

Strategies and Challenges in Using Multimedia to Facilitate Language Acquisition in English Teaching

Juan Wu¹, Yingying Du², Xiaohui Zhang³, Duo Zhang^{4*}

¹School of Tourism and Health Sciences, University of Sanya, Sanya 572000, Hainan, China

²School of National Security, Jilin University, Changchun 130012, Jilin, China

³Hainan College of Foreign Studies, Wenchang 571300, Hainan, China

⁴School of Humanities and Social Science, Hainan Tropical Ocean University, Sanya 572000, Hainan, China

Corresponding author: Duo Zhang (e-mail: xiaohui202604@126.com)

ABSTRACT In non-native English classrooms, multimedia resources often produce a cognitive gap of high input but low output: visual contexts and target-language forms fail to form stable semantic connections. To address this problem, this study proposes a three-stage multimedia teaching strategy of context anchoring, modal verification, and output transfer. First, a micro-context video library is constructed using word-vector semantic association, with target vocabulary anchoring the semantic field while 12-second clips control cognitive load. Second, a semi-structured virtual dialogue agent dynamically triggers recast and clarification feedback according to learners' spoken output, grammar breaks, and pragmatic deviations. Third, digital narrative reconstruction tasks require learners to convert video-based semantic information into written storyboards, promoting cross-modal syntactic transfer. Experiments show that the semantic priming effect size increases to 145.2 ms, speech-flow break density decreases to 12.3, and average clause nesting depth rises to 1.89. The framework turns multimedia into a cognitive scaffold and is compatible with networked multimedia transmission, wireless interactive classrooms, and electromagnetic-safe digital learning spaces.

INDEX TERMS multimedia language acquisition, virtual dialogue agent, syntactic transfer, wireless classroom, multimodal interaction

I. INTRODUCTION

IN the teaching practice of English as a second language or foreign language, the widespread application of multimedia technology has shifted the classroom from a single paper text to an era of audiovisual composite input[1]. However, the integration of multimedia resources in current non-native English classrooms generally presents an imbalance of "high sensory stimulation and low cognitive involvement": although students frequently come into contact with English film clips, animations and audio materials, their attention is often overwhelmed by rapidly changing images and background sound effects, the auditory language channel is suppressed, resulting in a fragile and superficial connection between the visual context and the target language form[2]. This cognitive gap of "redundant audiovisual input and barren language output" makes it difficult for multimedia stimulation to cross the threshold of comprehension input and transform into real communicative competence, seriously restricting the deep promotion potential of multimedia technology for language acquisition[3]. Therefore, how to reconstruct multimedia from a passive background display tool into a cognitive

medium that drives learners to conduct in-depth semantic processing and syntactic construction has become a core issue that needs to be addressed in current English teaching research.

Researchers at home and abroad have explored the interaction between multimedia and second language acquisition from multiple dimensions. Fievez et al. found through empirical research on annotated subtitled film and television materials that multimedia environments can promote learners' incidental vocabulary acquisition to a certain extent, but its effect is highly dependent on the input presentation method and learners' attention allocation pattern[4]; Puimège et al. further pointed out with eye-tracking technology that although text-enhanced audiovisual input can improve the attentional salience of multi-word units, there are still obvious individual differences in the depth of learners' processing of language forms[5]. At the same time, Dang et al. analyzed the impact of academic lecture videos on vocabulary acquisition from the perspective of input modality and vocabulary frequency, showing that different modal combinations have a modal effect on vocabulary retrieval efficiency[6]; Choi, starting from visual salience,

revealed that the allocation mechanism of visual attention resources in subtitled videos will directly affect the processing effect of language forms[7]. In addition, with the development of artificial intelligence technology, Xiao et al. began to explore the application potential of conversational agents in language learning, pointing out that dialogue systems can support language practice to a certain extent, but their interaction depth and cognitive scaffold function still need to be further expanded[8]. However, the above studies mostly remain at the level of static presentation of multimedia materials, or limit the interaction to simple interface operations, lacking a dynamic examination of the key mechanism of “how multimedia acts as a real-time cognitive scaffold to participate in meaning negotiation”, especially neglecting the design of structured transfer paths from input to output for learners in multimedia contexts.

To address the above limitations, researchers have gradually realized the need to introduce more cognitively profound technical means and task frameworks. Zhang et al.’s meta-analysis shows that language learning based on chatbots has a moderate effect on improving learner participation and language performance, but its effect is highly dependent on the design of task structure and feedback mechanism [9]; Lim et al. pointed out through a systematic review that digital narrative, as a cross-modal learning method, can effectively integrate language input and creative output, showing significant potential in improving language expression ability and learning motivation [10]. Furthermore, Tecedor’s empirical study found that digital narrative not only promotes the development of language skills, but also significantly improves learners’ self-efficacy and learning attitude, thus acting on the language acquisition process at both the emotional and cognitive levels [11]. However, such methods either focus on the static adaptation of front-end materials or only provide mechanical, one-way prompts, failing to establish a closed-loop linkage mechanism between multimedia contexts, learners’ real-time output, and dynamic system feedback. This paper proposes a three-stage multimedia integrated teaching strategy: “context anchoring—modal verification—output transfer.” First, a micro-context video library is constructed using word vector semantic associations to achieve precise activation of the semantic field. Then, a semi-structured virtual dialogue agent is introduced to dynamically intervene in syntactic breaks and pragmatic deviations in spoken output through multimodal feedback. Finally, a digital narrative reconstruction task necessitates cross-modal translation and internalization transfer, aiming to compensate for the dual deficiencies of “lack of interactive scaffolding” and “ambiguous transfer paths” in existing multimedia teaching research.

The main innovation of this paper can be summarized into the following three aspects: First, this paper establishes a relatively complete cognitive processing chain from semantic activation and interactive scaffolding to output transfer in multimedia contexts, which offers a new perspective

to explain the intrinsic mechanism by which multimedia benefits language acquisition; Second, this paper puts forward a concrete teaching implementation framework with natural language processing, virtual dialogue agent, and cross-modal task design, which has strong practicability in the classroom; Third, this paper validates the efficacy of the strategy with multi-dimensional language output indicators and finds out attention interference factors and corresponding coping ways in multimedia contexts.

II. CONSTRUCTION OF A SEMANTIC FIELD-ADAPTIVE MICRO-CONTEXT MATERIAL LIBRARY

The construction of the micro-context material library is founded on the word vector semantic association technology. The learners can obtain short audiovisual clips with a high match degree of semantic network of the target vocabulary, and enhance the accurate connection relationship between visual context and language form in the input process as follows.

First, based on the target vocabulary list set for the teaching unit, original dialogue scene clips containing the above vocabulary are extracted from a large-scale film and television corpus. The extraction process uses an automatic extraction script based on subtitle timeline alignment to ensure that each clip remains relatively complete at the semantic boundaries, with the initial extraction duration controlled between 20 and 40 seconds.

Second, the original clips undergo semantic purity screening and fine-tuning. A pre-trained GloVe word vector model is used to calculate the cosine similarity between the dialogue text within the clip and the target vocabulary. Only clips with a similarity higher than 0.65 are retained for the next processing stage. Subsequently, combined with subtitle timestamps and scene transition detection algorithms, redundant shots unrelated to the target semantics are removed, and the clips are uniformly cut into “micro-context” video units with a duration of 10 to 15 seconds. Each unit satisfies the following three constraints: First, the main subject of the image directly corresponds to the action, object, or relationship referred to by the target word; second, the sentence containing the target word is clearly identifiable; third, the segment conveys a “semantic kernel”—a complete, self-contained communicative act (e.g., a greeting, a specific request, or a localized conflict). While individual clips are brief to minimize cognitive load, they are curated as “narrative building blocks” belonging to the same thematic corpus. This allows learners to synthesize multiple 12-second “kernels” into a coherent chain during the subsequent digital narrative reconstruction task, bridging the gap between isolated stimulus and continuous storytelling[12]. Finally, the materials are graded, labeled, and stored according to the learner’s stage of semantic network development. By comparing the distance between the learner’s mastered vocabulary and the target vocabulary in the word vector space, the micro-contextual videos are divided into three levels: “high contextual support,” “medium contextual support,”

and “low contextual support,” corresponding to contextual types with semantic distances less than 0.4, between 0.4 and 0.7, and greater than 0.7, respectively. During teaching, the system prioritizes pushing high contextual support segments to establish initial anchoring. After the learner’s semantic activation threshold increases, it gradually transitions to low contextual support segments, thereby achieving gradual expansion and deep activation of the semantic field while maintaining an optimized cognitive balance. Although the three-stage framework introduces interactive and transmodal tasks, the use of filtered 12-second micro-contexts serves as a “cognitive scaffold” that prevents the sensory overload typically associated with raw, unedited multimedia. This strategic sequencing ensures that the increased cognitive engagement is channeled into productive language processing rather than being wasted on irrelevant visual noise. The core parameters of the above screening and grading process are summarized in Table 1.

TABLE 1. Micro-context Video Segment Selection and Grading Parameter Standards

Parameter	Value/Criterion	Description
Initial segment duration	20 – 40 s	Auto-captured based on subtitle timestamps
Semantic similarity threshold	≥ 0.65	Minimum cosine similarity using GloVe embeddings
Maximum final clip duration	≤ 12 s	Controls cognitive load
Image-target word correspondence	1	Verified by manual inspection
Audio clarity (SNR)	≥ 25 dB	Ensures target utterance intelligibility
Contextual independence	Pass rate $\geq 95\%$	Plot comprehensible without prior context
High contextual support tier	Semantic distance < 0.4	Delivered at initial anchoring stage
Medium contextual support tier	Semantic distance $0.4 - 0.7$	Delivered at transitional expansion stage
Low contextual support tier	Semantic distance > 0.7	Delivered at deep activation stage

III. SEMI-STRUCTURED INTERACTIVE FEEDBACK INTERVENTION IN VIRTUAL DIALOGUE AGENTS

The introduction of virtual dialogue agents for transforming multimedia scenarios from one-way input media into two-way interactive communication with cognitive scaffolding in real-time feedback. The agent is not conducting an open dialogue with the learner; however, it uses a hybrid interactive logic that combines predefined pedagogical templates with a Large Language Model (LLM) backend for real-time response generation.

The core of the agent system consists of three modules: a scenario anchoring module, an oral parsing module, and a feedback decision-making module. The “semi-structured” nature refers to the system’s constraints: while the pedagogical goals (recasting and clarification) are structured by decision trees, the linguistic realization of feedback is dynamically generated by the LLM based on the learner’s specific lexical errors. This ensures that the agent provides pedagogically targeted yet naturalistic conversational intervention rather than static, repetitive prompts. Its overall interactive process and feedback decision-making mechanism are shown in Figure 1.

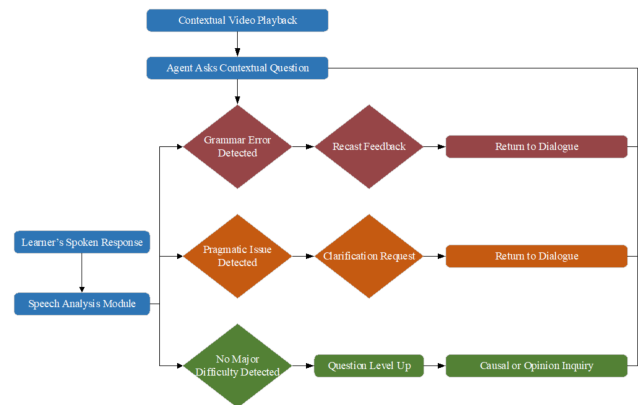


FIGURE 1. Interactive Feedback Decision-Making Mechanism of a Semi-structured Virtual Dialogue Agent

Show scenario anchoring module displays the micro-scenario video clip to the learner at the beginning of each round of dialogue. After playback of the video clip, the module asks a guiding question in the voice of a virtual character based on the scenario, e.g., “What do you think that character will do next?” or “What can you describe what just happened in your own words?”. Design of the question adopts the principle of “based on context but semantically open-ended questioning”, i.e., the context is confined in the just-presented visual semantic field and there is still enough space for learner’s expression.

The spoken language parsing module transcribes and analyzes learners’ real-time speech input, identifying two key output features: one is grammatical breaks, mainly referring to silent pauses of more than 0.5 seconds within a sentence, incomplete clauses, or incomplete subject-verb structures; the other is pragmatic deviations, encompassing unclear referents, conflicts between tense and context time-lines, and avoidance substitution of target vocabulary. This module does not pursue complete syntactic tree parsing, but rather quickly locates output weaknesses based on shallow dependency syntax rules and keyword matching [13]. When the feedback decision module detects positive feedback such as grammatical break, it triggers recasting feedback to restate the learner’s intended expression with the correct grammatical form, and adds a follow-up question that is semantically related to the expression to keep the conversation ongoing. For example, when the learner output is He... go to the office, the agent triggers feedback and says “You mean he went to the office? What did he do there?”. When pragmatic breaks are detected, the agent triggers clarification request and asks learners in a natural tone to reflect on the pragmatic deviation of their expression in relation to the context, e.g. “Did that happen before or after the meeting we just saw?”. If learners do not show obvious production difficulties in two consecutive rounds of dialogue, the agent will proactively raise the cognitive level of question, i.e. from factual descriptions to causal inferences or stance expressions, to push syntactic complexity up to the level.

IV. DESIGN OF MANDATORY OUTPUT TASK FOR CROSS-MODAL DIGITAL NARRATIVE RECONSTRUCTION

The initial design of the digital narrative reconstruction task is to encourage learners to reorganize and express the semantic information they have acquired through written modality after being removed from dynamic audiovisual stimuli, thereby completing the internalization and transfer from comprehension input to output skills. The task is carried out in the form of group collaboration, requiring learners to adapt the micro-context video materials they have watched in the previous stage in a cross-modal manner, and the output form is a storyboard of a graphic novel.

The task implementation is divided into three progressive stages. The first stage is semantic extraction and key frame capture. The group first reviews the micro-context video clips and determines the core nodes of the plot through group discussion. Each node corresponds to an indispensable narrative turning point. Then, learners use the screenshot tool of the player to capture the static picture that best represents the semantic content of the node. To capture the essential progression of the 12-second micro-context without inducing visual redundancy, the number of screenshots for each clip is strategically limited to one to three representative frames. This selection necessitates that learners identify only the most critical “semantic anchors”—typically the initial state, the core action, and the resulting state—thereby promoting precise information filtering and cognitive synthesis [14].

The second stage is subtitle reconstruction and syntactic transformation. After the key frame sequence is acquired, learners are required to write down the text description for each key frame. According to the video watching rules, the written part should be in complete formal sentences and is forbidden to copy the original video subtitle or use colloquial phrases. That is, the causal, conditional and concessive relationships reflected by the image and dialogue in the video should be transformed into complex sentences in writing based on subordinating conjunctions; actors' expression information and camera language information about characters' behaviors and psychological states should be transcribed into text. The focus of this stage is to empower learners to use syntactic knowledge to complete linear semantic encoding in a situation without auditory information and dynamic visual field.

The third stage is narrative coherence verification and inter-group peer review. After finishing the first draft of storyboards, each group should exchange their draft with other groups and evaluate each other. The peer review criteria include:

- 1) Is it possible to comprehend the complete plot in the narration and static images;
- 2) Are there any jumps or missing references in the text connection between two adjacent key frames.

Each group revises their first draft according to feedback information. The revision focus includes two aspects:

- 1) repairing the information which is implied due to modality switching in the video into the written script;
- 2) adjusting the connection between sentences.

V. EVALUATION OF THE EFFICACY OF LANGUAGE COGNITIVE PROCESSING UNDER STRATEGIC INTERVENTION

A. OVERALL EXPERIMENTAL SETUP

This study used pre- and post-test precision experimental design. Through quantitatively detecting and analyzing three indicators: syntactic complexity (ratio of subordinate clauses to T-units), lexical semantic extraction accuracy (mean cosine similarity between produced words and target corpus), and communicative repair behavior (proportion of joint triggering frequencies of self-correction and clarification requests) in speakers' and writers' spoken and written corpora before and after multimedia intervention, the study evaluated the effectiveness of three-stage strategy on depth of language cognitive processing comprehensively.

B. IMPACT TEST ON THE DEPTH OF LEXICAL SEMANTIC FIELD ACTIVATION OF THE MICRO-CONTEXTUAL VIDEO LIBRARY

Sixty learners, randomly assigned to experimental and control groups, watched 15 micro-contextual video clips (12 seconds each) chosen by word vectors. To ensure rigorous experimental control and eliminate variables such as boredom or cognitive fatigue caused by repetitive or lengthy stimuli, the control group watched 15 conventional video segments of the same 12-second duration. These control segments contained the same target words but lacked the semantic-vector-based micro-contextual filtering used for the experimental group. This study was approved by the Ethics Committee of Sanya University, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form. The total time of viewing was kept constant at 180 seconds for both groups to ensure that the observed effects were due to the quality of the context rather than the quantity of input. Immediately after the watching, both groups were administered with cross-modal semantic priming task, in which participants judged quickly the semantic relevance of subsequent images based on the auditory target word. The difference of reaction time between relevant trials and irrelevant trials was extracted to calculate the semantic priming effect size and used to measure the depth of activation of the mental representation of a lexical level. As shown in Figure 2:

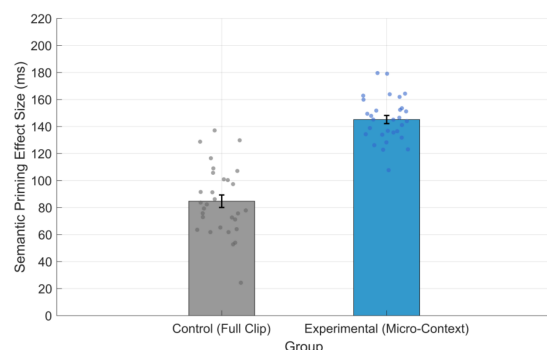


FIGURE 2. Assessment of the Impact of the Micro-Contextual Video Library on the Depth of Lexical Semantic Field Activation

The mean semantic priming effect size in the experimental group was 145.2 ms (SEM=3.0), significantly higher than the control group's 84.7 ms (SEM=4.6). This represents a 71% improvement. In cognitive terms, a larger priming effect (measured in milliseconds) indicates that the micro-contextual videos helped learners form a “fast-track” in their brains between the English word and its visual meaning. The 145.2ms result suggests that the experimental group could recognize and retrieve vocabulary meanings much more instinctively and with less mental effort than those watching standard full-length videos. Independent samples t-test results showed that the difference between groups reached a highly significant level ($t(58)=11.0$, $p<0.001$), with an effect size of Cohen's $d=2.8$, which falls within the large effect category. The above data indicate that the 12-second micro-contextual clips, filtered by word vector semantics, can effectively suppress the interference of redundant visual information on the language channel, promoting learners to accurately allocate cognitive resources to the semantic binding between target words and core contextual elements, thereby significantly enhancing the activation depth and retrieval efficiency of lexical mental representations.

C. INTERVENTION TEST EXPERIMENT ON FLUENCY AND REPAIR STRATEGIES OF SPOKEN LANGUAGE PRODUCTION BY VIRTUAL DIALOGUE AGENT

The experiment used single group pre- and post-test design, and randomly chose 40 learners with intermediate level of English as subjects. In the pre-test, the subjects were exposed to a pre-recorded character only played fixed video clips for 5 minutes of spoken dialogue. After a gap of 30 minutes, in the post-test, the subjects gave a dialogue on the same topic with a semi-structured virtual dialogue agent with clarification requests and refactoring feedback functions. Then the number of silent pauses and total number of syllables in the two tests were extracted with the speech analysis tool and the speech flow break density was calculated to measure the immediate intervention effect of feedback from virtual agent on spoken fluency. As shown in Figure 3:

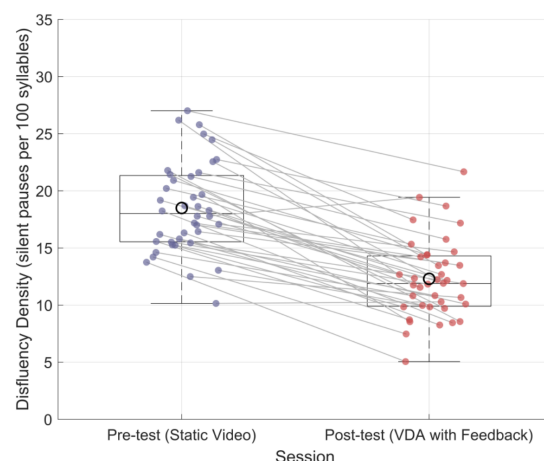


FIGURE 3. Intervention evaluation of virtual dialogue agent on spoken fluency and repair strategies

Figure 3: Intervention evaluation of virtual dialogue agent on spoken fluency and repair strategies. After intervention with the virtual dialogue agent, the speech flow break density of the subjects decreased from 18.5 (SD=3.97) in the pre-test to 12.3 (SD=3.45) in the post-test, with an average reduction of 6.2 percentage points. Paired-samples t-test results showed that the difference between the pre-test and post-test was highly significant ($t(39)=11.24$, $p<0.001$), with an effect size of Cohen's $d_z = 1.78$. The data indicate that the semi-structured virtual agent, through real-time clarification requests and reconstructive feedback mechanisms, can effectively shorten the cumulative duration of silent pauses for learners, reduce the cognitive load in the spoken output process, and thus significantly improve speech fluency and speech flow coherence under immediate interactive conditions.

D. A DELAYED-LASTING TEST OF THE EFFECT OF DIGITAL NARRATIVE RECONSTRUCTION TASK ON SYNTACTIC COMPLEXITY TRANSFER

This experiment extended the previous grouping method to select 60 intermediate-level English learners to form the experimental group and the control group, and there were 30 participants in each group. After watching the micro-contextual videos, the experimental group completed the cross-modal adaptation task in groups to use the video material to construct a storyboard of a graphic novel. The control group only watched the same video material but did not complete the reconstruction task. One week later, the two groups completed the propositional writing test simultaneously without watching video materials. We counted the number of subordinate clauses and T-units in the text and calculated the nesting depth index of subordinate clauses to measure the delayed-latitude preservation effect of digital narrative reconstruction on syntactic complexity transfer. The details are shown in Table 2:

TABLE 2. Delayed Preservation Assessment of Syntactic Complexity Transfer in Digital Narrative Reconstruction Task

Group	n	Mean (M)	Std. Deviation (SD)	Std. Error Mean (SEM)
Control	30	1.42	0.31	0.057
Experimental	30	1.89	0.28	0.051

Note: Independent samples t-test results: $t(58)=6.18$, $p<0.001$, Cohen's $d=1.59$.

Delayed writing test data showed that the mean nesting depth index of clauses in the experimental group was 1.89 ($SD=0.28$), while that in the control group was 1.42 ($SD=0.31$), with the experimental group being approximately 33% higher than the control group. Independent samples t-test results indicated a highly significant difference between groups ($t(58)=6.18$, $p<0.001$), with an effect size of Cohen's $d=1.59$, indicating a large effect. These results confirm that the digital narrative reconstruction task, by forcing learners to translate visual semantic information into linear written syntactic structures after being removed from the multimedia context, effectively promoted the internalization and consolidation of complex syntactic representations. Furthermore, this promoting effect remained stable one week after the task ended, demonstrating a lasting intervention advantage in the transfer of syntactic complexity.

VI. CONCLUSION

This paper reports on the proposal and demonstration of the promoting effect of a three-stage multimedia fusion strategy of “contextual anchoring—modal verification—output transfer” on English acquisition. The results revealed that semantic anchoring of micro-contextual video libraries greatly enhanced the activation of lexical mental representations, the immediate scaffolding of virtual dialogue agents greatly compressed the cognitive pauses of spoken output, and digital narrative reconstruction drove the cross-contextual transfer of syntactic complexity. These results suggested that fusing a presentation tool into a continuous field of cognitive processing and meaning negotiation could narrow the gap between input and output. This study also had the following limitations: a small sample size and lack of exploration of the long-term maintenance effect. Future research could integrate the eye-tracking and longitudinal tracking designs to further explore the attention allocation mechanism and delayed stability of strategic interventions in multimedia environments.

AUTHOR CONTRIBUTIONS

Conceptualization – Juan Wu, Yingying Du, Xiaohui Zhang and Duo Zhang; methodology – Juan Wu, Yingying Du, Xiaohui Zhang and Duo Zhang; investigation – Juan Wu, Yingying Du, Xiaohui Zhang and Duo Zhang; writing-original draft preparation – Juan Wu, Yingying Du, Xiaohui Zhang and Duo Zhang. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research was supported by,

1. Special Research Projects on Foreign Languages at the Hainan Provincial Research Base for Applied Foreign Languages: Research on the Industry-Education Integration Talent Cultivation Mechanism of “Multilingual + High-end Equipment Manufacturing”, Project Number: (HNWYJD25-02);
2. Sanya Social Science Fund Project: “A Study on Optimization Pathways for Multilingual Government Service Platforms in the Development of Sanya's Digital Government” Project No: SYSK 2025-09

HUMAN RESEARCH SUBJECTS

This study was approved by the Ethics Committee of Sanya University, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

- [1] R. Zhang and D. Zou, “A state-of-the-art review of the modes and effectiveness of multimedia input for second and foreign language learning,” *Computer Assisted Language Learning*, vol. 35, no. 9, pp. 2790-2816, 2022, doi: 10.1080/09588221.2021.1896555.
- [2] M. Montero Perez, “Second or foreign language learning through watching audio-visual input and the role of onscreen text,” *Language Teaching*, vol. 55, no. 2, pp. 163-192, 2022, doi: 10.1017/S0261444821000501.
- [3] M. F. Teng, “The effectiveness of multimedia input on vocabulary learning and retention,” *Innovation in Language Learning and Teaching*, vol. 17, no. 3, pp. 738-754, 2023, doi: 10.1080/17501229.2022.2131791.
- [4] I. Fievez, M. Montero Perez, F. Cornillie, and P. Desmet, “Promoting incidental vocabulary learning through watching a French Netflix series with glossed captions,” *Computer Assisted Language Learning*, vol. 36, no. 1-2, pp. 26-51, 2023, doi: 10.1080/09588221.2021.1899244.
- [5] E. Puimège, M. Montero Perez, and E. Peters, “Promoting L2 acquisition of multiword units through textually enhanced audiovisual input: An eye-tracking study,” *Second Language Research*, vol. 39, no. 2, pp. 471-492, 2023, doi: 10.1177/02676583211049741.
- [6] T. N. Y. Dang, C. Lu, and S. Webb, “Open access academic lectures as sources for incidental vocabulary learning: Examining the role of input mode, frequency, type of vocabulary, and elaboration,” *Applied Linguistics*, vol. 44, no. 4, pp. 747-770, 2023, doi: 10.1093/applin/amac044.
- [7] S. Choi, “Visual saliency in captioned digital videos and learning of English collocations: An eye-tracking study,” *Language Learning & Technology*, vol. 27, no. 1, pp. 1-21, 2023, doi: 10.64152/10125/73536.
- [8] F. Xiao, P. Zhao, H. Sha, D. Yang, and M. Warschauer, “Conversational agents in language learning,” *Journal of China Computer-Assisted Language Learning*, vol. 4, no. 2, pp. 300-325, 2024, doi: 10.1515/jccall-2022-0032.
- [9] S. Zhang, C. Shan, J. S. Y. Lee, S. Che, and J. H. Kim, “Effect of chatbot-assisted language learning: A meta-analysis,” *Education and Information Technologies*, vol. 28, no. 11, pp. 15223-15243, 2023, doi: 10.1007/s10639-023-11805-6.
- [10] N. Z. L. Lim, A. Zakaria, and V. Aryadoust, “A systematic review of digital storytelling in language learning in adolescents and adults,” *Education and Information Technologies*, vol. 27, no. 5, pp. 6125-6155, 2022, doi: 10.1007/s10639-021-10861-0.
- [11] M. Tecedor, “Digital storytelling: Changing learners' attitudes and self-efficacy beliefs,” *Applied Linguistics*, vol. 45, no. 1, pp. 65-87, 2024, doi: 10.1093/applin/amad002.

- [12] N. Y. Kim and D. Kim, "Digital storytelling: Facilitating learning and identity development," *Educational Technology Research and Development*, vol. 70, no. 5, pp. 2201-2223, 2022, doi: 10.1007/s11423-022-10145-2.
- [13] H. Yang, H. Kim, J. H. Lee, and D. Shin, "Implementation of an AI chatbot as an English conversation partner in EFL speaking classes," *ReCALL*, vol. 34, no. 3, pp. 327-343, 2022, doi: 10.1017/S0958344022000039.
- [14] A. F. Alserhan, T. E. I. Arar, Z. Abu-Laila, et al., "New Marketing Strategies for Promoting Language Teaching Courses: The Case of Spanish and English Languages," *Journal of Ecohumanism*, vol. 3, no. 8, 2025, doi: 10.62754/joe.v3i8.5715.