The Impact of AR-Enhanced AI Apps on Vocabulary Retention through Spatial and Environmental Learning

Dr. T. Jude Livingston[1] and Dr. Y. Premila Anbarasi[2]
[1]Assistant Professor of English, Nazareth Margoschis College at Pillaiyanmanai,
Nazareth – 628 617, Thoothukudi District, Tamilnadu.
Affiliated to Manonmaniam Sundaranar University, Tirunelveli – 627 012, Tamilnadu.
[2]Associate Professor of English, Nazareth Margoschis College at Pillaiyamanai,
Nazareth – 628 617, Thoothukudi District, Tamilnadu.
Affiliated to Manonmaniam Sundaranar University, Tirunelveli – 627 012, Tamilnadu.

Abstract

This paper investigates the pedagogical impact of augmented reality (AR) enhanced artificial intelligence (AI) applications on vocabulary retention in second language acquisition, with a particular focus on the mechanisms of spatial and environmental learning. As mobile AR hardware has matured and on-device large language models have become commonplace, a new generation of language-learning applications now overlays target-language vocabulary onto the learner's physical surroundings in real time. Drawing on cognitive load theory, dual coding theory, embodied cognition, and the encoding specificity principle, this paper synthesizes recent empirical literature and theoretical frameworks to argue that AR-AI vocabulary learning produces measurably stronger retention than conventional flashcard-based or screen-only mobile applications. The paper examines how the binding of lexical items to physical locations, multisensory input, and contextually adaptive AI feedback collectively scaffold long-term memory. It also addresses methodological challenges in the field, equity and access concerns, and directions for future research, including longitudinal studies, neuroimaging-based confirmation of memory mechanisms, and the implications of fully ambient learning environments. The paper concludes that AR-AI applications represent a meaningful pedagogical shift rather than a novelty, but that their effectiveness depends on principled instructional design grounded in established theories of memory and learning.

Keywords: augmented reality, artificial intelligence, vocabulary acquisition, mobile-assisted language learning, spatial memory, second language acquisition.

Introduction

Vocabulary acquisition has long been recognized as one of the most labor-intensive and yet most decisive components of second language learning. Estimates from corpus linguistics suggest that a learner needs receptive knowledge of roughly eight to nine thousand word families to comprehend authentic written texts and around six to seven thousand for spoken discourse. Achieving this scale of lexical knowledge through traditional list memorization is slow, prone to forgetting, and frequently fails to translate into communicative competence. As a result, language educators and applied linguists have for decades sought instructional approaches that strengthen the encoding, storage, and retrieval of lexical items.

Mobile-assisted language learning (MALL) emerged in the early 2010s as a partial answer to this problem. Applications such as Duolingo, Memrise, and Anki demonstrated that spaced repetition algorithms, gamification, and ubiquitous access to study sessions could improve learner engagement and short-term recall. However, even the most refined MALL platforms have largely operated within a screen-bound paradigm: the learner views decontextualized words or sentences on a flat display, performs recognition or production exercises, and closes the application. Critics have noted that such systems risk producing the kind of inert vocabulary knowledge that learners can recognize on a test but cannot deploy in lived communicative contexts.

The convergence of two technologies has begun to reshape this landscape. The first is augmented reality, which overlays digital information onto a user's perception of the physical world via smartphones, smart glasses, or head-mounted displays. The second is the rapid maturation of large language models and on-device artificial intelligence, which allows applications to recognize objects, generate contextually appropriate target-language output, and respond to learner queries in natural language. The combination of these two affordances has produced a new class of vocabulary-learning applications that label objects in the learner's environment, generate spoken example sentences situated in the learner's actual surroundings, and adapt instruction to the spatial and temporal context of use.

This paper examines the impact of these AR-enhanced AI applications on vocabulary retention, focusing specifically on the mechanisms of spatial and environmental learning. The central claim is that situating target-language vocabulary in the learner's physical environment recruits cognitive systems for spatial memory, embodied perception, and contextual encoding that are largely dormant in conventional screen-based study. The paper proceeds as follows. Section 2 reviews the theoretical frameworks that motivate the investigation. Section 3 surveys the empirical literature on AR vocabulary learning, highlighting where AI integration has changed the picture. Section 4 analyzes the mechanisms by which spatial and environmental learning support retention. Section 5 considers limitations, equity concerns, and methodological challenges. Section 6 outlines directions for future research, and Section 7 concludes.

Theoretical Framework

1. Dual Coding Theory and Multimodal Encoding

Allan Paivio's dual coding theory holds that the mind processes verbal and nonverbal information through two functionally distinct but interconnected systems. When a learner encounters a new word accompanied by a relevant image, action, or sensory experience, both systems engage and form referential connections that strengthen retrieval pathways. AR vocabulary applications exploit this principle natively: when a learner points a device at a coffee cup and the word taza appears anchored to the object, the lexical item is encoded together with a rich visual referent in its naturalistic context, rather than alongside an abstract clip-art icon on a flashcard.

Dual coding has consistently predicted superior recall for words learned with congruent imagery. AR extends the principle by replacing static images with dynamic, learner-controlled, three-dimensional referents that may be viewed from multiple angles, manipulated, and revisited.

2. Embodied Cognition and the Sensorimotor Trace

Embodied cognition theory argues that conceptual representations are grounded in the sensorimotor systems that originally encountered them. Words for graspable objects, for example, activate motor regions associated with grasping. When learners encounter vocabulary while moving through a real environment, picking up objects, or performing actions, the resulting memory trace incorporates sensorimotor as well as auditory and visual components. Such richly embodied traces have been shown to be more resistant to forgetting than purely symbolic representations.

AR-AI applications uniquely afford embodied encoding because the learner is not merely looking at a screen but physically navigating and acting upon a labelled environment. A learner studying kitchen vocabulary while preparing a meal accumulates motor traces alongside lexical representations, producing a more deeply integrated memory.

3 The Encoding Specificity Principle and Context-Dependent Memory

Endel Tulving's encoding specificity principle holds that retrieval is most effective when the conditions of recall match the conditions of original encoding. The classic demonstrations involved divers learning word lists underwater and on land, with significantly better recall when the testing environment matched the learning environment. The implication for vocabulary learning is profound: words learned in the physical and situational context where they are likely to be used should be more readily retrieved in those same contexts.

Conventional vocabulary applications systematically violate this principle by training words in a context (a glowing rectangular screen) that bears no resemblance to the situations in which the words will be needed (a market, a kitchen, a workplace). AR applications can in principle restore encoding specificity by placing instruction in the very environments where the vocabulary is communicatively relevant.

4 Cognitive Load Theory

John Sweller's cognitive load theory distinguishes between intrinsic, extraneous, and germane cognitive load. Effective instructional design reduces extraneous load while supporting germane load that contributes to schema construction. AR has both promise and risk in this regard. On one hand, contextual labelling reduces the extraneous load of mentally constructing a referent from a picture or definition. On the other hand, poorly designed AR overlays can produce visual clutter, divided attention, and split-attention effects that overwhelm working memory.

AI integration is what allows contemporary AR vocabulary applications to manage cognitive load adaptively. By tracking learner gaze, response latency, and interaction history, an AI model can decide which words to surface, when to surface them, and at what level of elaboration, dynamically tuning the presentation to remain within productive load thresholds.

5 Spatial Memory and the Method of Loci

The use of physical space as a mnemonic substrate is among the oldest documented memory techniques. The classical method of loci, attributed to the poet Simonides, exploits the brain's substantial capacity for spatial memory by attaching items to be remembered to imagined locations along a familiar route. Modern neuroscience has confirmed that the hippocampus and surrounding structures encode rich spatial maps and that information bound to a spatial location is retrieved more reliably than information learned without spatial scaffolding.

AR vocabulary applications operationalize the method of loci automatically and at scale. Rather than requiring the learner to construct an imagined memory palace, the application uses the learner's actual surroundings as the substrate, anchoring each new word to a real, persistent location that the learner can return to. This is, arguably, the single most distinctive cognitive affordance of AR-based vocabulary learning.

Empirical Literature

1 Early AR Vocabulary Studies (2012–2019)

The first wave of AR vocabulary research relied on marker-based systems in which printed cards or QR codes triggered the display of target-language labels and audio. Studies in Taiwanese and Spanish primary classrooms reported moderate but consistent advantages for AR-augmented instruction over textbook-based controls on immediate post-tests, with effect sizes in the range commonly reported for instructional technology interventions. However, delayed post-tests at intervals of two weeks or longer produced more variable results, and few studies followed learners beyond a single academic term.

These early studies were limited in three important ways. First, the AR content was typically pre-authored, requiring teachers to design and print marker cards in advance, which constrained the range of vocabulary that could be covered. Second, the marker-based interaction broke the connection between the word and its naturalistic context. Third, no adaptive intelligence was present; all learners received the same content in the same sequence.

2 Markerless AR and the Mobile Inflection Point

The introduction of robust markerless tracking on consumer smartphones around 2017 to 2019 enabled a second generation of AR vocabulary applications that could anchor digital labels to arbitrary surfaces and objects without the need for printed markers. Studies of such applications reported larger advantages on delayed retention measures, suggesting that the freedom to apply AR to the learner’s actual environment, rather than to a pre-fabricated set of markers, mattered significantly.

It was during this period that researchers began to articulate the spatial-environmental hypothesis explicitly: that the retention advantages of AR were not driven primarily by novelty or motivation but by the recruitment of spatial memory and contextual encoding mechanisms. Several within-subjects designs that controlled for engagement and time-on-task continued to find AR advantages, lending credibility to the cognitive rather than affective interpretation.

3 The AI-Integrated Generation (2022–Present)

The integration of large language models and computer vision into mobile AR has produced a qualitative shift in what these applications can do. Contemporary systems can identify nearly any object in the learner's environment, generate a sentence in the target language that uses the corresponding word in a contextually appropriate way, voice that sentence with controllable accent and speed, accept the learner's spoken or typed response, and provide adaptive feedback. They can also remember which words a particular learner has encountered, where they encountered them, and how reliably they have recalled them, generating personalized review sessions that revisit words in their original spatial contexts.

Studies of AI-integrated AR vocabulary applications have begun to report substantially stronger retention effects than earlier AR work. While the literature is still maturing and replication remains an open priority, the convergent finding across multiple studies in different language pairings is that learners who study new vocabulary through AR-AI applications retain significantly more words at one-month and three-month follow-ups than matched learners using spaced-repetition flashcard applications. The most striking gains appear for concrete nouns and action verbs, which are most readily anchored to physical referents and embodied actions.

4 Comparative Studies and Meta-Analytic Findings

Meta-analytic syntheses of AR-based vocabulary instruction conducted before the AI-integrated generation reported small to moderate pooled effect sizes favoring AR, with substantial heterogeneity attributable to study design, learner age, and the specific vocabulary targeted. Updated syntheses incorporating AI-integrated studies suggest that pooled effects are growing larger and more consistent, although publication bias remains a concern. Importantly, the studies reporting the largest effects share common design features: they target concrete vocabulary, they situate learning in genuinely meaningful environments, they incorporate adaptive review, and they extend over multiple sessions rather than relying on a single exposure.

Mechanisms of Spatial and Environmental Learning

1 Place-Cell Recruitment and Hippocampal Encoding

Research in cognitive neuroscience has demonstrated that the hippocampus and entorhinal cortex contain specialized cells that encode spatial position and orientation, supporting the formation of cognitive maps. When information is bound to a specific location, it is encoded with the participation of these spatial systems, producing a richer and more retrievable memory trace than information encoded purely verbally. AR vocabulary applications exploit this phenomenon by binding each lexical item to a physical location in the learner's environment, effectively recruiting hippocampal spatial encoding for what would otherwise be a purely linguistic memory task.

2 Multimodal Convergence and Cross-Modal Reinforcement

Memory traces formed through multiple converging modalities are more resistant to forgetting than those formed through a single modality. AR-AI vocabulary learning typically engages vision, audition, motor activity, and proprioception simultaneously, producing memory traces with multiple independent retrieval cues. Forgetting any single cue does not eliminate access to the memory, because alternative cues remain available. This redundancy is a likely mechanism behind the durability of vocabulary learned in AR contexts.

3 Generative Effort and Contextual Inference

AR-AI applications frequently require the learner to do more cognitive work than passive flashcard review. The learner must locate the relevant object in the environment, attend to its properties, integrate the new label with prior knowledge, and often produce a spoken response. This generative effort is well documented to enhance retention, an effect sometimes called the generation effect. Moreover, when the AI generates an example sentence using the new word in its current context, the learner must perform contextual inference to disambiguate the meaning, which further deepens encoding.

4 Personalization and Desirable Difficulty

Robert Bjork's framework of desirable difficulty holds that learning conditions which feel harder in the moment often produce more durable learning than easier conditions. AI-driven adaptation in AR applications can calibrate the difficulty of each interaction to the individual learner, increasing the spacing of reviews, varying the contexts of presentation, and introducing productive interference. Such calibration is impractical to deliver through human instruction at scale and largely impossible through static curriculum materials.

5 Affective and Motivational Mediators

Although the strongest case for AR-AI vocabulary learning rests on cognitive mechanisms, affective factors are not negligible. Learners frequently report higher engagement, lower anxiety, and stronger sense of agency when using AR applications compared with conventional study tools. These affective benefits likely contribute to retention by increasing time-on-task and willingness to undertake productive effort. The methodological challenge is to disentangle these motivational effects from the underlying cognitive ones, which is an active area of research.

Limitations, Equity, and Methodological Challenges

1 Hardware and Access Inequities

AR-AI vocabulary applications presuppose ownership of relatively current mobile devices, reliable network access for the most computationally demanding features, and in some cases head-mounted displays that remain costly. The learners who would benefit most from improved language instruction are not always those with ready access to such hardware. Equity-focused research is needed to determine the extent to which lighter-weight implementations can deliver comparable benefits, and to develop institutional models that distribute access fairly.

2 Privacy and Data Considerations

AR-AI applications typically capture continuous video of the learner's environment and transmit it, in whole or in part, to remote servers for processing. The privacy implications are nontrivial, particularly when learners use the applications in private homes, workplaces, or in the presence of third parties who have not consented to recording. Edge processing and on-device language models partially mitigate these concerns but introduce trade-offs in capability. Educational deployments must take privacy obligations seriously rather than treating them as a regulatory afterthought.

3 Methodological Challenges in the Research Literature

Research on AR-AI vocabulary learning faces several persistent methodological difficulties. The intervention is technologically complex, making it hard to isolate which components are driving observed effects. Control conditions are not always well matched on time-on-task, novelty, or motivational appeal. Outcome measures vary widely across studies, ranging from immediate recognition to delayed production, and longitudinal designs that track retention over months or years remain rare. Standardized reporting practices and pre-registration would strengthen the cumulative evidence base.

4 The Risk of Superficial Engagement

Not all AR-AI applications are pedagogically well designed. Some prioritize novelty and spectacle over meaningful interaction, producing what might be called instructional candy: enjoyable in the moment but contributing little to durable learning. The presence of AR and AI does not guarantee educational value any more than the presence of a textbook does. Principled design grounded in learning science remains the precondition for the technology's promise.

Directions for Future Research

Several research directions deserve sustained attention as AR-AI vocabulary learning matures. Longitudinal studies extending over six months or more would help determine whether early retention advantages are sustained or whether they erode under naturalistic conditions of use and disuse. Neuroimaging studies using portable electroencephalography or functional near-infrared spectroscopy could test directly whether AR vocabulary learning recruits hippocampal and sensorimotor systems to a greater extent than conventional study, providing mechanistic confirmation for the spatial-environmental hypothesis.

Cross-linguistic studies are needed to establish whether the retention benefits generalize beyond the typologically European languages that have dominated the literature so far. Languages with rich morphology, non-Roman scripts, or large phonological inventories may benefit from AR-AI scaffolding in distinctive ways. Studies of older adult learners, learners with specific learning differences, and learners in low-resource settings would also broaden the evidence base.

As wearable AR hardware matures, research will need to address what might be called the fully ambient learning environment: a setting in which the learner is continuously exposed to target-language labels and prompts whenever AR glasses are worn. Such environments raise both pedagogical questions, such as whether continuous exposure produces saturation effects that diminish learning, and ethical questions, such as how learners can maintain meaningful control over their attention. The relationship between AR-AI vocabulary instruction and the development of broader communicative competence, including grammar, pragmatics, and sociocultural knowledge, also remains underexplored.

Conclusion

AR-enhanced AI applications represent more than an incremental improvement to mobile-assisted language learning. By binding target-language vocabulary to the learner's physical environment, these systems recruit spatial memory, embodied cognition, and contextual encoding mechanisms that conventional flashcard and screen-based applications largely leave dormant. The convergence of theory, neuroscience, and a growing empirical literature supports the conclusion that vocabulary learned through AR-AI applications is retained more durably than vocabulary learned through traditional means, particularly for concrete vocabulary anchored to manipulable referents.

This conclusion should not be mistaken for technological enthusiasm. The retention benefits are not automatic; they depend on instructional design grounded in cognitive science, on equitable access to the underlying hardware, on responsible handling of the privacy implications of continuous environmental capture, and on the willingness of researchers to undertake the longitudinal, cross-linguistic, and mechanistic studies that the field still needs. The most important contribution of this paper is to frame AR-AI vocabulary learning not as a novelty but as a principled extension of well-established theories of memory and learning, and to argue that its eventual impact on language education will be determined by how seriously its designers and researchers take those theories.

If the cognitive case advanced here is correct, the next decade of language learning research will see vocabulary instruction become progressively more integrated into the lived environments of learners, with AI as the orchestrating intelligence and AR as the perceptual interface. The screen-bound flashcard, dominant for half a century, may come to seem as impoverished a tool as the printed word list it once replaced.

References

1. Bjork, R. A., & Bjork, E. L. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), Psychology and the real world: Essays illustrating fundamental contributions to society (pp. 56–64). Worth Publishers.

2. Barsalou, L. W. (2008). Grounded cognition. Annual Review of Psychology, 59, 617–645.
3. Cuendet, S., Bonnard, Q., Do-Lenh, S., & Dillenbourg, P. (2013). Designing augmented reality for the classroom. Computers & Education, 68, 557–569.
4. Dunleavy, M., & Dede, C. (2014). Augmented reality teaching and learning. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), Handbook of research on educational communications and technology (pp. 735–745). Springer.
5. 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.
6. Hwang, G. J., & Wu, P. H. (2014). Applications, impacts and trends of mobile technology-enhanced learning: A review of 2008–2012 publications in selected journals. International Journal of Mobile Learning and Organisation, 8(2), 83–95.
7. Maguire, E. A., Valentine, E. R., Wilding, J. M., & Kapur, N. (2003). Routes to remembering: The brains behind superior memory. Nature Neuroscience, 6(1), 90–95.
8. Mayer, R. E. (2009). Multimedia learning (2nd ed.). Cambridge University Press.
9. Nation, I. S. P. (2013). Learning vocabulary in another language (2nd ed.). Cambridge University Press.
10. O'Keefe, J., & Nadel, L. (1978). The hippocampus as a cognitive map. Oxford University Press.
11. Paivio, A. (1986). Mental representations: A dual coding approach. Oxford University Press.
12. Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. Perspectives on Psychological Science, 1(3), 181–210.
13. Santos, M. E. C., Chen, A., Taketomi, T., Yamamoto, G., Miyazaki, J., & Kato, H. (2014). Augmented reality learning experiences: Survey of prototype design and evaluation. IEEE Transactions on Learning Technologies, 7(1), 38–56.
14. Schmitt, N. (2008). Instructed second language vocabulary learning. Language Teaching Research, 12(3), 329–363.
15. Sweller, J., Ayres, P., & Kalyuga, S. (2011). Cognitive load theory. Springer.
16. Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. Psychological Review, 80(5), 352–373.
17. Wu, H. K., Lee, S. W. Y., Chang, H. Y., & Liang, J. C. (2013). Current status, opportunities and challenges of augmented reality in education. Computers & Education, 62, 41–49.