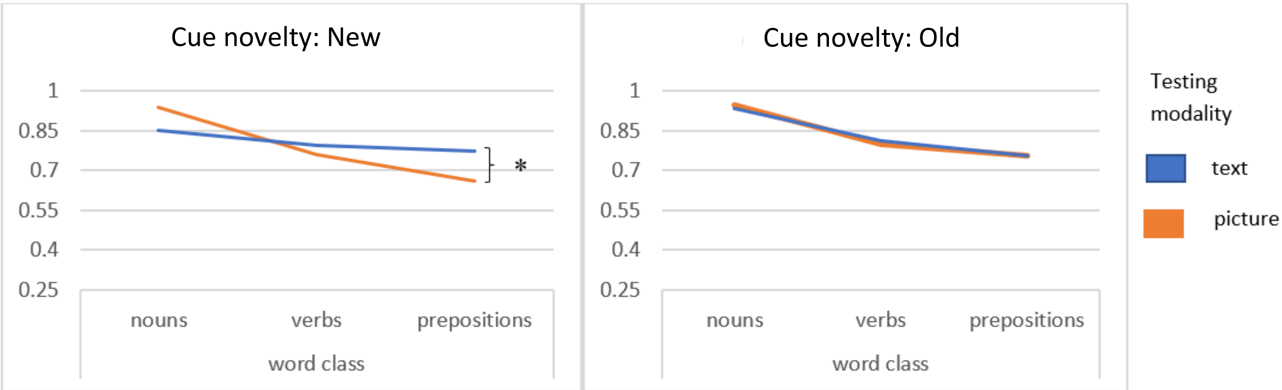

Figure 15. The interaction between testing modality, cue novelty and word class in the immediate recognition task. The baseline on the graph is set at 0.25 to indicate chance level performance (25%) on the recognition task.

2.2.4.2.2. Recognition: Delayed

A significant two-way interaction between the learning modality and the testing modality was found in the delayed recognition task, χ^2= 7.051, df=2, p=.008 (Figure 16; for details see Appendix A). Further analysis showed the difference between the picture-picture (picture modality at learning and picture modality at testing; M = 0.784, SD = 0.411) and text-picture (text modality at learning and picture modality at testing; M = 0.699, SD = 0.459) conditions, as well as between text-text (M = 0.799, SD = 0.401) and text-picture modality (M

= 0.699, SD = 0.459) combinations of modalities, showing that in this case the congruency between the modalities led to a significant advantage (β = 1.926, p = .023) in both cases (β = 0.470, p = .009).

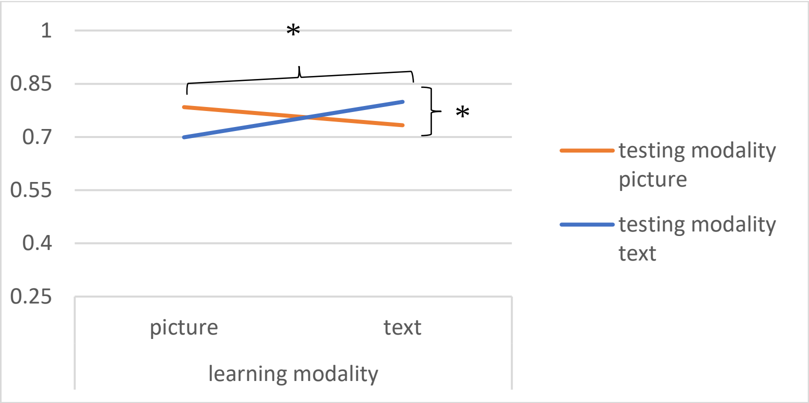

Figure 16. Plot showing the interaction between learning and testing modalities in delayed recognition task. The baseline on the graph is set at 0.25 to indicate chance level performance (25%) on the recognition task.

2.2.4.2.3. Production: Immediate

The analysis of the immediate production task showed significant interactions between learning modality and word class (χ^2= 6.382, df=2, p=.041), as well as between testing modality and word class (χ^2=7.545, df=2, p=.023). However, post-hoc analysis identified that the significant differences were driven by the expected differences between word classes (discussed earlier).

2.2.4.2.4. Production: Delayed

The analysis of the delayed production condition produced a significant three-way interaction between testing modality, cue novelty and word class, χ^2= 6.214, df=2, p=.045 (Figure 17). There was a significant difference in the learning outcomes for two of the three word classes depending on the testing modality. Firstly, the learning outcomes for prepositions tested with the novel (different) cues were significantly higher (β=0.373, p=.007) when tested with text (M=0.339, SD=0.474) than with pictures (M=0.185, SD=0.389). The same difference was found in verbs (β = 0.459, p = 0.034) where again text as a testing modality (M=0.321, SD=0.467) outscored pictures (M=0.195, SD=0.397), repeating the trend found in the initial study.

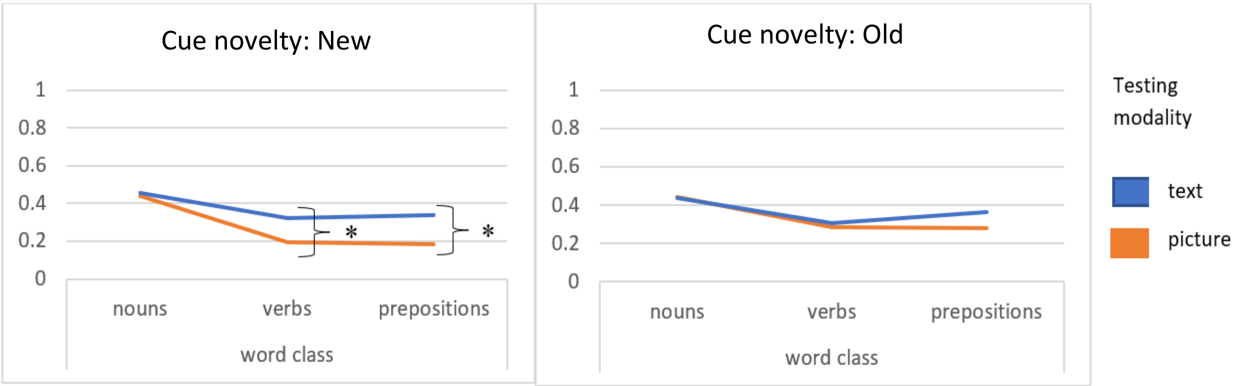

Figure 17. Plot showing the interaction between testing modality, cue novelty and word class in the delayed production data

2.2.4.2.5. The analysis of individual differences

To control for individual differences, verbal and visuospatial tests (Mill Hill Vocabulary Test, MHVT; Raven & Court, 1998, The Matrix Reasoning test, MR; Wechsler, 2009) were again analysed using a one-way ANOVA. Groups did not differ significantly neither on verbal scores, $F(3, 158) = 1.57$, $p = .199$, nor on visuospatial scores $F(3, 158) = 1.015$, $p = .388$, indicating that the groups were matched on these measures. The detailed values of descriptives and analyses outputs can be found in the Appendix A.

2.2.4.2.6. Summary of results for Experiment 2

The results partially confirmed the initial hypotheses. The first hypothesis was related to the congruency of modalities at learning and testing. Based on the results from the first experiment, it was assumed that in the condition where the cues are different between learning and testing, the picture learning – text testing combination would lead to the highest score, whereas in the condition where the cues are the same the congruent cues would have an advantage. However, there were no significant three-way interactions between the modalities at learning and testing and the cue novelty. This indicates that the novelty of the cues does not explain the effect seen in the first experiment. Instead, an interaction between Learning modality and Testing modality showed a general advantage of congruent conditions over the text-learning/picture-testing condition was found in the delayed recognition condition.

The second hypothesis proposing that verbs and prepositions would benefit more from the text modality at testing than the picture modality was confirmed. In the immediate recognition task, the text modality seemed to bring an advantage to prepositions in the different (novel) cue condition. In the delayed production task, the text modality at testing improved the scores for both verbs and prepositions compared to picture modality, thus further confirming the initial findings.

2.2.5. General Discussion

These experiments set out to investigate the effects of learning modality in second language vocabulary learning through inclusion of key variables often overlooked in previous research. A review of previous studies showing inconclusive results on the benefits of text versus picture modality on L2 word learning outcomes motivated a comprehensive examination of the interplay between learning modality, testing modality, word class and word ambiguity in L2 vocabulary learning. In two experiments learning outcomes were tested at two time points (immediate and delayed) and with two task types (recognition and production), resulting in four models per experiment. We first consider the results across the experiments before exploring implications for theories of L2 vocabulary learning.

Across all experiments that included ambiguity and word class variables, significant main effects of these variables were found, indicating that nouns are learned easier than verbs and prepositions, and that unambiguous words are learned easier than ambiguous words. These main effects confirmed previous findings both in L1 and L2 language research showing that nouns are learned easier than other word classes (Gentner, 1982, 2006; Ellis & Beaton, 1993; Mestres-Missé et al., 2010; Waxman et al., 2013), while ambiguous words are more difficult to learn than unambiguous words (Degani & Tokowicz, 2010; Tokowicz & Kroll, 2007).

As hypothesized, testing modality was found to be a significant factor that modifies the effect of the learning modality variable in unambiguous words in both experiments. Likewise, testing modality interacted with the word class variable, showing an advantage of text modality at testing for verbs and prepositions across both experiments, confirming both the roles of the testing modality and different word properties in the process of initial second language learning. The majority of these effects were expressed in the recognition tasks, although they were found in two production tasks as well. The reason for that could be in the overall lower learning scores in the production task, which possibly masked the effects.

However, some of the results were not consistent across experiments. When looking into the effect of the testing modality, it was observed that in the first experiment the incongruent picture-learning/text-testing combination led to the best learning outcomes, whereas in the second experiment the congruent condition was more advantageous for performance. Also, the learning modality-testing modality interaction appeared to be significant in only one of the models in the second experiment, whereas it was significant in three out of four models in experiment one.

In the following sections, we examine how these complex patterns of interactions between learning modality, testing modality and word class inform our theoretical understanding of L2 vocabulary learning mechanisms, as well as the implications for practice.

2.2.5.1. The influence of learning modality revisited

Across both experiments, none of the analyses showed a main effect of learning modality on the learning outcomes, indicating that no single modality is inherently superior at facilitating L2 word learning. Instead, the effect of learning modality was always expressed through interactions with testing modality, thus emphasizing its crucial role.

However, with respect to the direction of interactions involving learning and testing modalities, the results were not consistent across the two experiments. Rather than observing the expected advantage of modality congruence, as suggested by the encoding specificity principle (Tulving & Thomson, 1973) and related research (Lotto & de Groot, 1998; Jones, 2004), the incongruent combination of picture-based learning and text-based testing resulted in best learning outcomes in the first experiment. These results were consistent across three out of four analyses conducted. On the other hand, the second experiment did show a positive effect of congruency, with the text-text and picture-picture conditions resulting in highest learning scores. Still, this effect was only found in the delayed recognition task.

The results of the first experiment could be interpreted in line with the Revised Hierarchical Model (Kroll & Stewart, 1994), which posits that in the initial stages of L2 vocabulary learning, learning is mediated by the lexical connections to the L1 vocabulary, whereas the conceptual link is formed only as the learner becomes more proficient. Therefore, the activation path in beginner learners takes longer when using pictures as a cue because the information still has to go through the L1 lexicon in order to retrieve the target L2 word. This could potentially mean that while pictures can work well as a tool to enhance encoding by providing learners with more concrete and easily interpretable visual representation, in the very early stages of adult L2 learning they might not work well as a cue for retrieval. In this particular group retrieval might still be guided more strongly by the lexical connections, which is in this study reflected in the accuracy scores. Therefore, staying within the linguistic system would be less demanding than retrieving the meaning via conceptual representations which are not fully formed in beginner learners yet (Chen & Leung, 1989; De Groot & Hoeks, 1995; Kroll & Curley, 1988). The results of the second experiment align with the initial hypotheses and the encoding specificity principle (Tulving & Thomson, 1973), which holds that memory retrieval is more successful when the learning and testing conditions match. However, the results were found in only one out of the four analyses conducted. This potentially indicates that the nature of the interaction between learning modality and testing modality plays a larger role in more challenging learning conditions, as the learners in the first experiment had more words to learn, as it also involved ambiguous words.

The mixed findings here are similar to observations from previous research on modality congruence effects between encoding and retrieval where congruency advantage was also not universally present (Nelson et al., 2005, Wolf, 2018). For example, Nelson et al. (2005) found that the congruency benefit depends on individual learners’ characteristic. Faster learners showed reduced modality dependence, while slower learners gained more from the modality congruence. Likewise, Wolf et al. (2019) failed to find congruency effects in their study investigating written and auditory modalities at L2 vocabulary learning. These results were explained as emerging from the fact that adult participants that are proficient readers automatically recode novel word forms both into their phonological and orthographic forms, regardless of the input stimuli (Wolf et al., 2019).

While it is clear from the results that the testing modality is emerging as an important factor in second language vocabulary learning (Jelani & Boers, 2018), more research is needed to fully understand its role.

2.2.5.2. How do different word types shape L2 vocabulary learning?

In contrast to the interaction between retrieval and modality effects across experiments, there were strong and consistent effects of word type. As hypothesised, different word types led to different learning outcomes. This was evident both with word ambiguity (Experiment 1) and with word class (Experiments 1 and 2), suggesting that different cognitive processes are engaged between ambiguous and unambiguous words and across different word classes. Ambiguous words have been shown to be significantly more difficult to learn compared to unambiguous words, due to the cognitive complexity involved in (re)creating concepts, as opposed to linking the L2 forms to existing concepts when learning unambiguous words (Barcroft & Sunderman, 2008; Jarvis & Pavlenko, 2009). The low learning scores in this study led to reduced variability that potentially masked the existing effects of other variables on learning of those words. Nevertheless, these results confirmed the inherent challenges of learning ambiguous words (Degani & Tokowicz, 2010), underscoring the need for further examination of the effects specific for learning this type of word.

Furthermore, this study highlights the advantage of text modality when testing prepositions and verbs, thus supporting the idea that different word classes are grounded and processed in distinct ways (Coventry & Garrod, 2004; Vigliocco et al., 2011; Zhang et al., 2013). Work on spatial prepositions (Coventry et al., 1994; Coventry et al., 2004) provides evidence that purely visual representations are insufficient for capturing the complexity of spatial terms, with use determines also by subtle distributional cues and linguistic usage patterns (Figure 1) that makes it particularly challenging to acquire native-type fluency for prepositions in an L2 (see fo example Damayanti & Sundari, 2022 for a recent review). Moreover, it has been proposed that different grounding of words stems from their levels of abstractness, with verbs and especially prepositions being more abstract compared to concrete nouns (Bird et al., 2000; Borghi et al., 2019). While concrete words rely more on sensory-motor experiences, abstract concepts activate the linguistic system. They often need to be understood in terms of the relationship they describe (e.g. actions or spatial relations), rather than through perceptual grounding. As a result, Borghi et al. (2019) propose that these word types may be better processed through linguistic cues, which are better supported by text-based modality during testing, thus explaining the present results. Since word class effects have rarely been studied in L2 vocabulary learning research, only a few studies so far have investigated the relationship between word class and learning modality (García-Gámez & Macizo, 2019; Rosenthal-von der Putten & Bergmann, 2020). The present study is the first to our knowledge that has examined and found interactions between word class and testing modality in L2 word learning. This confirms that not only do learning modalities differently affect word encoding across word classes as previously found (Rosenthal-von der Putten & Bergmann, 2020), but this is the case for the word retrieval process as well.

In the broader context, these results contribute to the debate between distributional and embodied accounts. While distributional approaches emphasize the role of statistical patterns and regularities in how words co-occur in language input, the embodied approach posits that language learning is deeply connected to sensory and motor experiences, arguing that understanding and producing language involves simulating physical actions and perceptual states. The importance of word class in the present findings is in line with hybrid approaches discussed by Andrews et al. (2009; 2014), and further debated through the Words as Social Tools theory (WAT) (Borghi & Binkofski, 2014) and Language is an Embodied Neuroenhancement and Scaffold (LENS) (Dove et al., 2020), where there is integration of distributional and the embodied approaches by recognizing the role of both the sensorimotor and linguistic networks during language processing, as well as the word-level properties. For instance, concrete nouns might activate embodied representations more strongly, while abstract words, such as prepositions and verbs, engage linguistic processing. The involvement of each network depends on the type of mediation and representation required for a specific word. However, although the hybrid approaches point to relevant factors in L2 word learning, their application to the field of L2 vocabulary research is still rare.

Moving forward, we suggest that current models of L2 vocabulary learning should be modified to accommodate effects of word class and testing modality on performance. Building on word class findings, we suggest that different word classes involve both grounded and distributional components that both need to be learned, but that prepositions have a higher weighting for distributional information at one end of a continuum with nouns at the other end (see Figure 1). Furthermore, at retrieval, word classes with the biggest distributional weights are easier to access when tested with distributional (text cues), and hence performance is better for these terms (compared to nouns).

This speaks to modification of existing models, such as the Revised Hierarchical Model (RHM) (Kroll & Stewart, 1994) mentioned above. This model served as a foundation for investigation of encoding and retrieval processes across many studies of multimodality in L2 vocabulary learning (Comesaña et al., 2009; García-Gámez & Macizo, 2019). Its initial version was further developed to account for the effect of pictures at input (Yoshii, 2006) as well as cross-linguistical ambiguity (Eddington & Tokowicz, 2013; Pavlenko, 2009). We propose that the present findings could further extend the model account for the effect of different word classes and testing modalities, which were found to affect L2 word learning outcomes. The underlying concepts for words consist of both distributional and embodied information, but for some words distributional information is more informative than for others (Borghi &

Binkofski, 2014; Bultena et al., 2013; Markostamou & Coventry, 2022; Pulvermüller et al, 2010; Vigliocco et al., 2011; Zhang et al., 2013), which has implications for the L2 language retrieval, as demonstrated in the present studies. Similarly as Pavlenko (2009) in her Modified RHM model assigned different weights to different concepts/representations of words based on cross-linguistical ambiguity, we propose a similar differentiation, but based on word class. In accordance with that, the use of different testing cues at retrieval (text or picture) results in varying retrieval outcomes for specific words, depending on whether they are more embodied or distributional, with prepositions and verbs at retrieval benefitting more from text modality compared to picture modality.

Future studies would do well to consider an even wider range of word classes. The predictions are that the greater the weighting of distributional relative to grounded representations for specific terms, the greater should be the advantage of text modality at testing. Even when comparing within the same word class, for example, concrete versus abstract nouns - one would predict that abstract noun learning performance may be better when tested with text, compared to concrete nouns. Furthermore, languages that make more fine-grained distributional distinctions for the same types of terms should exhibit greater testing modality effects compared to languages with fewer distributional distinctions.

2.2.5.3. Practical implications and future directions

Apart from theoretical implications, the findings from this study also inform teaching and assessment, as well as digital learning and software design. Understanding that different word classes may benefit from different modalities suggests that tailored instructional strategies could enhance vocabulary acquisition. In most online language tools initial vocabulary teaching is based on word-to-word and picture-to-word learning and testing.

However, our results and modified model indicate that the focus of both learning and testing should be modified as a function of word class. For instance, incorporating more text-based activities for verbs and prepositions might optimize learning outcomes. Additionally, the impact of testing modality on learning suggests that assessments should be designed while considering the modalities used during learning to maximize retrieval success. This should also be considered in research that concerns the effects of different learning strategies to ensure that the conclusions are not confounded by the modality of retrieval or its (in)congruency with the learning method.

In general, future research on this topic should continue to explore the interplay between learning and testing modalities by considering a broader range of word classes, semantic classes and varying levels of word concreteness as there is overlap between those concepts (Borghi, 2019). For example, a grammatical class can encompass different semantic classes, from highly concrete to abstract concepts, each potentially engaging different learning mechanisms. The words included in this study were concrete nouns, action words and spatial prepositions, which could also be aligned with semantic classes of objects, actions and spatial relations. To get a better understanding of the borders between these concepts in terms of L2 vocabulary learning, future studies should consider investigating a broader range of words within each of the word classes considered. Although Martin and Tokowicz (2019) have found that the grammatical class effect is separable from the concreteness effect in L2 language learning, including concreteness as an additional variable with both semantic and grammatical classes could help in understanding what are the drivers of the effects observed in the present studies (Borghi et al., 2019).

Investigating these variables in real-world educational settings would enhance the ecological validity of the current design. Moreover, including more diverse learner populations, such as children or more advanced learners would enhance the generalizability of the findings and provide a better insight into the dynamics behind the language learning processes. By addressing these factors, future studies can provide more comprehensive insights into effective L2 vocabulary learning strategies and contribute to the development of more tailored and effective instructional and assessment practices.

Chapter 3 | The role of pointing gestures and eye gaze in second language vocabulary learning

3.1. Introduction to Chapter 3

In addition to multimodal and distributional accounts, social approaches have recently begun to gain traction in the context of language learning, as described in the previous chapters. Much like the sensory-motor cues proposed by the embodied theories of L2 vocabulary learning, social interaction is also seen as a type of "enrichment" (Çetinçelik et al., 2021; Kreysa et al., 2018; Li & Jeong, 2020).

Although social-based learning has often been associated with children, findings suggest that adult learners can also benefit from the socially enriched approaches compared to traditional methods. For example, in an fMRI study by Jeong et al. (2021) adult participant learned L2 vocabulary in two conditions – one group learned through simulated social interactions (e.g. short video clips of communicative scenarios), whereas the other group learned via L1 translations. The findings showed that social interaction increased activation in the brain regions that are also engaged during native language processing, such as social, affective, and perception–action-related processing (bilateral superior temporal sulcus, posterior middle temporal gyri, right inferior parietal lobule, right temporal parietal junction, right hippocampus). On the other hand, the activation in the same areas in the control group that used traditional (text-based) learning methods was weaker. Similarly, another fMRI study by Verga and Kotz (2019) showed that the mere presence of a social partner during word learning task resulted in increased activation of the brain regions involved in visuospatial learning and sensorimotor processing that are also active during first language learning.

Building on those findings, Li and Jeong (2020) argue that social interaction fosters active, multimodal and emotionally enriched processing which strenghtens the mapping between L2 word forms and their underlying semantic representations. However, social aspects have been understudied in the context of L2 learning (Hardison & Pennington, 2021) and differently operationalized across the studies. This leaves uncertainties about the exact factors that drive the learning gains and indicates a need for further research (Hardison & Pennington, 2021; Li & Jeong, 2020;).

To initiate this investigation, the experiments in this chapter were designed to examine deictic cues as the key components of social interaction, joint attention and theory of mind (Coventry & Diessel, 2024) and probe their influence in screen-based L2 vocabulary learning. The experiments focused on pointing gestures and eye-gaze and their importance in screen-based environment.

While the studies described in the previous chapter expand an existing topic from a new point of view, the experiments presented below open up new avenues by investigating the influence of deictic cues in digital second language vocabulary learning. As one of the early explorations of a topic, different factors were considered as relevant and different methodological choices were made compared to those considered in the previous chapter. For example, given that deictic cues typically refer to objects, the target words were concrete nouns. Their properties were carefully considered to account both for psycholinguistic properties of words, as well as sensory-motor properties of objects, which will be thoroughly described in the rest of the chapter. However, several key elements were maintained from the previous experiments. Participants were recruited from the same population and via the same recruitment methods. Pseudowords were again used to minimise the influcence of language background and enable a better control of properties of word form. Both production and recognition tasks were used to test learning outcomes.

Systematically examining how gaze and gesture influence vocabulary learning outcomes can reveal which aspects of social interaction engages the mechanisms underlying social-learning in adults. This would help in identifying the most effective instructional techniques for adult L2 learners, ultimately informing the design of digital language learning platforms.

3.2. The role of pointing gestures and eye gaze in second language vocabulary learning

This chapter has been published as:

Janjić, P., Arikan, G., Gudde, H. B., Murphy, J. J. C., Sivaram, L., & Coventry, K. R. (2024). The role of pointing gestures and eye gaze in second language vocabulary learning. *Discourse Processes*, 1–19. https://doi.org/10.1080/0163853X.2024.2343625

The numbering will be adjusted to fit in with the thesis format.

3.2.1. Abstract

Learning a second language is recognized as a necessity for social, political, and economic development. However, the processes contributing to initial vocabulary learning have not been explicated. In a series of experiments, this study examines the role of deictic gestures and gaze in second language vocabulary learning. Such cues have been shown to be fundamental in first language learning, but their efficacy in second language learning has not been established. In three experiments participants (total N=435) learned pseudowords by watching images of a teacher naming objects placed on a table while systematically manipulating pointing and gaze. Moreover, manipulating the position of the object relative to the teacher (within or out of reach), served to establish the possible importance of these cues as social versus attentional constructs in second language vocabulary learning. Results show that gaze and gesture did not affect vocabulary learning, but object position did. We discuss implications of these results for theories of L1 and L2 vocabulary learning.

3.2.2. Introduction

One of the first challenges learners come across in second language learning is acquiring vocabulary. The dominant linguistic approach to vocabulary learning treats language as an informationally-encapsulated system, in which learners need to acquire distributional information within a language (Mitchell et al., 2019). However, recent language theories recognize language as social, embodied, and multimodal (Barsalou, 2008; Coventry & Garrod, 2004; Pulvermüller et al., 2005), emphasizing how language is deeply situated in a physical and communicative context (Murgiano et al., 2021). A prerequisite to developing a functional communicative context is establishing joint attention between interlocutors. This is typically achieved using pointing gestures and eye-gaze, cues that are regarded as fundamental to language processing and learning (Goldin-Meadow, 2007; Tomasello et al., 2005).

It has been established that shared attention and presence of multimodal cues such as eye-gaze affect language processing (Grzyb & Vigliocco, 2020; Kreysa et al., 2018). In development, understanding pointing gestures and eye gaze is tightly coupled to the emergence of joint attention, which sets the stage for early language acquisition (Brooks & Meltzoff, 2008; Goldin-Meadow, 2007). Numerous studies have shown that gaze and pointing in infancy are correlated with later language and vocabulary growth (e.g., Brooks & Meltzoff; 2008; Çetinçelik et al., 2021; Law et al., 2012; McGillion et al., 2017).

The role of attention and the accompanying sub-processes of alertness, orientation and detection have also been identified as crucial in the context of second language learning (Tomlin & Villa, 1994). Likewise, theories of second language learning have become more situated, multimodal, and embodied (Atkinson, 2010; Douglas Fir, 2016; Mitchell et al., 2019). However, the topic of eye gaze and pointing gestures in second language vocabulary learning has not been much considered. To date research in L2 word learning relating to gestures has mostly been focused on iconic gestures (García-Gámez & Macizo, 2020; Macedonia et al., 2011; Morett, 2014).

The main goal of the present study is to examine the role of eye gaze and pointing gestures in second language vocabulary learning. Specifically, we systematically vary gaze and pointing gestures of a 'teacher' naming objects placed on a table in front of them in a vocabulary learning task, where participants learn pseudowords they hear in combination with presented objects. To tease apart the mechanisms involved in possible effects of gaze and gesture on word learning, this series of studies will vary where the named object is positioned on the table with respect to the teacher (within peripersonal (reachable) versus extrapersonal (non-reachable) space) in order to establish whether the social/interactional component of deictic communication facilitates word learning, or if gesture/eye gaze merely draw attention to an object (Caldano & Coventry, 2019; Coventry et al., 2008; Rocca et al., 2019). We also vary the number of objects presented on the table to examine whether non-linguistic cues are especially helpful in word learning when the referent is potentially ambiguous. Next, we briefly review evidence for the role of non-verbal cues in first and second language learning prior to detailing the proposed experiments.

3.2.2.1. The role of eye gaze and pointing gestures in first language (L1) processing and learning

The ability to establish joint attention in infancy by using and following eye-gaze and pointing gestures is a crucial step in children's language acquisition (Goldin-Meadow, 2007; Tomasello et al., 2005).

Such 'deictic cues' have been hypothesised to promote learning by enhancing the child's attention to the referent, and/or by enhancing the child's understanding of the speaker's intentions (Booth et al., 2008). Çetinçelik et al. (2021) support the notion that eye gaze plays a greater role than simply acting as a mere deictic pointer, such as arrows. In a systematic review, they argue that it facilitates infants' learning by enhancing their alertness, memory, and attentional capacities, which potentially benefit the encoding process during word learning in a unique way.

The importance of multimodal cues varies as a function of stage of development (Golinkoff & Hirsh-Pasek, 2006). For example, Paulus and Fikkert (2014) found that 14–24-month-old infants paid more attention to an agent's eye gaze when learning to map a novel word to a referent, whereas older children and adults relied more on pointing cues, suggesting there is a developmental change in the use of different cues during word learning. In 28-31-month-olds, however, Booth et al. (2008), showed that gaze and gesture together led children to learn more new words compared to the single cue condition.

The number of studies relating to gaze and pointing gestures in vocabulary learning beyond infancy is sparse. Bang and Nadig (2020) tested children aged 6-11 comparing gaze to a non-social directional cue (e.g., a pointing arrow) in word learning. Children learned from both cues equally well, but with qualitative differences in the way they attended to the gaze as opposed to the arrow cue. In the gaze condition, children spent more time looking at the cued area and engaged in more back-and-forth looking between the referent and the cue, compared to the arrow cue condition.

In language processing in adults, indexical cues such as eye-gaze and pointing gestures appear together with speech, with evidence that the combination of non-verbal and verbal cues facilitates language comprehension by reducing potential ambiguity and cognitive effort (Holler et al., 2014; Holler & Levinson, 2019; Murgiano et al., 2021). Knoeferle et al. (2018) propose three possible mechanisms by which gaze may enhance language comprehension. First, it could serve the role of a simple pointer to a relevant location in the same way that an arrow directs attention to a specific object/location (‘deictic account’). Alternatively, the presence of social attention in the form of speakers’ gaze could help listeners more successfully link spoken words to their referents, by making object representations more grounded and salient (‘referential account’). In that way gaze does not only direct attention, but also contributes to encoding processes through enriching the listener’s mental representations. Finally, the ‘syntactic-thematic account’ suggests that gaze informs not only deictic and/or referential processes, but also other comprehension processes such as syntactic structure building and thematic role assignment by highlighting certain parts of the utterance.

While eye gaze and pointing are both clearly fundamental in early L1 vocabulary learning, their role in second language learning has received less attention (Cooperrider & Mesh, 2022). If L2 learning is based on similar processes as L1 (the ‘fundamental continuity hypothesis’; Ellis & Wulff, 2015; Krashen, 1988; MacWhinney, 2012), it would be expected that eye-gaze and gesture are instrumental in L2 learning as well.

3.2.2.2. Eye gaze and pointing gesture in second language (L2) learning

Newly emerging socio-cognitive and socio-cultural approaches to L2 learning place stronger emphasis on multimodality, embodiment, and the situated nature of second language learning (Atkinson 2010, Douglas Fir Group, 2016; Ortega, 2013), leading to a greater focus on non-linguistic cues during L2 learning. Multiple second language learning studies (Comesaña et al., 2012; García-Gámez & Macizo, 2020; Poarch et al., 2015) have explored aspects of multimodality in L2 vocabulary learning within the framework of the Revised Hierarchical model (RHM) (Kroll & Stewart, 1994). In this model, learning of new words in L2 is mediated by words in L1, which are used as a bridge to reach the underlying concepts. As learners increase their proficiency, they access concepts directly from L2, without the need for mediation from L1 words, thus enhancing retrieval and understanding.

Accordingly, a focus on strengthening the links between new words and concepts should enhance initial vocabulary learning. This can be achieved through presenting multimodal materials, such as pictures (Morett, 2014) and gestures (García-Gámez & Macizo, 2020; Macedonia et al., 2011) that enrich the linguistic stimuli, consistent with Dual-coding theory (Paivio, 1971) or through mental manipulations of words in activities like categorizing and describing, in line with the Levels of Processing framework (Craik & Lockhart, 1972). For example, Paivio proposed that learners create mental representations through both linguistic and non-linguistic channels. Therefore combining linguistic elements with additional cues enriches the learning experience and facilitates better retention of vocabulary.

However, eye-gaze and pointing gestures – key cues in L1 vocabulary learning and processing that guide attention – have thus far not been considered, although attention itself has been shown to be a crucial part of second language learning (Tomlin & Villa, 1994).

In their classification of second language teachers' non-verbal behavior, Allen (1999) points out that 82% of all teachers' communication is nonverbal. Consistent with these findings, native English speakers spontaneously produce more deictic and representational gestures when speaking to L2 English learners, as compared to English L1 speakers (Adams, 1998 as cited in Morett, 2014). Iconic gestures have been found to facilitate instructed L2 vocabulary learning across different age groups and settings (Allen, 1995; Macedonia et al., 2011; Morett, 2014).

Deictic gestures in L2 learning have been studied thus far only in classroom L2 grammar learning (Matsumoto & Dobs, 2017; Stam & Tellier, 2022) and in the context of second language vocabulary learning where different non-human agents (e.g. virtual agents, robots, avatars) act as tutors (Belpaeme et al., 2018; Bergmann & Macedonia, 2013; Demir-Lira et al., 2020; Tsuji et al., 2021; Vogt et al., 2019). However, findings have been inconsistent, with the final emphasis put on the need to further investigate the role of specific cues and their interactions.

3.2.2.3. The present studies

In this series of experiments the goal was to examine the role of pointing gestures and gaze in adult foreign language vocabulary learning. Participants learned new vocabulary (nouns in a pseudo-language) through watching an instructor in varying conditions of gaze and gesture. Specifically, the experiments included the impact of eye gaze, pointing, and the combined effects of pointing plus gaze on noun learning compared to a baseline condition.

Based on the relevance of the social cues during processing and learning observed in past (L1) studies (Booth et al., 2008; Caruana et al., 2021), the hypothesis was that word learning performance will be optimal in the gaze and pointing condition. Both in adults (Caruana et al., 2021) and in children (Booth et al., 2008) the appearance of these two cues simultaneously was shown to enhance language processing and/or learning compared to when they would appear individually. Differences were hypothesized to arise between each individual cues as well. It was predicted that pointing gestures might lead to better outcomes than gaze only (Paulus & Fikkert, 2014). Since gaze can be used to both send and receive social information, it is more ambiguous than a pointing gesture which indicates both the communicative intent and the referent more explicitly (Caruana et al., 2021). Finally, the condition with no cue was expected to lead to the poorest outcomes.

The series of experiments was planned in advance, with later experiments contingent on the results of the previous one. The decision tree with the main experiment and associated follow-up experiments, as well as the data and stimuli are available on OSF (https://osf.io/q5uzw/).

3.2.3. Experiment 1

The first experiment was set up to investigate the role of gaze and pointing gestures in the basic learning condition where each item to be learned was individually placed in front of an instructor (see Figure 1). The procedure was based on different vocabulary learning studies that employ similar methodology (García-Gámez & Macizo, 2020; Morett, 2014).

3.2.3.1. Materials and methods

3.2.3.1.1. Participants

Participants were monolingual English-speaking adults between 18 and 30 years old, with normal or corrected-to-normal vision and no record of learning disabilities or hearing impairments. Participants were recruited through online recruitment platforms. Student participants recruited through SONA were rewarded in SONA credits in accordance with the university's regulations. Participants from the Prolific platform were compensated financially, following the Prolific guidelines. The target sample size (n=145) was determined based on Brysbaert & Stevens' (2018) sample size recommendations for power of 90%, alpha= 0.05 and effect size of d=.4 (r=.2). This effect size has consistently been shown to be the average effect size in psychological science and is therefore considered to be a reliable guide when setting up new research (Brysbaert & Stevens, 2018). We note that this target sample size is noticeably larger than similar studies in the field.

3.2.3.1.2. Design and procedure

Participants learned pseudowords in an online experimental setting (using the Gorilla Experiment Builder; www.gorilla.sc). The experiment used a within-subjects design with 1 factor (cue-type) consisting of four levels (1x4): no cue, gaze, gesture, gaze+gesture.

Participants saw pictures of target items, with a simultaneous auditory presentation of the pseudoword followed by its translation to English (as in previous studies; Macedonia et al., 2011; Morett, 2014). Both the pseudoword and the translation were presented in the auditory modality, images were presented with no text. Across four within-participant conditions, the instructor's cue(s) towards the target object in the image were varied to show either: gaze, pointing gesture (eyes closed), both gaze and a pointing gesture or no cue at all (closed eyes and straight head direction from the teacher) (Figure 18). In the gaze-only condition, the head is positioned to be aligned with the direction of gaze, thereby adding to the saliency, visibility, and ecological validity of the cue, since the perception of gaze depends also on the orientation of the head (Böckler et al., 2015). Participants learned 32 words, distributed across 4 conditions of interest, each condition consisting of 8 words. To avoid effects of confounding variables and potential carryover effects, the order of conditions was randomised and words were not tied to a certain condition, but appeared in all four conditions across participants. Also, there were two instructors and different pointing hands. Throughout the whole experiment each participant would see a randomly assigned instructor that would point with both hands.

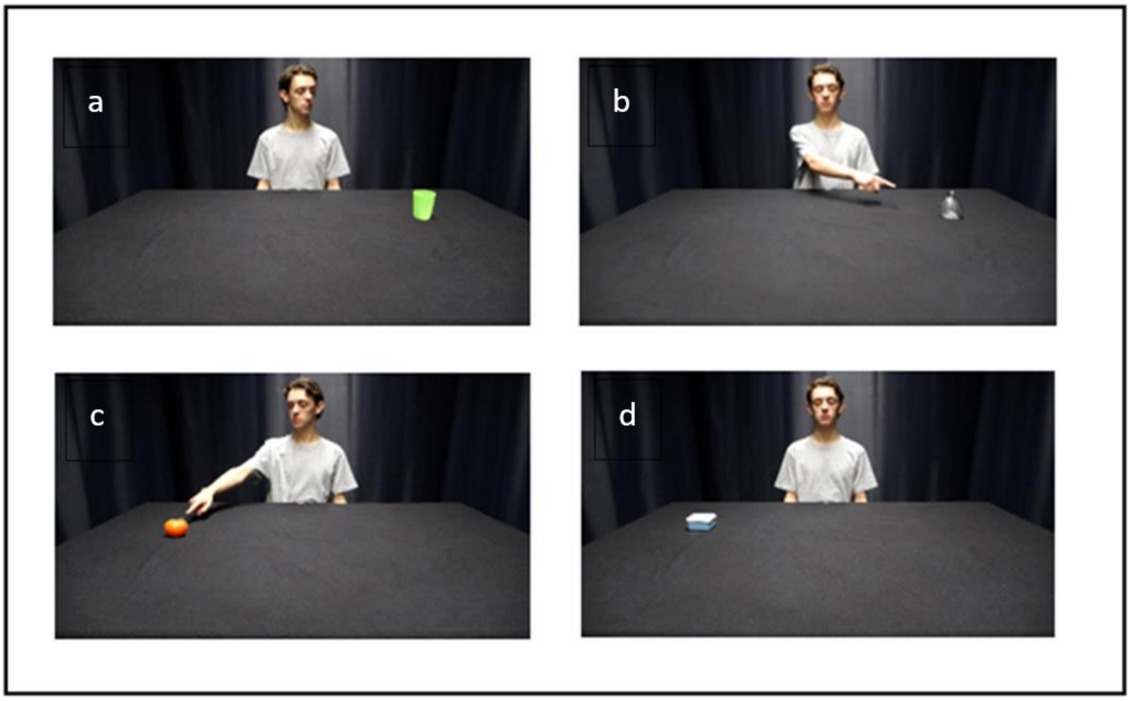

Figure 18. Examples of stimuli illustrating the four main conditions: a) gaze; b) pointing only (eyes closed); c) gaze + pointing; d) no cue (eyes closed)

3.2.3.1.3. Experimental stimuli

Picture stimuli. 32 stimuli were selected based on two existing datasets of objects (Brodeur et al., 2014; Geusebroek et al., 2005), initially resulting in 250 objects across 4 categories and 10 subcategories. The list of objects was initially narrowed down by having two independent native English-speaking raters remove the items that stood out in terms of size (objects that would be too large to fit on a table or objects that would potentially be too small to identify, e.g. blueberry), asymmetry (asymmetrical objects or objects with a handle and tools have been removed due to their effect on visual processing) and familiarity (less familiar objects or objects that are named by using two words (e.g. hair dryer, toilet paper)) leaving a total of 108 items. The final selection was made based on psycholinguistic properties of the 108 remaining words, matching them based on frequency, age of acquisition, familiarity, imageability and concreteness taken from available norms (SUBTLEX-UK; van Heuven et al., 2014) for frequency and The Glasgow Norms (Scott et al., 2019) (Table 6).

Table 6. Psycholinguistic properties of the selected stimuli

	MEAN	SD	MIN	MAX
Frequency	4.334	0.470	3.318	5.210
Age of acquisition	2.553	0.542	1.529	3.382
Familiarity	6.199	0.412	5.286	6.783
Imageability	6.553	0.237	5.971	6.933
Concreteness	6.542	0.231	5.912	6.903

Pictures were taken of the 32 resulting items, half of which belonged to the 'food' category and the other half to 'household items' (Appendix B). The instructors (one male, one female) and item images were created separately and merged in Photopea (Kutskir, 2022) editing software. Items were positioned to the left and right of the instructor, equally distanced from the middle point of the table, where the instructor was seated. The items were positioned so that they are always within the reach of the pointing hand. The pointing hand would correspond to the side on which the item was positioned on the table. The dimensions of the table were 122 cm x 166 cm, and the distance of the items from the middle point of the table edge was dependent upon the height and arm length of the instructors (60 cm/67 cm). The camera was centred 30 cm from the other side of the table, at a height of 97 cm. For the purpose of counterbalancing, all different condition combinations we created resulting in 512 different pictures that show the 32 items across 2 positions (left; right), with 4 different cue conditions (gaze; gesture; gaze + gesture; no cue) and 2 instructors (male/left-handed; female/right-handed). The images that included gaze as a condition included also a congruent head turn (such that gaze and head direction were identical). This was done to ensure visibility/salience of the gaze cue, thus minimising ambiguity related to the cue direction. Moreover, having the head follow the gaze is more natural/ecologically valid than having only the eyes moving from left to right, while having the head face forward. This decision was supported by the literature showing that head turn is an essential part of eye gaze (Atkinson et al., 2018; Bockler et al., 2015., Langton et al., 2000).

Pseudowords stimuli. A list of 32 pseudowords was created in the Wuggy software (Keuleers & Brysbaert, 2010). The words are 4-7 letters long and contained 2-3 syllables. Three native English raters ensured they do not bear obvious resemblance to any existing English words. The final 32 pseudowords were randomly assigned to the target items (see Appendix B for a final list). The audios were created in Google's text-to-speech software.

3.2.3.1.4. Procedure

Participants were informed that they were taking part in an experiment targeting learning of words in a new language. They were briefed on the set up of the experiment, its duration and structure which consisted both of a learning and a testing part. Before starting, the participants were presented with practice trials. At the beginning of the learning phase participants were presented with a list of the words to be learned and asked to read through and retype them. The goal of this familiarization task was to facilitate learning outcomes during the following short learning session. The learning phase of the experiment took approximately 40 minutes in total, with each item presented for 6 seconds. The procedure was repeated twice within each of the four blocks corresponding to four cue conditions. The whole cycle across the blocks was presented three times with a 90-second break in between (see trial procedure in Figure 19). As an additional task to control for the attention, the participants were asked to press the space bar every time they spot a mismatch between the word and the image. For every dozen (real) trials, a mismatch trial would appear once, at a random point. The items or pseudowords that would appear in the attention checks were different from the target stimuli to avoid interference with learning.

Figure 19. Outline of the trial procedure: introducing instructor with a straight gaze (0.5s), followed by a target appearing in one of 2 positions on the lateral plane along with the instructor's cue (depending on condition) and audio naming the item and its translation to L1

The testing phase took place immediately after the learning phase and consisted of two tasks, the first one targeting production of newly learned words and the second one testing their recognition. In the production task participants were shown the target items and were asked to name the stimuli by typing their answer into the bracket below the picture. In the recognition task, instead of writing down the response, participants chose between four possible options. The incorrect options were the other pseudowords that were included in the learning session. The experiment ended with a brief questionnaire asking about the learning strategies used during the experiment.

The outcome measure was accuracy score (binary: correct/incorrect). To account for potential spelling mistakes in the production tasks, answers which contain a spelling error of one Levenshtein's distance from the target answer (addition, substitution, or omission of one letter) were still scored as correct.

3.2.3.2. Results and discussion

Participants' scores were included in the analysis only if a) passed the attention checks during the experiment with a score of over 50%, b) scored above chance level in the recognition task (over 25% correct answers), c) did not self-report use of any additional strategies in the final questionnaire (e.g. writing words down), d) they were monolingual. Of the 195 recruited participants, 26 (13.3%) were excluded for failing attention checks by scoring 6 or fewer correct responses out of 12. The remaining 169 participants showed adequate attention to the task, with attention scores ranging from 7 to 12 (M=9.8, SD=1.6). The cutoff of >6 correct responses (>50% accuracy) was chosen as it represents above-chance performance and indicates basic task engagement. After applying the exclusion criteria to the remaining participants, the final sample consisted of 145 monolingual English-speaking participants between 18 and 30 years old (M=24.07 ; SD= 5.11), who were exposed to 32 pseudowords during the learning session. The table below shows the learning outcomes on the production and the recognition tasks (Table 7).

Table 7. Table showing the mean percentage correct scores, standard deviations and standard errors on each of the four conditions for both production and recognition tasks in the first experiment

	Production task			**Recognition task**		
Condition	**Mean**	**SD**	**SE**	**Mean**	**SD**	**SE**
No cue	36.9%	0.483	0.015	75.1%	0.433	0.013
Gaze only	33.2%	0.471	0.015	74%	0.439	0.013
Pointing only	35.7%	0.479	0.015	74.3%	0.437	0.013
Gaze + pointing	36.4%	0.481	0.015	74.5%	0.436	0.013

To analyse the data and investigate the effect of different cues on vocabulary learning, Generalized Linear Mixed Models (GLMM; Baayen, 2008) with binomial error structure and logit link function was used. Pointing, gaze, and their interaction were included into the model as fixed effects with two levels (absent and present) and the intercepts of items and participants were included as random effects factors to account for the repeated observations of the same individuals and items. The outcome measure was the accuracy score (correct/incorrect). The contrast coding method used was sum coding.

P-values were estimated using the significance criterion of $\alpha < 0.05$. In analysis of each experiments' data, two separate models were run for two testing tasks (production task and recognition task).

The analysis showed no effects of pointing (production: $F(1, 4464)= .701$, $p = .403$; recognition: $F(1, 4465)= .191$, $p = .662$), gaze (production: $F(1, 4463)= .860$, $p = .354$; recognition: $F(1, 4464)= .417$, $p = .519$) and their interaction (production: $F(1, 4463)= 2.849$, $p = .092$; recognition: $F(1, 4464)= .520$, $p = .471$) on L2 word learning outcomes.

Table with the summary statistics for both models can be found in the Appendix B.

This result did not confirm the hypotheses, which expected to find effects of cue. One potential explanation for the absence of effects of cues could be the fact that a single object was used in the first experiment in each trial. Some studies on first language processing and learning found that the effect of gaze differs depending on whether it is used to disambiguate a message (Kreysa et al., 2018; Macdonald & Tatler, 2013; Jachman et al., 2019). Having only one item present in the scene potentially made the cues redundant for the adult L2 vocabulary learners, since the attention is already directed towards the target object that stands out in the scene even without the cue. For that reason, the second experiment presented two objects on each trial.

3.2.4. Experiment 2

Following the null results from Experiment 1, where one item was presented during learning, the planned modification for the second experiment was to include an additional item in the learning set up. This would create ambiguity, where receiving a cue would potentially bring a more tangible advantage to the learner. The hypotheses remained the same.

3.2.4.1. Design and procedure

3.2.4.1.1. Participants

A total of 172 participants were recruited through the SONA and Prolific recruitment platforms. The participants were excluded based on the same exclusion criteria as in the previous experiment. Of the 172 recruited participants, 25 (14.5%) were excluded. The remaining participants showed adequate attention to the task, with attention scores ranging from 7 to 12 (M=10.5, SD=1.43). the final sample consisted of N=146 monolingual English-speaking participants between 18 and 30 years old (M=24.6; SD=4.98).

3.2.4.1.2. Experimental stimuli

The target items (n=32) remained the same as in the previous experiment. To create ambiguity, distractor items were added into the learning set up. The distractor item was always one of the objects appearing in the same condition as the target object. They were counterbalanced to always appear for the same number of trials. The distractor items always appeared simultaneously as the target items and at the same distance from the instructor (Figure 20). The cues were always congruent with the position of the target item.

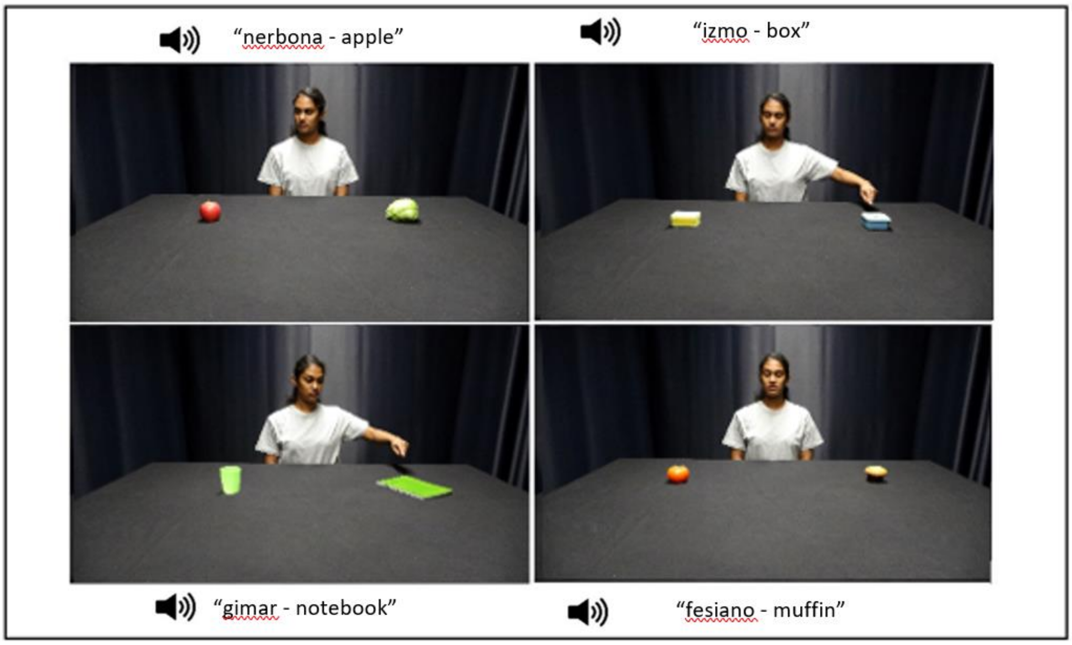

Figure 20. Images in four modified conditions for Experiment 2: a) gaze only; b) pointing only; c) gaze + pointing; d) no cue

3.2.4.1.3. Procedure

The procedure was the same as described in the first experiment, only this time participants were told that there will be two items on the table accompanied by audio indicating the correct word. They were also informed about the testing phase afterwards, as well about the task of identifying mismatches that will serve as an attention check in the analysis. Before starting, participants were presented with practice trials. The structure and the duration of the procedure remained the same as in the first experiment.

3.2.4.2. Results and discussion

The model was built in the same way as in the previous experiment. It included gaze and pointing and their interaction as fixed factors and their random intercepts. The table below shows the learning outcomes on the production and the recognition tasks (Table 8).

Table 8. Table showing the mean percentage scores, standard deviations and standard errors on each of the four conditions for both production and recognition tasks in the second experiment

	Production task			**Recognition task**		
Condition	**Mean**	**SD**	**SE**	**Mean**	**SD**	**SE**
No cue	30.8%	0.013	0.014	73.4%	0.442	0.013
Gaze only	34%	0.474	0.014	72%	0.449	0.013
Pointing only	34.4%	0.475	0.014	75.5%	0.431	0.013
Gaze + pointing	35.9%	0.48	0.014	71.4%	0.452	0.013

The generalized linear mixed model was built in the same way as for the first experiment. The production data showed that none of the predictors had an impact on the dependent variable measuring the learning outcome (pointing, $F(1, 4326) = 1.086$, $p = .297$; gaze, $F(1, 2886) = 1.568$, $p = .211$; interaction pointing*gaze, $F(1, 4326) = .515$, $p = .473$). Please refer to Appendix B for the detailed results.

However, the analysis of the data from the recognition test showed that there was a main effect of gaze, $F(1, 2283) = 11.12$, $p<0.001$, $\eta 2 = 0.005$. The presence of gaze had a negative effect on the learning outcome, meaning that the presence of gaze cue resulted in lower accuracy compared to the baseline (see Table 9). There was no observed effect of pointing, $F(1, 4125) = .846$, $p = .358$ or the interaction, $F(1, 4122) = .018$, $p = .893$.

Table 9. Summary of the generalized linear mixed-effects model for word learning outcomes measured by the recognition test (Experiment 2)

Model equation: model_r <- glmer(score ~ gaze*pointing + (1|id) + (1|item), data = total_recognition2, family = binomial())						
Fixed Effects						
	Estimate	Std. Error	Z value	p value	Lower CI (2.5%)	Upper CI (97.5%)
Intercept	1.570	0.190	8.268	<0.001	1.199	1.957

Gaze	**-0.330**	**0.106**	**-3.106**	**0.002**	-0.542	-0.119
Pointing	-0.081	0.090	-0.886	0.376	-0.259	0.099
Gaze * pointing	0.01	0.181	0.055	0.956	-0.351	0.372
Random Effects						
	Variance	SD				
ID	2.633	1.623				
Item	0.483	0.695				
Model fit						
	R^2 marginal	R^2 conditional				
	0.005	0.489				

Although an effect of deictic cue on L2 vocabulary learning was expected, it was surprising to find only an effect of gaze, because among the three conditions involving cues (pointing, gaze, pointing+gaze), gaze was hypothesised to have the weakest effect. Although small (d=-0.04), the direction of the effect was also unexpected, since we assumed that all cues would have a positive effect on learning.

One possible explanation for the hindering effect of gaze is that gaze may capture the attention of the learner, perhaps therefore distracting rather than facilitating attention towards the target object. Support for this possible interpretation comes from an an eye-tracking study of word learning in school-age children (6-11 years of age), in which gaze condition produced a different pattern of eye movements compared to an arrow-cue condition (although it did it not produce a different learning outcome) (Bang & Nadig, 2020). In the gaze condition, children engaged in more back-and-forth looking between the referent and the cue. The different pattern observed requiring additional time and drawing attention away from the target might explain the negative effect that it had on learning in the current study. When comparing to the other three conditions, in the gaze condition the cue was at the farthest point from the target item. To further explore the mechanisms behind the unexpected negative effect of the gaze, a third experiment was conducted manipulating the distance of an object from instructors and the congruence/incongruence of cues. By varying the distance, it is possible to test if the cost of gaze indeed comes from the position of the target object compared to the cue, as speculated. Manipulating congruence would help to unpack the hindering effect of gaze and to test if it can be replicated in a similar experimental set up.

3.2.5. Experiment 3

The third experiment introduced incongruency between the deictic cue and the position of the target item to further elucidate the results from the previous experiments. Based on the previous results, we expected to find no difference between learning conditions in the number of words learned with congruent or absent cues, which would confirm that having gaze and gestures as cues during L2 word learning does not bring a facilitative effect compared to no cue at all. However, while our previous results indicate that gaze and pointing cues do not facilitate word learning, we wanted to establish if there might be a cost to word learning if cues are incongruent with the target object. If gaze is distracting when congruent aligned with the target object (congruent), then we expected that such distracting effects might be magnified when the cues are pointing to the distractor object, i.e. incongruent with the target object. Thus, in line with the results of the previous experiment, it would be expected that the effect of congruency would show that the learning outcomes would be best in the condition with no cues where there is no distracting effect of gaze and the worse when the gaze is present and pointing to the distractor item (incongruent).

Following the results of Experiment 2 where the gaze cue had a negative effect on word learning outcomes, one explanation could be that the distance of the target objects from the cues was also expected to affect affecting word learning. The effect of pointing did not come across because the hand cue is much closer to the target object than the gaze cue, thereby avoiding the need to look back-and-forth. With the addition of the distance manipulation (within-reach vs outside-reach), it could be established if the negative effect of gaze can be attributed to the close proximity of the target object that is drawing the learner's attention. Finding an interaction between distance and condition in this study could confirm whether the negative effect of gaze is due to distance between the face and target object. Moreover, whether an object is within reach of a speaker (in their peripersonal space) or not has been shown to be integral to spatial communication (Caldano & Coventry, 2019; Coventry et al., 2008). When an object is out of reach, we argue that it is more detached from the speaker, and hence it is more distal both physically and mentally (Piwek et al., 2008).

In summary, we expected to find a difference between the number of words learned depending on different cue-related conditions (congruent, incongruent, absent) and different item positions (within-reach or outside-reach).

3.2.5.1. Design and procedure

3.2.5.1.1. Participants

Participants were recruited and compensated in the same way as in the previous two experiments, with the same exclusion criteria applied. Given the increased number of trials, attention checks were expanded from 12 to 18 items. Of the 188 total participants, 153 passed the attention check by achieving scores above 50% (M=16.37, SD=2.01). After applying all exclusion criteria, the final sample consisted of 145 monolingual English-speaking participants between 18 and 30 years old (M=22.9; SD=6.69).

3.2.5.1.2. Experimental stimuli

The third experiment used a within-subjects design with 2 factors (3x2): cue-type (with 3 levels – congruent, incongruent, absent) and distance (with 2 levels – within-reach and outside-of-reach). Gaze and pointing appeared together, but when present, they were either both congruent or incongruent with the target object. In the first two experiments, the objects always appeared within reach, whereas in Experiment 3, they appeared both within and outside of reach. Other than in the new (distal) position of the objects (Figure 21), the images were similar to those used in previous experiments.

The number of items was increased to fit the new 3x2 design with an equal number of items per cell. The new number of items was 36. The additional target items, as well as 20 new distractor items, were chosen from the same selection of words that was matched on different psycholinguistic properties for the purposes of the first two experiments.

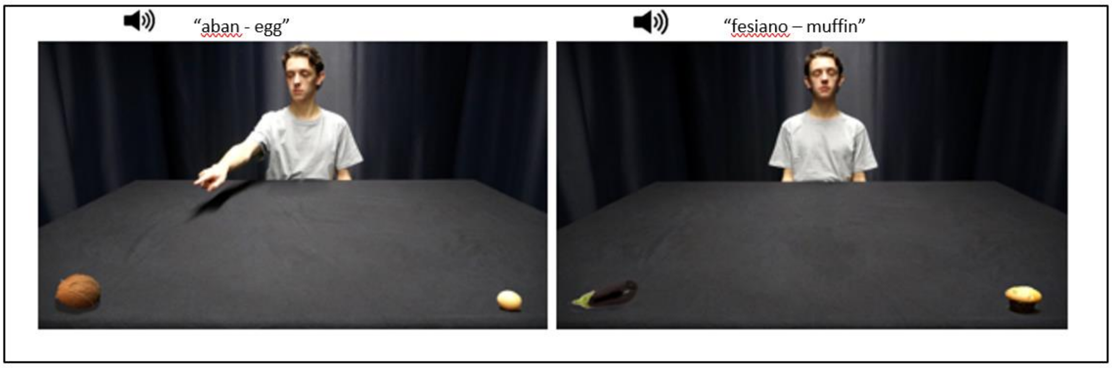

Figure 21. Figure showing the new (distal) position of the items combined with the incongruent and absent cues

3.2.5.1.3. Procedure

The instructions were modified to clarify that there will be trials with a mismatch occurring between the cue and the target item. Participants were told that ultimately the audio cue always correctly indicated the target item. For the attention check task, instead of spotting a mismatch as in previous experiments, participants pressed the space bar whenever they saw a blue dot over one of the items. To avoid the interference with learning of the target items, separate trials were introduced with additional words/items. The rest of the procedure remained the same as described before.

3.2.5.2. Results and discussion

N = 145 participants were exposed to 36 pseudowords during the learning session. The model included cue-type and distance, as well as their interaction as fixed factors and participants and items as random factors. The table below shows the learning outcomes on the production and the recognition tasks (Table 10).

Table 10. Table showing the mean percentage scores, standard deviations and standard errors on each of the six conditions for both production and recognition tasks in the third experiment

Condition		**Production task**			**Recognition task**		
Cue type	**Position**	**Mean**	**SD**	**SE**	**Mean**	**SD**	**SE**
Absent	Within	32.8%	0.470	0.016	74.4%	0.437	0.015
Congruent	Outside	30.2%	0.460	0.016	75.6%	0.430	0.015
Incongruent	Within	33.4%	0.472	0.016	74.9%	0.434	0.015
Absent	Outside	30.1%	0.459	0.016	71.5%	0.452	0.015
Congruent	Within	34.6%	0.476	0.016	75.9%	0.428	0.015
Incongruent	Outside	28.9%	0.453	0.015	62.1%	0.485	0.017

In the third experiment the model was built with the fixed factors ‘cue type’ and ‘position’, as well as their interaction. The random effects remained the same as in the previous model, accounting for the variation in the intercept of items and participants. Sum coding was utilized to contrast the categorical variables, allowing us to interpret the model coefficients as deviations from the overall mean of the dependent variable. As can be seen in Table 10, the Incongruent-Outside condition produced poorer learning outcomes in both testing tasks. However, the analysis of the production task showed no significant effects (cue type: $F(2, 3248) = .360$, $p = .698$; position: $F(1, 1410) = 1.050$, $p = .306$) (see Appendix B), while the analysis of the recognition task data showed a main effect of position, $F(1, 2114) = 6.547$, $p = .011$, $\eta 2 = 0.003$ (see Table **4**11). Cue type (congruence) as a predictor was not found to be significant, $F(1, 3959) = 1.413$, $p = .244$.

Table 4. Summary of the generalized linear mixed-effects model for word learning outcomes measured by the recognition test (Experiment 3)

Model equation: model_r <- glmer(score ~ cue_type*position+ (1|id) + (1|item), data = total_recognition3, family = binomial())						
Fixed Effects						
	Estimate	Std. Error	Z value	p value	Lower CI (2.5%)	Upper CI (97.5%)
Intercept	1.441	0.234	6.165	<0.001	0.983	1.899
Cue Type (congruent)	0.184	0.115	1.610	0.107	-0.040	0.409
Cue Type (incongruent)	0.143	0.122	1.171	0.242	-0.096	0.382
Position (outside)	**0.330**	**0.127**	**2.602**	**0.009**	**0.081**	**0.579**
Cue_type*Position (Congruent*Outside)	0.209	0.269	0.777	0.437	-0.318	0.735
Cue_type*Position (Incongruent*Outside)	0.156	0.237	0.667	0.511	-0.308	0.619
Random Effects						
	Variance	SD				
ID	2.44	1.56				
Item	1.27	1.13				
Model fit						
	R^2 marginal	R^2 conditional				
	0.005	0.532				

The results of Experiment 3 help to partially elucidate findings from the previous experiment. Although the expected interaction between the type of cue and position didn't come out, there was a main effect of position showing that the words were learned better when they were in the instructor's peripersonal space. This could be related to the negative effect of gaze from the previous experiment. Since the gaze cue was most distant from the object (the pointing hand was in the previous experiment always close to the object), the distance of the cue from the target item might play a role in explaining the learning outcomes. Having an on-screen cue that is placed away from the object, might require additional time to switch and direct attention from the cue to the target. An alternative possibility is that this is a result of the effect of peripersonal versus extrapersonal space, which was already shown to affect language processing (Caldano & Coventry, 2019), but might also affect language learning. Further experiments would be needed to explore this avenue and disentangle the effects from distance from the effects of peripersonal versus extrapersonal space. It would be interesting to see whether the effect of distance would translate into non-social cues, such as arrows or pointers, that are often compared to social cues in word learning studies (Barry et al., 2015; Michel et al., 2019). Unexpectedly, the effect of congruence was not found to be significant. This could be due to the incongruent condition leading the participants to dismiss the cues altogether after realizing that they are unreliable. Thus, one possible future modification could be to only use the congruent and the absent condition, without using the incongruent condition. This would still serve to obtain the information if the negative effect of gaze can be replicated and lead to poorer learning outcomes compared to no cue condition.

3.2.6. General discussion

In the context of overwhelming evidence showing the importance of gaze and pointing gestures in first language learning (Booth et al., 2008; Golinkoff & Hirsh-Pasek, 2006), we set out to systematically investigate if the same attentional cues would also have an impact on adult second language vocabulary learning. Across three experiments, participants had the task of learning (pseudo)words in a new language by viewing an instructor in an image naming, pointing and gazing (not at all, congruently, incongruently) at a target object. Despite the arguments behind the fundamental continuity hypothesis that similar processes are at play both in L1 and L2 learning (Ellis & Wulff, 2015; MacWhinney, 2012), the overall results of the experiments failed to find support for a facilitative effect of pointing gesture and gaze in L2 word learning.

The starting point for the series of studies was an experiment set up with only one target item present. In this experiment with null results, we suggested that the absence of an effect might be due to cues having a different effect when there is ambiguity in the learning situation (i.e. where the referent is unambiguous), in line with findings from studies on language processing (Kreysa et al., 2018). Accordingly, Experiment 2 introduced ambiguity with the presentation of two objects on each trial. However, this experiment once again confirmed the previous finding that gaze and pointing gesture do not provide an advantage for L2 word learning. On the contrary, the results showed an unexpected negative impact of gaze on word learning (compared to the no cue condition).

Experiment 3 manipulated target object distance from the instructor and congruence of cues (congruent, incongruent, absent) to further examine the (negative) effect of gaze found in Experiment 2, and to test if incongruence of cues might lead to a drop in word learning performance. No effects of cue type were found, failing to replicate the (negative) effect of congruent gaze in the previous experiment. However, the experiment did find a main effect of object distance, with word learning rates lower overall for objects distal to the instructor in the images. Below we explore explanations for the effects we found across the studies (distance, negative effect of gaze) before considering why we failed to find facilitative effects of gaze and pointing overall on second language vocabulary learning.

The effect of distance in Experiment 3, taken together with the negative effect of gaze in Experiment 2, speak in favour of divided attention between the instructor in the image and the target object in the present experiments. When the object was further away from the instructor, the object was bigger on the screen and hence more visually salient, yet word learning rates were nevertheless lower than when the object was smaller and positioned within the reach of the instructor. It is well known that faces generally grab a lot of attention in images (e.g. Gliga et al., 2010; Ro et al., 2001; Theeuwes & Van der Stigchel, 2006). Therefore, it is possible that it was easier for participants to process objects placed closer to the instructor, compared to dividing attention between the face of the instructor and the target object. The results of Experiment 2 are consistent with the view that the eyes are one of the most attention grabbing facial elements, with open eyes diverting attention away from the target object. Indeed, there is much evidence for selective attention to eyes from early infancy, forming one of the bases of intention understanding (see for example Langton et al., 2000). So open eyes attract attention more than closed eyes, taking away attention from processing the target object and hence reducing the likelihood of binding the new object name to the object presented. Of course, one way of testing this hypothesis would be to rerun the studies using eye tracking to see if distance and eye gaze indeed mediate looking time to the target object.

We now turn to the failure to find facilitative effects of instructor cues. One possibility is that gaze and gesture are important cues to second language vocabulary learning, but only when learning takes place in more ecologically realistic situations. The use of static images in the present experiments meant that the dynamics of eye gaze and pointing were not captured. Still images omit important information about the dynamics of gaze and pointing that may influence action perception and therefore looking behaviour (Haxby et al., 2020; Welke & Vessel, 2022). For example, the trajectory (pre-stroke, stroke, post-stroke hold and retreat) of a referential gesture is both rich and elaborate making it a reliable and strong cue to a referent (Kendon, 1980; McNeil, 1992; Norris, 2011; Shattuck-Hufnagel et al., 2016). So, it might be predicted that using dynamic gesture and gaze might divert attention away from the face of the instructor to focus on the target object more decisively, thus potentially enhancing word learning. All that said, many online second language learning apps use static images, and therefore the present findings indicate that pointing and eye gaze may offer little aid to word learning in that context.

The alternative possibility is that the present results may be the first evidence that gesture and eye gaze – so central to first language acquisition – may not facilitate second language acquisition. There are many possible reasons why this might be the case. First, one of the key differences between learning vocabulary in first versus second language is that the object name is known in the first language when learning a second language, but that is of course not the case in first language acquisition. A child acquiring language does not know a priori what a new word refers to. In that case a caregiver can use pointing and eye gaze to make it to the child what the word does refer to. For instance, ‘car’, ‘wheel’, ‘bonnet’, etc. can all be used to refer to a car, and pointing and eye gaze can be essential to disambiguate what is being referred to. In contrast, in second language acquisition the word ‘car’ can be used to make it clear what is being referred to, obviating the need for gaze and gesture to do that work. Indeed in the present experiments, the object name was always given in English at the start of each trail, making it clear what the new word to be learned referred to. While participants processed the images (performing above chance on the catch trials and affected by distance in Experiment 3), the heard word may have reduced the need for attentional visual processing during the task.

Second, it may be the case that cues are only effective for some types of processing. For example, in an eye-tracking study that looked at the effects of speaker’s gaze on first language comprehension and processing (Kreysa et al., 2018), gaze was found to help direct attention to the target object more rapidly, but did not lead to improved memory performance. This might suggest that a speaker’s gaze might be of limited importance for slower processes such as consolidating word-object relations in memory, or might even come with a cost of dividing attention while learning.

3.2.6.1. Limitations and future directions

The results of the present study could potentially be accounted for by its design. Most notably, the presentation of social cues in static images may have affected their influence on L2 word learning. Although the sequencing of images was used to resemble the dynamic nature of social cues, their repetitiveness might have stripped away crucial dynamic elements that occur in natural use of social cues (Kendon, 1980; McNeil, 1992; Norris, 2011; Shattuck-Hufnagel et al., 2016). Social cues derive their communicative power from their dynamic unfolding in real-time, which is why the static images might have altered the processing mechanisms used. To address this limitation, future research should investigate the role of social cues using video stimuli or live interaction paradigms to better capture the temporal aspects of gaze and gesture that may be critical for their facilitative effects to occur.

While we focused on gaze and pointing based on their established role in first language acquisition, adults may rely on different or additional social cues when learning new vocabulary. Further research should explore a broader range of social signals that may be relevant for adult learners, including facial expressions, beat gestures, iconic gestures or multimodal combinations of signals. The gesture literature distinguishes between several distinct categories that may differentially impact L2 learning processes in certain contexts. Iconic gestures have already been found to complement verbal input by depicting semantic content through hand shape and movement, thus facilitating L2 vocabulary learning (Macedonia et al., 2011). Beat gestures, which provide rhytmic emphasis and mark discourse structure (McNeill, 1992), may facilitate learning by creating temporal anchors for new lexical items (Pi et al., 2021). These rhytmic movements have been shown to enhance memory consolidation through their coupling with speech rhytm (Kelly et al., 2020), but have been less studied in the L2 vocabulary learning contexts. Eye tracking studies could provide valuable insights into which visual cues adults naturally attended to and prioritized during vocabulary learning tasks, potentially revealing social signals that were not captured in the present investigation.

Overall, the results across experiments did not find clear support for the effect of attentional cues in adult second language vocabulary learning. The results of this study contribute to the body of research discovering the role of attention processes in second language learning (Tomlin & Villa, 1994) by adding a comprehensive and novel investigation of gaze and pointing gesture in this context. The results could be used to further motivate the discussion on the continuity of the learning processes across L1 and L2 learning. Finally, the unexpected adverse effect of features such as gaze and position that were discovered could be applied to the industry setting to serve as a reminder for creators of digital language learning materials on the importance of basing their design-related decisions for second language learning apps on research findings.

Chapter 4 | Digital Language Learning Resources: Analysis of Software Features and Usage Patterns in UK Schools

4.1. Introduction to Chapter 4

In this chapter the topic of the multimodal L2 vocabulary learning will be transferred into the practical context of language learning technologies and their classroom applications. While controlled studies can demonstrate the potential of specific multimodal features, teacher evaluations provide insights into how these principles perform when embedded within complex software packages used in authentic classroom settings. Apart from considering multimodality as one of the important components of effective language learning software design, this chapter will also address other crucial pedagogical and language learning principles.

The next section examines challenges related to research, evaluation and use of second language learning technologies. These challenges will be further addressed with the introduction of a theory-based evaluation tool applied to the most frequently used language learning software to evaluate their strengths and weaknesses. As in the previous chapters, the introduction will provide a broader context and will be followed by the presention of a study (in the form of a self-contained manuscript).

4.1.1. Technology in second language vocabulary learning

As early as 1960. McLuhan argued for the importance of understanding learning in the context of new technological advances which still holds today: "*The aim is to develop awareness about print and the newer technologies of communication (...) so that we can get* *the best out of each in the educational process*" (Culkin, 1967). Technology-based learning offers the potential to help second language learners integrate modality-specific information with newly acquired L2 word forms to achieve a more native-like understanding (Li & Jeong, 2020; Li & Lan, 2022). The capacity of digital software to present visuo-auditory experiences or mimic social interactions creates unique opportunities for multimodal learning that traditional classroom methods cannot easily replicate. However, the extent to which current language learning software capitalizes on the multimodal and other affordances remains underexplored. The rising interest in L2 learning combined with an increased accessibility of digital technologies has resulted in a rapidly expanding number of digital products intended for second language learning (Gangaiamaran, 2017; Rosell-Aguilar, 2017). The number of possibilities that they offer make them an attractive tool for teaching and learning (Hirsh-Pasek et al., 2015). Their characteristics support ubiquitous learning, individualisation and personalisation, communication and collaboration (Persson & Nouri, 2018) as well as enhance learner's motivation, autonomy and confidence (Kacetl & Klimova, 2019).

As views on cognition and learning have evolved, so have language learning technologies (Li & Lan, 2022). With the dominant behaviorist approaches computer-assisted language learning focused on drill and practice methods. The 1980s and 1990s saw a shift toward cognitive approaches, emphasizing communicative exercises and fluency over grammar. Newer technologies focus on social perspectives, highlighting authentic learning contexts and social interaction. Nevertheless, Li & Lan (2022) point to gaps between theories and practice that are still present and urge for more effective knowledge transfer between academia and industry.

4.1.1.1. Research on second language learning technology

While many agree on the great potential that digital technology holds for education, there is an apparent lack of empirical research on this topic (Heil et al., 2016; Reber & Rothen, 2018; Sung et al., 2015). Even though many studies report a generally positive impact of the digital technologies on language learning (Mihaylova et al., 2020; Persson & Nouri, 2018), others point out the reasons why these results cannot be taken for granted. Some argue that the observed positive effect is short-term (Burston, 2015; Reber & Rothen, 2018) and could be based predominantly on the temporary initial motivation of the users (Reber & Rothen, 2018). Others mention studies that found no significant difference between the outcomes of learning with traditional methods versus using mobile applications (Chen et al, 2020), or only limited evidence for their efficacy (Golonka et al., 2012).

Comprehensive reviews of the available literature on this topic often warn about validity issues that they come across, especially with regards to the number of participants, variables observed or the methods reported. For example, an analysis of 291 studies on mobile assisted language learning by Burston (2015) showed that only 19 were deemed reliable enough to determine the learning outcomes of mobile-based language learning technologies. Likewise, in their review of mobile-assisted language learning applications Mihaylova et al. (2020) claim to have found insufficient power in over half of the included studies. In their meta-analysis of mobile assisted language learning research Sung and collagues (2015) do report a moderate effect of using mobile-based language learning applications, but still warn about publication bias present in the field which might skew the results of this type of research.

These results underline the need for further investigation (Kamasak et al., 2021). Rapidly changing trends in digital technologies, with new technologies emerging and disappearing within a matter of few years (Zhang & Zou, 2020), make this gap in research increasingly obvious. Despite the growing number of available 'educational' applications, their educational quality is questionable (Hirsh-Pasek et al., 2015; Kolak et al., 2021; Li & Lan, 2022). Sung and colleagues (2015) report designs choices that encourage distracting behaviours or offering irrelevant materials during learning periods, thus hindering the learning process. Furthermore, Kolak et al. (2021) found that not even a higher price of the app can indicate better quality. This chaotic landscape is frequently referred to as the "Digital Wild West" (Shing & Yuan, 2016), describing a market where quality control and clear guidelines are lacking, leaving users overwhelmed by the options.

As a result, it happens that despite the globally raised awareness of the importance of second language learning and an increasing number of language learning tools available to learners, the end-results seem to be unsatisfactory. The analysis of the European Survey on Language Competences supports this outlook by reporting a low level of competence (Costa & Albergaria-Almeida, 2015). Among 54,000 European students only 42% reached the level of independent user in the first foreign language, and close to half of adult Europeans report that they are unable to hold a conversation in any foreign language. This discrepancy between the number of possibilities for language learning and the disappointing outcomes leads to the assumption that more should be done to use the existing potential more efficiently.

4.1.1.2. Evaluating existing technologies

When looking back at the development of digital applications, the first educational apps were essentially simple digital reproductions of existing games and tasks. Sweeney and Moore (2012) estimate that this was the case for around 90% or more of the language learning apps 20 years ago, which was one of the reasons for many of their shortcomings. At that time there was little to no explicit consideration of learning principles. As the field evolved, the importance of collaboration between educational science and technology development became increasingly recognized, although this collaboration is still far from the norm.

According to Hirsh-Pasek and colleagues (2015), addressing the issue of educational software quality requires two key approaches. The first approach emphasizes the need to bridge the gap between the existing research and the process of software design by improving communication in the early phases of the development. The other approach is oriented towards the already existing applications and their evaluation. Given the large number of second language applications already available, it is important for the learners and educators to be made aware of their effectiveness, as well as the level to which they correspond to language learning theories. This could help users bring a more informed decision about using them, as well as to encourage researchers to compare and further explore the differences and similarities between existing applications.

A consistent and wide-ranging approach to evaluation can enhance our understanding of the ways in which digital technology resources can support the language teaching and learning process (Reinders & Pegrum, 2015). Classifications and frameworks that look at the applications used for second language learning from different perspectives bring a useful overview and serve as guides through a large number of technologies available (Rosell-Aguilar, 2017). However, most of the recent classifications and frameworks have focused primarily on mobile applications for language learning (Chen, 2016; Heil et al., 2016; Reinders & Pegrum, 2015; Rosell-Aguilar, 2017), leaving out the language learning software that can be used in different contexts, such as classrooms.

4.1.2. From individual learning to classroom implementation: The teacher perspective

Although technology has the potential to bridge the gap between traditional and digital language education, the effectiveness of technology-assisted language learning ultimately depends on the context in which it is used and the specific needs it addresses (Burston, 2015; Kamasak et al., 2021; Sung et al., 2015). The same learning tool can be used in different ways and across different contexts, which needs to be taken into consideration when attempting to evaluate them (Jamieson & Chapelle, 2010). Therefore, to expand our understanding of the learning tools, it is important to include user perspectives and experiences as well.

While the previous chapters of this thesis examined multimodal vocabulary learning through controlled experiments with individual participants, real-world second language learning typically occurs in classroom environments where the language is taught under the guidance of teachers. This shift from individual to classroom contexts represents a crucial transition from laboratoey-controlled conditions to the realities of educational practice. Understanding how teachers, as the primary mediators of second language learning, select, evaluate and implement digital technologies provides essential insights into the practical application of research findings. While individual and adult learners may be more likely to choose applications based on marketing appeal, teachers' professional expertise in child development and language makes them particularly qualified to evaluate educational software.

Still, despite many types of language learning resources available in education, few empirical data exist about what resources are actually being used by teachers. Many of the current classroom studies have focused on a single technology type, thereby missing the opportunity to provide a more comprehensive overview (Zhang & Zou, 2020). However, when creating an overview, it is important to keep the context of use in mind (Jamieson & Chapelle, 2010). Different teachers, languages, institutions and countries where the teaching is taking place all influence the teaching habits and the use of resources (Kaimara et al., 2021; Timotheou et al., 2022; Vermeulen et al., 2017; Wang et al., 2019). For example, research has shown that one of the major barriers to the technology use is the lack of resources at the institutional level (Kaimara et al., 2021). The availability of software for specific languages potentially introduces large differences between the resources available depending on countries/languages taught. Also, an extensive part of the current literature refers to the use of digital technologies in EFL teaching (English as a Foreign Language) (Yang & Shadiev, 2019; Zhang & Zou, 2020), with other foreign languages significantly less represented.

4.1.2.1. Second language learning in the UK context

The language learning context of the UK, as well as the other English-speaking countries, presents specific challenges. Due to English being predominantly taught as a second language in other countries (Busse, 2017), the motivation for native English speakers to learn foreign languages is often lacking (Lanvers & Coleman, 2017). The systematic teaching of languages across primary schools in the United Kingdom is rather recent. The official requirement to incorporate language learning into the primary school came into force in 2014 after its importance has been highlighted by experts in the field. Before that, as Myles et al. (2019) report, the implementation tended to be "*localised, vulnerable to change and variable in quality*". Although the policy introduction was an improvement, the guidelines on implementation were lacking, leading to unsatisfactory learning outcomes (Tinsley & Doležal, 2018). The main issues seem to be lack of time, lack of progress tracking, and low levels of staff language proficiency and confidence that together lead to students' lack of motivation and loss of interest (Myles et al., 2019).

One of the proposed solutions to address these issues is to "develop effective use of digital technology to support learning, training and reporting" (Myles et al., 2019). To develop a truly effective approach to implementing L2 learning technologies, it is essential to integrate learning principles with a deep understanding of contextual factors. As Zhang and Zou (2020) suggest, a comprehensive evaluation of L2 learning technologies should consider not only the technical capabilities of the tools but also how they fit within the educational environment. By taking into account the specific learning objectives, user preferences, and institutional limitations, more effective digital resources can be designed, that bridge the gap between theoretical research and real-world application.

To that end, the following study was designed to evaluate the alignment of the current language learning software with the theoretical principles of second language learning. Furthermore, it will examine how the use of language software is applied in the classroom context.

4.2. Digital Language Learning Resources: Analysis of Software Features and Usage Patterns in UK Schools

This chapter has been published in Frontiers in Education as:

Janjić, P. & Coventry, K. R. (2025). Digital language learning resources: Analysis of software features and usage patterns in UK schools. *Frontiers in Education*, 10, Article 1532802. https://doi.org/10.3389/feduc.2025.1532802

The numbering will be adjusted to fit in with the thesis format.

4.2.1. Abstract

The increasing availability of digital technology for second language (L2) learning is transforming traditional teaching methods, yet the quality of these resources remains unclear. A survey was conducted among a stratified sample of second language teachers (N=118) from UK primary and secondary schools to evaluate the use and theoretical background of digital tools used in classrooms. A rating tool, grounded in educational and language learning theories, was developed to assess individual software features. Results showed extensive use of digital resources, with 89% of teachers utilizing digital devices and over half employing more than three different software types. However, evaluations revealed varying adherence to language learning principles. Strengths were identified in the components of ‘engagement’ and ‘input’, whereas opportunities for improvement are observed in the aspects of 'output' and ‘social learning’. Additionally, higher software ratings did not correlate with usage frequency or price. These findings highlight the integration of digital tools in UK language learning pedagogy, but underscore the need for ongoing evaluation to improve software quality and effectiveness.

4.2.2. Introduction

The rising interest in language learning combined with increased access to digital technologies has resulted in a rapidly expanding number of digital products intended for language learning (Gangaiamaran & Pasupathi, 2017; Rosell-Aguilar, 2017). However, while the use of digital technologies in language classrooms is generally encouraged, there is a lack of empirical research on the use and effectiveness of these technologies (Heil et al., 2016; Reber & Rothen, 2018; Sung et al., 2015). Rapidly changing trends, with new technologies emerging and disappearing within a matter of a few years (Zhang & Zou, 2020), make this gap in research increasingly obvious. Although there is an increase in the number of available "educational" applications, most of them remain unregulated and untested (Hirsh-Pasek et al., 2015), resulting in designs that could hinder instead of promote learning (e.g. encouragement of distracting behaviours by presenting irrelevant materials during critical learning periods; see Hirsh-Pasek and colleagues, 2015 for discussion).

Currently, the levels of competence in a second language remain (on average) low (Costa & Albergaria-Almeida, 2015). This is especially the case in the UK, where it was found that only 32% of 15-30-year-olds feel confident reading and writing in two or more languages (European Commission, 2019). This discrepancy between the increasing number of language learning tools available and the disappointing outcomes may be in part due to the challenges related to *creating*, *evaluating* and *using* language learning software. We consider each of these in turn.

4.2.2.1. Creating language learning software

When it comes to *creating* the digital content, around a decade ago it was estimated that 90% of language learning software was simply a digital reproduction of existing non-digital materials (Sweeney & Moore, 2012). Most of the software created around that time did not use the full potential of digitalization or consider the language learning principles that are specific for the digital content (Hirsch-Pasek et al., 2015). Therefore, it became increasingly apparent that to create educational software of high quality, the design process should include relevant scientific research and collaboration between different stakeholders. Although the need to understand the relationship between theory and the design of technology-based materials was already identified three decades ago (Garret, 1991; Shaughnessy, 2003), this has not yet become standard. According to a recent review by Yang and Shadiev (2019) there is still a lack of theoretical grounding present in the field of computer-assisted language learning ('CALL'). Additionally identified issues include insufficient collaboration between classroom teachers, researchers, and technology specialists (O'Brien, 2020) and the lack of consistent use of evaluation tools (Kolak et al., 2021). Together, these shortcomings result in a market saturated with software of variable educational quality (Kolak et al., 2021).

4.2.2.2. Evaluating language learning software

To enable navigation of the overwhelming offer of language learning software, it is critical to develop subject-specific *evaluation* tools. Essential for evaluating the CALL technologies is the theoretical grounding of the evaluation tool (Chapelle, 2001; O'Brien, 2021).

In their review Hubbard & Levy (2016) point out that there is no single established theory that uniquely characterises CALL. Rather, this field uses and combines theories from different sources and traditions. Additionally, some theoretical approaches to L2 learning can sometimes be difficult to transform into practical implications for language learning software (Chapelle, 2009; Hubbard & Levy, 2016). For example, perspectives that attribute learning mostly to learners' internal mechanisms are less concerned with the properties of the input, which is crucial in designing language software. On the other hand, psycholinguistic approaches are more focused on processing the input and the interaction, thereby offering potential guidelines that can be used to solve practical issues related to the software design (Chapelle, 2009).

Therefore, one solution to creating a theoretically grounded evaluation tool is to provide a holistic view that would include theories that explain critical aspects of language learning (Hubbard & Levy, 2016).

4.2.2.2.1. Introducing theoretical principles for evaluating language learning software

When talking both about the design and the evaluation of language learning software, understanding the linguistic and educational principles of second language learning has been identified as an important starting point (Chapelle, 2001; Reinders & Pegrum, 2015; Starke et al., 2020). To do so, it is useful to distinguish between *general learning principles* and *language-specific learning principles*.

General learning principles. Hirsh-Pasek et al. (2015) defined a set of general learning principles to advise the current design of educational apps. The "four pillars of learning" as set out by Hirsh-Pasek and colleagues are: *active learning*, *meaningful learning*, *engaging learning* and *social learning*. Including them as a cornerstone of educational software design helps in aligning the software to children's natural inclination towards learning.

The *active* learning principle (pillar I) draws on constructivist theories that propose that children play an active role in their own learning, which results in knowledge being constructed rather than simply transmitted (Phillips, 1995). This translates into software that is "minds-on" and require challenge, thinking and mental manipulation of the content, as opposed to automatic tapping or swiping or merely observing the content. The principle of *engagement* (pillar II) in the learning process embodies behavioural, emotional and cognitive engagement. Engagement allows the learner to stay on task, avoid distractions and benefit more from the learning content. To achieve this, learning software should reduce the features that draw attention away from the learning content (sounds, flashy animations, etc), that ultimately disrupt the learning process (Hirsh-Pasek et al., 2015). The third principle characterises *meaningful* learning (pillar III) as learning that is relevant and connects to the learner's experience and existing knowledge. This is where software have great potential to offer examples and multimodal connections that relate to the real world and learner's existing concepts. Finally, the principle of *social* interaction (pillar IV) emphasizes the value of learning from social and communicative partners, which is in line with Vygotsky's Sociocultural theory (1978). Although social interactions are difficult to represent through software, this can be achieved through inclusion of multiple users to interact and collaborate or through parasocial interactions with on-screen characters.

Similarly, key areas that contribute to the educational value of learning apps in general have been explored by others as well, highlighting some of the same principles, with the addition of "feedback", "narrative", "language" and "adjustable content" (Kolak et al., 2021). In the context of language learning software, these additional areas are best considered in relation to language learning principles.

Language-specific learning principles. Second language learning methods have shifted in recent decades from behaviourist to constructivist learning methods. Constructivists hold that individualised drilling and reinforcement practices (staples of behavioral methods) are insufficient for language learning. Instead, the learner learns by blending pre-existing with new knowledge encountered in social contexts, thus making language learning a social activity (Heil et al., 2016). From the language learning and teaching perspective, there is much agreement on pivotal aspects of language learning research that should be included in the process of language learning (Dubiner, 2018; Reinders & Pegrum, 2015); sufficient *input* through reading or hearing, opportunities to produce *output*, *interaction* through conversations or simulations and *rehearsal* where the contextualized material is repeated and practiced. Through the inclusion of these aspects of language use, software can move away from traditional reinforcement practices and come closer to the naturalistic and well-rounded language learning experience associated with constructivist theories.

Receiving *input* is a first step and a crucial element of language learning as it provides a basis for creating form-meaning mappings, according to Input Processing theory (Van Patten, 1993). However, the nature of that exposure makes a difference for the learner. Input that is comprehensible, contextualized and adapted to the learner's level leads to the best learning outcomes (Krashen, 1994). Another important aspect of the input is the modality through which it is presented. The multimodal nature of digital input deserves particular attention in the context of screen-based learning. While the well-established theories such as Dual coding theory (Paivio, 1971) and Multimedia learning theory (Mayer, 2009) propose that learning should ideally occur through a combination of different modalities (vision, audition, etc.), the practical implementation of multimodal input in language learning software can vary considerably. Complementary to input, *output* or production is also a fundamental part of language acquisition process, as described by the Output hypothesis (Swain, 2005). Production, both spoken and written, gives the learner the opportunity to test out their hypotheses about transforming ideas into linguistic messages, to notice the gaps in knowledge and to reflect on their language use (Dubiner, 2018). The Interaction account (Long, 1996) integrates Input and Output by emphasizing the role of *interaction*, which supports language learning by engaging the learner to consolidate form-meaning relationships. One of the crucial aspects that occurs as a product of interaction is feedback. The feedback information would typically be received from the interactional partner and would serve as a tool for the speaker to evaluate their own production. This relates back to the well-known Vygotsky's (1978) sociocultural theory, according to which learning stems from social experience and from navigating interactions with peers. Therefore, to make up for the absence of the real social interactions that are important in language learning, software should be able to provide meaningful and constructive feedback information that can mirror the ones normally obtained through conversations. Finally, with the practical constrictions within the classroom contexts in terms of the learning outcomes, the proponents of the Skill Acquisition Theory (SAT; DeKeyser, 2007) and the associate-cognitive framework (Ellis, 2007) argue how it is important to include *rehearsal* opportunities. These theories conceptualize language learning as being similar to general learning mechanisms, suggesting that learning occurs through repeated exposure. This has practical implications for the design of language learning software. While acknowledging the importance of context, the advocates of this theory support a mixed approach achieved through retaining automatization activities, but modified to be meaningful and contextualized (Gatbonton & Segalowitz, 2005).

Using only one of the theoretical approaches mentioned to guide the design or evaluation of comprehensive language learning software would be limited. Therefore, there is a need to combine these approaches to be able to thoroughly study second language learning (Chapelle, 2009; Hubbard & Levy, 2016). The principles derived from these approaches, together with the previously mentioned general learning principles form a theoretical, pedagogical and empirical base that should be built into language learning software to make it comprehensive and well-rounded, thereby optimizing learning.

4.2.2.3. Using software in language teaching

The properties of the technologies on their own are not sufficient to bring positive language learning effects (Sung et al., 2015). The way that the technology is being *used* is also critical (Burston, 2015). The same learning tool can be used in different ways and across different contexts, which need to be taken into consideration when attempting evaluation (Egbert et al., 2009; Jamieson & Chapelle, 2010). Much of the research on digital language learning tools has focused on software for individual use (Mihaylova et al., 2020; Reber & Rothen, 2018). However, despite many types of language learning resources available, exactly what and how resources are actually being used by teachers in the classroom is not well understood. Moreover, understanding the classroom context and the viewpoints of teachers who are both the users of these technologies, but also the experts in the educational aspect of language learning, is a valuable and often overlooked source of information in CALL research (Egbert, 2009).

4.2.2.3.1. Predictors of software use in school context

Studies have examined teachers' perspectives to discover what drives and prevents the use of educational software within classrooms. Kaimara et al. (2021) groups the found barriers within the Greek education system into external and internal category. The major external barriers found were lack of financial resources, lack of infrastructure and lack of policy and framework, whereas the internal barriers were the preference to traditional teaching methods and lack of training. Similarly, in their study of teaching practice across primary and secondary schools in China, Wang et al. (2019) divide the predictors of digital educational resource use into school-level and teacher-level factors. However, here none of the school-level factors were associated with the use of digital educational resources, whereas teacher-level factors such as attitudes, knowledge and skills of teachers predicted the use of digital resources. In their review on the impact of digital technologies on education, Timotheou et al. (2022) highlight lack of digital competencies as one of the most common challenges appearing in the literature. This has been reported as one of the barriers both on the side of the students, as well as teachers.

Overall, the predictors of use are context-dependent and differ between countries (Kaimara et al., 2021; Wang et al., 2019), as well as different school subjects (Rončević

Zubković et al., 2022), thus emphasizing the need to examine the impact of different factors influencing language learning classrooms.

4.2.2.4. The present study

Classroom studies to date have focused on a single technology or technology type (Jamieson & Chapelle, 2010; Zhang & Zhou, 2022), and/or focus only on their use of in teaching English as a foreign language (Yang & Shadiev, 2020; Zhang & Zhou, 2022). The main aim of the present study is to understand the use and quality of digital language learning software within the context of UK primary and secondary school classrooms. By approaching language teachers, we aimed to gain comprehensive insights into their use and quality, benefiting from teachers' valuable experience.

To address these gaps, the main aim of the present study was to understand the patterns of use and quality of digital language learning software within the specific context of UK primary and secondary school classrooms. By approaching language teachers, we aimed to gain comprehensive insights into their use and quality, benefiting from teachers' experience and expertise.

Our research questions related both to (1) identifying the patterns of use of language learning software across the UK classrooms and (2) to assessing the quality of software used.

In relation to the first goal, we aimed to outline the patterns of use through:

a) understanding the nature of the digital environment in modern foreign language classrooms (e.g. how equipped are the language classrooms across the UK, are there prerequisites to the use of software?),

b) identifying the specific digital language learning resources used in the language classrooms and the dynamics of their use,

c) identifying specific teacher- and school-level factors that explain the use of digital educational resources.

In relation to the second goal, our aim was to rate language learning software based on the theoretical principles of learning (i.e. how do the features of digital language learning tools used in the classrooms relate to the theories of language learning?).

Capturing a snapshot of the current state of the use of digital language learning resources in the UK provides valuable and updated insights, with the ultimate goal of informing educational policies. Given that investing in technology does not automatically increase the use of these tools by educators (Yang & Huang, 2008, in Wang et al., 2019), nor do markers such as price or popularity of a software often correlate with its quality (Callaghan & Reich, 2018; Kolak et al., 2021), we hoped that evaluation guided by the theoretical principles of four pillars of learning (Hirsh-Pasek et al., 2015) and informed by language learning theory, would inform practice and future software development.

4.2.3. Methodology

We employed a questionnaire to collect data from second language learning teachers in the UK. The questionnaire included a rating tool that was internally constructed to reflect the theoretical frameworks relating to the principles of learning.

4.2.3.1. Participants

To ensure an appropriate and representative sample, a stratified random sampling method was employed. Since there are no exhaustive lists of second language teachers available in the UK (due to data protection rules), lists of schools available from government websites were used as the sample frame. The total number of primary and secondary schools listed across the UK (N=24,916) was divided into strata according to school type and location, including both private and state schools. Appropriate relative sizes were calculated for each stratum by using Cochran's sample size formula, with additional modification for strata of a smaller size where applicable (Bartlett et al., 2001). The total stratified sample of primary and secondary schools obtained in this way was N=3,294. The schools for this sample were randomly selected from each strata. In addition, two replacement schools were randomly chosen for each of the sampled schools during the initial sampling to account for missing responses. Schools were contacted with the information about the study to be forwarded to their language teacher(s). Based on previous studies that recruited in a similar way, the expected return rate was 10% (Dörnyei &Taguchi, 2010), which would be 329 teachers. In case of no reply, two follow-up emails were sent as reminders.

4.2.3.2. Materials

The questionnaire used was developed for the purpose of this study and consisted of five parts: (a) Demographic questions, (b) Digital environment in modern foreign language teaching, (c) Digital language learning resources**,** (d) Software rating and (e) Vocabulary teaching methods. The questionnaire is stored in an online repository (https://osf.io/2w8fk/files/osfstorage).

In the present study we focus on the first four parts. After the first section with introductory demographic questions, sections two and three were designed to inform on the situation in the classrooms relating to the available equipment and the habits of use of digital language learning resources. The fourth section consisted of a software rating instrument. The following paragraphs describe each of the questionnaire sections in more detail.

Demographic questions. The first section consisted of eight demographic questions which capture the properties of the sample and provide information for comparisons between different groups. The questions were designed to inform about the school- or teacher-level variables, such as the type and location of school or age and experience of the teacher. The questions were structured as multiple-choice questions.

Digital environment in modern foreign language teaching. The questions in the second section related to the digital language learning environment in modern foreign language classrooms and the equipment available to language teachers. Studies have shown that lack of resources in terms of institutional and infrastructure issues are consistently regarded as one of the major barriers to implementation of digital tools in education (Kaimara et al., 2021; Sánchez-Mena & Martí-Parreño, 2017; Timotheou et al., 2023). Therefore, through nine multiple-choice and Likert-type questions this section explored the language classroom equipment, as well as the frequency of teachers' and students' access to particular tools.

Digital language learning resources. The third part narrowed down to specific language learning resources and their use. There are different ways of categorizing educational software used in second language learning and teaching, ranging from the age of the intended user, to their content or type of software. For the purpose of this questionnaire, a part of the taxonomy by Rosel-Aguilar (2017) and Reinders and Pegrum (2015) was adopted, in which differentiation is made between applications designed specifically for language learning ("dedicated") and applications designed for a different purpose that serve as a tool for language learning ("generic"). Furthermore, to achieve a more fine-grained distinction, the group of software designed for language learning was divided into those intended for individual use and the those intended specifically for classroom use. A separate category was formed around the use of social media apps for language learning as they have also been explored as a useful tool both within language learning research and practice (Reinhardt, 2019). For each of the categories there was a section consisting of four open-ended questions where the respondents were asked to name different software that they use, as well as Likert-type question about the frequency of use. Additionally, there was an open-ended question about the software from different categories that the participants discontinued to use and the reason for discontinuing. Those answers were coded to analyse the most common obstacles to software use in teaching.

Software rating. The fourth part consisted of the rating tool and focused on the quality ratings of individual applications or software used. Participants were instructed to choose from language learning software that they listed in the previous sections the ones that they are most familiar with. Since it has been estimated that the rapidly evolving market makes it practically impossible to evaluate every educational app (Starke, 2020), this study focused on the ones that are most relevant in terms of their actual use among practitioners.

Creation of this evaluation tool was guided by McMurry et al. (2016), who issued a call to improve the evaluation tools in CALL by following standards set by the field of formal evaluation. In their work they introduce the framework derived from the formal evaluators such as those of the American Evaluation Association (AEA) and use it to review the two of the most prominent CALL evaluation frameworks by Hubbard (2011) and Chapelle (2001). They suggest that in creating a CALL evaluation tool the following steps should be considered: (a) identifying the evaluand, (b) identifying stakeholders, (c) determining the purpose of the evaluation, (d) selecting an evaluation type, (e) setting evaluation criteria, (f) asking evaluation questions, (g) collecting and analyzing the data, (h) reporting findings and implications, and (i) evaluating the evaluation.

In setting the evaluation criteria of the tool, the principle-based approach was followed (Jamieson et al., 2004), meaning that the development of the evaluation tool was guided by the pre-defined theoretical principles. These were operationalised through a set of questions forming specific variables. The questions were based on the theoretical accounts of language learning and educational software design which was described previously. The structure reflecting the learning principles is outlined in Figure 22.

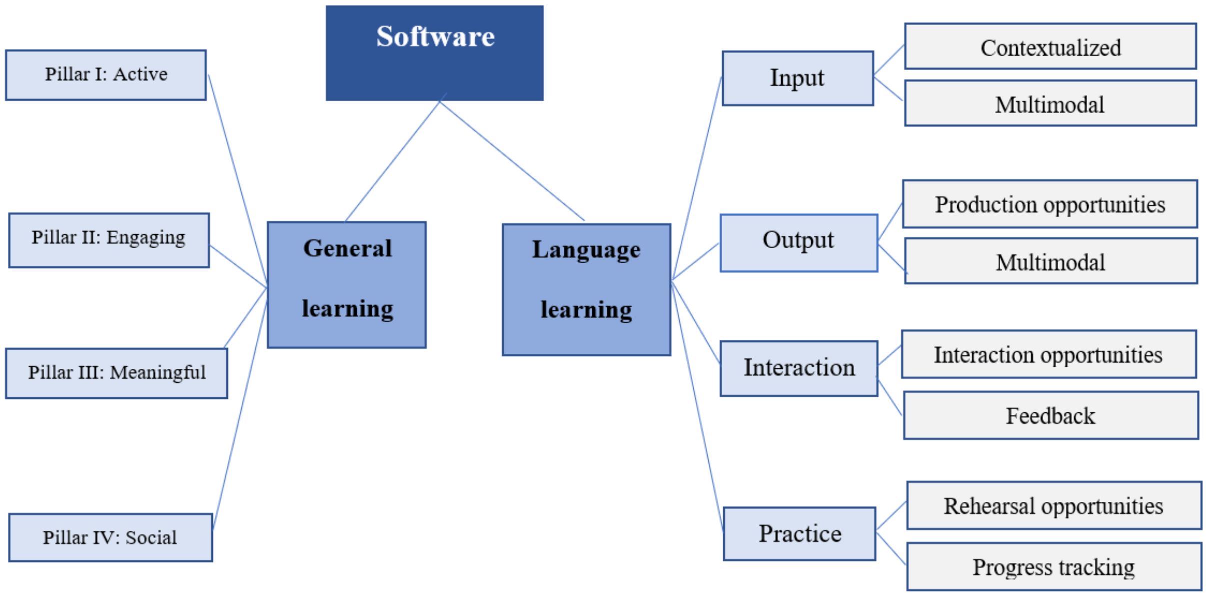

Figure 18. The overview of the subcomponents of the two main sections that contribute to the final rating score

In terms of educational design, the questionnaire investigated whether the applications follow the principles defined by Hirsh-Pasek et al (2015). In this questionnaire they were addressed through eight Likert-type questions corresponding to the four pillars of learning; whether the child is active during the use of the software, if the software promotes engagement, if the learning is meaningful and connects to childrens' experiences and if it promotes interactions with characters or people. The accuracy of statements describing how the software reflects the four principles was rated (e.g. '*The topics in the software are relevant to learner's experience*').

Language learning principles relating to input, output, interaction and practice were assessed through Likert-type questions that reflected the theoretical background of each of the components. The goal was to evaluate how comprehensively the software manages to address the concepts crucial for successful language learning. For input, the questions addressed contextualization of the language input (whether words are presented in isolation or in context), appropriate complexity (if there is a gradual increase in difficulty) and input modalities (if the input is presented through multiple modalities). Output was addressed through questions about the software providing opportunities for producing output and whether this is represented in different forms (speaking, writing). Regarding interaction, it was examined whether the software enables interactions (real or simulated) and feedback that is meaningful and constructive. Finally, rehearsal was rated through questions about the opportunities to rehearse and recycle the learned materials, as well as tracking of the learners' progress.

To evaluate and ensure the appropriateness of the questionnaire, language teachers and researchers in the field of L2 learning were consulted throughout the process. In the preparation stages we conducted two experts to obtain their insights. After constructing the questionnaire, pre-testing validation through expert reviews (N=4) was conducted to ensure that the items are clear, relevant and representative of the intended constructs. This process resulted mostly in changes regarding wording that improved the representatives of questions for specific contexts.

4.2.3.3. Procedure

Recruitment took place through online communication. Initial emails with Invitation Letters and survey links attached were sent out to selected schools' and headteachers' email addresses, which were publicly available in the gov.uk online registries or on the schools' websites. They were asked to forward the emails to all the foreign language teachers in their school, who were invited to take part by following an attached link. Two follow-up emails with reminders were sent. The questionnaire was hosted on the Qualtrics platform. The questionnaire took 30 minutes or less to complete.

4.2.4. Results

4.2.4.1. Demographic information

The final number of the respondents included in the analysis was 118 language teachers from primary and secondary schools across the UK, representing a response rate of 3.6%. The data from participants who completed at least the first two sections were included in the analysis. The detailed demographic data are listed in the Table 12.

Table 12. Demographic information about the sample (teachers completing the questionnaire; N=118)

Variable	Category	Frequency	Percent
Age	20-29	10	8.5%
	30-39	33	28.0%
	40-49	49	41.5%
	50-59	24	20.3%
	>60	2	1.7%
Gender	Female	94	17.2%
	Male	20	81.0%
	Other	2	1.7%
Years of teaching	<5	21	17.8%
	5-10	23	19.5%
	11-20	48	40.7%
	21-30	24	20.3%
	30+	2	1.7%
Type of school	Primary	46	39.0%
	Secondary	71	60.2%
	Mixed	4	3.4%
Type of school (funding)	State	105	88.98%
	Private	11	9.32%
Location	England	102	86.4%
	Scotland	7	5.9%
	Wales	2	1.7%
	Northern Ireland	7	5.9%
Languages taught	French	98	83.1%
	Spanish	57	48.3%
	German	33	28.0%
	Italian	8	6.7%
	Other	13	11.0%

4.2.4.2. The nature of the digital environment in language classrooms

It was first investigated if the classrooms have the basic infrastructure to support the use of technology in teaching, as this is commonly mentioned as one of the frequently occurring barriers to use (Sánchez-Mena & Martí-Parreño, 2017; Timotheou et al., 2022). The results showed that the classrooms in our sample are well-equipped, containing the basic prerequisites for implementing software for classroom activities (see Figure 23 and Table). The average number of digital devices reported is between 5 and 6 and there were no classrooms in the sample not equipped with at least some kind of technology accessible to teachers. Most of the respondents reported that their classrooms have Internet connection (94.1%) and sound speakers (89%) and over half of the participants indicated that they also have an interactive whiteboard, projector and a PC computer (Figure 23). These items together create a basic set-up for the classroom use of software.

The availability of necessary prerequisites for the classroom use of software is reflected in high levels of usage. In over 85% of cases the main equipment has been reported to be in use "every day". In contrast, classrooms are not equipped at a similar level with technologies for individual use, such as tablets and smartphones, resulting naturally in their lower general frequency of use. However, even when available, they are used less frequently. Particular devices and their frequencies of use can be found in the Table 13.

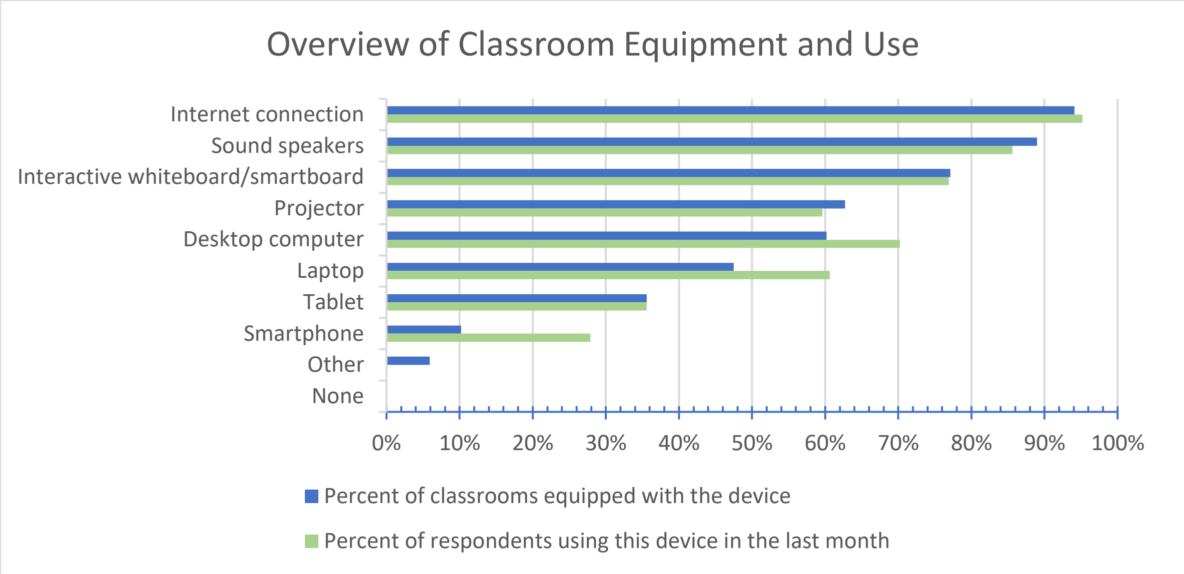

Figure 23. Bar chart showing the percent of classrooms that have access to particular digital equipment and if they had been used over the last month

Table 13. Table showing frequency of use of technical equipment within the classroom

	Internet connection	Interactive whiteboard /smartboard	Desktop computer	Sound speakers	Laptop	Tablet(s)	Smartphone(s)
Every day	92.80%	91.20%	87.30%	85.70%	58.20%	34.10%	25.00%
3-4 times a week	1.80%	2.20%	1.40%	12.40%	12.70%	14.60%	16.70%
1-2 times a week	4.50%	2.20%	1.40%	1.00%	7.30%	19.50%	41.70%
Up to 3 times a month	0.90%	1.10%	2.80%	1.00%	9.10%	7.30%	8.30%
Once per month	0%	1.10%	0%	0%	1.80%	12.20%	0.00%
A few times per year	0%	0%	4.20%	0%	7.30%	4.90%	8.30%
Never	0%	2.20%	2.80%	0%	3.60%	7.30%	0.00%
N Respondents that have access to particular equipment	111	91	71	105	55	41	12

4.2.4.2.1. The digital language learning resources used in language teaching

Participants were asked to list the names of the software they use. The results showed that language teachers actively use and are familiar with a large number of digital resources. Collectively, teachers identified over 100 different types of software that they use now or have used in the past. Out of the participants who use digital learning resources, more than a half use at least 3 different types of learning software. One third use 4 different software types and

10% listed that they use more than 5. Given a great diversity of software utilised, the questions were grouped to focus on two large categories of software. One group of questions addressed specific language learning software and the other non-specific types of software that are being used for language learning, but it is not their main purpose. The most used language learning software was Quizlet, with 27 occurrences, followed by Linguascope (N = 24). When asked about the generic resources used in the lessons by far the most frequent response was YouTube. The top 10 digital language learning resources from each category are shown in Table 14. The responses illustrate some overlap in the software listed in both groups (marked by an asterisk; see Discussion). While until recently the classification of software and applications used in second language learning grouped them into language specific and generic (Reinders & Pegrum, 2015; Rosel-Aguilar, 2017), responses in this survey indicate that this binary classification might no longer be applicable. Rather, educational platforms appear to be a new tool that supports learning in a different way than traditional apps and software, by combining aspects of both language-specific and generic software. They offer customizable templates, that can be used for any subject, but also include language-specific features. They also allow teachers to modify and create content for language learning, while maintaining general educational frameworks. For example, while Kahoot was not designed specifically for language learning, it is widely used in language classrooms possibly because teachers can create language-focused games and activities using its general quiz format. Similarly, Quizlet began as a flashcard tool, but evolved to include language-specific features which can be implemented into classrooms.

Table 14. Top 10 most frequently named software used in language classrooms

Language learning software		
Rank	**Software**	**Frequency**
1	Quizlet*	27
2	Linguascope	24
3	Blooket*	19

Generic software		
Rank	**Software**	**Frequency**
1	Youtube	69
2	Kahoot*	6
3	BBC	6

4	Kahoot*	15		4	Online dictionaries	5
5	Active Learn	14		5	Blooket*	4
6	Language Gym	9		6	Quizlet*	2
7	Kerboodle*	8		7	Kerboodle*	2
8	Languages online	7		8	Powerpoint	2
9	Memrise	7		9	Purplemash	2
10	Wordwall	7		10	Vimeo	2

When it comes to specific language learning software, around one quarter of respondents (24.5%) use them every day and 31.1% use them on a weekly basis – from 1 to 4 times per week (Figure 24). Almost 20% of teachers use them on a monthly basis (between 1 and 3 times per month). 14.2% of teacher use them a few times per year and only 10.4% never use them.

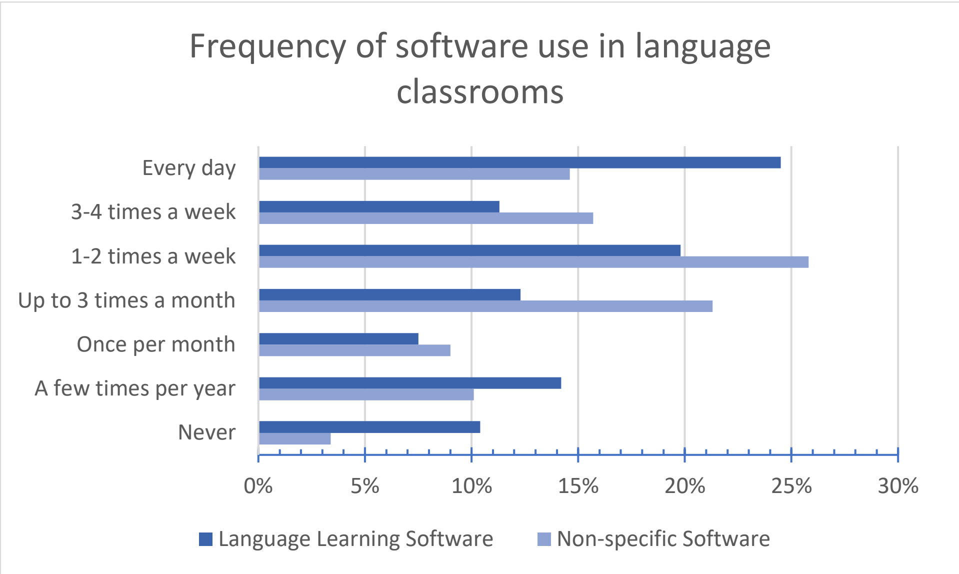

Figure 24. Comparison of the frequency of use of software classified as specific language learning software as opposed to non-specific (generic) software

The second group of software, language non-specific software, is used slightly less frequently. Compared to specific language learning software, less respondents answered that they use them on a daily basis. Mostly they are used several times per week, ranging from 1-2 times per week (25.8%) to 3-4 times per week (15.7%). A large number of participants use them up to 3 times a month (21.3%). A smaller number of participants uses them less frequently, ranging from once per month (9%), over few times per year (10.1%) to never (3.4%).

When comparing the distributions of frequencies, it can be noticed that the language learning software are more frequently used on a daily basis, whereas the non-specific software are usually used on a weekly or bi-weekly basis.

4.2.4.2.2. Reasons for discontinuing the use of software.

For every type of app, the participants were also asked about the software that they started to use, but discontinued using, and the main reasons for this. The answers were open-ended and analysed in a qualitative way through the use of codes and themes. The results showed that the main obstacles to their use were expenses, quality of the content and the overall usability of the software (Table 15).

Table 15. Table showing most frequent reasons for discontinuing the use of language learning software that respondents named

Reason for discontinuing	**N**
Cost	21
Content	14
Usability	12
Student engagement level	6
Preparation time	4
Outdated	2
Effectiveness	1
Privacy issues	1

4.2.4.3. Predictors of digital language learning resources use in teaching

To understand what drives the use of software in language learning classrooms, an ordinal logistic regression model was employed. The examined predictors included respondents’ age, years of experience, availability of equipment, and school type (primary or secondary), which were all found to be related to software use in previous studies (Kaimara et al, 2021; Timotheou et al., 2022; Vermeulen et al, 2017; Wang et al.,2019). The response variable was the response to question about the frequency of use of language learning software, scored on a 7-point Likert scale. Overall, the full model was significant when compared to the null model (likelihood ratio test comparing full and null model: $\chi2 = 36.28$, *df* = 4, $p < .001$). With the significance criterion of $\alpha < 0.05$ for p-values, the results indicated that age, experience and equipment did not affect the use of language learning software (see Table 16). The only variable that emerged as a significant predictor was school type, showing that software is used significantly more in secondary than in primary schools. The size of this effect was estimated using the odds ratio (*OR* = 4.68), indicating a large effect (Chen et al., 2010).

Table 16. Table showing the model and the outcomes of ordinal logistic regression

Model: Frequency of use ~ Age + Years of experience + Classroom equipment + School type				
Coefficients:	Estimate	Std. Error	Z value	p-value
Age	0.195	0.249	0.781	0.435
Experience	0.056	0.222	0.250	0.803
Equipment	0.054	0.116	0.464	0.642
School type	1.544	0.42399	3.642	<0.01 ***

The reason behind age and teaching experience not influencing technology adoption might be due to the institutional and pedagogical expectation playing a more dominant role in driving software use than individual characteristics. Additionally, growing immersion of digital tools in education may be diminishing age-related and experience-related barriers. As digital resources become more intuitive, standardized and widely integrated, teachers across all age groups and experience levels may feel increasingly comfortable using them, reducing the previously observed generational divide. On the other hand, the reason might be the opposite – despite the frequent assumption that the younger users will be more advanced when it comes to new digital technologies, Reddy et al. (2022a) warn that a gap exists between personal and academic use of technology. Although younger educators might be more frequent users of digital technologies in their personal time, this might not be reflected in their adoption of digital tools in their classrooms. Regarding availability of equipment, it may not be a strong factor in contexts where the infrastructure is already well-developed. Since most schools in our study had adequate technological resources, it might have not come up as a differentiating factor driving the software use.

4.2.4.4. The ratings of the digital language learning resources

In the software rating section participants chose one language learning software type that they are most familiar with from the list that they provided in the previous sections. Eighty-one participant responded to this section. Since this rating instrument was developed with language learning software in mind, excluded from the list were software non-specific to language learning (e.g. YouTube), resulting in 74 ratings.

Since this part of the questionnaire was scale-based, we conducted validity analysis upon collecting the data. Validity and internal consistency of the questionnaire were assessed using Cronbach's Alpha. Three out of eight scales (scales measuring output, interaction and practice) demonstrated good reliability, exceeding the α=0.7 threshold (α=0.75, α=0.73 and α=0.73). Three additional scales (scales measuring input, meaningfulness and social components) fell close below the threshold, indicating moderate consistency (α=0.59, α=0.66 and α=0.59). Two scales (scales 'active' and 'engagement') fell further from the threshold (α=0.3 and α=0.26), indicating the need for further refinement of this part in future research.

Grouped software ratings. First of all, the ratings for all of the rated software were combined to show an overview of the properties of currently used language learning software. The mean ratings for each of the components for both general and language learning principles are shown in the Appendix C and the overview shown in Figure 25.

The evaluation of the software revealed a range of mean scores for the rated categories from 3.20 to 5.53 out of a possible 7. When analyzing the results within the framework of language and general learning principles, distinct patterns emerged. In the domain of language principles Input produced the highest ratings (M = 5.53, SD = 1.04), while both Output (M = 4.30, SD = 1.68) and Interaction (M = 4.30, SD = 1.56) received poor ratings. Within the general learning principles, Engagement stood out (M = 5.53, SD = 1.30) as a particularly strong aspect, while the Social learning principle scored the lowest (M = 3.20, SD = 1.92).

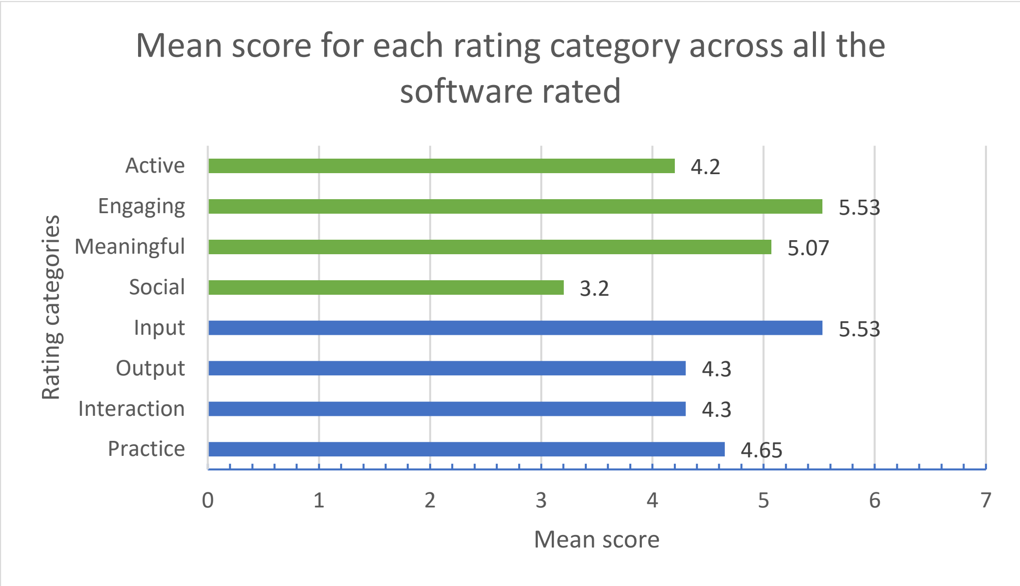

Figure 25. Bar plot showing how were the software rated on each of the subcomponents

Individual software ratings. In this section, the ratings of all the individual software which were chosen by at least two respondents were analysed, which resulted in 62 ratings of 11 different software types.

In the final table (Table 17), the ratings are summarised to create a final total score. The software with the highest overall score was Language Angels.

Table 17. Table showing total ratings according to individual software in descending order by the final score

TOTAL RATING				
Digital tool	**N raters**	**General learning principles**	**Language learning principles**	**Final (total) score**
1. Language Angels*	4	21.78	25.8	47.58
2. Languagenut*	4	18.84	23.17	42.01
3. Language Gym	4	19.42	20.83	40.25
4. This Is Language*	3	20.84	18.41	39.25
5. Duolingo*	3	14	24.68	38.68
6. Activelearn	9	18.41	20.07	38.48
7. Kerboodle	4	18.22	18.29	36.51
8. Quizlet	16	17.83	18.02	35.85
9. Memrise	4	16.34	18.79	35.13
10. Kahoot	2	18.75	15.13	33.88
11. Lingua Scope	9	16.97	14.47	31.44

* the asterisks indicate the apps which are not among the top 10 most frequently used

Software ratings – grouped by software price. The study also compared the software according to their pricing, since previous studies showed that price did not guarantee quality among the educational apps downloaded from the mobile app stores (Callaghan & Reich, 2018; Kolak et al., 2021). To establish if there is a significant difference between the software according to their cost the software was grouped into free versus paid and a Wilcoxon rank-sum test was used. The outcome variables were the scores on the subcomponents of the questionnaire that were relating either to the theory of general learning principles, language learning principles or teachers' satisfaction. The results showed that the cost of the software had no effect on the rating on any of the subcomponents (all $p > .05$). Detailed analysis outputs can be found in the Appendix C. This is in line with findings from Kolak et al. (2021), who found that the difference between free and paid educational apps was not present on the measures of content quality, but in the frequency of animations and elements on screen. This suggests that the price cannot be seen as an indicator of the educational quality and it might instead be driven by branding, marketing or additional features focused on entertainment.

4.2.5. Discussion

The results of the survey provide insights into the state of the use of language learning software and their quality vis-a-vis ratings of dimensions pertaining to theoretical principles of language learning and learning in general. In relation to our aims, the study identified the patterns of use of language learning software across the UK classrooms in terms of highlighting specific software used, their frequency of use, as well as their quality.

When it comes to understanding the drivers of digital software use, the results are in contrast with previous findings. Although it was expected that both the school-level and teacher-level variables would predict the frequency of use, this was not the case among this sample. The only predictor that emerged as significant was type of school, but not in the direction expected based on the previous research. Whereas Wang (2019) reports how, in China, the teachers in primary schools used digital tools more frequently and believed in their efficiency more than their secondary school colleagues, results of this study were in the opposite direction, with higher reports of software use in the secondary schools compared to primary schools. This might be due to the fact that in the UK, and to contrast to education policy in many countries, there is generally less emphasis put on language teaching in primary schools. These findings confirm the importance of understanding different contexts when evaluating use and quality of language learning software (Jamieson & Chapelle, 2010). At the same time the findings reveal that classrooms in our sample are well-equipped with the infrastructure necessary for integrating technology into language teaching, potentially indicating that one of the major technical barriers to use found in other contexts and countries

(Sánchez-Mena & Martí-Parreño, 2017; Timotheou et al., 2022) does not apply to the primary and secondary schools in the UK.

The availability of technologies for language learning in the classroom is also reflected in the high frequencies of use, suggesting a consistent incorporation of digital tools into language education. With this growing number of CALL resources available, stakeholders need to become increasingly critical when choosing the most appropriate ones. Many parameters should be considered when deciding on which resources to use in language classrooms, such as cost, technical requirements or usability. However, one of the crucial aspects that is frequently not sufficiently explored, is the pedagogical and linguistic soundness of the software (O'Brien, 2020). By creating a theory-based rating tool, this study enabled a fine-grained understanding of the strengths and weaknesses of digital language learning software. This becomes particularly relevant in light of the mismatch found between the quality of software and the frequency of use. Specifically, the software which obtained the highest score on these principles was not among the top ten most used software, and the same is true for four out of the top five rated software. This is in line with results from Kolak et al. (2020), who also discovered that most used software were not necessarily the best ones in terms of their educational design. A further mismatch that was found between quality of software and price, which is also in line with other studies on educational applications (Callaghan & Reich, 2018; Kolak et al., 2021), showing that price did not guarantee quality. These findings serve as a reminder to exercise caution and be mindful about the strengths and weaknesses when choosing the language learning software to use. Both educational institutions and teachers could benefit from structured guidelines or workshops focused on assessing the pedagogical soundness of digital language learning tools to ensure that high-quality and research-based tools are more widely recognised and adopted.

The ratings in this study indicated that current software used shows strength in the areas of Input and Engagement, but weaknesses in the domains of Output, Interaction and Social learning. The high scores for Input and Practice are in line with the Chapelle's (2008) observation that second language theories relating to the nature of input and practice, such as Input Processing (Krashen, 1994) and Skill Acquisition theory (DeKeyser, 2007) offer clearer and more applicable guidelines that are easier to implement into software design. The low scores obtained in the area of Social learning and Output corresponds to drawbacks identified 5-10 years ago (Burston, 2014; Heil et al., 2016), and demonstrate that the majority of software still remain predominately behaviourist in nature, despite the pedagogical shift from behaviourism towards more communicative-based and learner-centred approaches such as constructivism. To change this, the software should aim to provide more opportunities for students to create output and contextualise language through including more components that mimic social interactions. The current state of the language learning software can still be summarised in an observation that the software are not entirely living up to their potential (Burston, 2014; Heil et al., 2016). The key to achieving their full potential would be in becoming more innovative and collaborative, thus aligning better with language and educational principles.

The wide array of language learning software now available in a rapidly changing market also marks changes to how one can conceptualise software. For example, up until recently the software and applications used in second language learning have been classified into language specific and generic (Reinders & Pegrum, 2015; Rosel-Aguilar, 2017). This distinction is questioned by the responses in this survey with educational platforms emerging as comprehensive systems that support learning on a broader scale than traditional apps and software. The recently emerged platforms, situated at the intersection of language-specific and generic software, offer adaptive and engaging learning environments, empowering teachers to tailor content while maintaining a generic core. The emergence of template-sharing communities providing pre-made language learning products that can be easily adapted further complicates the distinction between the two traditional types of software, as well as evaluation of their quality. This results in less adaptation requirements and almost a ready-made language learning product that still retains the possibility of modifying the content. The flexibility with the content creation could potentially be one of the major reasons for their success, especially seeing how the participants in the survey listed "low quality of content" as one of the main reasons against the use of some software. Notably, the most frequently used software all offer users the ability to create or modify content, underscoring the importance of customization in effective language learning tools.

One limitation of the present study is the low response rate from a large cohort of teachers sampled across the UK. This may be due to self-selection and great demands that teachers have in their workloads, limiting time. Dörnyei and Taguchi (2010) propose complementing questionnaire studies with follow-up interviews to address and further explore the findings from the questionnaire by using a "sequential exploratory design". Moreover, in addition to the judgemental evaluation analysis done in this study, Chapelle (2001) suggests that software evaluation can benefit from being complemented by the analysis of learning outcomes.

Overall, the results of the survey highlight the imperative for future software development to align more closely with social and interactive learning principles, emphasizing the need for innovation to bridge the gap between theory and current software offerings. Insights from those delivering language learning in the classroom merit careful consideration by relevant stakeholders to establish quality standards for educational software, ensuring alignment with pedagogical principles. Moreover, developers need to work hand-in-hand with teachers to ensure that applications are developed with maximum efficacy, and in line with leading pedagocial and educational principles. Furthermore, understanding teachers' habits and preferences, as well as challenges related to specific software types, can inform policies aimed at optimizing resource allocation and training for teachers.

Chapter 5 | Discussion

This dissertation set out to address several research objectives related to multimodality in second language vocabulary learning, examining it from multiple perspectives. The first aim was to deepen our understanding of the cognitive mechanisms that underpin L2 word learning. In particular, this work focused on how specific word properties and testing tasks influence learning outcomes in the two most common instructional modalities: text and pictures. Secondly, this thesis explored the processes involved in screen-based L2 vocabulary learning by setting the studies in a digital environment. A special emphasis was put on the role of social cues, to understand how the importance of social interactions transfers into screen-based L2 vocabulary instruction. Lastly, it was investigated how cognitive principles of word learning are applied in practice by developing a software evaluation tool grounded in evidence-based insights on L2 vocabulary learning. This tool was then used to assess the quality of language-learning software currently employed within UK language learning classrooms. The overarching goal of this thesis was to contribute to optimizing word learning through expanding the theoretical insights while keeping them connected to the real-world application. Thereby, scientific insights are used to develop evidence-informed instruction methods and guide users into choosing the optimal products. In turn, the effects of these methods in practice would provide new insights for research contributing to the iterative process between research, software design and real-world application, as portrayed in the Introduction chapter (Figure 1). While this approach makes research questions addressed through this thesis broad, there are multiple common threads that tie this work into a whole.

Across the studies presented in this dissertation, one of the key themes is the focus on screen-based learning. Since research already established that the mere presence of technology does not guarantee improved learning outcomes compared to traditional methods (Basöz & Can, 2016; Hao, 2021; Taghizadeh & Porkar, 2018; Young & Wang, 2014), understanding its specific impact on vocabulary learning remains essential. By situating these studies within a digital environment, this dissertation critically examined how screen-based learning is currently used, as well as how it can be leveraged to understand and enhance vocabulary acquisition. With one of the crucial possibilities offered by the digital technologies being the inclusion of multimodal learning experiences, a second unifying element across the studies was the investigation of multimodal, situated, and embodied learning approaches applied to this context. The experimental studies contrasted different modalities during the learning process and examined how they interact with other factors. They also investigated the social aspects of language use in the digital setting by unpacking the effects of fundamental deictic cues and their relation to the position of target objects on screen. The final study integrated this with investigating how is multimodality represented in the input provided by the available language learning software.

Within the experimental studies the scope was focused on vocabulary learning, as a crucial first step in L2 learning, and an important predictor of comprehension, listening, speaking, reading, and writing skills (Grabe, 2009; Li & Zhang, 2019; Stahr, 2009). At the same time, expanding initial L2 vocabulary remains one of the greatest challenges students encounter, primarily due to the limited classroom time and the often limited exposure to the target language outside the classroom (Hao, 2021). Although different theories on L2 vocabulary learning have been proposed, there is still no definitive answer on the nature of background processes involved. Considering these factors, the present studies investigated factors that interact during vocabulary learning process while also exploring how technology can support vocabulary acquisition.

5.1. Implications for understanding of cognitive processes

The findings across the studies presented contribute to understanding cognitive processes in multimodal second language vocabulary learning. The results underscore the complexity and nuanced nature of vocabulary learning. By examining how modality, word type, social cues, and ambiguity shape L2 learning, this work contributes to theories of embodied, distributional, and hybrid cognitive processes in language learning.

The first key finding highlights the importance of previously unconsidered factors of testing modality and word class in L2 vocabulary learning. The results presented in chapter 2 consistently showed that it is not only the learning modality that influences word learning outcomes, but also how the modalities at learning interact with other variables, such as testing modality and cross-linguistical ambiguity. Futhermore, results showed that text-based testing was more advantageous for verbs and prepositions, suggesting that these word types might rely more heavily on linguistic than sensorimotor grounding. This aligns with previous research suggesting that different word classes are grounded in different ways (Bultena et al., 2013; Markostamou & Coventry, 2022; Vigliocco et al., 2011), and confirms these implications in the context of second language learning. In a practical sense, the findings indicated that verbs and prepositions, which often convey more abstract relationships, benefit more from linguistic context compared to nouns.

From a theoretical perspective, these findings support a hybrid approach to cognitive processing in language learning, which acknowledges how both embodied experiences (e.g., pictures) and linguistic context (e.g., text-based exposure) play significant roles in language processing and learning (Borghi, 2019; Borghi & Binkofski, 2014; Dove et al., 2020). Crucially, these approaches emphasize how different words engage different processes in their representation. Specifically, conceptualization of words that tend to be more abstract than (concrete) nouns, such as verbs and prepositions, is influenced more by language contexts and distributional learning, whereas concrete words, most typically nouns, are more influenced by sensory-motor experiences, activating embodied representations more intensely (Borghi & Binkofski, 2014). The difference in concreteness also contributes to more abstract words having more pronounced cross-linguistical differences compared to concrete words, which are more heavily constrained by the environment and therefore conceptualized in a similar way across languages (Gentner & Boroditsky, 2001). This nuanced view emphasizes that word type and context greatly affect how meaning is processed and learned, underscoring the importance of implementing hybrid approaches into the domain of second language acquisition, where both embodied interactions and linguistic exposure are essential.

The debate about semantic representations has evolved beyond the dichotomy between purely linguistic and purely embodied views (Vigliocco et al. 2009; Zwaan, 2009). The updated hybrid views hold that concepts are represented both at the linguistic level and through sensory-motor and affective experiences of the world (Barsalou et al., 2008; Borghi, 2019; Dove, 2011; Vigliocco et al., 2009; Zwaan, 2014). These accounts acknowledge that different types of information - sensorimotor, affective and linguistic – contribute to meaning representation in varying degrees. This aligns with Zwaan's (2009) proposal to move past examining whether cognition is embodied and to start answering the question "how is cognition embodied?", to encourage researchers to develop more detailed accounts of embodiment. The current thesis extend this proposal into the domain of second language vocabulary learning. The results demonstrate that the impact of multimodaly is relative and that the learning outcomes also depend on both word type and testing modality.

Another important aspect of L2 learning apart from multimodality is that it is inherently situated in a social context (Çetinçelik et al., 2021; Kreysa et al., 2018; Li & Jeong, 2020).

Introducing social context into L2 vocabulary learning is thought to promote and enrich mental representations. This has also been confirmed in neuroimaging studies showing that partner-based learning or learning L2 vocabulary through simulated social interactions resulted in neural activation similar to L1 language processing (Li & Jeong, 2020; Verga & Kotz, 2019). However, as noted by Li and Jeong (2020) it is important to unpack this premise to show which aspects of social context and in which learning environments bring benefit. A step in that direction was to examine the role of social cues such as gesture and gaze, in digital L2 learning environments.

While gesture and gaze cues did not impact learning in this setup (chapter 3), the proximity of the target object to the agent was found to be relevant. This finding suggests that while certain social cues on their own might not directly affect L2 vocabulary acquisition in a digital experimental setting, the presence of an interactive agent can still impact learning. The position of referent relative to the interlocutor has already been found to affect language processing dependent on peripersonal vs extra-personal object placement (Coventry et al., 2008). The research shows that how we use spatial language (e.g. demonstratives like 'this' and 'that') is strongly linked to whether objects are within our reach (Coventry et al., 2008; Coventry et al., 2014). Furthermore, the concept of peripersonal space has also been demonstrated to have an effect on attention. Objects that are near our hands receive more attention and are processed more quickly by our visual system compared to objects that are farther away (Reed & Park, 2021). Crucially for the digital contexts, these effects of peripersonal space on language and attention extend beyond one's own peripersonal space onto the interlocutor's peripersonal space (Brozzoli et al., 2013; Coventry et al., 2014). In their fMRI study, Brozzoli and colleagues (2013) revealed similar neural responses to objects near participants' own hand or another person's hand.

The present results align with these findings. By showing how the proximity of objects to the virtual agent in the study can have a positive effect on the learning outcomes, this study expands previous findings to consider the influence of agent-object distance in the context of digital second language learning. From a broader perspective, this could be related to theories that underscore the importance of the social aspect of learning. For example, the Social Agency Theory (Mayer et al., 2003) posits that including social cues in multimedia learning can activate the social conversation schema in learners, thereby creating a sense of partnership between the learner and the instructional agent. More recent findings extend this to second language vocabulary learning by showing that adding social context into the learning situation is associated with deeper cognitive processing and better learning outcomes (Jeong et al., 2020; Verga & Kotz, 2019). Therefore, in this case the presence of the agent, in combination with their proximity to the target object might create a learning environment that activates aspects of social learning, resulting in a positive learning effect.

Finally, although this was not the focal point of the present work, it was confirmed that word class is a significant factor in L2 vocabulary learning, with nouns being easier to learn than verbs or prepositions (Gentner, 1982; Waxman et al., 2013). Likewise, the findings support the well-known notion that ambiguous words – those with multiple meanings and translations - are more difficult to learn (Degani & Tokowicz, 2010). In the first study these findings served as a positive control to confirm the validity and reliability of the experimental methods. Positive controls are conditions or groups within an experiment that are expected to produce a known and predictable outcome. In addition to confirming previous findings, by demonstrating established effects such as the greater ease of learning nouns over verbs and the added difficulty of learning ambiguous words, the study confirmed that the experimental setup could reliably detect expected patterns in L2 vocabulary learning. By verifying that the study conditions could replicate known effects, the positive controls helped confirm the validity and reliability of the experiment, lending greater confidence to the novel insights investigated in this research.

Overall, the findings from this work reinforce the importance of hybrid cognitive models in explaining L2 vocabulary learning. Purely embodied or distributional approaches may each fall short of fully capturing the complexities observed in the learning process. Hybrid accounts, on the other hand, suggest that both embodied and linguistic elements are necessary to create L2 vocabulary representations. For example, while embodied learning strategies (such as visual representations) may effectively support concrete nouns, distributional strategies (such as repeated exposure to text-based contexts) are better suited for words that tend to be more abstract, such as prepositions. This layered and complex structure connecting factors relevant in L2 vocabulary learning underscores the need for an integrative model that incorporates both sensorimotor and linguistic grounding and remains flexible and adaptive to different types of learning content.

5.2. Implications for software design and practical application in L2 vocabulary learning

In addition to informing theoretical approaches, it is important that the insights obtained through research find their way into practice. When integrating cognitive and linguistic theories with the results observed in the present studies, the overarching theme is the nuanced nature of second language vocabulary learning. The observed interaction between modality, word class, and ambiguity suggests that vocabulary acquisition is not a one-size-fits-all process. Therefore, the emphasis is put on the importance of flexible approach that combines different learning methods and adapts to specific learning content, as well as learning contexts.

The hybrid approach—integrating embodied and distributional perspectives—presents a comprehensive framework for understanding vocabulary learning that aligns with the present results. Concrete words, with their strong sensory-motor ties, benefit from embodied experiences like images, or gestures. In contrast, abstract terms, such as verbs and prepositions, benefit more from contextual and text-based exposure, as these words depend on distributional and linguistic cues to establish meaning. This could be utilized both in software design and in teaching to adapt the input type and retrieval cues to specific target words. For example, pairing new vocabulary with physical actions or context-rich imagery can enhance learners' retention or retrieval of concrete nouns. On the other hand, incorporating sentence-level or usage-based examples helps learners grasp more abstract words such as prepositions by providing the linguistic context needed to form and access the existing meaning representations. The importance of acknowledging individual word properties in L2 word learning has already been recognised by cognitive linguistic approaches. They particularly focus on the types of words that are generally more difficult to learn, such as prepositions, modal verbs, conditionals and cross-linguistically ambiguous words (Jacobsen, 2018; Tyler et al., 2010; Tyler et al., 2011). To enhance their learning outcomes in L2 learning, they recommend using specifically-tailored activities that capture the embodied, physical and spatial meanings of particular words to help in creating their representations or image schemas (Zhao, 2020). Furthermore, the present findings can also be associated with the Dual Code Theory, which emphasizes that concepts are processed through both sensory-motor and language-based systems, which calls for a need to incorporate multiple input modalities to meet diverse learning needs. However, in case of screen-based L2 vocabulary learning, more is not necessarily better (Zhang et al., 2013). The effect of cognitive load that emerges when too much diverse input is introduced can mask the multimodality benefit. For example, de Nooijer et al. (2014) showed that enacting and imitating gestures during L2 verb learning had a positive effect on students with good language skills, whereas it hindered the learners with poor language skills. On the other hand, when it comes to virtual reality (VR), struggling students seemed to benefit more from it than successful learners (Legault et al., 2019). Therefore, the introducing of different modalities should be carefully curated to match specific word properties, but also cognitive demands and individual differences.

The same is true for the use of social cues and virtual agents in digital-based L2 vocabulary learning. Pedagogical agents are digital characters that show human-like appearance and behaviours, such as of speech, gesture and movement (Park, 2015). They have become increasingly popular in educational software as they are thought to facilitate learning through the positive effect of social enrichment and embodiment (Frachette & Moreno, 2010; Gulz & Haake, 2006; Park, 2015). However, other studies have found that they might have no effect or even contribute to the cognitive load by diverting the learner’s attention from relevant information (Frechette & Moreno, 2010; Mayer & Dapra, 2012). To disentangle these effects, the present work considered the crucial social elements of virtual agents related to deictic cueing in L2 vocabulary learning, namely pointing gesture and eye-gaze. The findings indicate that while isolated social cues do not significantly impact learning in controlled experimental settings, the proximity of interactive agents to the object representing the target L2 word can influence learning. These insights align with Social Agency Theory (Mayer et al., 2003), which posits that social elements in multimedia learning can encourage cognitive engagement. Therefore, software design might be enhanced by incorporating elements that create a social presence, as the inter-personal space of the agent seemed to have a positive effect on remembering the target words. However, more research should be done to establish if the social cues would have an effect in slightly different experimental setting that would create a more realistic social presence by using dynamic and immersive contexts.

The findings in this thesis draw attention to important, yet unconsidered factors in second language vocabulary learning. However, they cannot serve as exhaustive or comprehensive guidelines, as the research into these areas is still in its early stages and requires further exploration to build a more complete understanding. The optimal way to build and improve software would be through an iterative process with research findings informing software design and software use further informing research (Li & Lan, 2022). There are examples of language learning software companies that follow this approach and benefit from using the information about software use and learning outcomes to further enhance their product, but also to contribute to a deeper understanding of language learning process.

However, while the market is still saturated with software of low quality, it is crucial for users to be able to evaluate the available software (Kolak et al., 2021). As shown in chapter 4, there is a mismatch between the most used and the highest rated software, indicating how more should be done to control the quality of educational software used across different settings. To make informed choices, evaluation frameworks should include both linguistic and cognitive criteria and most importantly, be made available for end-users. Likewise, integrating user-centred feedback into the research and design processes is essential to enhance the quality at all three levels.

The insights across the present research offer a foundation for designing software that supports nuanced, flexible, and research-informed L2 vocabulary learning. As digital language learning tools evolve, embracing hybrid models and collaborating with educators will be key to creating learning experiences that address the complex cognitive processes involved in second language vocabulary learning.

5.3. Synthesis of findings: Towards an integrated understanding of multimodal L2 vocabulary learning

While each study in this thesis examined a different aspect of multimodal L2 vocabulary learning, together they begin to reveal a coherent picture of the complex, interconnected factors that influence learning outcomes. The findings across studies show how are fundamental principles of vocabulary language learning applied to a screen-based setting and motivate further research.

One of the emerging patterns across the studies is the complexity of the modality effects. First study demonstrated that the effectiveness of specific modalities is dependent on interacting variables of word class and testing modality, confirming previous suggestions that word class moderates the effects of multimodality (Rosenthal-von der Putten et al., 2020). Second study revealed that social cues do not universally enhance learning in all settings. Instead, the spatial proximity between virtual agents and target objects plays a role in learning, aligning with research on peripersonal space effects in language processing (Coventry et al., 2008; Coventry et al., 2014). The need for further research and the complexity of transfering the research findings is also shown through the final study. Exploring real-life application of language learning software showed that the most widely used language learning software often lacks the nuanced, adaptive features needed for successful language learning, consistent with concerns raised about the quality of eduactional technology (Kolak et al., 2021). The complex nature of multimodality in L2 vocabulary learning could be one of the reasons explaining mixed results in previous research (for the overview see Chapter 2).

The findings together speak to three different levels of L2 word learning. At the representational level, different word types engage different cognitive systems (Andrews et al., 2009; 2014; Borghi & Binkofski, 2014; Dove et al., 2020). Abstract words like verbs and prepositions rely more heavily on distributional patterns and linguistic contexts, as opposed concrete words with their strong sensorimotor associations (Borghi et al., 2019). This aligns with previous research suggesting that different words are grounded in different ways (Barsalou, 2008; Pulvermüller, 2013; Vigliocco et al., 2011). At the contextual level, learning occurs within spatially and socially situated environments (Murgiano et al., 2021). The proximity effect observed in the second study suggests that even in digital contexts, the spatial arrangement of elements on screen could activate peripersonal space mechanisms that influence attention and encoding. This finding connects individual cognitive processes with environmental factors, showing how spatial cognition extends into virtual learning environments. At the systematic level, these processes must be coordinated within learning systems. The software evaluation suggests that current educational technology fails to implement important aspects derived from the current literature and research, highlighting the research-practice gap noted by other researchers (Li & Lan, 2022).

The synthesis of findings across this thesis reveals that the multimodality in L2 vocabulary learning cannot be described in unversals. Instead, it emerges from complex interactions between modalities, word properties and learning contexts. Understanding and optimizing this process requires integrated approaches that consider these multiple factors simultaneously. Moving forward, the goal should not be to identify the "best" modality, but rather to understand how different elements can be combined to support the complex process of L2 vocabulary learning.

5.4. Limitations and implications for future research

Despite the new findings provided by this research (or precisely because of them), there is still more to uncover. Each of these studies, while offering answers, also opens up new questions and directions, ranging from theoretical to methodological and practical. Both limitations and future directions will be discussed in the following chapter, as they can be considered linked to each other.

One of the primary implications for future L2 vocabulary learning research is the need to consider additional variables such as testing modality, word properties and cross-linguistic differences. These variables have previously received limited attention in L2 vocabulary learning studies, yet the research presented here has consistently shown that they influence the learning outcomes. There were previous studies that discussed some of these effects in hindsight, as a possible explanation for their unexpected findings (Altarriba & Knickerbocker, 2011; Boers et al., 2017; Comesaña et al., 2012; Li & Jeong, 2021), but their interactions were never systematically tested.

With the present studies confirming the influence of these variables, there are new questions forming. When it comes to modality, this thesis looked at the text and picture modalities, which were most frequently tested. However, there is also a growing body of research looking into iconic gestures as an input that potentially benefits L2 vocabulary learning through mechanisms of embodiment. Since Allen's pioneering study in 1995, both behavioural (Morett et al., 2016; Repetto et al., 2017; Tellier, 2008) and neuroimaging studies (Macedonia et al., 2011; Macedonia et al., 2019, Mayer et al., 2020) have continued to find indications of the positive gesture effect in L2 vocabulary learning. Studies have contrasted observing and performing gestures (for a review, see Oppici, 2023) as well as different types of gestures and found that including meaningful as opposed to meaningless gestures leads to enhanced L2 vocabulary learning outcomes. In that way, gestures lead to deeper encoding on two levels, by involving multiple sensory and motor systems in their representations and second by inducing semantic processing (Macedonia et al., 2019). Kelly and colleagues (2009) explain the effect of gestures by suggesting that their representational meaning helps index a newly learned word to an established concept, which strengthens the connection thus enabling it to endure in memory for a longer time. The mechanism behind this enhancement is the same as proposed for the pictures (Yoshii, 2006). However, similarly as with pictures, the positive impact of gestures has not always been consistent, suggesting that while gestures are a powerful tool, their integration into L2 learning can depend on various factors (Macedonia, 2019; Morett, 2019; de Nooijer et al., 2014). The factors at play could be identified in a similar way as in the present studies. Exploring the moderating effects that different word classes have on the modality effect would be a good starting point. Considering how action verbs are associated with motor processing areas, they might benefit more from gestures as an input modality. Neuroimaging studies have demonstrated involvement of motor cortex area in learning and retrieval of L2 action verbs, but not in words for objects (Branscheidt et al., 2018; de Grauwe et al., 2014), which could suggest that also different mechanisms at learning might benefit nouns versus verbs retention in L2 learning. So far, only a few studies looked into the effect of gestures on different word classes with mostly mixed results (García-Gámez et al., 2018; Lewis & Kirkhart, 2018; Rosenthal-von der Putten et al., 2020). Therefore, a comprehensive examination of the relationship between different word classes and different modalities would help further refine our understanding of how to tailor multimodal learning strategies to specific vocabulary types.

Although this research contributed to unpacking the effect of specific word classes, it also opened up new avenues. While the present study included rarely considered verbs, prepositions and ambiguous words as a factor, there are still other word classes, as well as different types of ambiguity to examine. In their work, Degani et al. (2016) distinguish between several types of translational ambiguity, such as the direction of ambiguity between the languages – one-to-many or many-to-one mappings. Also, the ambiguity can be in word form or in word meaning, with the overlap in meaning being total or only partial. These different types of ambiguity might possibly also have a unique effect on L2 learning based on word class. For instance, a study by Bultena et al. (2013) examined effects cognates and ambiguity simultaneously and found that both noun and verb processing in bilinguals is affected by within- and between-language overlap. Results showed that verbs displayed facilitatory effect from both types of overlap, while nouns only experienced cognate effect, but no ambiguity effect. The authors explain the reason for this difference as arising from the greater complexity of verb representations as opposed to simpler noun representations, which are activated slower, thereby leaving less space for the facilitatory effects to occur. These differences could transfer also into the second language learning, which is one of the reasons to consider different types of translational ambiguity in relation to different word classes. Uncovering the nuanced relationships between these factors also contributes to a better understanding of the cognitive processes governing language learning, which would help in refining the current relevant theories.

To enable the inclusion of ambiguous words into the L2 vocabulary learning research, it is necessary to modify the existing experimental paradigms. Current research studies on these topics usually present words in isolation with their one-word L1 translations. However, in ambiguous conditions where one word has potentially multiple translations between L1 and L2, it is necessary to provide context to adequately represent their nuanced meanings (Borghi, 2019). Therefore, previous studies looking into translation ambiguity used definitions (Degani & Tokowicz, 2010; Degani et al., 2016) instead one-word translations that are typical in studies that investigate the effect of different input modalities. To address this issue, studies in chapter 2 provided both the one-word translation, but also embedded the target words into short and balanced sentences that provided context that helped represent their meanings. To be able to do that in the beginner-learners group, a mixed-language procedure was followed, where a target word is presented in L2, but the context is presented in L1 (for similar mixed-language procedure see Beatty Martínez & Dussias, 2017; García-Gámez & Macizo, 2020; Moreno et al., 2002; Proverbio et al., 2004). Future research could consider expanding this work to intermediate learners where the context could also be provided in L2, leading to a more naturalistic learning setting. This could also enable the use of a naturalistic language which has its benefits, as well as drawbacks. However, the current study introduced pseudolanguage, which in this case had the benefit of providing a control for participants' previous language exposure and psycholinguistic word properties.

In the second series of studies relating to the effects of social cues, future studies could expand the current design to also test the effects of gestures and gaze compared to no agent present. The current experiment had a baseline condition which was labeled "no cue" that had an agent who used neither gaze nor pointing to indicate the object. The reason for this was to prevent a confound that would obscure the effects of gesture and gaze in a screen-based setting. However, seeing how the current conditions were quite matched, it would be informative to compare them to no agent present or to an arrow on the screen acting as a cue. This would help reveal the nature of the cues and their role in digital language learning, whether they facilitate learning by directing attention, enriching the representation or disrupt learning by increasing cognitive load (Kreysa et al., 2018). If the results were to show that gesture and gaze are matched with an arrow, that would indicate that in this context these social cues act as pointers, without providing further enrichment. However, the arrow condition proving to be superior in relation to the learning outcome would indicate a distracting effect of social agent or the effect of cognitive load. The opposite outcome would mean that the presence of the agent potentially acts as an enrichment in the learning situation. The results could have practical implication for future software design and inclusion of agents as facilitators of L2 vocabulary learning, as research with gestures suggested that the efficiency of virtual agents can be matched with human instructors (Macedonia, Groher & Roithmayr, 2014; Macedonia, Kern & Roithmayr, 2014).

Similarly, the observed effect of proximity to the agent could be further disentangled. For example, by again replacing the agent with an inanimate pointer in one of the conditions, it would be revealed if the proximity benefits comes from the object being in the agent's interpersonal space or simply from the shorter distance between the pointer and the object. If the former is true, that would further confirm the effect of interpersonal space on language processing and learning (Coventry et al., 2008). On the other hand, if the latter is correct, that would suggest that the proximity effect benefits learning because the learner is using less effort and time to indicate the relevant object.

Taken together, both of the series of experimental studies presented in this work could be easily expanded to include different ages and different populations, including bilingual participants or language learners at different stages of their L2 learning process. Different patterns of modality benefit were shown at different ages (Chen & Leung, 1989; Comesaña et al, 2008; Poarch et al., 2015), indicating that age might be one of the relevant factors in fully understanding the modality effect and the current research findings. Although young children are considered to be better at language learning and use distinctive processes, recent work by Hartshorne and colleagues (2018) obtaining data from thousands of participants showed that this benefit extends further than previously thought, into adolescence. Therefore, exploring the modality effects at different timepoints from childhood to adulthood would help paint a complete picture of this effect and contribute to the fundamental continuity debate that aims to understand the change in language learning mechanisms between first and second language (Ellis & Wulff, 2015; Lee et al., 2009; MacWhinney, 2012).

The role of modality might also change between beginner and advanced learners (Abutalebi et al., 2001; Dijkstra & Van Heuven, 2002; Kroll & Tokowicz, 2005; Kroll et al., 2010; Ma et al., 2014; Potter et al., 1984). In a review of functional neuroimaging literature Abutalebi et al. (2001) found that as L2 proficiency increases, learners engage different cognitive control areas of the brain, visible both during comprehension, production and lexical access. Potter et al. (1984) suggested that beginner learners connect L2 words to their L1 equivalents, while more proficient learners operate their two languages independently, having both L1 and L2 word forms connect directly to their underlying meaning, making their access more efficient. On the other hand, BIA+ model of bilingual processing proposes a shared lexicon for both L1 and L2, meaning that words in both languages are stored in the same mental system (Dijkstra & Van Heuven, 2002). Beginner learners may experience greater difficulty suppressing interference from their dominant L1 because the inhibitory control mechanism is less efficient. As L2 proficiency increases, bilinguals become better at managing interference and retrieving words more fluently. Proficiency has also been shown to influence how nouns and verbs are processed in L2. Chan et al. (2008) found that early Chinese-English bilinguals processed both nouns and verbs similarly in L1 but displayed heightened neural activity in motor areas when processing L2 verbs. Late bilinguals, however, did not show such distinctions, suggesting that proficiency plays a crucial role not only for L2 word processing in general, but also differently interacts with specific word classes, affecting how they are represented and processed (Yang et al., 2011).

Taken together, it is important to consider learners' characteristics, learning methods and words learned, as research has shown that these factors interact with the L2 learning process (Tonzar et al., 2009). As Zwaan (2021) recently proposed, the way to advance research on embodied cognition is through doing direct replication studies, but also crucially by complementing them with conceptual replications. Conceptual replications would extend the original experiment by involving variations in the populations, task type, stimuli, thus providing a more comprehensive and thorough understanding of the phenomena under study.

Besides from the already described possible variations, another one would be to place the experimental studies presented in a more interactive and social communicative context. One of the frequently raised points recently is the importance of studying language in the social and interactive contexts, as opposed to the traditional experimental methods focused on the individual (Li & Jeong, 2021; Zwaan, 2014). This could be done by developing methodologies that would involve interaction with people and objects, but also through virtual environments (Li & Jeong, 2021). The present work takes steps in that direction by embedding words into contexts rather than predominantly used word pairs, as well as by studying social cues. Still, based on these findings, the present research questions can always be adapted to more interactive and social contexts, which would certainly bring additional valuable insights.

Finally, although the online recruiting services have enabled an access to a large pool of participants and allowed this research to take place at a time when testing in person was not possible, there are certain benefits from testing in-person. Most notably, the conditions of testing are strictly controlled, ensuring similar experience for all participants. However, recent research comparing online and in-person testing established that results across different psycholinguistic tasks have been replicated in the online setting, proving the general validity and potential of online testing (Fairs & Strijkers, 2021; Hilbig, 2016; McConnell et al., 2024; Nguyen et al., 2024). Still, future research might consider implementing these studies in person or in a classroom context. This would be another step in bringing this research closer to the relevant stakeholders in education and software design.

5.5. Conclusion

Even in the short time span of a few years since the beginning of this project, the landscape of available technologies has already changed significantly. Most notably, large language models (LLMs) and AI became accessible, widely used and integrated into many existing software, including software for language learning (Li & Lan, 2022). This underlines the growing need for comprehensive research in this area to support the design and use of the available technologies. As the gap between what is technically possible and the knowledge of what is educationally useful increases, clearer guidelines are needed to help navigate the increasing offer of technological solutions.

This thesis aimed to address this gap by pointing to important, yet unconsidered factors relevant in second language vocabulary learning. Advancing the understanding of the core processes that take place in language learning will make it easier to implement them to any new and upcoming technologies. Additionally, applying the findings into the classroom context and making them accessible to language teachers could serve to support teachers with concrete guidelines.

Considering the evidence from the current literature, the future of second language learning goes towards integrating linguistic, social and embodied aspects of language, approaching contextualised and natural learning. The rapid growth of technological possibilities shows promise and potential to support this approach. What remains is the communication and integration between research findings, technologies and practice that would enable iterative processes to take place that would deepen our understanding of language and contribute towards optimizing the language learning outcomes.

References

Abutalebi, J., Cappa, S. F., & Perani, D. (2001). The bilingual brain as revealed by functional neuroimaging. *Bilingualism: Language and Cognition*, *4*(2), 179–190. doi:10.1017/S136672890100027X

Allen, M., Witt, P. L., & Wheeless, L. R. (2006). The role of teacher immediacy as a motivational factor in student learning: Using meta-analysis to test a causal model. *Communication Education*, *55*(1), 21-31. https://doi.org/10.1080/03634520500343368

Allen, L.Q. (1995). The Effects of Emblematic Gestures on the Development and Access of Mental Representations of French Expressions. *The Modern Language Journal*, *79*, 521-529. https://doi.org/10.1111/j.1540-4781.1995.tb05454.x

Allen, L.Q. (1999). Functions of Nonverbal Communication in Teaching and Learning a Foreign Language. The French Review, 72, 469-480.

Altarriba, J., & Knickerbocker, H. (2011). Acquiring second language vocabulary through the use of images and words. In P. Trofimovich & K. McDonough (Eds.), Applying priming methods to L2 learning, teaching, and research: Insights from psycholinguistics (pp. 21-48). John Benjamins. https://doi.org/10.1075/lllt.30.06alt

American Academy of Arts and Sciences, British Academy, Academy of the Social Sciences in Australia, Australian Academy of the Humanities, & Royal Society of Canada. (2020). *The importance of languages in global context: An international call to action*. https://www.amacad.org/sites/default/files/media/document/2020-11/Joint-Statement-on-Languages.pdf

Andrews, M., Frank, S., & Vigliocco, G. (2014). Reconciling embodied and distributional accounts of meaning in language. *Topics in cognitive science*, *6*(3), 359–370. https://doi.org/10.1111/tops.12096

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review, 116*(3), 463–498. https://doi.org/10.1037/a0016261

Atkinson, D. (2002). Toward a sociocognitive approach to second language acquisition. *The Modern Language Journal, 86*(4), 525-545. https://doi.org/10.1111/1540-4781.00159

Atkinson, D. (2010). Extended, Embodied Cognition and Second Language Acquisition. *Applied Linguistics, 31*(5), 599-622. https://doi.org/10.1093/applin/amq009

Atkinson, M. A., Simpson, A. A., & Cole, G. G. (2018). Visual attention and action: How cueing, direct mapping, and social interactions drive orienting. *Psychonomic Bulletin & Review, 25*(5), 1585–1605. https://doi.org/10.3758/s13423-017-1354-0

Baayen, R. H. (2008*). Analyzing Linguistic Data: A Practical Introduction to Statistics using R.* Cambridge University Press. https://doi.org/10.1017/CBO9780511801686

Baldwin, D.A., & Moses, L.J. (2001). Links between Social Understanding and Early Word Learning: Challenges to Current Accounts. *Social Development, 10*, 309-329. https://doi.org/10.1111/1467-9507.00168

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English Lexicon Project. *Behavior Research Methods, 39*, 445-459. https://doi.org/10.3758/BF03193014

Bang, J.Y. & Nadig A. (2020). An investigation of word learning in the presence of gaze: evidence from school-age children with typical development or autism spectrum disorder. *Cognitive Development*, *54*, 100847. https://doi.org/10.1016/j.cogdev.2020.100847

Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning*, *52*(2), 323-363. https://doi.org/10.1111/0023-8333.00186

Barcroft, J. (2015). *Lexical input processing and vocabulary learning*. John Benjamins Publishing Company. https://doi.org/10.1075/lllt.43

Barcroft, J., & Sunderman, G. (2008). Learning new words for objects and nonobjects: Theoretical and methodological implications. *The Mental Lexicon*, *3*(3), 315–340. https://doi.org/10.1075/ml.3.3.05bar

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), https://doi.org/10.1016/j.jml.2012.11.001

Barry, R. A., Graf Estes, K., & Rivera, S. M. (2015). Domain general learning: Infants use social and non-social cues when learning object statistics. *Frontiers in Psychology*, *6*, 551. https://doi.org/10.3389/fpsyg.2015.00551

Barsalou, L. W. (1999). Perceptual symbol systems. *Brain and Behavioural Sciences*, *22*, 577–660. https://doi.org/10.1017/s0140525x99002149

Barsalou L. W. (2010). Grounded cognition: past, present, and future. *Topics in cognitive science*, *2*(4), 716–724. https://doi.org/10.1111/j.1756-8765.2010.01115.x

Barsalou, L. W., Santos, A., Simmons, W. K. and Wilson, C. D. (2008) Language and simulation in conceptual processing. In: de Vega, M., Glenberg, A. and Graesser, A. (eds.) *Symbols and Embodiment Debates on meaning and cognition*. Oxford

University Press, pp. 245-284. ISBN 9780199217274 https://doi.org/10.1093/acprof:oso/9780199217274.003.0013

Bartlett, J.E., Kotrlik, J.W. & Higgins, C.C. (2001). Organizational Research: Determining Appropriate Sample Size in Survey Research. *Information Technology, Learning, and Performance Journal*, *19*, 43-50.

Basöz, T., & Can, D. T. (2016). The effectiveness of computers on vocabulary learning among preschool children: A semiotic approach. *Cypriot Journal of Educational Sciences*, *11*(1), 02–08. https://doi.org/10.18844/cjes.v11i1.266

Bates, J., & Son, J. B. (2020). English vocabulary learning with simplified pictures. *TESL-EJ: The Electronic Journal for English as a Second Language*, *24*(3), 1-20.

Baxter, P., Bekkering, H., Dijkstra, T., Droop, M., van den Hurk, M., & Leoné, F. (2022). Contrasting orthographically similar words facilitates adult second language vocabulary learning. *Learning and Instruction*, *80*, Article 101582. https://doi.org/10.1016/j.learninstruc.2022.101582

Beatty-Martínez, A. L., & Dussias, P. E. (2017). Bilingual experience shapes language processing: Evidence from codeswitching. *Journal of Memory and Language*, *95*, 173–189. https://doi.org/10.1016/j.jml.2017.04.002

Belpaeme, T., Vogt, P., van den Berghe, R., Bergmann, K., Göksun, T., de Haas, M., Kanero, J., Kennedy, J., Küntay, A. C., Oudgenoeg-Paz, O., Papadopoulos, F., Schodde, T., Verhagen, J., Wallbridge, C. D., Willemsen, B., de Wit, J., Geçkin, V., Hoffmann, L., Kopp, S., Krahmer, E., … Pandey, A. K. (2018). Guidelines for Designing Social Robots as Second Language Tutors. *International Journal of Social Robotics*, *10*(3), 325–341. https://doi.org/10.1007/s12369-018-0467-6

Bergmann, K., & Macedonia, M. (2013). A virtual agent as vocabulary trainer: Iconic gestures help to improve learners' memory performance. In R. Aylett, B. Krenn, C. Pelachaud, & H. Shimodaira (Eds.), *Intelligent virtual agents: 13th international conference, IVA 2013*. Springer. https://doi.org/10.1007/978-3-642-40415-3_12

Bialystok, E. (1978). A theoretical model of second language learning. *The Modern Language Journal*, *62*(3), 69-83. https://doi.org/10.1111/J.1467-1770.1978.TB00305.X

Bird, H., Howard, D., & Franklin, S. (2000). Why is a verb like an inanimate object? Grammatical category and semantic category deficits. *Brain and Language*, *72*(3), 246–309. https://doi.org/10.1006/brln.2000.2292

Bisson, M. J., van Heuven, W. J., Conklin, K., & Tunney, R. J. (2015). The role of verbal and pictorial information in multimodal incidental acquisition of foreign language vocabulary. *Quarterly Journal of Experimental Psychology*, *68*(7), 1306–1326. https://doi.org/10.1080/17470218.2014.979211

Bley-Vroman, R. (2009). The evolving context of the fundamental difference hypothesis. *Studies in Second Language Acquisition*, *31*(2), 175–198. https://doi.org/10.1017/S0272263109090275

Boddaert, G., Casalis, S., & Mahé, G. (2021). Photograph method fosters direct access to second-language word meaning: Direct evidence from a word-picture matching task. *The British journal of developmental psychology*, *39*(3), 407–423. https://doi.org/10.1111/bjdp.12375

Boers, F., Piquer Píriz, A. M., Stengers, H., & Eyckmans, J. (2009). Does pictorial elucidation foster recollection of figurative idioms? *Language Teaching Research*, *13*, 367–388. https:// 10.1177/1362168809341505

Boers, F., Warren, P., Grimshaw, G., and Siyanova-Chanturia, A. (2017). On the benefits of multimodal annotations for vocabulary uptake from reading. *Computer Assisted Language Learning*, *30*(7), 709-725. https://doi.org/10.1080/09588221.2017.1356335

Booth, A., McGregor, K., & Rohlfing, K. (2008). Socio-Pragmatics and Attention: Contributions to Gesturally Guided Word Learning in Toddlers. *Language Learning and Development*, *4*, 179-202. https://doi.org/10.1080/15475440802143091

Borghi, A. M., Barca, L., Binkofski, F., Castelfranchi, C., Pezzulo, G., & Tummolini, L. (2019). Words as social tools: Flexibility, situatedness, language and sociality in abstract concepts: Reply to comments on "Words as social tools: Language, sociality and inner grounding in abstract concepts". *Physics of Life Reviews*, *29*, 178–184. https://doi.org/10.1016/j.plrev.2019.06.004

Borghi, A. M., & Binkofski, F. (2014). *Words as social tools: An embodied view on abstract concepts*. Springer. https://doi.org/10.1007/978-1-4614-9539-0

Bowerman, M., & Choi, S. (2001). Shaping meanings for language: Universal and language-specific in the acquisition of spatial semantic categories. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 476-511). Cambridge University Press.

Böckler, A., van der Wel, R.P.R.D., Welsh, T. N. (2015). Eyes only? Perceiving eye contact is neither sufficient nor necessary for attentional capture by face direction. *Acta Psychologica*, *160*, 134-140. https://doi.org/10.1016/j.actpsy.2015.07.009

Branscheidt, M., Hoppe, J., Zwitserlood, P., & Liuzzi, G. (2018). tDCS over the motor cortex improves lexical retrieval of action words in poststroke aphasia. Journal of neurophysiology, 119(2), 621–630. https://doi.org/10.1152/jn.00285.2017

Brecht, R. D. (2015). America's languages: Challenges and promise. American Academy of Arts and Sciences. https://www.amacad.org/multimedia/pdfs/AmericasLanguagesChallengesandPromise.pdf

Brodeur, M.B., Guérard K., & Bouras M. (2014). Bank of Standardized Stimuli (BOSS) Phase II: 930 New Normative Photos. *PLoS ONE*, *9*(9), e106953. https://doi.org/10.1371/journal.pone.0106953

Brojde, C. L., Ahmed, S., & Colunga, E. (2012). Bilingual and monolingual children attend to different cues when learning new words. *Frontiers in Psychology*, *3*, Article 155. https://doi.org/10.3389/fpsyg.2012.00155

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: a longitudinal, growth curve modeling study. *Journal of Child Language*, *35*(1), 207–220. https://doi.org/10.1017/s030500090700829x

Brozzoli, C., Gentile, G., Bergouignan, L., & Ehrsson, H. H. (2013). A shared representation of the space near oneself and others in the human premotor cortex. *Current Biology*, *23*(18), 1764-1768. https://doi.org/10.1016/j.cub.2013.07.004

Brysbaert, M., & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition*, *1*(1), 9. https://doi.org/10.5334/joc.10

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*, 904-911. https://doi.org/10.3758/s13428-013-0403-5

Bultena, S., Dijkstra, T., & van Hell, J. G. (2013). Cognate and word class ambiguity effects in noun and verb processing. *Language and Cognitive Processes, 28*(9), 1350-1377. https://doi.org/10.1080/01690965.2012.718353

Burston, J. (2015). Twenty years of MALL project implementation: A meta-analysis of learning outcomes. *ReCALL: Journal of Eurocall, 27*(1), 4–20. https://doi.org/10.1017/S0958344014000159

Busse, V. (2017). Plurilingualism in Europe: Exploring attitudes toward English and other European languages among adolescents in Bulgaria, Germany, the Netherlands, and Spain. *The Modern Language Journal, 101*(3), 566–582. https://doi.org/10.1111/modl.12415

Butler, Y. G. (2019). Teaching vocabulary to young second- or foreign-language learners: What can we learn from the research? *Language Teaching for Young Learners, 1*(1), 4-33. https://doi.org/10.1075/ltyl.00003

Caldano, M., & Coventry, K. R. (2019). Spatial demonstratives and perceptual space: To reach or not to reach? *Cognition, 191*, 103989. https://doi.org/10.1016/j.cognition.2019.06.001

Callaghan, M. N., & Reich, S. M. (2018). Are educational preschool apps designed to teach? An analysis of the app market. *Learning, Media and Technology, 43*(3), 280–293. https://doi.org/10.1080/17439884.2018.1498355

Carpenter, S. K., & Geller, J. (2020). Is a picture really worth a thousand words? Evaluating contributions of fluency and analytic processing in metacognitive judgements for pictures in foreign language vocabulary learning. *Quarterly Journal of Experimental Psychology, 73*(2), 211-224. https://doi.org/10.1177/1747021819879416

Carpenter, S. K., & Olson, K. M. (2012). Are pictures good for learning new vocabulary in a foreign language? Only if you think they are not. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(1), 92-101. https://doi.org/10.1037/a0024828

Caruana, N., Inkley, C., & Nalepka, P. (2021). Gaze facilitates responsivity during hand coordinated joint attention. *Scientific Reports, 11*, 21037. https://doi.org/10.1038/s41598-021-00476-3

Çetinçelik, M., Rowland C.F., & Snijders, T.M. (2021). Do the Eyes Have It? A Systematic Review on the Role of Eye Gaze in Infant Language Development. *Frontiers in Psychology, 11*, 589096. https://10.3389/fpsyg.2020.589096

Chan, A. H., Luke, K. K., Li, P., Yip, V., Li, G., Weekes, B., & Tan, L. H. (2008). Neural correlates of nouns and verbs in early bilinguals. *Annals of the New York Academy of Sciences, 1145*, 30–40. https://doi.org/10.1196/annals.1416.000

Chapelle, C. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing and research.* Cambridge University Press. https://doi.org/10.1017/CBO9781139524681

Chapelle, C.A. (2008). Computer Assisted Language Learning. In B. Spolsky and F.M. Hult (Eds.), The Handbook of Educational Linguistics. Blackwell Publishing Ltd. https://doi.org/10.1002/9780470694138.ch41

Chen, H., Cohen, P., & Chen, S. (2010). How Big is a Big Odds Ratio? Interpreting the Magnitudes of Odds Ratios in Epidemiological Studies. *Communications in Statistics - Simulation and Computation, 39*(4), 860–864. https://doi.org/10.1080/03610911003650383

Chen, H.C. & Leung, Y.S. (1989) Patterns of Lexical Processing in a Non-Native Language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 316-325. https://doi.org/10.1037/0278-7393.15.2.316

Chomsky, N. (1959). Review of B.F. Skinner *Verbal behavior*. *Language 35*, 26–8. https://doi.org/10.2307/411334

Chomsky, N. (1965). *Aspects of the theory of syntax.* MIT Press.

Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *Modern Language Journal*, *80*(2), 183–198. https://doi.org/10.2307/328635

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*, 335-359.

Collen, I. (2020). Language trends 2020: Language teaching in primary and secondary schools in England. *British Council*. https://www.britishcouncil.org/sites/default/files/language_trends_2020_0.pdf

Comesaña, M., Perea, M., Piñeiro, A., & Fraga, I. (2009). Vocabulary teaching strategies and conceptual representations of words in L2 in children: Evidence with novice learners. *Journal of Experimental Child Psychology*, *104*, 22-33. https://doi.org/10.1016/j.jecp.2008.10.004

Comesaña, M., Soares, A. P., Sánchez-Casas, R., & Lima, C. (2012). Lexical and semantic representations of L2 cognate and noncognate words acquisition in children: Evidence from two learning methods. *British Journal of Psychology*, *103*, 378-392. https://doi.org/10.1111/j.2044-8295.2011.02080.x

Cook, G. (2010). *Translation in language teaching: An argument for reassessment.* Oxford University Press.

Cooperrider, K., & Mesh, K. (2022). Pointing in Gesture and Sign. In A. Morgenstern & S. Goldin-Meadow (Eds.), *Gesture in Language: Development across the Lifespan* (pp. 21-46). De Gruyter Mouton; American Psychological Association. https://doi.org/10.1037/0000269-002

Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics,* 5, 161–169. http://dx.doi.org/10.1515/iral.1967.5.1-4.161

Costa, P., & Albergaria-Almeida, P. (2015). The European survey on language competences: Measuring foreign language student proficiency. *ProcediaSocial and Behavioral Sciences*, *191*, 2369-2373. https://doi.org/10.1016/j.sbspro.2015.04.255

Coventry, K. R., Carmichael, R., & Garrod, S. C. (1994). Spatial prepositions, object-specific function and task requirements. Journal of Semantics, 11(4), 289-309. https://doi.org/10.1093/jos/11.4.289

Coventry, K. R., & Diessel, H. (2024). Spatial communication systems and action*. Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2024.10.002

Coventry, K. R., & Garrod, S. C. (2004). *Saying, seeing and acting: The psychological semantics of spatial prepositions.* Psychology Press. https://doi.org/10.4324/9780203641521

Coventry, K. R., Griffiths, D., & Hamilton, C. J. (2014). Spatial demonstratives and perceptual space: Describing and remembering object location. *Cognitive Psychology*, *69*, 46-70. https://doi.org/10.1016/j.cogpsych.2013.12.001

Coventry, K. R., Guijarro-Fuentes, P., & Valdés, B. (2011). Spatial language and second language acquisition. In V. Cook & B. Bassetti (Eds.), *Language and bilingual cognition* (pp. 263-286). Psychology Press.

Coventry, K. R., Pedro Guijarro-Fuentes & Berenice Valdés (2012) On the First and Second Language Acquisition of Spatial Language. *Spatial Cognition & Computation: An Interdisciplinary Journal*, *12*(4), 219-230, https://doi.org/10.1080/13875868.2012.713058

Coventry, K. R., Valdes, B., Castillo, A., & Guijarro-Fuentes, P. (2008). Language within your reach: Near-far perceptual space and spatial demonstratives. Cognition, 108(3), 889–895. https://doi.org/10.1016/j.cognition.2008.06.010

Culkin, J. M. (1967, March). A schoolman's guide to Marshall McLuhan. The Saturday Review, 51-53, 70-72. Retrieved from http://www.unz.org/Pub/SaturdayRev-1967mar18-00051

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671-684. https://doi.org/10.1016/S0022-5371(72)80001-X

Damayanti, D., & Sundari, H. (2022). The use of prepositions among EFL learners: A systematic review. SCOPE. Journal of English Language Teaching, 6, 56-61.

De Grauwe, S., Willems, R. M., Rueschemeyer, S. A., Lemhöfer, K., & Schriefers, H. (2014). Embodied language in first- and second-language speakers: neural correlates of processing motor verbs. *Neuropsychologia*, *56,* 334–349. https://doi.org/10.1016/j.neuropsychologia.2014.02.003

De Groot, A. M. B., & Hoeks, C. J. (1995) The development of bilingual memory: Evidence from word translation by trilinguals. Language Learning, 45, 683-724.

De Groot, A. M. B. (1993). Word-type effects in bilingual processing tasks: Support for a mixed-representational system. In R. Schreuder & B. Weltens (Eds.), *The Bilingual Lexicon* (pp. 27–51). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.6.04gro

Degani, T., Prior, A., Eddington, C. M., Arêas da Luz Fontes, A. B., & Tokowicz, N. (2016). Determinants of translation ambiguity: A within and cross-language comparison. *Linguistic Approaches to Bilingualism,* *6*(3), 290-307. https://doi.org/10.1075/lab.14013.deg

Degani, A., & Tokowicz, N. (2010). Ambiguous words are harder to learn. *Bilingualism: Language and Cognition,* *13*(3), 299-314. https://doi.org/10.1017/S1366728909990411

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*(4), 499–533. https://doi.org/10.1017/S0272263100004022

DeKeyser, R. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 97-113). Erlbaum. Demir-Lira, Ö.E., Kanero, J., Oranç, C., Koskulu, S., Franko, I., Göksun, T., & Küntay, A.C. (2020). L2 Vocabulary Teaching by Social Robots: The Role of Gestures and On-Screen Cues as Scaffolds. *Frontiers in Education, 5*, 599-636. https://10.3389/feduc.2020.599636

Demir-Lira, Ö. E., Kanero, J., Oranç, C., Koşkulu, S., Franko, I., Göksun, T., & Küntay, A. C. (2020). L2 vocabulary teaching by social robots: The role of gestures and on-screen cues as scaffolds. *Frontiers in Education*, 5, Article 599636. https://doi.org/10.3389/feduc.2020.599636

Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(3), 175–197. https://doi.org/10.1017/S1366728902003012

Dijkstra, T., Wahl, A., Buytenhuijs, F., Van Halem, N., Al-Jibouri, Z., De Korte, M., & Rekké, S. (2019). Multilink: a computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition*, *22*(4), 657–679. https://doi.org/10.1017/S1366728918000287

Douglas Fir Group. (2016). A transdisciplinary framework for SLA in a multilingual world. Modern Language Journal, 100, 19– 47. https://doi.org/10.1111/modl.12301

Dove, G. O. (2022). Rethinking the role of language in embodied cognition. *Philosophical Transactions of the Royal Society B*, 377(1866). https://doi.org/10.1098/rstb.2021.0375

Dove, G., Barca, L., Tummolini, L. & Borghi, A. M. (2020). Words have a weight: language as a source of inner grounding and flexibility in abstract concepts. *Psychological Research*, *86*, 2451–2467. https://doi.org/10.1007/s00426-020-01438-6

Dörnyei, Z., & Taguchi, T. (2010). *Questionnaires in second language research: Construction, administration, and processing* (2nd ed.). Routledge.

Dronjic, V. (2019). How (Not) to Teach English Vocabulary. *The CATESOL Journal*, 31(1), 29-54.

Dubiner, D. (2018). Second language learning and teaching: From theory to a practical checklist. *TESOL Journal*, *9*(4), Article e398. https://doi.org/10.1002/tesj.398

Durrant, P., Siyanova-Chanturia, A., Kremmel, B., & Sonbul, S. (2022). Research methods in vocabulary studies. John Benjamins Publishing Company.

Eddington, C. M., & Tokowicz, N. (2013). Examining English–German translation ambiguity using primed translation recognition. *Bilingualism: Language and Cognition*, *16*(2), 442–457. https://doi.org/10.1017/S1366728912000387

Egbert, J., Huff, L., McNeil, L., Preuss, C. and Sellen, J. (2009), Pedagogy, Process, And Classroom Context: Integrating Teacher Voice And Experience Into Research On Technology-Enhanced Language Learning. *The Modern Language Journal*, *93*, 754-768. https://doi.org/10.1111/j.1540-4781.2009.00971.x

Ellis, N. C. (2007). The associative-cognitive CREED. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 77–95). Erlbaum. https://doi.org/10.4324/9781410615299

Ellis, N. C. (2008). The emergentism of language: What's learned, what's innate, and what emerges? Lingua, 118(4), 620-644. https://doi.org/10.1016/j.lingua.2006.12.005

Ellis, N. C. (2008). Usage-based and form-focused language acquisition: The associative learning of constructions, learned attention, and the limited L2 endstate. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 372-405). Routledge.

Ellis, N. C. (2019). Essentials of a theory of language cognition. The Modern Language Journal, 103(S1), 39-60. https://doi.org/10.1111/modl.12532

Ellis, N. C., & Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning*, *43*(4), 559–617. https://doi.org/10.1111/j.1467-1770.1993.tb00627.x

Ellis, N. C., & Wulff, S. (2015). Second language acquisition. In E. Dabrowska & D. Divjak (Eds.), *Handbook of cognitive linguistics* (pp. 409–431). De Gruyter Mouton.

Emirmustafaoğlu A., Gökmen D. U. (2015). The effects of picture vs. translation mediated instruction on l2 vocabulary learning. *Procedia - Social and Behavioral Sciences*, *199*, 357–362. https://doi.org/10.1016/j.sbspro.2015.07.559

European Commission. (2018). *Commission staff working document accompanying the document proposal for a Council recommendation on a comprehensive approach to the teaching and learning of languages*. Publications Office of the European Union.

European Commission, Directorate-General for Education, Youth, Sport and Culture. (2019). *The European education area: Report*. Publications Office of the European Union.

Fairs, A., & Strijkers, K. (2021). Can we use the internet to study speech production? Yes we can! Evidence contrasting online versus laboratory naming latencies and errors. *PLOS ONE*, *16(10),* Article e0258908. https://doi.org/10.1371/journal.pone.0258908

Farley, A. P., Ramonda, K., & Liu, X. (2012). The concreteness effect and the bilingual lexicon: The impact of visual stimuli attachment on meaning recall of abstract L2 words. *Language Teaching Research*, *16*(4), 449-466. https://doi.org/10.1177/1362168812436910

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences of the United States of America, 103*(32), 12203–12208. https://doi.org/10.1073/pnas.0602173103

Ferré, P., García, T., Fraga, I., Sánchez-Casas, R., & Molero, M. (2009). Memory for emotional words in bilinguals: Do words have the same emotional intensity in the first and in the second language? *Cognition and Emotion*, *24*(5), 760–785. https://doi.org/10.1080/02699930902985779

Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of motor system in language comprehension. *The Quarterly Journal of Experimental Psychology*, *61*(6), 825–850. https://doi.org/10.1080/17470210701623605

Flynn, S. (1983). *A study of the effects of principal branching direction in second language acquisition: The generalization of a parameter of Universal Grammar from first to second language acquisition* [Doctoral dissertation, Cornell University]

Fodor, J. A. (1983). *The modularity of mind*. MIT Press. https://doi.org/10.7551/mitpress/4737.003.0009

Fraisse, P. (1960). Recognition time measured by verbal reaction to figures and words. *Perceptual and Motor Skills, 11*, 204. https://doi.org/10.2466/PMS.11.6.204-204

Frechette, C., & Moreno, R. (2010). The roles of animated pedagogical agents' presence and nonverbal communication in multimedia learning environments. *Journal of Media Psychology: Theories, Methods, and Applications*, *22*(2), 61–72. https://doi.org/10.1027/1864-1105/a000009

Gangaiamaran, R., & Pasupathi, M. (2017). Review on use of mobile apps for language learning. *International Journal of Applied Engineering Research, 12*(21), 11242-11251.

García-Gámez, A. B., & Macizo, P. (2019). Learning nouns and verbs in a foreign language: The role of gestures. *Applied Psycholinguistics, 40(2),* 473–507. https://doi.org/10.1017/S0142716418000656

García-Gámez, A. B., & Macizo, P. (2020). The way in which foreign words are learned determines their use in isolation and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46(2),* 364–379. https://doi.org/10.1037/xlm0000721

Garrett, N. (1991). Technology in the Service of Language Learning: Trends and Issues. *Modern Language Journal*, 75(1), 74-101. https://doi.org/10.1111/j.1540-4781.1991.tb01085.x

Gatbonton, E., Segalowitz, N. (2005). Rethinking communicative language teaching: A focus on access to fluency. *The Canadian Modern Language Review 61(3),* 325-353. https://doi.org/10.3138/cmlr.61.3.325

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. *Language, 2*, 301-334.

Gentner, D., & Boroditsky, L. (2001). Individuation, relativity and early word learning. In M. Bowerman, & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 215-256). Cambridge University Press.

Geusebroek, J. M., Burghouts, G. J., & Smeulders, A. W. M. (2005). The Amsterdam library of object images. *International Journal of Computer Vision, 61*(1), 103–112. https://doi.org/10.1023/B:VISI.0000042993.50813.60

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*(2), 135-176. https://doi.org/10.1016/S0010-0277(99)00036-0

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard Words. *Language Learning and Development, 1*(1), 23-64. https://doi.org/10.1207/s15473341lld0101_4

Glenberg, A. M., Witt, J. K., & Metcalfe, J. (2013). From the Revolution to Embodiment: 25 Years of Cognitive Psychology. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science, 8*(5), 573–585. https://doi.org/10.1177/1745691613498098

Gliga, T., Elsabbagh, M., Andravizou, A., & Johnsn, M. (2010). Faces attract infants' attention in complex displays. *Infancy, 15*(5), 550-562. https://doi.org/10.1080/15250000903144199

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology, 66*(3), 325–331. https://doi.org/10.1111/j.2044-8295.1975.tb01468.x

Goldin-Meadow, S. (2007), Pointing Sets the Stage for Learning Language—and Creating Language. *Child Development, 78,* 741-745. https://doi.org/10.1111/j.1467-8624.2007.01029.x-

Goldin-Meadow, S. (2014). Widening the lens: What the manual modality reveals about language, learning and cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), https://doi.org/10.1098/rstb.2013.0295

Golinkoff, R. M., & Hirsh-Pasek, K. (2006). The Emergentist Coalition Model of Word Learning in Children Has Implications for Language in Aging. In E. Bialystok & F. I. M. Craik (Eds.), *Lifespan cognition: Mechanisms of change* (pp. 207–222). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195169539.003.0014

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2012). Technologies for foreign language learning: a review of technology types and their effectiveness. *Computer Assisted Language Learning*, *27*(1), 70–105. https://doi.org/10.1080/09588221.2012.700315

Grabe, W. (2009). *Reading in a Second Language: Moving from Theory to Practice*. Cambridge University Press. https://doi.org/10.1017/CBO9781139150484

Grant, H. M., Bredahl, L. C., Clay, J., Ferrie, J., Groves, J. E., McDorman, T. A., & Dark, V. J. (1998). Context-dependent memory for meaningful material: Information for students. Applied Cognitive Psychology, 12(6), 617–623. https://doi.org/10.1002/(SICI)1099-0720(1998120)12:6<617::AID-ACP542>3.0.CO;2-5

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, *114*(2), 211–244. https://doi.org/10.1037/0033-295X.114.2.211

Grzyb, B. J., & Vigliocco, G. (2020). Beyond robotic speech: Mutual benefits to cognitive psychology and artificial intelligence from the study of multimodal communication. In S. Muggleton & N. Chater (Eds.), *Human-like machine intelligence* (pp. 274-294). Oxford Academic. https://doi.org/10.1093/oso/9780198862536.003.0014

Gulz, A., & Haake, M. (2006). Design of animated pedagogical agents--A look at their look. *International Journal of Human-Computer Studies*, *64*(4), 322–339. https://doi.org/10.1016/j.ijhcs.2005.08.006

Hald, L. A., de Nooijer, J., van Gog, T., & Bekkering, H. (2016). Optimizing word learning via links to perceptual and motoric experience. *Educational Psychology Review*, *28*, 495-522. https://doi.org/10.1007/s10648-015-9334-2

Hao, T., Wang, Z., & Ardasheva, Y. (2021). Technology-Assisted Vocabulary Learning for EFL Learners: A Meta-Analysis. *Journal of Research on Educational Effectiveness, 14(3),* 645–667. https://doi.org/10.1080/19345747.2021.1917028

Hardison, D. M., & Pennington, M. C. (2021). Multimodal Second-Language Communication: Research Findings and Pedagogical Implications. *RELC Journal, 52(1),* 62-76. https://doi.org/10.1177/0033688220966635

Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition, 177,* 263-277. https://doi.org/10.1016/j.cognition.2018.04.007

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron, 41(2),* 301-307. https://doi.org/10.1016/S0896-6273(03)00838-9

Hauk, O., & Tschentscher, N. (2013). The Body of Evidence: What Can Neuroscience Tell Us about Embodied Semantics? *Frontiers in Psychology, 4,* 50. https://doi.org/10.3389/fpsyg.2013.00050

Haxby, J. V., Gobbini, M. I., & Nastase, S. A. (2020). Naturalistic stimuli reveal a dominant role for agentic action in visual representation. *NeuroImage, 216,* 116561. https://doi.org/10.1016/j.neuroimage.2020.116561

Heil, C.L., Wu, J.S., Lee, J.J., Schmidt, T. (2016). A Review of Mobile Language Learning Applications: Trends, Challenges, and Opportunities. *The EuroCALL Review*, 24(2), 32-50. https://doi.org/10.4995/eurocall.2016.6402

Hernandez, A., Li, P., & MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends in Cognitive Sciences, 9,* 220-225. https://doi.org/10.1016/j.tics.2005.03.003

Hilbig, B. E. (2016). Reaction time effects in lab- versus Web-based research: Experimental evidence. *Behavior Research Methods, 48(4),* 1718–1724. https://doi.org/10.3758/s13428-015-0678-9

Hirsh-Pasek, K., Zosh, J.M., Golinkoff, R.M., Gray, J.H., Robb, M.B., Kaufman, J. (2015). Putting Education in "Educational" Apps: Lessons From the Science of Learning. *Psychological Science in the Public Interest. 16(1),* 3-34. https://doi.org/10.1177/1529100615569721

Holler, J., Schubotz, L., Kelly, S., Hagoort, P., Schuetze, M., & Özyürek, A. (2014). Social eye gaze modulates processing of speech and co-speech gesture. *Cognition, 133(3),* 692–697. https://doi.org/10.1016/j.cognition.2014.08.008

Holler, J., & Levinson, S. C. (2019). Multimodal Language Processing in Human Communication. *Trends in Cognitive Sciences, 23(8),* 639–652. https://doi.org/10.1016/j.tics.2019.05.006

Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., Hennon, E., & Rocroi, C. (2000). Breaking the language barrier: an emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development, 65(3),* i–123. https://doi.org/10.1111/1540-5834.00090

Hubbard, P. (2011). Evaluation of courseware and websites. In L. Ducate & N. Arnold (Eds.), *Present and future promises of CALL: From theory and research to new directions in foreign language teaching* (pp. 407-440). CALICO.

Hubbard, P., & Levy, M. (2016). Theory in computer-assisted language learning research and practice. In F. Farr & L. Murray (Eds.), *The Routledge Handbook Of Language Learning And Technology*. Routledge. https://doi.org/10.4324/9781315657899

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological science*, *16*(5), 367–371. https://doi.org/10.1111/j.0956-7976.2005.01542.x

Jachmann, T. K., Drenhaus, H., Staudte, M., & Crocker, M. W. (2019). Influence of speakers' gaze on situated language comprehension: Evidence from Event-Related Potentials. *Brain and Cognition, 135,* 103571. https://doi.org/10.1016/j.bandc.2019.05.009

Jahangard, A. (2022). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation (A conceptual replication study). *MEXTESOL Journal*, *46(3)* https://doi.org/10.61871/mj.v46n3-9

Jacobsen, N. D. (2018). The best of both worlds: Combining cognitive linguistics and pedagogic tasks to teach English conditionals. *Applied Linguistics, 39*(5), 668–693. https://doi.org/10.1093/applin/amw030

Jaeger T. F. (2008). Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of memory and language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Jamieson, J., & Chapelle, C. A. (2010). Evaluating CALL use across multiple contexts. *System, 38(3),* 357–369. https://doi.org/10.1016/j.system.2010.06.014

Jamieson, J., Chapelle, C., & Preiss, S. (2004). Putting principles into practice. *ReCALL, 16(2),* 396-415. https://doi.org/10.1017/S0958344004001028

Jarvis, S., & Pavlenko, A. (2007). *Crosslinguistic influence in language and cognition.* Routledge. https://doi.org/10.4324/9780203935927

Jelani, N. A. M., & Boers, F. (2018). Examining incidental vocabulary acquisition from captioned video. *ITL-International Journal of Applied Linguistics, 169(1),* 169-190. https://doi.org/10.1075/itl.00011.jel

Jeong, H., Li, P., Suzuki, W., Sugiura, M., & Kawashima, R. (2021). Neural mechanisms of language learning from social contexts. *Brain and Language, 212,* Article 104874. https://doi.org/10.1016/j.bandl.2020.104874

Jones, L. (2004). Testing L2 vocabulary recognition and recall using pictorial and written test items. *Language Learning & Technology, 8(3),* 122-143. http://dx.doi.org/10125/43998

Kacetl, J., & Klímová, B. (2019). Use of smartphone applications in English language learning—A challenge for foreign language education. Education Sciences, 9(3), Article 179. https://doi.org/10.3390/educsci9030179

Kaimara, P., Fokides, E., Oikonomou, A., & Deliyannis, I. (2021). Potential barriers to the implementation of digital game-based learning in the classroom: Pre-service teachers' views. *Technology, Knowledge and Learning, 26,* 825-844. https://doi.org/10.1007/s10758-021-09512-7

Kamasak, R., Özbilgin, M., Atay, D., & Kar, A. (2021). The effectiveness of mobile-assisted language learning (MALL): A review of the extant literature. In A. Moura, P. Reis, & M. Cordeiro (Eds.), *Handbook of research on determining the reliability of online assessment and distance learning* (pp. 194-212). IGI Global. https://doi.org/10.4018/978-1-7998-4769-4.ch008

Kaschak, M. P., & Borreggine, K. L. (2008). Temporal dynamics of the action-sentence compatibility effect. *Quarterly journal of experimental psychology (2006)*, *61*(6), 883–895. https://doi.org/10.1080/17470210701623852

Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review*, *99*(2), 349–364. https://doi.org/10.1037/0033-295X.99.2.349

Kelly, S. D., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes, 24*(2), 313–334. https://doi.org/10.1080/01690960802365567

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *The relationship of verbal and nonverbal communication* (pp. 207-227). De Gruyter Mouton. https://doi.org/10.1515/9783110813098.207

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press. https://doi.org/10.1017/CBO9780511807572

Kendon, A. (2014). Semiotic diversity in utterance production and the concept of 'language'. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*, 6529757. https://doi.org/10.1098/rstb.2013.0293

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods 42(3)*, 627-633. https://doi.org/10.3758/BRM.42.3.627

Kim, M., Crossley, S. A., & Kyle, K. (2018). Lexical sophistication as a multidimensional phenomenon: Relations to second language lexical proficiency, development, and writing quality. *Modern Language Journal*, *102*(1), 120–141. https://doi.org/10.1111/modl.12447

Kirk, E., Donnelly, S., Furman, R., Warmington, M., Glanville, J., & Eggleston, A. (2022). The relationship between infant pointing and language development: A meta-analytic review. *Developmental Review, 64,* Article 101023. https://doi.org/10.1016/j.dr.2022.101023

Klimesch, W. (1987). A connectivity model for semantic processing. *Psychological Research, 49(1),* 53-61. https://doi.org/10.1007/BF00309203

Knoeferle, P., Kreysa, H. & Pickering, M. (2018). Effects of a speaker's gaze on language comprehension and acquisition: Studies on the role of eye gaze in dialogue. In Brone G. & Oben, B. (Eds), *Advances in Interaction Studies: Eye-Tracking in Interaction* (pp. 47-66). https://doi.org/10.1075/ais.10.03kno

Kolak, J., Norgate, S. H., Monaghan, P., & Taylor, G. (2021). Developing evaluation tools for assessing the educational potential of apps for preschool children in the UK. *Journal of Children and Media, 15(3),* 410–430. https://doi.org/10.1080/17482798.2020.1844776

Krashen, S.D. (1981) *Second Language Acquisition and Second Language Learning*. Pergamon Press Inc., Oxford.

Krashen, S. (1994). The input hypothesis and its rivals. In N. Ellis (Ed.). *Implicit and Explicit Learning of Languages* (pp. 45-77). Academic Press.

Kreysa, H., Nunnemann, E. M., & Knoeferle, P. (2018). Distinct effects of different visual cues on sentence comprehension and later recall: The case of speaker gaze versus depicted actions. Acta Psychologica, 188, 220–229. https://doi.org/10.1016/j.actpsy.2018.05.001

Kroll, J. F., & Curley, J. (1988). Lexical memory in novice bilinguals: The role of concepts in retrieving second language words. *Practical Aspects Of Memory, 2(389-395),* 8.

Kroll, J. F., & Linck, J. A. (2007). Representation and skill in second language learners and proficient bilinguals. In I. Kecskes & L. Albertazzi (Eds.), *Cognitive aspects of bilingualism* (pp. 237-269). Springer. https://doi.org/10.1007/978-1-4020-5935-3_12

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33(2),* 149-174. https://doi.org/10.1006/jmla.1994.1008

Kroll, J. F., & Tokowicz, N. (2005). Models of bilingual representation and processing. In J. F. Kroll & De Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches*, pp. 531–553. Oxford University Press.

Kroll, J. F., van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism, 13(3),* 373–381. https://doi.org/10.1017/S136672891000009X

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior research methods*, *44*(4), 978–990. https://doi.org/10.3758/s13428-012-0210-4

Kutskir, I. (2022). Photopea: Free online photo editor. https://www.photopea.com/

Lado, R. (1964). *Language teaching: A scientific approach*. McGraw Hill.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4,* 863. https://doi.org/10.3389/fpsyg.2013.00863

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* University of Chicago Press.

Langton, S. R. H., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social cognition. *Trends in Cognitive Sciences, 4(2),* 50-59. https://doi.org/10.1016/s1364-6613(99)01436-9

Lantolf, J. P. (1995). Sociocultural theory and second language acquisition. *Annual Review of Applied Linguistics, 15*, 108-124. https://doi.org/10.1017/S0267190500002646

Lanvers, U., & Coleman, J. A. (2017). The UK language learning crisis in the public media: a critical analysis. *The Language Learning Journal, 45(1),* 3–25. https://doi.org/10.1080/09571736.2013.830639

Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics, 29(4),* 694-716. https://doi.org/10.1093/applin/amn018

Laufer, B., & Rozovski-Roitblat, B. (2015). Retention of new words: Quantity of encounters, quality of task, and degree of knowledge. *Language Teaching Research, 19(6),* 687-711. https://doi.org/10.1177/1362168814559797

Law, B., Houston-Price, C. & Loucas, T. (2012). Using gaze direction to learn words at 18 months: relationships with later vocabulary. *Language Studies Working Papers, 4*, 3-14.

Legault, J., Zhao, J., Chi, Y. -A., Chen, W., Klippel, A., & Li, P. (2019). Immersive Virtual Reality as an Effective Tool for Second Language Vocabulary Learning. *Languages, 4(1),* 13. https://doi.org/10.3390/languages4010013

Lenneberg, E. (1967). *Biological foundations of language*. Wiley.

Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *369*(1651), 20130302. https://doi.org/10.1098/rstb.2013.0302

Lewis, T. & Kirkhart, M. (2020). Effect of iconic gestures on second language vocabulary retention in a naturalistic setting . *International Review of Applied Linguistics in Language Teaching*, *58*(3), 263-287. https://doi.org/10.1515/iral-2016-0125

Li, P., Jeong, H. (2020). The social brain of language: grounding second language learning in social interaction. *npj Science of Learning*, *5*, 8. https://doi.org/10.1038/s41539-020-0068-7

Li, P., & Lan, Y.-J. (2022). Digital Language Learning (DLL): Insights from Behavior, Cognition, and the Brain. *Bilingualism: Language and Cognition*, *25*(3), 361–378. https://doi.org/10.1017/S1366728921000353

Li, P., & MacWhinney, B. (2013). Competition Model. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics*. Wiley.

Li, Y., & Zhang, X. (2019). L2 vocabulary knowledge and L2 listening comprehension: A structural equation model. Canadian Journal of Applied Linguistics, 22(1), 85-102. https://doi.org/10.7202/1060907ar

Li, W., Yu, J., Zhang, Z. and Liu, X. (2022). Dual Coding or Cognitive Load? Exploring the Effect of Multimodal Input on English as a Foreign Language Learners' Vocabulary Learning. *Frontiers in Psychology 13*, 834706. https://doi.org/10.3389/fpsyg.2022.834706

Locker, L., Hoffman, L. & Bovaird, J.A. (2007). On the use of multilevel modeling as an alternative to items analysis in psycholinguistic research. *Behavior Research Methods 39*, 723–730. https://doi.org/10.3758/BF03192962

Long, M. H. (1996). The role of linguistic environment in second language acquisition. In W. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). Academic Press. https://doi.org/10.1016/B978-012589042-7/50015-3

Lotto, L., & De Groot, A. M. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning, 48(1),* 31-69.

Louwerse M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science, 3(2),* 273–302. https://doi.org/10.1111/j.1756-8765.2010.01106.x

Ma, T., Chen, B., Lu, C., & Dunlap, S. (2017). Proficiency and sentence constraint effects on second language word learning. *Journal of Memory and Language*, *94*, 1-17. https://doi.org/10.1016/j.jml.2016.11.002

Macedonia, M. (2019). Embodied learning: Why at school the mind needs the body. *Frontiers in Psychology*, *10*, 2098. https://doi.org/10.3389/fpsyg.2019.02098

Macedonia, M., Groher, I., & Roithmayr, F. (2014). Intelligent virtual agents as language trainers facilitate multilingualism. *Frontiers in Psychology, 5,* 295. https://doi.org/10.3389/fpsyg.2014.00295

Macedonia, M., Kern, R., & Roithmayr, F. (2014). Do children accept virtual agents as foreign language trainers? *International Journal of Learning, Teaching and Educational Research, 7(1),* 131-137.

Macedonia, M., Müller, K., & Friederici, A. D. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping, 32(6),* 982–998. https://doi.org/10.1002/hbm.21084

Macdonald, R. G., Tatler, B. W. (2013). Do as eye say: Gaze cueing and language in a real-world social interaction. *Journal of Vision, 13(4),* 6. https://doi.org/10.1167/13.4.6

MacWhinney, B. (2012). The Logic of the Unified Model. In Gass, S. & Mackey, A. (Eds.), *Handbook of Second Language Acquisition* (pp. 211-227). Routledge.

Mahowald, K., Ivanova, A. A., Blank, I. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2024). Dissociating language and thought in large language models. *Trends in Cognitive Sciences, 28(6),* 517-540. https://doi.org/10.1016/j.tics.2024.01.011

Majid, A., Jordan, F. M., & Dunn, M. (2015). Semantic systems in closely related languages. *Language Sciences, 49,* 1-18. https://doi.org/10.1016/j.langsci.2014.11.002

Malt, B. C., Sloman, S. A., & Gennari, S. P. (2003). Universality and language specificity in object naming. *Journal of Memory and Language, 49(1),* 20-42. https://doi.org/10.1016/S0749-596X(03)00021-4

Malt, B. C., Gennari, S. P., Imai, M., Ameel, E., Saji, N., & Majid, A. (2015). Where are the concepts? What words can and can't reveal. In E. Margolis & S. Laurence (Eds.), *The conceptual mind: New directions in the study of concepts*. MIT Press. https://doi.org/10.7551/mitpress/9383.003.0019

Markostamou, I., & Coventry, K. R. (2022). Naming spatial relations across the adult lifespan: At the crossroads of language and perception. *Neuropsychology, 36(3),* 216-230. https://doi.org/10.1037/neu0000789

Martin, K. I., & Tokowicz, N. (2020). The grammatical class effect is separable from the concreteness effect in language learning. *Bilingualism: Language and Cognition, 23(3),* 554–569. https://doi.org/10.1017/S1366728919000233

Martín-Luengo, B., Hu, Z., Cadavid, S., & Luna, K. (2023). Do pictures influence memory and metamemory in Chinese vocabulary learning? Evidence from Russian and Colombian learners. *PloS one*, *18*(11), e0286824. https://doi.org/10.1371/journal.pone.0286824

Matsumoto, Y. & Dobs, A.M. (2017). Pedagogical Gestures as Interactional Resources for Teaching and Learning Tense and Aspect in the ESL Grammar Classroom. *Language Learning, 67,* 7-42. https://doi.org/10.1111/lang.12181

Mayer, K. M., Yildiz, I. B., Macedonia, M., & von Kriegstein, K. (2015). Visual and motor cortices differentially support the translation of foreign language words. *Current Biology, 25(4),* 530–535. https://doi.org/10.1016/j.cub.2014.11.068

Mayer, R. E. (2009). *Multimedia Learning.* Cambridge University Press. https://doi.org/10.1017/CBO9780511811678

Mayer, R.E., & Dapra, C. (2012). An embodiment effect in computer-based learning with animated pedagogical agents. *Journal of experimental psychology. Applied, 18 3,* 239-52.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38(1),* 43–52. https://doi.org/10.1207/S15326985EP3801_6

Mayer, R. E., Sobko, K., & Mautone, P. D. (2003). Social cues in multimedia learning: Role of speaker's voice. *Journal of Educational Psychology, 95(2),* 419-425. https://doi.org/10.1037/0022-0663.95.2.419

Mayer, K. M., Yildiz, I. B., Macedonia, M., & von Kriegstein, K. (2015). Visual and motor cortices differentially support the translation of foreign language words. *Current biology : CB*, *25*(4), 530–535. https://doi.org/10.1016/j.cub.2014.11.068

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review, 88(5),* 375–407. https://doi.org/10.1037/0033-295X.88.5.375

McConnell, K., Hintz, F., & Meyer, A. S. (2024). Individual differences in online research: Comparing lab-based and online administration of a psycholinguistic battery of linguistic and domain-general skills. *Behavior research methods*, *57*(1), 22. https://doi.org/10.3758/s13428-024-02533-x

McGillion, M., Herbert, J. S., Pine, J., Vihman, M., dePaolis, R., Keren-Portnoy, T., & Matthews, D. (2017). What Paves the Way to Conventional Language? The Predictive Value of Babble, Pointing, and Socioeconomic Status. *Child development, 88(1),* 156–166. https://doi.org/10.1111/cdev.12671

McLaughlin, B. (1987). *Theories of second-language learning*. Edward Arnold.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

McNeill, D. (2012). *How language began: Gesture and speech in human evolution.* Cambridge University Press. https://doi.org/10.1017/CBO9781139108669

Mestres-Missé, A., Münte, T. F., & Rodriguez-Fornells, A. (2014). Mapping concrete and abstract meanings to new words using verbal contexts. Second Language Research, 30(2), 191-223. https://doi.org/10.1177/0267658313512668

Meteyard, L., Bahrami, B., & Vigliocco, G. (2007). Motion detection and motion verbs: Language affects low-level visual perception. *Psychological Science*, *18*(11), 1007–1013. https://doi.org/10.1111/j.1467-9280.2007.02016.x

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: a review of embodiment and the neuroscience of semantics. *Cortex, 48(7),* 788–804. https://doi.org/10.1016/j.cortex.2010.11.002

Meteyard, L., & Davies, R. (2020, March 4). Best practice guidance for linear mixed-effects models in psychological science. https://doi.org/10.17605/OSF.IO/BFQ39

Meteyard, L., Zokaei, N., Bahrami, B., & Vigliocco, G. (2008). Visual motion interferes with lexical decision on motion words. *Current Biology, 18(17),* 732–733. https://doi.org/10.1016/j.cub.2008.07.016

Michel, C., Wronski, C., Pauen, S., Daum, M. M., & Hoehl, S. (2019). Infants' object processing is guided specifically by social cues. *Neuropsychologia, 126*, 54–61. https://doi.org/10.1016/j.neuropsychologia.2017.05.022

Mihaylova, M., Gorin, S., Reber, T. P., & Rothen, N. (2022). A Meta-Analysis on Mobile-Assisted Language Learning Applications: Benefits and Risks. *Psychologica Belgica, 62(1),* pp. 252–271. DOI: https://doi.org/10.5334/pb.1146

Mitchell, R., Myles, F., & Marsden, E. (2019). Second language learning theories: Fourth Edition. Milton: Routledge. https://doi.org/10.4324/9781315617046

Miyakoda, H., Kaneko, K., & Ishikawa, M. (2011). Effective learning material for mobile devices: Visual data vs. Aural data vs. Text data. *Journal for Language Technology and Computational Linguistics*, *26*(1), 39–51. https://doi.org/10.21248/jlcl.26.2011.135

Monaghan, P., Christiansen, M. H., Farmer, T. A., & Fitneva, S. A. (2010). Measures of phonological typicality: Robust coherence and psychological validity. *The Mental Lexicon, 5*(3), 281–299. https://doi.org/10.1075/ml.5.3.02mon

Monaghan, P., Christiansen, M. H., & Fitneva, S. A. (2011). The arbitrariness of the sign: learning advantages from the structure of the vocabulary. *Journal of experimental psychology. General*, *140*(3), 325–347. https://doi.org/10.1037/a0022924

Montero Perez, M. (2020). Multimodal input in SLA research. *Studies in Second Language Acquisition*, *42*(3), 653–663. https://doi.org/10.1017/S0272263120000145

Moreno, E.M., Federmeier, K.D., & Kutas, M. (2002). Switching Languages, Switching Palabras (Words): An Electrophysiological Study of Code Switching. Brain and Language, 80, 188-207. 10.1006/brln.2001.2588

Morett, L. (2014). When hands speak louder than words: The role of gesture in the communication, encoding, and recall of words in a novel second language. *The Modern Language Journal*, *98(3),* 834-853. https://doi.org/10.1111/j.1540-4781.2014.12125.x

Morett, L. M. (2019). The power of an image: Images, not glosses, enhance learning of concrete L2 words in beginning learners. *Journal of Psycholinguistic Research, 48,* 643-664. https://doi.org/10.1007/s10936-018-9623-2

Morett, L., Gibbs, R., & MacWhinney, B. (2012). The Effects of Mental Imagery and Embodied Action on L2 Word Learning. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 34. Retrieved from https://escholarship.org/uc/item/4rb6111j

Morgan, J. L., Shi, R., & Allopenna, P. (1996). Perceptual bases of grammatical categories. In J. L. Morgan & K. Demuth (Eds.), Signal to syntax: Bootstrapping from speech to grammar in early acquisition (pp. 263–283). Lawrence Erlbaum Associates.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. Journal of Verbal Learning & Verbal Behavior, 16(5), 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

Murgiano, M., Motamedi, Y. S., & Vigliocco, G. (2021). Situating Language in the Real-World: The Role of Multimodal Iconicity and Indexicality. *Journal of Cognition, 4(1),* 38, 1–18. https://doi.org/10.5334/joc.113

Myles, F. (2010). The development of theories of second language acquisition. *Language Teaching, 43(3),* 320-332. https://doi.org/10.1017/S0261444810000078

Myles, F., Tellier, A., & Holmes, B. (2019). *Embedding languages policy in primary schools in England: summary of the RiPL White Paper proposing solutions.* https://doi.org/10.17863/CAM.41901

Nation, I. S. P. (2001). *Learning vocabulary in another language.* Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

Nation, P. (2003). The role of the first language in foreign language learning. Asian EFL Journal 1, 35–39.

Nelson, J. R., Balass, M., & Perfetti, C. A. (2005). Differences between written and spoken input in learning new words. *Written Language & Literacy, 8(2),* 25–44. DOI: https://doi.org/10.1075/wll.8.2.04nel

Nguyen, D. K.-L., Fitzpatrick, N., & Floccia, C. (2024). Adapting language development research paradigms to online testing: Data from preferential looking, word learning and vocabulary assessment in toddlers. *Journal of Child Language*, 1–33. https://doi.org/10.1017/S0305000924000035

de Nooijer, J. A., van Gog, T., Paas, F., & Zwaan, R. A. (2014). Words in action: Using gestures to improve verb learning in primary school children. Gesture, 14(1), 46-69. https://doi.org/10.1075/gest.14.1.03noo

Norris, S. (2011). Three hierarchical positions of deictic gesture in relation to spoken language: A multimodal interaction analysis. *Visual Communication, 10(2),* 1-19. https://doi.org/10.1177/1470357211398439

O'Brien, M. G. (2020). *Facilitating language learning through technology: A literature review on computer-assisted language learning*. Canadian Association of Second Language Teachers. https://www.caslt.org/wp-content/uploads/2021/12/sample-call-lit-review-en.pdf

Oppici, L., Mathias, B., Narciss, S., & Proske, A. (2023). Benefits of Enacting and Observing Gestures on Foreign Language Vocabulary Learning: A Systematic Review and Meta-Analysis. *Behavioral sciences, 13(11),* 920. https://doi.org/10.3390/bs13110920

Ortega, L. (2013). SLA for the 21st Century: Disciplinary Progress, Transdisciplinary Relevance, and the Bi/multilingual Turn. *Language Learning, 63,* 1-24. https://doi.org/10.1111/j.1467-9922.2012.00735.x

Paivio, A. (1971). *Imagery and Verbal Processes (1st ed.).* Psychology Press. https://doi.org/10.4324/9781315798868

Paivio, A. (1990). Mental representations: A dual coding approach. *Oxford University Press.* https://doi.org/10.1093/acprof:oso/9780195066661.001.0001

Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology, 5(2),* 176-206. https://doi.org/10.1016/0010-0285(73)90032-7

Park, S. (2015). The effects of social cue principles on cognitive load, situational interest, motivation, and achievement in pedagogical agent multimedia learning. *Educational Technology & Society*, 18(4), 211–29.

Paulus, M., & Fikkert, P. (2014). Conflicting Social Cues: Fourteen- and 24-Month-Old Infants' Reliance on Gaze and Pointing Cues in Word Learning. *Journal of Cognition and Development, 15(1),* 43-59. https://doi.org/10.1080/15248372.2012.698435

Pavlenko, A. (2009). Conceptual Representation in the Bilingual Lexicon and Second Language Vocabulary Learning. In A. Pavlenko (Ed.), *The Bilingual Mental Lexicon: Interdisciplinary Approaches* (pp. 125-160). Blue Ridge Summit: Multilingual Matters. https://doi.org/10.21832/9781847691262-008

Perniss, P. (2018). Why we should study multimodal language. *Frontiers in Psychology, 9,* Article 1109. https://doi.org/10.3389/fpsyg.2018.01109

Persson, V., & Nouri, J. (2018). A Systematic Review of Second Language Learning with Mobile Technologies. *International Journal of Emerging Technologies in Learning (iJET), 13*(02), 188–210. https://doi.org/10.3991/ijet.v13i02.8094

Phillips, D.C. (1995). The Good, the Bad, and the Ugly: The Many Faces of Constructivism. *Educational Researcher*, 24(7), 5-12. https://doi.org/10.3102/0013189X0240070

Pi, Z., Zhu, F., Zhang, Y. & Yang, J. (2021). An instructor's beat gestures facilitate second language vocabulary learning from instructional videos: Behavioral and neural evidence. *Language Teaching Research, 28*(5). https://doi.org/10.1177/13621688211039023

Piwek, P., Beun, R.J., & Cremers, A. (2008). 'Proximal' and 'distal' in language and cognition: Evidence from deictic demonstratives in Dutch. *Journal of Pragmatics, 40(4*), 694–718. https://10.1016/j.pragma.2007.05.001

Poarch, G. J., Van Hell, J. G., & Kroll, J. F. (2015). Accessing word meaning in beginning second language learners: Lexical or conceptual mediation? *Bilingualism: Language and Cognition, 18(3),* 357–371. https://doi.org/10.1017/S1366728914000558

Potter, M. C., & Faulconer, B. A. (1975). Time to understand pictures and words. *Nature, 253(5491),* 437–438. https://doi.org/10.1038/253437a0

Potter, M. C., So, K.-F., von Eckardt, B., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior, 23(1),* 23–38. https://doi.org/10.1016/S0022-5371(84)90489-4

Prior, A., MacWhinney, B. & Kroll, J. F. (2007). Translation norms for English and Spanish: The role of lexical variables, word class, and L2 proficiency in negotiating translation ambiguity. *Behavior Research Methods, 39(4),* 1029–1038. https://doi.org/10.3758/bf03193001

Proverbio, A. M., Leoni, G., & Zani, A. (2004). Language switching mechanisms in simultaneous interpreters: An ERP study. *Neuropsychologia*, 42, 1636-1656.

Pulvermüller F. (2013). How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends in cognitive sciences, 17(9),* 458–470. https://doi.org/10.1016/j.tics.2013.06.004

Pulvermüller, F., Cappelle, B., Shtyrov, Y., & Alexopoulou, T. (2010). The word processing deficit in semantic dementia: All categories are equal, but some categories are more equal than others. *Journal of Cognitive Neuroscience, 22(9),* 2027-2041. https://doi.org/10.1162/jocn.2009.21339

Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience, 21,* 793–797. https://10.1111/j.1460-9568.2005.03900.x

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with "The language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language, 41*(3), 416–426. https://doi.org/10.1006/jmla.1999.2650

Raven, J. C., & Court, J. H. (1998). *Raven's progressive matrices and vocabulary scales.* Oxford pyschologists Press.

Reber, T. P., & Rothen, N. (2018). Educational app-development needs to be informed by the cognitive neurosciences of learning and memory. *npj Science of Learning, 3,* Article 22. https://doi.org/10.1038/s41539-018-0039-4

Rebuschat, P., & Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics, 33*(4), 829–856. https://doi.org/10.1017/S0142716411000580

Reed, C. L., & Park, G. D. (2021). Functional actions of hands and tools influence attention in peripersonal space. In F. de Vignemont, A. Serino, H. Y. Wong, & A. Farné (Eds.), *The world at our fingertips: A multidisciplinary exploration of peripersonal space* (pp. 101–116). Oxford University Press. https://doi.org/10.1093/oso/9780198851738.003.0006

Reinders, H., and Pegrum, M. (2015). Supporting Language Learning on the Move. An evaluative framework for mobile language learning resources. In Tomlinson, B., *Second Language Acquisition Research and Materials Development for Language Learning*(Eds.), (pp.116-141). Taylor & Francis.

Reinhardt, J. (2019). Social media in second and foreign language teaching and learning: Blogs, wikis, and social networking. *Language Teaching, 52(1),* 1-39. https://doi.org/10.1017/S0261444818000356

Repetto, C., Pedroli, E., & Macedonia, M. (2017). Enrichment Effects of Gestures and Pictures on Abstract Words in a Second Language. *Frontiers in Psychology, 8,* 2136. https://doi.org/10.3389/fpsyg.2017.02136

Ro, T., Russell, C. & Lavie, N. (2001). Changing faces: A detection advantage in the flicker paradigm. *Psychological Science, 12,* 94 – 99. https://doi.org/10.1111/1467-9280.00317

Rocca, R., Wallentin, M., Vesper, C. & Tylen, K. (2019). This is for you: Social modulations of proximal vs. distal space in collaborative interaction. *Scientific Reports, 9,* 14967. https://doi.org/10.1038/s41598-019-51134-8

Romero Lauro, L., Mattavelli, G., Papagno, C., & Tettamanti, M. (2013). She runs, the road runs, my mind runs, bad blood runs between us: Literal and figurative motion verbs: An fMRI study. *NeuroImage, 83,* 361-371. https://doi.org/10.1016/j.neuroimage.2013.06.050

Rončević Zubković, B., Pahljina-Reinić, R., & Kolić-Vehovec, S. (2022). Predictors of ICT use in teaching in different educational domains. *Humanities Today: Proceedings, 1*(1), 75.

Rosell-Aguilar, F. (2017). State of the app: A taxonomy and framework for evaluating language learning mobile applications. *CALICO Journal, 34(2),* 243–258. https://doi.org/10.1558/cj.27623

Rosenthal-von der Pütten, A. M., & Bergmann, K. (2020). Non-verbal enrichment in vocabulary learning with a virtual pedagogical agent. *Frontiers in Psychology, 11*, Article 533839. https://doi.org/10.3389/fpsyg.2020.533839

Rueschemeyer, S.-A., van Rooij, D., Lindemann, O., Willems, R. M., & Bekkering, H. (2010). The function of words: Distinct neural correlates for words denoting differently manipulable objects. *Journal of Cognitive Neuroscience, 22(8),* 1844–1851. https://doi.org/10.1162/jocn.2009.21310

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Sánchez-Mena, A., & Martí-Parreño, J. (2017). Drivers and Barriers to Adopting Gamification: Teachers' Perspectives. *Electronic Journal of e-Learning, 15(5),* 434-443.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research, 12(3),* 329-363. https://doi.org/10.1177/1362168808089921

Schmitt, N., & González-Fernández, B. (2019). Word knowledge: exploring the relationships and order of acquisition of vocabulary knowledge components. *Applied Linguistics, 41(4),* 481-505. https://doi.org/10.1093/applin/amy057

Scott, G. G., Keitel, A., Becirspahic, M., Yao, B., & Sereno, S. C. (2019). The Glasgow Norms: Ratings of 5,500 words on nine scales. *Behavior Research Methods,* 1258-1270. https://doi.org/10.3758/s13428-018-1099-3

Shapiro, L. (Ed.). (2014). *The Routledge handbook of embodied cognition.* Routledge/Taylor & Francis Group

Shattuck-Hufnagel, S., Ren, A., Mathew, M., Yuen, I., & Demuth, K. (2016). Non-referential gestures in adult and child speech: Are they prosodic. In *Proceedings from the 8th international conference on speech prosody* (pp. 836-839). Boston University. https://doi.org/10.21437/speechprosody.2016-171

Shaughnessy, M. (2003). CALL, commercialism and culture: inherent software design conflicts and their results. *ReCALL 15 (2),* 251–268. https://doi.org/10.1017/S0958344003000922

She, H.-C., & Fisher, D. (2002). Teacher communication behavior and its association with students' cognitive and attitudinal outcomes in science in Taiwan. *Journal of Research in Science Teaching*, *39(1)*, 63-78. https://doi.org/10.1002/tea.10009

Shing, S., & Yuan, B. (2016). Apps developed by academics. *Journal of Education and Practice, 7(33),* 1.

Skinner, B. F. (1957). *Verbal behavior*. Appleton-Century-Crofts. https://doi.org/10.1037/11256-000

Slobin, D. I. (1996). From "thought and language" to "thinking for speaking." In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70-96). Cambridge University Press.

Smith, L., & Gasser, M. (2005). The development of embodied cognition: six lessons from babies. *Artificial Life, 11(1-2)*, 13–29. https://doi.org/10.1162/1064546053278973

Smith, Y., Walters, J., & Prior, A. (2012). Translation norms for Hebrew and English. *Edmond J. Safra Technical Report*, *1*, 96-125.

Stahr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in english as a foreign language. *Studies in Second Language Acquisition, 31(4),* 577–607. https://doi.org/10.1017/S0272263109990039

Stam, G., & Tellier, M. (2022). Gesture helps second and foreign language learning and teaching. In A. Morgenstern & S. Goldin-Meadow (Eds.), *Gesture in language: Development across the lifespan* (pp. 335-364). De Gruyter Mouton. https://doi.org/10.1515/9783110567526-014

Starke, A., Leinweber, J. & Ritterfeld, U. (2020). Designing apps to facilitate first and second language acquisition in children, in K. J. Rohlfing & C. Müller-Brauers (Eds.), *International Perspectives on Digital Media and Early Literacy: The Impact of Digital Devices on Learning, Language Acquisition and Social Interaction*. Routledge. https://doi.org/10.4324/9780429321399-11

Sung, Y., Chang, K., & Yang, J.M. (2015). How effective are mobile devices for language learning? A meta-analysis. *Educational Research Review, 16,* 68-84. https://doi.org/10.1016/j.edurev.2015.09.001

Swain, M. (2005). *The output hypothesis: theory and research*. Mahwah, Lawrence Erlbaum.

Sweeney, P., & Moore, C. (2012). Mobile Apps for Learning Vocabulary: Categories, Evaluation and Design Criteria for Teachers and Developers. *International Journal of Computer-Assisted Language Learning and Teaching, 2,* 1-16. https://doi.org/10.4018/ijcallt.2012100101

Taghizadeh, M., & Porkar, P. (2018). Tablet, flashcard and SMS and their effects on EFL learners' attitudes and vocabulary Knowledge. *International Journal of English Language and Translation Studies, 6(1),* 105–118.

Talmy, L. (2000). *Towards a cognitive semantics (Vols. 1-2).* MIT Press.

Taylor, J.E., Beith, A. & Sereno, S.C. (2020). LexOPS: An R package and user interface for the controlled generation of word stimuli. *Behavior Research Methods. 52*, 2372–2382. https://doi.org/10.3758/s13428-020-01389-1

Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture*, *8*(2), 219–235. https://doi.org/10.1075/gest.8.2.06tel

Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. J., Shah, R. J., Malle, B. F., & Morgan, J. L. (2015). Attention to the mouth and gaze following in infancy predict language development. *Journal of child language*, *42*(6), 1173–1190. https://doi.org/10.1017/S0305000914000725

Timotheou, S., Miliou, O., Dimitriadis, Y., Sobrino, S. V., Giannoutsou, N., Cachia, R., Monés, A. M., & Ioannou, A. (2023). Impacts of digital technologies on education and factors influencing schools' digital capacity and transformation: A literature review. *Education and information technologies, 28(6),* 6695–6726. https://doi.org/10.1007/s10639-022-11431-8

Tinsley, T., & Board, K. (2013). Languages: The State of the Nation: Demand and supply of language skills in the UK. *British Academy.*

Tinsley, T., & Doležal, N. (2018). Language Trends 2018: Language teaching in primary and secondary schools in England. *British Council.*

Theeuwes, J., & Van der Stigchel, S. (2006). Faces capture attention: Evidence from inhibition of return. *Visual Cognition,* *13*(6), 657–665. https://doi.org/10.1080/13506280500410949

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Harvard University Press. https://doi.org/10.2307/j.ctv26070v8

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behavioral and Brain Sciences, 28*, 675– 735. https://10.1017/S0140525X05000129

Tomlin, R.S., & Villa, V.M. (1994). Attention in Cognitive Science and Second Language Acquisition. *Studies in Second Language Acquisition, 16*, 183 - 203. https://doi.org 10.1017/S0272263100012870

Tonzar, C., Lotto, L., & Job, R. (2009). L2 vocabulary acquisition in children: Effects of learning method and cognate status. *Language Learning, 59(3)*, 623-646. https://doi.org/10.1111/j.1467-9922.2009.00519.x

Tsuji, S., Fiévet, A., & Cristia, A. (2021). Toddler word learning from contingent screens with and without human presence. *Infant behavior & development, 63*, 101553. https://doi.org/10.1016/j.infbeh.2021.101553

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80(5)*, 352–373. https://doi.org/10.1037/h0020071

Tyler, A., Mueller, C., & Ho, V. (2010). Applying cognitive linguistics to instructed L2 learning: The English modals. *AILA Review, 23*, 30–49.

Tyler, A., Mueller, C., & Ho, V. (2011). Applying cognitive linguistics to learning the semantics of English prepositions to, for, and at: An experimental investigation. *International Journal of Applied Linguistics, 8*, 180–205

Valenzeno, L., Alibali, M. W., & Klatzky, R. (2003). Teachers' gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology, 28(2),* 187–204. https://doi.org/10.1016/S0361-476X(02)00007-3

Verga, L., & Kotz, S. A. (2019). Spatial attention underpins social word learning in the right fronto-parietal network. *NeuroImage, 195,* 165–173. https://doi.org/10.1016/j.neuroimage.2019.03.071

Vermeulen, M., Kreijns, K., van Buuren, H. & Van Acker, F. (2017). The role of transformative leadership, ICT-infrastructure and learning climate in teachers' use of digital learning materials during their classes. *British Journal of Educational Technology.* 48, 1427-1440. https://doi.org/10.1111/bjet.12478

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, *1*(2), 219–247. doi:10.1515/LANGCOG.2009.011

Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 369(1651)*, 20130292. https://doi.org/10.1098/rstb.2013.0292

Vigliocco, G., Vinson, D. P., Druks, J., Barber, H., & Cappa, S. F. (2011). Nouns and verbs in the brain: a review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neuroscience and biobehavioral reviews, 35(3),* 407–426. https://doi.org/10.1016/j.neubiorev.2010.04.007

Vogt, P., van den Berghe, R., de Haas, M., Hoffman, L., Kanero, J., Mamus, E., Montanier, J., Oranç, C., Oudgenoeg-Paz, O., García, D. H., & Pandey, A. K. (2019). Second language tutoring using social robots: A large-scale study. In *Proceedings of the 14th*

ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 497-505). IEEE.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Wang, S., & Lee, C. I. (2021). Multimedia gloss presentation: Learners' preference and the effects on EFL vocabulary learning and reading comprehension. *Frontiers in Psychology, 11*, Article 602520. https://doi.org/10.3389/fpsyg.2020.602520

Wang J., Tigelaar D.E. & Admiraal W. (2019). Connecting rural schools to quality education: rural teachers' use of digital educational resources. *Computers and Human Behaviour, 101*, 68-76. https://doi.org/10.1016/j.chb.2019.07.009

Waxman, S., Fu, X., Arunachalam, S., Leddon, E., Geraghty, K., & Song, H.-j. (2013). Are nouns learned before verbs? Infants provide insight into a long-standing debate. *Child Development Perspectives*, *7(3)*, 155–159. https://doi.org/10.1111/cdep.12032

Webber, N. E. (1978). Pictures and Words as Stimuli in Learning Foreign Language Responses. *The Journal of Psychology, 98(1)*, 57–63. https://doi.org/10.1080/00223980.1978.9915946

Wechsler, D. (2009). Wechsler memory scale (4th ed.). Pearson Assessment.

Welke, D., & Vessel, E. A. (2022). Naturalistic viewing conditions can increase task engagement and aesthetic preference but have only minimal impact on EEG quality. *NeuroImage*, *256*, 119218. https://doi.org/10.1016/J.NEUROIMAGE.2022.119218

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*, 625–636. https://doi.org/10.3758/BF03196322

Witt, P. L., Wheeless, L. R., & Allen, M. (2004). A meta-analytical review of the relationship between teacher immediacy and student learning. *Communication Monographs, 71(2),* 184–207. https://doi.org/10.1080/036452042000228054

Wolf, M.C., Smith, A.C., Rowland, C.F., & Meyer, A.S. (2019). Modality Effects in Vocabulary Acquisition. *Annual Meeting of the Cognitive Science Society.*

Yang, M.K. & Shadiev, R. (2019). Review of Research on Technologies for Language Learning and Teaching. *Open Journal of Social Sciences, 7,* 171-181. https://doi.org/10.4236/jss.2019.73014

Yang, J., Tan, L. H., & Li, P. (2011). Lexical representation of nouns and verbs in the late bilingual brain. *Journal of Neurolinguistics*, *24*(6), 674-682. https://doi.org/10.1016/j.jneuroling.2011.07.002

Yoshii, M. (2006). L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning & Technology, 10*(3), 85-101.

Young, S. S. C., & Wang, Y. H. (2014). The game embedded CALL system to facilitate English vocabulary acquisition and pronunciation. *Journal of Educational Technology & Society, 17(3),* 239–251.

Yu, C., & Ballard, D.H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing, 70,* 2149-2165. https://10.1016/j.neucom.2006.01.034

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition, 125(2),* 244-262. https://doi.org/10.1016/j.cognition.2012.06.016

Zhang, P., & Zhang, S. (2024). Multimedia enhanced vocabulary learning: The role of input condition and learner-related factors. *System, 122*, Article 103275. https://doi.org/10.1016/j.system.2024.103275

Zhang, R, Zou, D. (2020): Types, purposes, and effectiveness of state of-the-art technologies for second and foreign language learning, Computer Assisted Language Learning. https://doi.org/10.1080/09588221.2020.1744666

Zhang, X., Han, Z., & Bi, Y. (2013). Are abstract and concrete concepts organized differently? Evidence from the blocked translation paradigm. *Applied Psycholinguistics, 34(5)*, 1059–1092. https://doi.org/10.1017/S0142716412000124

Zhao, H., Huang, S., Zhou, Y., & Wang, R. (2020). Schematic diagrams in second language learning of english prepositions: a behavioral and event-related potential study. *Studies in Second Language Acquisition*, *42*(4), 721–748. doi:10.1017/S027226311900069X

Zwaan, R. (2009). Theory development n embodiment research: getting the timing right. *European Journal of Social Psychology, 7*, 1160-1161. https://doi.org/10.1002/ejsp.677

Zwaan, R. A. (2014). Embodiment and language comprehension: Reframing the discussion. *Trends in Cognitive Sciences, 18(5)*, 229-234. https://doi.org/10.1016/j.tics.2014.02.008

Zwaan, R. A. (2021). Two challenges to “embodied cognition” research and how to overcome them. *Journal of Cognition, 4(1)*, Article 14. https://doi.org/10.5334/joc.151

Zwaan, R. A., & Kaschak, M. (2006). *Language, visual cognition and motor action.* In K. Brown (Ed.), *Encyclopedia of language & linguistics* (2nd ed., pp. 648–651). Elsevier. https://doi.org/10.1016/B0-08-044854-2/00865-8

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shape of objects. *Psychological Science, 13*(2), 168–171. https://doi.org/10.1111/1467-9280.00430

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General*, *135*(1), 1–11. https://doi.org/10.1037/0096-3445.135.1.1

Appendices

1. Appendix A – Supplemental materials for chapter 2

1.1. Stimuli list

Table A.1. Table showing a list of words and pseudowords, organized by ambiguity and word class categories

Unambiguous					
Nouns		**Verbs**		**Prepositions**	
Pseudo-word	**L1**	**Pseudo-word**	**L1**	**Pseudo-word**	**L1**
Tarpal	Window	Rillar	Sit	Chig	With
Laguni	Bus	Whans	Eat	Zain	Near
Havo	Rabbit	Bick	Stop	Yop	Between
Nells	Shoe	Brone	Learn	Ari	Up
Firang	Rain	Pusk	Carry	Gelf	Under
Kander	House	Teane	Pull	Prem	Down
Brip	Chocolate	Cruck	Catch	Mork	Outside
Blace	Bird	Hoag	Pick	Lusk	Toward
Ambiond	Ball	Prush	Throw	Mish	Across
Vorter	Car	Pheep	Write	Pook	Among
Ambiguous					
Nouns		**Verbs**		**Prepositions**	
Pseudo-word	**L1**	**Pseudo-word**	**L1**	**Pseudo-word**	**L1**
Mudge	Fish (animal)	Dertex	Cut (with scissors)	Vulk	On (a flat surface)
Plunt	Fish (meat)	Rulk	Cut (with a blade)	Apen	On (vertical support)
Fimit	Flower (planted)	Zob	Remove (from loose fit)	Blat	In (tight fit)
Wack	Flower (picked)	Skows	Remove (from tight fit)	Sif	In (loose fit)
Rerry	Vehicle (on land)	Dital	Break (by smashing)	Elho	Behind (objects moving)
Dereve	Vehicle (on water)	Whirp	Break (of brittle and long things)	Crog	Behind (objects still)
Clerry	Snow (in the air)	Pord	Open (a container)	Zat	Over (objects in contact)
Zore	Snow (on the ground)	Dros	Open (spread out flat)	Nary	Over (objects not in contact)
Lommon	Water (in a container)	Jore	Hold (fabric)	Ul	At (point of coincidence)
Verking	Water (natural body of water)	Bems	Hold (long object)	Sely	At (functional interaction)

1.2. The outputs of generalized linear mixed-effects models across Experiment 1

Table A.2. Overview of the generalized linear mixed-effects model for the immediate recognition data. Significant values are presented in bold.

Fixed Effects

	X^2	df	p
learning_modality	0.180	1.00	.672
ambiguity	**60.388**	**1.00**	**< .001**
wordclass	**14.20**	**2.00**	**< .001**
task_modality	**5.876**	**1.00**	**.015**
learning_modality * ambiguity	**7.982**	**1.00**	**.005**
learning_modality * wordclass	3.428	2.00	.180
ambiguity * wordclass	4.383	2.00	.112
learning_modality * task_modality	3.068	1.00	.080
ambiguity * task_modality	2.786	1.00	.095
wordclass * task_modality	**36.091**	**2.00**	**< .001**
learning_modality * ambiguity * wordclass	2.155	2.00	.340
learning_modality * ambiguity * task_modality	**26.906**	**1.00**	**< .001**
learning_modality * wordclass * task_modality	3.707	2.00	.157
ambiguity * wordclass * task_modality	5.200	2.00	.074

Model specification: score ~ 1 + learning_modality + ambiguity + wordclass + task_modality + learning_modality:ambiguity + learning_modality:wordclass + ambiguity:wordclass + learning_modality:task_modality + ambiguity:task_modality + wordclass:task_modality + learning_modality:ambiguity:c_wordclass + learning_modality:ambiguity:task_modality + c learning_modality:wordclass:task_modality + ambiguity:wordclass:task_modality + (1 | word_L2) + (1 | ID)

Table A.3. Overview of the generalized linear mixed-effects model for the delayed recognition data. Significant values are presented in bold.

Fixed Effects

	X^2	df	p
learning_modality	0.232	1.00	.630
ambiguity	**57.995**	**1.00**	**< .001**
wordclass	**13.403**	**2.00**	**.001**
task_modality	3.753	1.00	.053
learning_modality* ambiguity	**9.469**	**1.00**	**.002**
learning_modality*wordclass	4.381	2.00	.112
ambiguity*wordclass	4.311	2.00	.116
learning_modality*task_modality	3.570	1.00	.059
ambiguity*task_modality	**5.048**	**1.00**	**.025**
wordclass*task_modality	**46.522**	**2.00**	**< .001**
learning_modality*ambiguity*wordclass	3.555	2.00	.169
learning_modality*ambiguity*task_modality	**51.480**	**1.00**	**< .001**
learning_modality*wordclass*task_modality	4.714	2.00	.095
ambiguity*wordclass*task_modality	**11.999**	**2.00**	**.002**

Model specification: score ~ 1 + learning_modality + ambiguity + wordclass + task_modality + learning_modality:ambiguity + learning_modality:wordclass + ambiguity:wordclass + learning_modality:task_modality + ambiguity:task_modality + wordclass:task_modality + learning_modality:ambiguity:c_wordclass + learning_modality:ambiguity:task_modality + c learning_modality:wordclass:task_modality + ambiguity:wordclass:task_modality + (1 | word_L2) + (1 | ID)

Table A.4. Overview of the generalized linear mixed-effects model for the immediate production data. Significant values are presented in bold.

Fixed Effects

	X²	df	p
learning_modality	1.311	1.00	0.252
ambiguity	**25.820**	**1.00**	**< .001**
wordclass	**34.868**	**2.00**	**< .001**
task_modality	1.839	1.00	0.175
learning_modality*ambiguity	**28.123**	**1.00**	**< .001**
learning_modality*wordclass	**13.633**	**2.00**	**0.001**
ambiguity*wordclass	0.505	2.00	0.777
learning_modality*task_modality	0.146	1.00	0.702
ambiguity*task_modality	0.097	1.00	0.755
wordclass*task_modality	3.743	2.00	0.154
learning_modality* ambiguity*wordclass	0.5284	2.00	0.768
learning_modality*ambiguity*task_modality	**12.827**	**1.00**	**< .001**
learning_modality*wordclass*task_modality	2.483	2.00	0.289
ambiguity*wordclass*task_modality	5.966	2.00	0.051

Model specification: score ~ 1 + learning_modality + ambiguity + wordclass + task_modality + learning_modality:ambiguity + learning_modality:wordclass + ambiguity:wordclass + learning_modality:task_modality + ambiguity:task_modality + wordclass:task_modality + learning_modality:ambiguity:c_wordclass + learning_modality:ambiguity:task_modality + c learning_modality:wordclass:task_modality + ambiguity:wordclass:task_modality + (1 | word_L2) + (1 | ID)

Table A.5. Overview of the generalized linear mixed-effects model for the delayed production data. Significant values are presented in bold.

Fixed Effects

	X²	df	p

learning_modality	0.031	1.00	0.860
ambiguity	**15.710**	**1.00**	**< .001**
wordclass	**43.69**	**2.00**	**< .001**
task_modality	1.245	1.00	0.264
learning_modality*ambiguity	**28.505**	**1.00**	**< .001**
learning_modality*wordclass	**11.326**	**2.00**	**0.003**
ambiguity*wordclass	0.657	2.00	0.720
learning_modality*task_modality	0.365	1.00	0.546
ambiguity*task_modality	**10.813**	**1.00**	**0.001**
wordclass*task_modality	**14.904**	**2.00**	**< .001**
learning_modality*ambiguity*wordclass	**9.193**	**2.00**	**0.010**
learning_modality*ambiguity*task_modality	**3.975**	**1.00**	**0.046**
learning_modality*wordclass*task_modality	2.834	2.00	0.242
ambiguity*wordclass*task_modality	4.268	2.00	0.118

Model specification: score ~ 1 + learning_modality + ambiguity + wordclass + task_modality + learning_modality:ambiguity + learning_modality:wordclass + ambiguity:wordclass + learning_modality:task_modality + ambiguity:task_modality + wordclass:task_modality + learning_modality:ambiguity:c_wordclass + learning_modality:ambiguity:task_modality + c learning_modality:wordclass:task_modality + ambiguity:wordclass:task_modality + (1 | word_L2) + (1 | ID)

1.3. The outputs of generalized linear mixed-effects models across Experiment 2

Table A.6. Overview of the generalized linear mixed-effects model for the immediate recognition data. Significant values are presented in bold.

Fixed Effects

	X^2	df	p
learning_modality	0.338	1.00	0.561
testing_modality	0.095	1.00	0.757
cue_novelty	2.715	1.00	0.099
word_class	**53.327**	**2.00**	**<.001**
learning_modality*testing_modality	**4.335**	**1.00**	**0.037**
learning_modality*cue_novelty	1.861	1.00	0.172
testing_modality*cue_novelty	0.217	1.00	0.641
learning_modality*word_class	3.022	2.00	0.221
testing_modality*word_class	**14.09**	**2.00**	**<.001**
cue_novelty*word_class	3.809	2.00	0.149
learning_modality*testing_modality*cue_novelty	0.06	1.00	0.806
learning_modality*testing_modality*word_class	2.707	2.00	0.258
learning_modality*cue_novelty*word_class	1.448	2.00	0.485
testing_modality*cue_novelty*word_class	**9.426**	**2.00**	**0.009**

Model specification: score ~ 1 + learning_modality + testing_modality + cue_novelty + word_class + learning_modality:testing_modality + learning_modality:cue_novelty + testing_modality:cue_novelty + learning_modality:word_class + testing_modality:word_class + cue_novelty:word_class + learning_modality:testing_modality:cue_novelty + learning_modality:testing_modality:word_class + learning_modality:cue_novelty:word_class + testing_modality:cue_novelty:word_class + (1 | word) + (1 | ID)

Table A.7. Overview of the generalized linear mixed-effects model for the delayed recognition data. Significant values are presented in bold.

Fixed Effects	X^2	df	p
learning_modality	0.308	1.00	0.579
testing_modality	1.084	1.00	0.298
cue_novelty	1.216	1.00	0.270
word_class	**42.835**	**2.00**	**<.001**
learning_modality*testing_modality	**7.051**	**1.00**	**0.008**
learning_modality*cue_novelty	3.525	1.00	0.060
testing_modality*cue_novelty	0.014	1.00	0.906
learning_modality*word_class	1.397	2.00	0.497
testing_modality*word_class	5.084	2.00	0.079
cue_novelty*word_class	0.167	2.00	0.920
learning_modality*testing_modality*cue_novelty	0.119	1.00	0.730
learning_modality*testing_modality*word_class	1.269	2.00	0.530
learning_modality*cue_novelty*word_class	0.970	2.00	0.616
testing_modality*cue_novelty*word_class	2.02	2.00	0.364

Model specification: score ~ 1 + learning_modality + testing_modality + cue_novelty + word_class + learning_modality:testing_modality + learning_modality:cue_novelty + testing_modality:cue_novelty + learning_modality:word_class + testing_modality:word_class + cue_novelty:word_class + learning_modality:testing_modality:cue_novelty + learning_modality:testing_modality:word_class + learning_modality:cue_novelty:word_class + testing_modality:cue_novelty:word_class + (1 | word) + (1 | ID)

Table A.8. Overview of the generalized linear mixed-effects model for the immediate production data. Significant values are presented in bold.

Fixed Effects	X^2	df	p
learning_modality	0.835	1.00	0.361
testing_modality	0.891	1.00	0.345
cue_novelty	3.069	1.00	0.080

word_class	**13.05**	**2.00**	**0.001**
learning_modality*testing_modality	0.818	1.00	0.366
learning_modality*cue_novelty	0.301	1.00	0.583
testing_modality*cue_novelty	1.051	1.00	0.305
learning_modality*word_class	**6.382**	**2.00**	**0.041**
testing_modality*word_class	**7.545**	**2.00**	**0.023**
cue_novelty*word_class	0.225	2.00	0.894
learning_modality*testing_modality*cue_novelty	0.038	1.00	0.845
learning_modality*testing_modality*word_class	1.430	2.00	0.489
learning_modality*cue_novelty*word_class	3.800	2.00	0.150
testing_modality*cue_novelty*word_class	1.787	2.00	0.409

Model specification: score ~ 1 + learning_modality + testing_modality + cue_novelty + word_class + learning_modality:testing_modality + learning_modality:cue_novelty + testing_modality:cue_novelty + learning_modality:word_class + testing_modality:word_class + cue_novelty:word_class + learning_modality:testing_modality:cue_novelty + learning_modality:testing_modality:word_class + learning_modality:cue_novelty:word_class + testing_modality:cue_novelty:word_class + (1 | word) + (1 | ID)

Table A.9. Overview of the generalized linear mixed-effects model for the delayed production data. Significant values are presented in bold.

Fixed Effects

	X²	df	p
learning_modality	0.045	1.00	0.832
testing_modality	1.629	1.00	0.202
cue_novelty	0.141	1.00	0.707
word_class	**22.228**	**2.00**	**<.001**
learning_modality*testing_modality	3.274	1.00	0.070
learning_modality*cue_novelty	1.554	1.00	0.213
testing_modality*cue_novelty	1.871	1.00	0.171
learning_modality*word_class	0.657	2.00	0.720
testing_modality*word_class	**8.512**	**2.00**	**0.014**
cue_novelty*word_class	2.504	2.00	0.286
learning_modality*testing_modality*cue_novelty	0.056	1.00	0.813
learning_modality*testing_modality*word_class	2.003	2.00	0.367
learning_modality*cue_novelty*word_class	4.920	2.00	0.085
testing_modality*cue_novelty*word_class	**6.214**	**2.00**	**0.045**

Model specification: score ~ 1 + learning_modality + testing_modality + cue_novelty + word_class + learning_modality:testing_modality + learning_modality:cue_novelty + testing_modality:cue_novelty + learning_modality:word_class + testing_modality:word_class + cue_novelty:word_class + learning_modality:testing_modality:cue_novelty + learning_modality:testing_modality:word_class +

learning_modality:cue_novelty:word_class + testing_modality:cue_novelty:word_class + (1 | word) + (1 | ID)

1.4. Individual differences across Experiments 1 and 2

Table A.10. Descriptive statistics values for Millhill test separated by group (experiment 1)

Group	n	Mean	SD	Median	Min	Max
picture_picture	40	16.2	3.11	17	8	21
picture_text	40	17.4	3.41	17	11	27
text_picture	40	16.9	3.73	17	6	24
text_text	40	16.9	3.02	17	11	24

Table A.11. ANOVA Summary table for Millhill test scores testing differences between groups (experiment 1)

Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	3	29.8	9.94	0.898	0.444
Residuals	153	1693.6	11.07		

Table A.12. Descriptive statistics values for Matrix reasoning test separated by group (experiment 1)

Group	n	Mean	SD	Median	Min	Max
picture_picture	40	17.6	3.02	18	5	22
picture_text	40	19.2	3.25	19	12	24
text_picture	40	19.4	3.33	20	8	24
text_text	40	19.0	2.56	20	13	24

Table A.13. ANOVA Summary table for Matrix test scores testing differences between groups (experiment 1)

Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	3	81.7	27.240	2.923	0.0358
Residuals	156	1453.6	9.318		

Table A.14. Descriptive statistics values for Millhill test separated by group (experiment 2)

Group	**n**	**Mean**	**SD**	**Median**	**Min**	**Max**
picture_picture	40	17.0	3.21	17	9	29
picture_text	41	18.4	3.34	18	10	27
text_picture	40	17.2	3.06	17	9	28

text_text	41	16.4	4.14	16	3	24

Table A.15. ANOVA Summary table for Millhill test scores testing differences between groups (experiment 2)

Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	3	90.2	30.06	1.57	0.199
Residuals	158	3024.3	19.14		

Table A.16. Descriptive statistics values for Matrix reasoning test separated by group (experiment 2)

Group	n	Mean	SD	Median	Min	Max
picture_picture	40	18.0	3.21	18.0	11	25
picture_text	41	19.0	3.34	20.0	11	26
text_picture	40	19.3	3.06	19.5	13	24
text_text	41	18.8	4.14	20.0	8	25

Table A.17. ANOVA Summary table for Matrix test scores testing differences between groups (experiment 2)

Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	3	36.6	12.20	1.015	0.388
Residuals	158	1899.4	12.02		

Table A.18. Overview of the generalized linear mixed-effects model for the immediate recognition data with Matrix test score as a covariate (experiment 1). Significant values are presented in bold.

Fixed effects

Effect	**X²**	**df**	**p**
learning_modality	1.424	1.00	0.233
ambiguity	**60.493**	**1.00**	**< 0.001**
wordclass	**14.220**	**2.00**	**< 0.001**
task_modality	**4.522**	**1.00**	**0.033**
matrix_score	**28.491**	**1.00**	**< 0.001**
learning_modality × ambiguity	**8.092**	**1.00**	**0.004**
learning_modality × wordclass	3.403	2.00	0.182
ambiguity × wordclass	4.384	2.00	0.112
learning_modality × task_modality	0.985	1.00	0.321
ambiguity × task_modality	2.839	1.00	0.092
wordclass × task_modality	**36.025**	**2.00**	**< 0.001**
learning_modality × ambiguity × wordclass	2.142	2.00	0.343
learning_modality × ambiguity × task_modality	**27.711**	**1.00**	**< 0.001**
learning_modality × wordclass × task_modality	3.735	2.00	0.155
ambiguity × wordclass × task_modality	5.167	2.00	0.075

Model specification: score ~ 1 + learning_modality + ambiguity + wordclass + task_modality + **matrix_score** + learning_modality:ambiguity + learning_modality:wordclass + ambiguity:wordclass + learning_modality:task_modality + ambiguity:task_modality + wordclass:task_modality + learning_modality:ambiguity:c_wordclass + learning_modality:ambiguity:task_modality + c learning_modality:wordclass:task_modality + ambiguity:wordclass:task_modality + (1 | word_L2) + (1 | ID)

Table A.19. Overview of the generalized linear mixed-effects model for the delayed recognition data with Matrix test score as a covariate (experiment 1). Significant values are presented in bold.

Fixed effects

Effect	**X²**	**df**	**p**
learning_modality	1.95	1.00	0.162
ambiguity	**58.07**	**1.00**	**< 0.001**
wordclass	**13.43**	**2.00**	**0.001**
learning_modality	2.47	1.00	0.116
matrix_score	**39.32**	**1.00**	**< 0.001**
learning_modality × ambiguity	**9.62**	**1.00**	**0.002**
learning_modality × wordclass	4.33	2.00	0.115
ambiguity × wordclass	4.32	2.00	0.115
learning_modality × task_modality	1.08	1.00	0.298

Fixed effects

Effect	X²	df	p
ambiguity × task_modality	**5.07**	**1.00**	**0.024**
wordclass × task_modality	**46.51**	**2.00**	**< 0.001**
learning_modality × ambiguity × wordclass	3.52	2.00	0.172
learning_modality × ambiguity × task_modality	**51.61**	**1.00**	**< 0.001**
learning_modality × wordclass × task_modality	4.72	2.00	0.095
ambiguity × wordclass × task_modality	**12.00**	**2.00**	**0.002**

Model specification: score ~ 1 + learning_modality + ambiguity + wordclass + task_modality + **matrix_score** + learning_modality:ambiguity + learning_modality:wordclass + ambiguity:wordclass + learning_modality:task_modality + ambiguity:task_modality + wordclass:task_modality + learning_modality:ambiguity:c_wordclass + learning_modality:ambiguity:task_modality + c learning_modality:wordclass:task_modality + ambiguity:wordclass:task_modality + (1 | word_L2) + (1 | ID)

Table A.20. Overview of the generalized linear mixed-effects model for the immediate production data with Matrix test score as a covariate (experiment 1). Significant values are presented in bold.

Fixed effects

Effect	X²	df	p
learning_modality	2.223	1.00	0.136
ambiguity	**25.833**	**1.00**	**< 0.001**
wordclass	**34.854**	**2.00**	**< 0.001**
learning_modality	1.309	1.00	0.253
matrix_score	**5.987**	**1.00**	**0.014**
learning_modality × ambiguity	**28.171**	**1.00**	**< 0.001**
learning_modality × wordclass	**13.650**	**2.00**	**0.001**
ambiguity × wordclass	0.506	2.00	0.776
learning_modality × task_modality	4.42e-4	1.00	0.983
ambiguity × task_modality	0.100	1.00	0.752
wordclass × task_modality	3.731	2.00	0.155
learning_modality × ambiguity × wordclass	0.529	2.00	0.768
learning_modality × ambiguity × task_modality	**12.826**	**1.00**	**< 0.001**
learning_modality × wordclass × task_modality	2.482	2.00	0.289
ambiguity × wordclass × task_modality	5.969	2.00	0.051

Model specification: score ~ 1 + learning_modality + ambiguity + wordclass + task_modality + **matrix_score** + learning_modality:ambiguity + learning_modality:wordclass + ambiguity:wordclass + learning_modality:task_modality + ambiguity:task_modality + wordclass:task_modality + learning_modality:ambiguity:c_wordclass + learning_modality:ambiguity:task_modality + c learning_modality:wordclass:task_modality + ambiguity:wordclass:task_modality + (1 | word_L2) + (1 | ID)

Table A.21. Overview of the generalized linear mixed-effects model for the delayed production data with Matrix test score as a covariate (experiment 1). Significant values are presented in bold.

Fixed effects

Effect	**X²**	**df**	**p**
learning_modality	0.3045	1.00	0.581
ambiguity	**15.6995**	**1.00**	**< 0.001**
wordclass	**43.6621**	**2.00**	**< 0.001**
learning_modality	0.7593	1.00	0.384
matrix_score	**7.8602**	**1.00**	**0.005**
learning_modality × ambiguity	**28.5457**	**1.00**	**< 0.001**
learning_modality × wordclass	**11.3422**	**2.00**	**0.003**
ambiguity × wordclass	0.6582	2.00	0.720
learning_modality × task_modality	0.0192	1.00	0.890
ambiguity × task_modality	0.7762	1.00	0.378
wordclass × task_modality	**14.8647**	**2.00**	**< 0.001**
learning_modality × ambiguity × wordclass	**9.1925**	**2.00**	**0.010**
learning_modality × ambiguity × task_modality	**3.9714**	**1.00**	**0.046**
learning_modality × wordclass × task_modality	2.8283	2.00	0.243
ambiguity × wordclass × task_modality	4.2714	2.00	0.118

Model specification: score ~ 1 + learning_modality + ambiguity + wordclass + task_modality + **matrix_score** + learning_modality:ambiguity + learning_modality:wordclass + ambiguity:wordclass + learning_modality:task_modality + ambiguity:task_modality + wordclass:task_modality + learning_modality:ambiguity:c_wordclass + learning_modality:ambiguity:task_modality + c learning_modality:wordclass:task_modality + ambiguity:wordclass:task_modality + (1 | word_L2) + (1 | ID)

2. Appendix B – Supplemental materials for chapter 3

2.1. List of pseudowords

Table B.1. Table showing the pseudowords used as target stimuli form for experiments in chapter 2

WORDLIST A (4 letters, 2 syllables)	**WORDLIST B** (5 letters, 2 syllables)	**WORDLIST C** (6 letters, 3 syllables)	**WORDLIST D** (7 letters, 3 syllables)
saro	bueda	neveto	bafioro
bava	tromo	dekido	tavilco
abes	firpa	isalto	nerbona
ibre	yarza	conveo	deroxte
laga	conzo	tovasa	fesiano
izmo	milbo	seanos	almieso
poad	sigre	temiga	parensa
bipa	resel	pamatu	acrivar
luvo	gimar	civeno	sovosta
azme	zindo	revato	inzagio

2.2. List of items

Table B.1. Table showing the target items used for experiments in chapter 2.

WORD	**Category**
Egg	food
Bread	food
Muffin	food
Mushroom	food
Orange	food
Apple	food
Lemon	food

Banana	food
Pear	food
Potato	food
Pepper	food
Onion	food
Tomato	food
Carrot	food
Corn	food
Lettuce	food
Book	household_items
Tape	household_items
Ruler	household_items
Notebook	household_items
Box	household_items
Clock	household_items
Vase	household_items
Battery	household_items
Candle	household_items
Cup	household_items
Glass	household_items
Plate	household_items
Bowl	household_items
Bottle	household_items
Tray	household_items
Envelope	household_items

2.3. Model outputs across experiments in chapter 2

Table B.1. Summary of the generalized linear mixed-effects model for word learning outcomes measured by the production test (Experiment 1)

Model equation: model_p <- glmer(score ~ gaze*pointing + (1|id) + (1|item), data = total_production1, family = binomial())						
Fixed Effects						
	Estimate	Std. Error	Z value	p value	Lower CI (2.5 %)	Upper CI (97. 5%)
Intercept	-0.979	0.21	-4.658	<0.001	-1.392	-0.567
Gaze	-0.079	0.077	-1.022	0.307	-0.233	0.072
Pointing	0.077	0.077	0.999	0.318	-0.074	0.228
Gaze * pointing	0.257	0.154	1.663	0.096	-0.046	0.559
Random Effects						
	Variance	SD				
ID	3.524	1.877				
Item	0.563	0.750				

Table B.2. Summary of the generalized linear mixed-effects model for word learning outcomes measured by the recognition test (Experiment 1)

Model equation: model_r <- glmer(score ~ gaze*pointing + (1|id) + (1|item), data = total_recognition, family = binomial())						
Fixed Effects						
	Estimate	Std. Error	Z value	p value	Lower CI (2.5 %)	Upper CI (97. 5%)
Intercept	1.532	0.174	8.822	<0.001	1.191	1.872
Gaze	-0.039	0.077	-0.500	0.617	-0.190	0.113
Pointing	-0.036	0.077	-0.470	0.638	-0.188	0.115
Gaze * pointing	0.159	0.154	1.028	0.304	-0.144	0.461

Random Effects						
	Variance	SD				
ID	1.835	1.355				
Item	0.48	0.693				

Table B.3. Summary of the generalized linear mixed-effects model for word learning outcomes measured by the production test (Experiment 2)

Model equation: model_r <- glmer(score ~ gaze*pointing + (1|id) + (1|item), data = total_recognition, family = binomial())						
Fixed Effects						
	Estimate	Std. Error	Z value	p value	Lower CI (2.5 %)	Upper CI (97. 5%)
Intercept	-1.106	0.200	-5.534	<0.001	-1.505	-0.713
Gaze	0.142	0.107	1.325	0.185	-0.071	0.355
Pointing	0.086	0.085	1.006	0.314	-0.084	0.256
Gaze * pointing	-0.142	0.17	-0.834	0.405	-0.482	0.197
Random Effects						
	Variance	SD				
ID	2.869	1.694				
Item	0.577	0.759				

3. Appendix C – Supplemental materials for chapter 4

Table C.1. Mean ratings of each category for all the evaluated software

Rating components	**Mean score**	**SD**
Active	4.2	1.424
Engaging	5.53	1.30
Meaningful	5.07	1.46
Social	3.20	1.92
Input	5.53	1.04
Output	4.30	1.68
Interaction	4.30	1.56
Practice	4.65	1.61

Table C.2. Mean ratings across rating subsections and total scores for general learning principles

GENERAL LEARNING PRINCIPLES MEAN RATINGS						
Digital tool	**N raters**	**Pillar I: Active**	**Pillar II: Engaging**	**Pillar III: Meaningful**	**Pillar IV: Social**	**Total**
1. Language Angels	4	6.25	7	4.86	3.67	**21.78**
2. This Is Language	3	3.5	5.5	6.17	5.67	**20.84**
3. Language Gym	4	5.13	6	5.25	3.04	**19.42**
4. Languagenut	4	4.33	6.17	5.84	2.5	**18.84**
5. Kahoot	2	4.25	4.25	4	6.25	**18.75**
6. Activelearn	9	4.18	6.29	4.86	3.08	**18.41**
7. Kerboodle	4	3.63	5	5.25	4.34	**18.22**
8. Quizlet	16	3.88	6	5.16	2.79	**17.83**
9. Lingua Scope	9	4.21	5.67	4.71	2.38	**16.97**
10. Memrise	4	4.67	5.34	4.5	1.83	**16.34**
11. Duolingo	3	4	5	5	/	**14**

Table C.3. Mean ratings across rating subsections and total scores for language learning principles

LANGUAGE LEARNING PRINCIPLES MEAN RATINGS						
Digital tool	**N raters**	**Input**	**Output**	**Interaction**	**Practice**	**Total**
1. Language Angels	4	6.92	6.78	5.68	6.42	**25.8**
2. Duolingo	3	6.13	6.17	6	6.38	**24.68**
3. Languagenut	4	6.25	6	5	5.92	**23.17**
4. Language Gym	4	5.75	4.58	4.75	5.75	**20.83**
5. Activelearn	9	6.19	4.47	4.53	4.88	**20.07**
6. Memrise	4	5.75	3.78	3.68	5.58	**18.79**
7. This Is Language	3	5.92	2.33	5.68	4.5	**18.41**
8. Kerboodle	4	5.46	4.83	3.79	4.21	**18.29**
9. Quizlet	16	5.07	4.20	4.26	4.49	**18.02**
10. Kahoot	2	3.88	1.5	5.25	4.5	**15.13**
11. Lingua Scope	9	5.10	3.22	3.04	3.11	**14.47**

Table C.4. Table showing ratings of software in free and paid categories on general and language subcomponents, as well as the total score

GENERAL LEARNING PRINCIPLES					
Digital tool	**Active**	**Engaging**	**Meaningful**	**Social**	**Total**
1. Free	3.93	5.46	4.89	2.62	**16.9**
2. Paid	4.47	5.69	5.18	3.53	**18.87**
LANGUAGE LEARNING PRINCIPLES					
Digital tool	**Input**	**Output**	**Interaction**	**Practice**	**Total**
1. Free	5.19	4.21	4.2	4.58	**18.18**
2. Paid	5.69	4.31	4.4	4.67	**19.07**

TOTAL SCORE			
Digital tool	**General learning principles**	**Language learning principles**	**Final score**
1. Free	17.04	18.18	**35.22**
2. Paid	18.92	19.07	**37.99**

Table C.5. Table showing the results of the Wilcoxon rank-sum test comparing free and paid software.

Area of interest	Subcomponent	Analysis results
		Cost
General learning principles	Active	W = 323, p = 0.3598
	Engaging	W = 416, p = 0.6808
	Meaningful	W = 298, p = 0.5316
	Social	W = 191, p = 0.198
Language learning principles	Input	W = 284, p = 0.0727
	Output	W = 418.5, p = 0.7883
	Interaction	W = 213.5, p = 0.179
	Practice	W = 338, p = 0.7126

Table C.6. Table showing the most frequently used and highest rated software in ascending order, from most frequently used to the less frequently used. Additionally, the general properties of each software are listed.

Software name	N users in the sample	Rating*	Software type	Business model	Target education level	Content origin
Quizlet	27	35.85	Platform	Free Individual/school-level subscription (bonus features)	Not defined	Software User Community
Lingua Scope	24	31.44	Platform App	School-level subscription Individual-level subscription	Primary and secondary	Software User
Blooket	19	/*	Platform	School-level subscription Individual-level subscription	Not defined	Software User Community
Kahoot	15	33.88	Platform	Individual-level subscription School-level subscription	Not defined	Software User
Activelearn	14	38.48	Platform	School-level subscription	Secondary	Software

Language Gym	9	40.25	Website	School-level subscription	Not defined	Software
Kerboodle	8	36.51	Platform	School-level subscription	Primary and secondary	Software
Languages online	7	/*	Website	Free	Primary and secondary	Software
Memrise	7	35.13	Platform App	Free/ Individual subscription for bonus features	Not defined	Software
Wordwall	7	/*	Platform	School-level subscription Individual-level subscription	Not defined	User Community
Languagenut	6	42.01	Platform App	School-level subscription	Primary and secondary	Software User
This Is Language	6	39.25	Platform	School-level subscription	Not defined	Software
Duolingo	6	38.68	App	Free	Not defined	Software
Language Angels	4	47.58	Platform	School-level subscription	Primary	Software

* The rating for this software is not included, as there were <2 raters per software