Proceedings of ICEIPI 2025 Symposium: Reimagining Society: AI's Role in Cultural Transformation and Learning Environments
DOI: 10.54254/2753-7048/2025.BO25065

The Implementation of Multimodal Input in French Learning Platforms and Its Impact on Vocabulary Acquisition: A Case Study of Duolingo

Ziqing Wang

School of Foreign Languages, East China Normal University, Shanghai, China
florencewang1109@163.com

Abstract. The advancement of digital technology has shifted language learning from text-based, unimodal models to multimodal approaches integrating diverse presentation methods. Despite exposure to multimodal resources, French learners still struggle with pronunciation and vocabulary retention. This study uses Duolingo as a case study to explore the impact of multimodal input on vocabulary acquisition, grounded in Cognitive Load and Dual Coding Theories. Findings show that multimodal input enhances vocabulary acquisition through cognitive reinforcement, memory consolidation, and application transfer. Limitations include a single-case design and insufficient discussion on cultural adaptability. Future research should explore cross-platform comparisons and leverage AIGC technologies for personalized solutions. This study supports multimodal theory in French acquisition and offers insights for platform design.

Keywords: Multimodal input, French, Vocabulary acquisition, Duolingo platform

1. Introduction

The digital shift in language education has moved from text-based methods to multimodal integration. For French learners, challenges such as phonetic nuances and grammatical gender complicate vocabulary acquisition. Despite increasing multimodal platforms, empirical evidence on their synergistic effects remains limited.

This study examines (a) how multimodal input is implemented in French learning platforms and (b) its impact on vocabulary acquisition. Using a systematic literature review of multimodal pedagogy and French Second Language Acquisition, combined with a case study of Duolingo's design and outcomes.

This study expands multimodal theory's applicability to French acquisition while addressing the underexplored vocabulary-input nexus, which represents a critical research gap concerning the relationship between vocabulary acquisition and multimodal input, offering insights for platform design and pedagogical strategies for lexical development.

DOI: 10.54254/2753-7048/2025.BO25065

2. Theoretical underpinnings of multimodal input

Multimodal input in language learning integrates text, audio, visuals, and interactive elements, simultaneously engaging multiple sensory systems. This cross-modal cognitive reinforcement, achieved by concurrently activating visual, auditory, and (where applicable) tactile pathways, thereby enhances information encoding and recall efficiency [1].

Cognitive Load Theory posits that multimodal input optimizes cognitive architecture by leveraging the modality effect to offload extraneous cognitive load. As empirically demonstrated in French vocabulary pedagogy, the synchronized presentation of orthographic representation ("la pomme"), visual referent (photographic depiction of an apple) and auditory input (native-speaker pronunciation) yields significantly greater information processing efficiency compared to unimodal (text-only) delivery [2].

Dual Coding Theory mechanistically explains these benefits through parallel verbal (text/audio) and nonverbal (image) processing streams [3]. Empirical evidence demonstrates that this coordinated multimodal input approach effectively addresses key pedagogical challenges in French language acquisition. These findings establish an evidence-based framework for developing digital language learning platforms.

3. Implementation pathways of multimodal input in Duolingo

3.1. Platform overview

Multimodal input in language learning is an instructional approach integrating multiple sensory channels for information presentation, including but not limited to textual (written form), auditory (spoken pronunciation), visual (static images or dynamic videos), and interactive elements (e.g., touch feedback, gamification mechanics). In digital language learning environments, MMI enhances cognitive processing by simultaneously engaging learners' visual, auditory, and tactile modalities, thereby improving information absorption and memory retention.

Duolingo utilizes a gamified, micro-learning model with 5-7 minute lessons, spaced repetition, and a motivational framework (e.g., knowledge maps, streak counters, leaderboards). Machine learning further personalizes content through placement testing and error analysis to optimize learning outcomes.

3.2. Exemplary functional cases of multimodal input in Duolingo

The platform deeply integrates linguistic and non-linguistic symbols through various media formats including text, images, and audio, exhibiting characteristic multimodal features that enhance vocabulary learning with greater intuitiveness and concreteness. This multimodal approach effectively stimulates learners' interest and more effectively meets the needs of contemporary foreign language learners [4]. Dual Coding Theory demonstrates that when text and images are simultaneously processed by the human brain, both the linguistic and visual channels are activated to process the information, thereby facilitating more efficient information storage and meaning construction [5]. Furthermore, learners can utilize the flexibility of mobile platforms to memorize vocabulary anytime and anywhere. This method effectively promotes the conversion of short-term memory into long-term retention.

DOI: 10.54254/2753-7048/2025.BO25065

3.2.1. The cognitive enhancement model of “voice input + visual identification + cultural context”

Duolingo integrates "image + word + voice" to enhance French vocabulary acquisition by coordinating auditory, textual, and visual stimuli, grounded in cognitive science. The system forms strong associations between vocabulary and its auditory/visual representations. Illustrations by professional artists for concrete nouns accurately convey core semantic features, complemented by audio playback to reinforce auditory input.

As shown in Figure 1, the system presents the target word “un bébé” (a baby) with standard Parisian French pronunciation, allowing users to replay the audio to reinforce phoneme-lexical connections. The interface also displays the text “un bébé” alongside four image options: the correct depiction (a blond Caucasian infant) and three distractors (a purple cat, the numeral "1," and a piece of cheese). These distractors represent grammatical gender (masculine nouns), phonological interference (homophony with "un"), and cultural stereotype (French cheese). This design applies Dual Coding Theory by requiring learners to verify connections between phonology, orthography, and imagery.

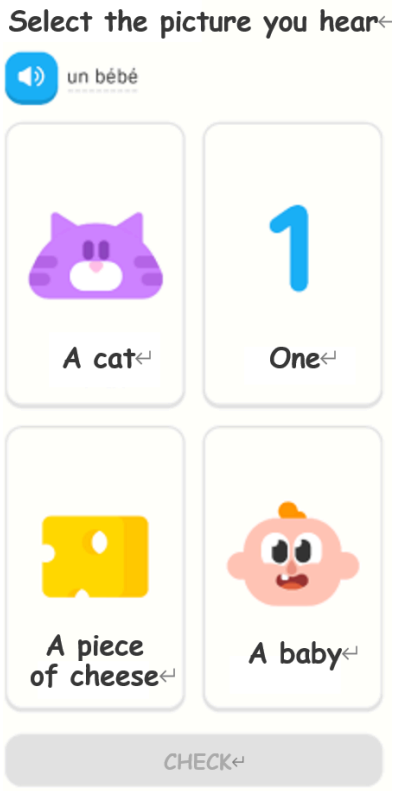

Figure 1. Duolingo French learning interface 1

3.2.2. The cognitive reinforcement model of “visual context + lexical selection”

The Duolingo platform uses a “visual context + lexical selection” strategy to enhance vocabulary acquisition through three phases: visual priming, contextual analysis, and discriminative selection. As shown in Figure 2, the interface presents a relevant image (a febrile patient in bed), where diagnostic markers (thermometer, bedclothes, nightwear) trigger illness-related semantic networks, forming a cognitive foundation for lexical processing.

DOI: 10.54254/2753-7048/2025.BO25065

Participants complete the French sentence “Je me sens trop mal pour sortir, je vais rester au_____ (I feel too unwell to go out, I’m going to stay in the_____)” through forced-choice selection among four lexico-semantic alternatives: jardin (garden), lit (bed), toilette (bathroom) and travail (work). This instructional design utilizes distractor contrast and contextual matching to optimize learning. It reinforces the expression “rester au lit” through multimodal consolidation and dual visual-contextual encoding, promoting deeper cognitive processing. The tripartite sequence—multimodal presentation, active retrieval, and immediate corrective feedback—aligns more effectively with human learning mechanisms than traditional memorization, as supported by psycholinguistic research [6].

Figure 2. Duolingo French learning interface 2

3.2.3. Cognitive reinforcement through “auditory input + lexical reconstruction”

This Duolingo module utilizes an “auditory input + lexical reconstruction” model, strengthening phoneme-lexeme connections through three key processes: auditory decoding, semantic comprehension, and active production. As shown in Figure 3, learners first hear the sentence “Un chat mange ton fromage” and then reconstruct it by selecting the correct lexical items from a word bank (aime, bébé, chat, deux, fromage, mange, ton, un, Un).

This process simulates language acquisition by: First, auditory processing builds bottom-up decoding skills, improving phonemic and prosodic perception. Next, lexical retrieval connects auditory and visual forms, enhancing sound-form mapping for word access. Finally, syntactic encoding advances learners to sentence-level production through drag-and-drop reconstruction. This scaffolded approach mirrors natural acquisition, with intentional sequencing to maximize efficiency.

DOI: 10.54254/2753-7048/2025.BO25065

Figure 3. Duolingo French learning interface 3

4. The impact of multimodal input on learners' vocabulary acquisition effectiveness

Multimodal input influences vocabulary acquisition through three interconnected dimensions: cognitive reinforcement, memory consolidation, and application transfer, forming a robust language acquisition mechanism.

At the cognitive level, multimodal input creates a three-dimensional lexical network through multisensory stimulation. Simultaneous phonological, orthographic, and visual information co-activates both linguistic and non-linguistic brain systems. Grounded in Dual Coding Theory, this approach effectively solidifies abstract grammatical features and phonological phenomena in French.

At the memory consolidation level, multimodal input promotes deep semantic processing through strategically designed distractors and contextualized practice. Duolingo employs three distractor types - (1) grammatical gender interference, (2) phonological interference, and (3) cultural interference - to enhance lexical discrimination, strengthening synaptic connectivity through pattern recognition and error-based learning. The integration of cultural context transforms lexical storage into a richly connected network, improving long-term vocabulary retention through situational schemata and episodic associations.

At the application transfer level, multimodal input forms a complete "input-processing-output" loop, simulating authentic language pathways from auditory decoding to lexical reconstruction and sentence completion [7]. Notably, real-time pronunciation feedback allows learners to calibrate articulatory patterns, developing accurate phonological representations. This output-focused design enhances lexical retrieval speed and facilitates the application of vocabulary in communication.

DOI: 10.54254/2753-7048/2025.BO25065

5. Conclusion

This study examines the impact of multimodal input design on Duolingo, focusing on auditory, visual, and textual modalities in lexical acquisition. Results show that multimodal input enhances learning through cognitive reinforcement, memory consolidation, and application transfer. Multisensory stimulation creates a multidimensional cognitive network, while distractors and contextual exercises promote deep semantic processing and long-term retention. The closed-loop model further fosters language automaticity.

However, this study has several limitations. It focused solely on the Duolingo platform, limiting case coverage and the generalizability of the findings. The use of qualitative methods also restricted the availability of experimental data and in-depth exploration of cultural factors, especially regarding non-Western learners. Future research should: (1) conduct cross-platform studies to analyze the effects of different multimodal designs across platforms (e.g., Quizlet), and (2) investigate cultural adaptation by exploring localized multimodal enhancements for non-European/American learners.

Multimodal language learning is rapidly evolving. Future research should investigate how AI-generated content (AIGC) can personalize input modalities to improve the efficiency and inclusivity of digital language education, leading to more adaptive systems that respond to diverse learner needs.

References

[1] Pellicer-Sánchez, A., Tragant, E., Conklin, K., Rodgers, M., Serrano, R., & Llanes, Á. (2020). Young learners' processing of multimodal input and its impact on reading comprehension: An eye-tracking study. Studies in Second Language Acquisition, 42(3), 577-598.

[2] Sweller, John., Ayres, Paul. author, Kalyuga, Slava. author, & SpringerLink. (2011). Cognitive Load Theory (1.), 129-140

[3] Paivio, A. (1990). Mental Representations: A Dual Coding Approach (1st ed., Vol. 9). New York: Oxford University Press, 53-83

[4] Liu, S. (2022). An exploration of the French vocabulary teaching model for second foreign languages based on the Quizlet platform. Linqu Teaching, (6), 78-81.

[5] Li, S., & Gao, Y. (2016). An empirical study on the effectiveness of mobile technology-assisted foreign language teaching in English vocabulary acquisition. Foreign Language World, (4), 73-81.

[6] Geng, Q. (2021). The integration of new media "Duolingo" with university foreign language listening and speaking teaching tests. Consumer Electronics, (10), 90-91, 87.

[7] Zhuang, N. (2023). A multimodal analysis of the 2019 edition of the Foreign Language Teaching and Research Press's high school English textbooks (Master's thesis). Southwest University.