VerbVenture: MR-powered Embodied Learning Experience for EFL Children

ANONYMOUS AUTHOR(S)*

This study explores the potential of Mixed Reality (MR) to facilitate English verb-noun collocation acquisition in children learning English as a Foreign Language (EFL). Grounding the learning methodology in theories of embodied cognition, VerbVenture integrates artificial intelligence and immersive technologies to transform everyday environments into engaging portals for learning collocations. The system presents two learning phases: 1) Vocabulary Acquisition: Learners interact with real-world objects through MR, encountering the object's name and pronunciation. 2) Collocation Practice: Animations guide learners to use real objects and perform actions representing target verb collocations in context. This active manipulation strengthens mental representations and situates learning in the real world. A pilot user study suggests VerbVenture encourages interest and comprehension of collocation meanings in children. Moving forward, VerbVenture aims to evaluate the efficacy of embodied MR for collocation acquisition, examining memory persistence and long-term learning outcomes.

CCS Concepts: • **Human-centered computing → Empirical studies in interaction design**.

Additional Key Words and Phrases: Mixed Reality, Tangible Interaction, Collocation Learning, Language Learning, Embodied Learning

ACM Reference Format:

Anonymous Author(s). 2024. VerbVenture: MR-powered Embodied Learning Experience for EFL Children. In *CHI '24: CHI Conference on Human Factors in Computing Systems, May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Learning English as a Foreign Language (EFL) presents distinct hurdles for young learners, particularly when navigating the intricate web of collocations. A collocation refers to a linguistic pattern marked by the consistent association of certain words, wherein one word frequently occurs near another [22]. Hoey[6] provided a definition of collocation as the connection a lexical item shares with other items that display a probability greater than random within its textual context. These inseparable word pairings, essential for fluent and idiomatic expression, pose unique challenges due to their context-dependent nature and susceptibility to native language (L1) interference [1, 8, 17, 20]. For children whose L1 diverges significantly from English, like Mandarin speakers for example, this often leads to frequent errors in collocation usage, impacting both comprehension and expression.

Traditional EFL instructional methods, often text-heavy and lacking dynamic engagement, can exacerbate these difficulties, resulting in low motivation and hindered learning outcomes. Research in cognitive science and education increasingly highlights the significance of embodied learning for knowledge acquisition [5, 12, 21]. By actively engaging with the environment through physical movement and gesture, learners foster deeper understanding and meaning construction. This resonates with Chomsky's observation [4] of natural language acquisition, where children learn through immersive exploration and interaction with the surrounding world. Recognizing these principles, we are exploring the potential of novel technologies like Mixed Reality (MR) to create engaging and embodied learning environments.

VerbVenture emerges as an innovative MR application specifically designed to address the EFL collocation challenge for young learners. Drawing inspiration from embodied learning and leveraging the immersive capabilities of MR technology, *VerbVenture*transforms everyday objects into interactive learning portals. Children, equipped with MR headsets, embark on a playful journey where AI object recognition identifies real-world objects, triggering the appearance of animated scenario and interactive tasks. These engaging cues guide learners through embodied interactions, such as carrying a bag or kicking virtual balls, reinforcing the connections between words and their corresponding actions or functions. This active manipulation of virtual elements within the context of real-world objects also aligns with situated learning principles [3], fostering authentic and meaningful engagement [14, 15] with the target language.

By exploring the design and effectiveness of *VerbVenture*, we aim to contribute to the evolving field of interactive EFL learning and shed light on the potential of embodied learning with MR technologies in fostering deeper language acquisition, particularly in the domain of collocations.

2 RELATED WORK

This section delves into relevant research areas intersecting with *VerbVenture*: EFL collocation acquisition, embodied language learning, and mixed reality and immersive learning.

2.1 EFL Collocation Acquisition

For elementary EFL learners, mastering collocations alongside target words poses a significant challenge. While teachers utilize example sentences and repetitive recitation, contextual comprehension and situational understanding play crucial roles in optimal collocation learning[23]. Nation[16] emphasized the importance of collocations for achieving language fluency and accuracy. Bahns and Eldaw [1] further pointed out that even advanced learners struggle with collocations, which constitute less than a quarter of vocabulary but contribute to over half of translation errors. Therefore, traditional vocabulary learning does not guarantee adequate collocation knowledge. *VerbVenture* seeks to address this gap by leveraging embodied learning principles within an engaging MR environment.

2.2 Embodied Language Learning

Embodied cognition theory posits that learning is deeply intertwined with physicality. Bodily movements and experiences directly impact cognitive processes like understanding and memory formation. This principle forms the foundation of embodied learning, a pedagogical approach that emphasizes the link between physical engagement and knowledge acquisition. It aligns with the cognitive theory of multimodal learning, emphasizing the selection, organization, and integration of information during learning [9, 15]. Numerous studies highlight the critical role of physical activity in children's development, not just physically but also cognitively [5, 12]. Embodiment, defined as the state of the body in relation to cognitive processes, plays a key role here. Research suggests that physically enacting concepts strengthens mental representations, leading to improved memory and learning [12].

The effectiveness of embodied learning hinges on aligning physical actions with learning objectives. Mavilidi et al.[13] compared physical movements and gestures in preschoolers learning Italian vocabulary. Their findings indicated that performing movements corresponding to the words led to the best learning outcomes, highlighting the positive impact of embodied learning on knowledge acquisition. Similarly, Schmidt et al.[19] found that physically active foreign language lessons enhanced vocabulary learning and attention performance in elementary school students.

Embodied learning provides a solution to the difficulties encountered by EFL learners in mastering verb-noun collocations. *VerbVenture* specifically targets this challenge by designing activities that link physical movements with specific collocations in elementary EFL learners, offering a fun and effective solution.

2.3 Mixed Reality and Immersive Learning

Mixed Reality (MR) seamlessly integrates the real and virtual worlds, providing immersive and interactive learning experiences that engage multiple senses. This transforms the learning process into a playful adventure, sparking curiosity and motivation in children. Tasks like visualizing 3D shapes in geometry class become effortless when students can manipulate them within a virtual space. This not only deepens understanding but also nurtures essential spatial reasoning skills crucial for various subjects. Numerous studies suggest that the objectives of MR and immersive learning include enhancing students' embodiment of concepts and spatial visualization, leading to the construction of cognitive representations and facilitating deep learning for long-term memory formation [11, 18].

While explicit guidance on MR methods for K-12 education remains limited [9], its unique ability to bridge physical spaces with learning content offers immense promise. Examples like the MEteor simulation [10], where students launch virtual asteroids through body movements, or the SMALLab collaborative learning environment [2], demonstrate how MR can connect with prior knowledge and engage learners across diverse subjects, including STEM and special education.

Essentially, successful MR learning relies on synchronizing physical actions with new concepts. Activities need meticulous design to empower learners to "embody" the intended understanding, reinforcing the link between their physical movements and the cognitive acquisition of knowledge. *VerbVenture* proposes an integration of artificial intelligence-based action recognition with mixed reality technology to offer a scalable and engaging learning experience in everyday scenarios.

3 SYSTEM DESIGN

Our research focuses on embodied learning for real-world collocations, seeking to seamlessly integrate verb-noun and verb-prep-noun collocations with relevant physical activities. This approach mimics the natural language acquisition process, encouraging children to freely explore real-world objects and engage in corresponding actions that exemplify verb-noun collocations.

3.1 Interactive flow

Figure 1 shows the initial storyboard of the system. The interaction flow consists of several steps as outlined below. First, the child wearing the MR headset explores interactive objects in a space by pointing at a physical object (Fig.2a). Second, once the physical object is selected, a vocabulary bubble appears above the object, pronouncing the target word. Simultaneously, an audio and visual cue prompt the child to repeat the word (Fig.2b). Third, upon the correct repetition of the target word, the bubble disappears. Subsequently, a sequenced simulated animation provides an intuitive context for the child to enact the verb-noun collocation action. For example, with the target word "umbrella," the sequential collocations could be "put up," "fold," and "roll." At the beginning of the animated scenario, depicting "Dark clouds gather, followed by rain," the system highlights the interactive physical object (umbrella) with a sparkle effect, prompting the child to pick up the object (Fig.2c). In this scenario, an animated icon of putting up the umbrella signals the child to perform the same action (Fig.2d). Subsequently, a collocation bubble containing "put up the umbrella" pops up, requiring the child to speak out while performing the action. Once the voice and action are recognized by the system, the next

Fig. 1. Storyboard depicting the sequential steps and interactions in the learning process of the target word 'Umbrella' and its associated collocations within the VerbVenture application.

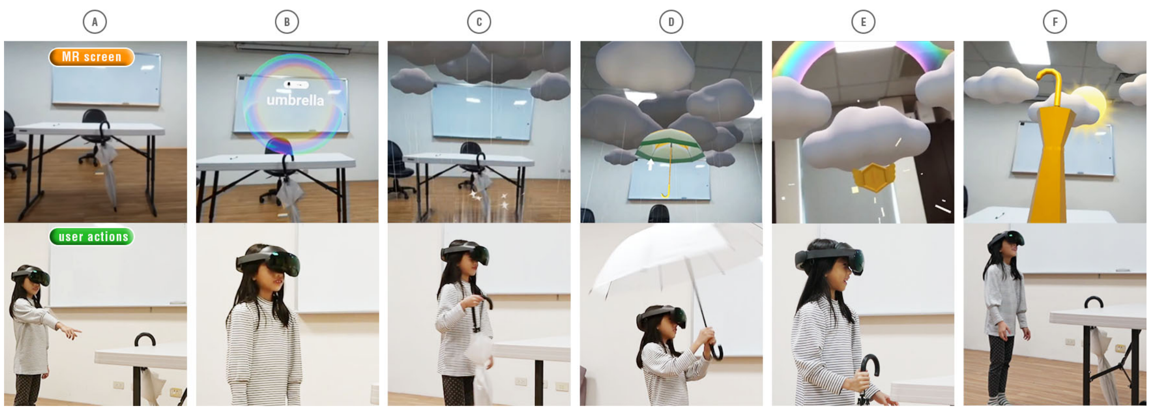

Fig. 2. Interactive flow within the VerbVenture prototype.

animated scenario, "Rain stops, rainbow appears," follows for folding the umbrella, and then "Clouds disperse, sun comes out" for rolling the umbrella. The child will collect a medal as a reward for each collocation word practice (Fig.2e) and receive a trophy after completing all collocation words related to the interactive object (Fig.2f). Finally, the system will display an indicator to remind the child to place back the object, concluding the session.

3.2 Prototype implementation

The section describes the technical detals of our prototype, built using the Meta Quest Pro headset and Unity game engine. While focused on Mixed Reality (MR) development, object and action recognition through the headset's front camera played a crucial role. Several Software Development Kits (SDKs) like Meta AI Segment Anything Model (SAM) were promising candidate for environmental object recognition. However, due to Meta's privacy regulations, direct access to real-world images captured by the Meta Quest Pro's Passthrough camera was initially unavailable. To ensure a seamless user testing experience, we employed the Wizard of Oz method. This technique involved researchers simulating object and action recognition functionalities through hotkeys and providing corresponding feedback, effectively emulating the intended MR experience. The initial setup featured a developer mode allowing the experiment facilitator to place a bounding box around the interactive physical object. This enabled children to trigger the object with a simple pointing gesture, recognized using Meta's Interaction SDK with hand tracking. For speech input, we implemented Automatic Speech Recognition (ASR) powered by the Oculus Voice SDK and supported by Wit.ai. Research, as highlighted by Kumar et al. and Yaniafari et al. [7, 24], demonstrates the positive impact of ASR on pronunciation improvement in language learning settings.

4 PILOT STUDY

Our pilot study aimed to evaluate whether VerbVenture could 1) effectively engage children with the interactive learning process, 2) enhance their interest and motivation in acquiring vocabulary and collocations in daily life, and 3) facilitate successful recall of learned collocations immediately after the learning session and after a two-day interval.

4.1 Participants and task design

In this phase, we invited a participation of an 8-year-old, second-grade elementary school girl. Notably, she had not been exposed to collocations in her previous learning experiences. While she had a fundamental understanding of the term "umbrella," her familiarity with collocational usage was limited. Due to her relatively short exposure to the English language, she also lacked experience in verbal communication. To assess the effectiveness of the learning outcome, we aligned with the curriculum guidelines and materials from elementary school English education in Taiwan. The target words and collocations were selected to correspond with the 2nd-grade level English curriculum.

4.2 Procedure

In the experimental design, three stages were executed.

Instruction Phase. Researchers briefly explained the game process and tasks to the *learners*, including initiating the game by pointing to objects with a finger, performing actions based on on-screen prompts, and ultimately achieving the winning condition of obtaining a trophy.

Learning Phase. Researchers assisted *learners* in donning MR headsets and autonomously initiating the learning of a word, "Umbrella," and three collocations, "put up," "fold," and "roll." During the learning process, *learners* triggered word bubbles by pointing to a real umbrella in the environment, recited the word following visual and auditory prompts, and then proceeded to the collocation learning segment. Each collocation's learning process was indicated by a contextual animation (e.g., rainfall), followed by dynamic cues (e.g., animation for "put up the umbrella") guiding *learners* to perform corresponding actions. Only upon correctly completing the action, the respective collocation bubble (e.g., put up the umbrella) appeared, requiring *learners* to repeat the pronunciation. Correct responses resulted in small rewards.

After completing the learning of three collocations, a semi-transparent umbrella icon prompted *learners* to return the umbrella to its place, ultimately earning the umbrella trophy as the final reward, concluding one learning experience. *learners* underwent two learning sessions, each lasting 4 minutes, and received a beverage reward after completing the tests.

Collocation Testing Phase. After the learning phase, *learners* rested for 10 minutes before entering two testing stages: Free-recall test and Cued recall test. The former required *learners* to freely describe the content they had just learned, including Chinese and English descriptions; the latter presented flashcards with images, requiring *learners* to correctly state the English collocation. Tests were repeated 2 days later to assess *learners*' retention of the content. Additionally, preferences and reasons for the learning process were assessed through filling out an emotional scale and conducting semi-structured interviews.

5 RESULT AND DISCUSSION

Through the collocation testing phase, it was observed that *learners* could distinctly comprehend the Mandarin meanings and associated actions for each collocation. However, challenges surfaced in terms of English oral expression, with *learners* frequently encountering difficulties in pronunciation. Throughout the MR learning process, multiple audio pronunciation prompts were required for precise retention.

Based on emotional scales and interviews, *learners* expressed a strong desire to learn more vocabulary using this method. They mentioned that the contextual animations prevented them from feeling bored with repetitive actions. In terms of interest ranking, the umbrella trophy, representing the end of the game and providing a visually impactful moment, was the *learners*' favorite. The subsequent rankings, from two to six, were the raining scenario, bubbles, self-performed actions, the appearance of a rainbow, and finally, the umbrella. The lower preference for the umbrella was attributed to it being the last option left in the ranking.

During the collocation learning phase, *learners* naturally followed audio prompts for recitation without needing auditory cues. However, their oral performance was suboptimal. Apart from the *learners*' young age, the fast pace of audio prompts hindered their ability to keep up. Additionally, *learners* had only one opportunity for oral expression for each collocation, limiting their chances for verbal practice in a single learning experience. Concerning the guidance from prompting animations, it was noted that *learners* struggled to comprehend the significance of the umbrella lighting up and the appearance of a semi-transparent umbrella.

Based on the results of pilot study, moving forward, we plan to refine the oral expression and prompting animation aspects. In the oral expression process, in addition to slowing down the audio speed, we will highlight differences in color based on syllables in the word to guide *learners* gradually in pronouncing complete collocations. We will also attempt to introduce different interactive elements to increase opportunities for *learners* to repeat and rehearse collocations. Regarding prompting animations, we will consider using dynamic hints to guide *learners* with "pick up" and "put down" symbols.

6 CONCLUSION

VerbVenture engages EFL learners, particularly elementary student, in developing an interest in understanding new vocabulary. They interact with both real and virtual objects in their daily environment, combining embodied learning methods to physically perform verb actions, thereby deepening their understanding and memory. Although it enhances children's learning interest and motivation overall, the effectiveness of the learning process requires extensive experimentation to yield more conclusive results. Furthermore, the AI object recognition functionality necessitates exploration of alternative approaches to overcome current limitations, enhancing the overall realism of the experience. In conclusion, we have discussed the integration of embodied learning, AI, and MR in teaching methods and processes, highlighting capabilities and challenges within VerbVenture. We believe that MR will prove to be a powerful tool for English collocation learning.

REFERENCES

[1] Jens Bahns and Moira Eldaw. 1993. Should we teach EFL students collocations? *System* 21, 1 (1993), 101–114. https://doi.org/10.1016/0346-251X(93)90010-E

[2] David Birchfield and Colleen Megowan-Romanowicz. 2009. Earth science learning in SMALLab: A design experiment for mixed reality. *Computer Supported Learning* 4 (2009), 403–421. https://doi.org/10.1007/s11412-009-9074-8

[3] John Seely Brown, Allan Collins, and Paul Duguid. 1989. Situated Cognition and the Culture of Learning. *Educational Researcher* 18, 1 (1989), 32–42. https://doi.org/10.3102/0013189X018001032

[4] Carol Chomsky. 1972. Stages in Language Development and Reading Exposure. *Harvard Educational Review* 42, 1 (1972), 1–33. https://doi.org/10.17763/haer.42.1.h78l676h28331480

[5] Michele Cox, Grant Schofield, and Gregory S. Kolt. 2010. Responsibility for children's physical activity: Parental, child, and teacher perspectives. *Journal of Science and Medicine in Sport* 13, 1 (2010), 46–52. https://doi.org/10.1016/j.jsams.2009.02.006

[6] M. Hoey. 1991. *Patterns of Lexis in Text.* Oxford University Press, Oxford, UK. https://books.google.com.tw/books?id=MBV6AAAAIAAJ

[7] Anuj Kumar, Pooja Reddy, Anuj Tewari, Rajat Agrawal, and Matthew Kam. 2012. Improving literacy in developing countries using speech recognition-supported games on mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1149–1158. https://doi.org/10.1145/2207676.2208564

[8] Batia Laufer and Tina Waldman. 2011. Verb-Noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English. *Language Learning* 61, 2 (2011), 647–672. https://doi.org/10.1111/j.1467-9922.2010.00621.x

[9] Robb Lindgren and Mina Johnson-Glenberg. 2013. Emboldened by Embodiment: Six Precepts for Research on Embodied Learning and Mixed Reality. *Educational Researcher* 42, 8 (2013), 445–452. https://doi.org/10.3102/0013189X13511661

[10] Robb Lindgren and Michael J. Moshell. 2011. Supporting children's learning with body-based metaphors in a mixed reality environment. In *Proceedings of the 10th International Conference on Interaction Design and Children* (Ann Arbor, Michigan) *(IDC '11)*. Association for Computing Machinery, New York, NY, USA, 177–180. https://doi.org/10.1145/1999030.1999055

[11] Robb Lindgren, Michael Tscholl, Shuai Wang, and Emily Johnson. 2016. Enhancing learning and engagement through embodied interaction within a mixed reality simulation. *Computers & Education* 95 (2016), 174–187. https://doi.org/10.1016/j.compedu.2016.01.001

[12] Christopher R. Madan and Anthony Singhal. 2012. Encoding the world around us: Motor-related processing influences verbal memory. *Consciousness and Cognition* 21, 3 (2012), 1563–1570. https://doi.org/10.1016/j.concog.2012.07.006

[13] Myrto-Foteini Mavilidi, Anthony D. Okely, Paul Chandler, Dylan P. Cliff, and Fred Paas. 2015. Effects of Integrated Physical Exercises and Gestures on Preschool Children's Foreign Language Vocabulary Learning. *Educational Psychology Review* 27, 3 (2015), 413–426. https://doi.org/10.1007/s10648-015-9337-z

[14] Richard E. Mayer. 2002. Multimedia learning. *Psychology of Learning and Motivation* 41 (2002), 85–139. https://doi.org/10.1016/S0079-7421(02)80005-6

[15] Richard E. Mayer. 2014. Incorporating motivation into multimedia learning. *Learning and Instruction* 29 (2014), 171–173. https://doi.org/10.1016/j.learninstruc.2013.04.003

[16] Ian Stephen Paul Nation. 2001. *Introduction* (3 ed.). Cambridge University Press, Cambridge UK, 1–7. https://doi.org/10.1017/9781009093873.001

[17] Tun pei Chan and Hsien-Chin Liou. 2005. Effects of Web-based Concordancing Instruction on EFL Students' Learning of Verb–Noun Collocations. *Computer Assisted Language Learning* 18, 3 (2005), 231–251. https://doi.org/10.1080/09588220500185769

[18] Mafor Penn and Umesh Ramnarain. 2023. *A Systematic Review of Pedagogy Related to Mixed Reality in K-12 Education.* Springer Nature Singapore, Singapore, 85–108. https://doi.org/10.1007/978-981-99-4958-8_5

[19] Mirko Schmidt, Valentin Benzing, Amie Wallman-Jones, Myrto-Foteini Mavilidi, David Revalds Lubans, and Fred Paas. 2019. Embodied learning in the classroom: Effects on primary school children's attention and foreign language vocabulary learning. *Psychology of Sport and Exercise* 43 (2019), 45–54. https://doi.org/10.1016/j.psychsport.2018.12.017

[20] Paweł Szudarski and Ronald Carter. 2016. The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics* 26, 2 (2016), 245–265. https://doi.org/10.1111/ijal.12092

[21] Feng (Mark) Teng. 2019. The effects of video caption types and advance organizers on incidental L2 collocation learning. *Computers & Education* 142 (2019), 103655. https://doi.org/10.1016/j.compedu.2019.103655

[22] S. Webb and P. Nation. 2017. *How Vocabulary is Learned.* Oxford University Press, Oxford, UK. https://books.google.com.tw/books?id=OwkjvgAACAAJ

[23] Shaoqun Wu, Margaret Franken, and Ian H. Witten. 2010. Supporting collocation learning with a digital library. *Computer Assisted Language Learnig* 23, 1 (2010), 87–110. https://doi.org/10.1080/09588220903532971

[24] Rahmati Putri Yaniafari, Viga Olivia, and Suharyadi. 2022. The Potential of ASR for Improving English Pronunciation: A Review. *KnE Social Sciences* 7, 7 (2022), 281–289. https://doi.org/10.18502/kss.v7i7.10670