Proceedings of CONF-SEML 2026 Symposium: Computational Analysis and Modeling in Complex Intelligent Systems
DOI: 10.54254/2755-2721/2026.GU33743

A Knowledge-Graph-Driven Adaptive Learning Pathway System for English Vocabulary Acquisition

Yaxuan Wang

Faculty of Education, Taylor University, Selangor, Malaysia
15653390177@163.com

Abstract. In college English teaching, vocabulary acquisition is often limited by the isolated learning of knowledge points and the discontinuity of learning paths. Traditional linear learning models fail to capture deep semantic connections between words, while existing online platforms often overwhelm learners with information overload. To address these issues, this paper proposes an adaptive learning path system for English vocabulary acquisition based on knowledge graphs. This system constructs an English domain ontology covering vocabulary, roots, derivatives, and test attributes, and uses the TransR entity embedding model to address the semantic sparsity problem. At the core algorithm level, this paper also proposes a hybrid recommendation model (KG-Rec) that integrates collaborative filtering and knowledge graphs. By calculating the semantic distance between the learner's cognitive state and knowledge nodes, the system dynamically generates personalized learning paths. Empirical studies and a 7-day controlled teaching experiment (sample size N=64) show that the experimental group significantly outperformed the traditional learning group in both vocabulary mastery (improvement of 29.7%) and learning efficiency. This study confirms that adaptive learning paths based on structured knowledge can effectively reduce cognitive burden and improve language learning efficiency.

Keywords: Knowledge Graph, Adaptive Learning, English Vocabulary Acquisition, Learning Path Recommendation, TransR Model

1. Introduction

Vocabulary is the cornerstone of language learning, but the sheer volume and semantic complexity of English words have long been a challenge in second language acquisition (SLA). Traditional vocabulary teaching often adopts the "vocabulary list" or "flashcard" approach, which is a linear and fragmented learning method that separates the intricate semantic connections between words (such as hierarchical relationships, synonyms and antonyms, etc.). This leads to learners being unable to build a systematic knowledge network and is prone to forgetting.

With the widespread integration of artificial intelligence (AI) technology in education, adaptive learning has emerged as a key approach to addressing the challenges of personalized instruction. However, the current mainstream adaptive learning systems mostly rely on collaborative filtering or simple rule matching, and have two major limitations: one is lack of semantic understanding, the system cannot perceive the logical dependencies between knowledge points, resulting in the

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.GU33743

recommended content being "knowing the result but not the reason". The other is the cold start problem, for new users or rare knowledge points, due to the lack of interaction data, the accuracy of recommendations drops significantly.

To overcome the aforementioned limitations, the introduction of Knowledge Graph (KG) technology is particularly necessary. Knowledge Graph can transform unstructured language knowledge into structured graph data, providing a "cognitive map" for the system. Previous research has shown that integrating knowledge graphs into educational recommendation systems can significantly enhance both the interpretability and the accuracy of the recommendations [1].

This research aims to develop an adaptive learning system capable of understanding semantics and understanding students. To achieve this, an English vocabulary domain ontology is constructed, featuring a graph structure tailored for language acquisition that encompasses multi-dimensional entities such as words, roots, prefixes, and keywords. Furthermore, a hybrid recommendation algorithm, termed KG-Rec, is proposed to address the semantic sparsity problem by integrating TransR graph embedding technology with collaborative filtering, thereby facilitating a transition from resource recommendation to path planning. Finally, the effectiveness of the system in enhancing vocabulary learning efficiency is evaluated through a controlled empirical validation.

2. Related work

2.1. The application of educational knowledge graphs

The application of knowledge graphs (KGs) in education has evolved beyond simple visualization to support in-depth reasoning and analysis. Scholars have systematically reviewed the construction technology of knowledge graphs and pointed out their core role in integrating heterogeneous educational data [2]. In the field of language teaching, some studies have used the combination of corpora and knowledge graphs to assist comprehensive English teaching. Through visual semantic networks, they help students establish connections between vocabulary and text topics [3]. In addition, for the automatic construction of question banks, Sun et al. proposed a knowledge point extraction method based on support vector machines, constructed a question knowledge graph, and realized the automatic association between test points and questions [4]. While these studies have confirmed the advantages of graphs for knowledge structuring, most have focused on static representation and lack mechanisms for generating dynamic learning paths.

2.2. Adaptive learning path recommendation

Path recommendation is the core of the adaptive learning system. Early path recommendation systems primarily relied on genetic or ant colony algorithms. However, these methods suffered from high computational complexity and could not respond in real time. In recent years, deep learning and reinforcement learning have been widely introduced. For instance, Ma et al. proposed a personalized English learning path recommendation system based on deep reinforcement learning (Deep Reinforcement Learning), using the PPO strategy to optimize long-term learning paths [5]. Hou et al. then proposed the KG-PLPPM model, which achieved more accurate remedial teaching path planning by evaluating the similarity and mastery degree of knowledge concepts [6].

DOI: 10.54254/2755-2721/2026.GU33743

2.3. Research gap

However, existing models remain deficient in handling the polysemy of English vocabulary and context dependence. This study attempts to introduce the TransR model to specifically handle the multiple semantic representations of lexical entities in different relationship spaces (such as "synonym relationship space" and "derivation relationship space"), in order to improve the semantic adaptability of path recommendations.

3. System design and methodology

3.1. Construction of the English vocabulary ontology

Knowledge graphs serve as the foundational data infrastructure for the proposed system [7]. Unlike general encyclopedic knowledge graphs, those designed for education must incorporate pedagogical attributes of knowledge, such as difficulty and frequency of appearance in assessments.

·Entity Layer:

Word (Word): Core entity, such as "Acquisition".

Root (Root): Constituent element, such as "quis" (seek).

Affix (Affix): Prefix/suffix, such as "ac-", "-tion".

Topic (Topic): Context in which the word belongs, such as "Education", "Science".

·Relation Layer:

Semantic relations: Synonym (synonym), Antonym (antonym), Derived_from (derived from).

Cognitive relations: Prerequisite (pre-requisite), for example, before learning the complex word "Deconstruction", the system suggests mastering "Construct" first.

The primary data sources are the WordNet database and the vocabulary lists specified in the syllabus for the College English Test Band 4 and Band 6 (CET-4/6). The construction process utilized a combination of automated tools and manual verification to ensure data quality. Specifically, the Python-based Natural Language Toolkit (NLTK) and Stanford CoreNLP were employed for initial tokenization, part-of-speech tagging, and lemmatization. Entity extraction followed a rule-based approach leveraging the lexical hierarchy of WordNet, while relationship alignment was executed through a semantic similarity threshold filtering mechanism. To ensure the precision of the graph, a random sample of 500 triplets was extracted for manual inspection by two English linguistics experts. The initial automated extraction achieved a precision of 91.5%, and all identified errors (such as ambiguous root-affix mappings) were manually corrected before the final domain graph—consisting of 4,800 core vocabulary nodes and 12,000 edges—was formally established.

3.2. Graph representation learning based on TransR

To calculate the semantic distance between words, entities and relations must be vectorized. Traditional models such as TransE struggle to handle the polysemy inherent in English words (manifested as 1-to-N relations) because they map all relations within a single, shared entity space. For instance, the word "Bank" can be linked to "River" via a Geographic relation and to "Finance" via an Economic relation. In TransE, this forces the vectors for "River" and "Finance" to be nearly identical, which is semantically incorrect. Therefore, this study adopts the TransR model. TransR addresses this by projecting entities into specific relation spaces. Using the same example, "Bank" is projected into an Economic space when considering its financial attributes, and into a Geographic

DOI: 10.54254/2755-2721/2026.GU33743

space for its topographical meaning. This allows the system to distinguish between different semantic manifestations of the same word depending on the context of the relation.

The core idea behind the TransR model is that entities are multifaceted, possessing multiple attributes, and different relations focus on different facets of an entity. For each triplet (h, r, t), the map entities h and t to the dedicated space of relation r through the projection matrix M_r, such that $h_r + r \approx t_r$. This enables the system to distinguish the different semantic manifestations of the same word in "synonymous relations" and "antonymous relations", thereby improving the accuracy of recommendations.

3.3. KG-REC hybrid recommendation algorithm

The system employs a hybrid recommendation strategy designed to integrate the semantic reasoning capability of the knowledge graph with the collective intelligence of collaborative filtering.

Semantic Path Score ($Score_{KG}$): This score is operationalized to quantify the cognitive proximity between a candidate word w and the learner's mastered vocabulary set S. Specifically, the system retrieves the entity vectors e_w and e_{s_i} from the TransR embedding space and projects them into a specific relation space $r \in R$ (e.g., synonymy or derivation) using a projection matrix M_r, calculated as $e_{w,r} = e_w M_r$. To focus on the "Zone of Proximal Development" (ZPD) and mitigate noise from unrelated terms, $Score_{KG}$ is defined as the inverse of the average Euclidean distance between w and its k most semantically similar neighbors in S across all pedagogical relations:

$$Score_{KG}(w) = \left[\frac{1}{k} \sum_{i=1}^{k} \min_{r \in R} \left| e_{w,r} - e_{s_i,r} \right|_2 \right]^{-1} \tag{1}$$

A higher score indicates that the word w possesses stronger semantic or logical dependencies with the learner's current knowledge, thereby facilitating more efficient acquisition.

Group Behavior Score ($Score_{CF}$): This score is determined using user-based collaborative filtering, where the "neighbor" groups are identified by calculating the Cosine Similarity between learners' vocabulary mastery vectors. Specifically, the system selects the Top-10 users with the highest similarity scores as neighbors to the current learner. $Score_{CF}(w)$ is then computed as the normalized frequency of word w within the successful learning paths of these identified neighbors, ensuring the recommendation reflects effective peer learning patterns.

Dynamic Path Generation: The final recommended score $R(w)$ is the weighted sum of the two (1):

$$R(w) = \alpha \cdot Score_{KG}(w) + (1 - \alpha) \cdot Score_{CF}(w) \tag{2}$$

Where $\alpha \in [0,1]$ denotes the harmonic weight coefficient used to balance the semantic structural constraints of the knowledge graph and the collective intelligence derived from learner groups. In this study, α is empirically set to 0.6, giving greater weight to the logical dependencies of the domain ontology. The system performs a descending sort on $R(w)$ to select the top-scoring candidate words for path generation, ensuring that the most cognitively appropriate content is presented first.

The generated learning path is restricted to 3 to 5 nodes. This design is grounded in Cognitive Load Theory, which suggests that human short-term memory can effectively process only a limited

DOI: 10.54254/2755-2721/2026.GU33743

number of information "chunks" (typically 5 ± 2) at once. A concise path helps prevent cognitive overload and maintains learner focus. Furthermore, the sequence of node types (e.g., root → basic word → derivative → discrimination question) implements a scaffolding instruction logic. This progression facilitates "incremental learning" by guiding students from foundational morphological recognition to complex semantic differentiation, effectively bridging the gap between passive vocabulary knowledge and active linguistic application.

4. Experimental design

4.1. Experimental setup

To verify the effectiveness of the system, this study conducted a comparative experiment at a certain university.

Participants: 64 non-English major sophomore students were selected and randomly divided into the experimental group (N=32) and the control group (N=32). Prior to the experiment, a pre-test evaluating vocabulary size and writing application was administered to ensure baseline equivalence. Independent sample t-test results indicated no significant difference between the experimental and control groups in both initial vocabulary mastery ($t = 0.45$, $p = 0.65 > 0.05$) and writing scores ($t = 0.38$, $p = 0.71 > 0.05$), confirming that the two groups were at a comparable starting level.

Experimental Tools:

Control Group: Utilized traditional vocabulary apps and learned in either alphabetical order or randomly shuffled sequence.

Experimental Group: Utilized the "Knowledge Graph Adaptive Learning System" developed in this study.

Procedure: The experiment lasted 7 days, with each participant completing a 30-minute study session per day. Both vocabulary knowledge and writing application tests were administered before (pre-test) and after (post-test) the intervention.

Additionally, upon completion of the 7-day learning period, a post-study questionnaire based on a 5-point Likert scale was administered to both groups. This questionnaire was designed to evaluate learners' "Path satisfaction," specifically assessing their subjective perception of the learning sequence's logical coherence, navigability, and overall helpfulness.

4.2. Experimental results

The experimental data were all subjected to independent sample t-test using SPSS 26.0. The main results are shown in Table 1.

DOI: 10.54254/2755-2721/2026.GU33743

Table 1. Comparison of learning effects between the experimental group and the control group

Evaluation Dimension	Control Group (Mean ± SD)	Experimental Group (Mean ± SD)	t -value	p
Vocabulary mastery rate	125.4 ± 15.2	162.8 ± 12.6	4.32	< 0.01
Writing application score	16.1 ± 3.1	22.9 ± 2.5	5.18	< 0.01
Path satisfaction	3.2 ± 0.8	4.5 ± 0.6	6.21	< 0.001

As shown in Table 1, the experimental group demonstrated superior performance across all evaluation dimensions compared to the control group. Specifically, in terms of vocabulary mastery rate, the experimental group achieved a mean score of 162.8 ± 12.6, which was significantly higher than the control group's 125.4 ± 15.2 ($t = 4.32$, $p < 0.01$). For the writing application score, the experimental group also significantly outperformed the control group (22.9 ± 2.5 vs. 16.1 ± 3.1, $t = 5.18$, $p < 0.01$). Furthermore, the subjective path satisfaction reported by the experimental group was notably higher (4.5 ± 0.6) than that of the control group (3.2 ± 0.8), displaying a highly significant statistical difference ($t = 6.21$, $p < 0.001$).

5. Discussion

5.1. In-depth analysis of findings

The experimental results provide several key insights regarding the effectiveness of the proposed system. First, there was a significant improvement in memory efficiency. The 29.7% higher vocabulary acquisition in the experimental group confirms that structured learning based on knowledge graphs effectively promotes memory retention [8]. By recommending semantically related word clusters (e.g., linking the root spect with inspect, prospect, and retrospect), the system serves as a cognitive scaffold, aiding learners in constructing mental "memory palaces." Second, the system proved effective in enhancing application skills. The superior writing test scores of the experimental group highlight that explicit derivation relationships (Derived_from) in the graph guide students not only to memorize definitions but also to internalize the rules of their grammatical changes. This transition from receptive recognition to productive vocabulary use is a crucial breakthrough in second language acquisition. Finally, the high path satisfaction underscores the algorithm's effectiveness. The KG-Rec algorithm successfully mitigates the "navigation confusion" typical of traditional platforms. Compared with traditional suggestions, the graph-based path logic is more consistent and aligns closer with human cognitive patterns. This finding corroborates those of Li et al. regarding the efficacy of AI-assisted personalized learning path optimization [9].

5.2. Limitations and future directions

Despite the promising results, this study has several limitations that should be addressed in future research.

DOI: 10.54254/2755-2721/2026.GU33743

Lack of Emotional Adaptability: The experiment revealed that recommending excessive prior knowledge could frustrate learners with weaker foundations. Future algorithm optimization must integrate more complex mechanisms to take into account emotional factors, such as dynamically adjusting cognitive load based on real-time affective feedback [10].

Limited Experimental Scope (Duration and Sample Size): The current study utilized a 7-day short-cycle experiment with 64 participants, which may not fully capture the long-term retention effects of the system. Future research will expand the sample size across multiple institutions and extend the experimental period to a full academic semester to evaluate longitudinal impacts.

Need for More Rigorous Validation (Control Group and Ablation Studies): While this study compared the system against a traditional vocabulary app, it lacked rigorous validation of individual module contributions. Future work will introduce comprehensive ablation experiments (e.g., disabling the TransR graph embedding module or the collaborative filtering module independently) within the traditional app environment to further isolate and verify the specific performance gains generated by the KG-Rec algorithm.

6. Conclusions

This study designed and implemented an adaptive learning path system for English vocabulary acquisition, leveraging knowledge graph technology to overcome the persistent challenges of information silos and fragmented learning trajectories. The research employed the TransR model to vectorize the semantic information of English vocabulary and proposed a KG-Rec hybrid recommendation algorithm. The system successfully transitioned from static resource accumulation to dynamic semantic navigation. Results show that the system significantly improved learners' vocabulary memorization and application abilities while maintaining high user satisfaction.

Future work will pursue two key directions to advance this line of research. First, the study will explore the integration of multimodal data (e.g., images, videos) to enrich the knowledge graph with perceptual and contextual information, constructing a more comprehensive and immersive learning ontology. Second, the study will investigate the synergy between knowledge graphs and large language models (LLMs). The objective is to leverage the powerful generative and reasoning capabilities of LLMs to automatically create contextualized learning scenarios and adaptive exercises that are dynamically aligned with the recommended learning paths. This integration aims to evolve the current system into a more intelligent and responsive educational ecosystem.

References

[1] Abu-Salih, B., & Alotaibi, S. (2024). A systematic literature review of knowledge graph construction and application in education. Heliyon, 10(3).
[2] Li, L. (2025). Artificial intelligence-driven personalized college English vocabulary learning path generation and optimization. Journal of Combinatorial Mathematics and Combinatorial Computing, 127, 279-293.
[3] Yu, W. (2025). Corpus and knowledge graph-assisted integrated English teaching. Sino-US English Teaching, 22(3), 113-117.
[4] Sun, Y., Tang, J., & Zhu, Z. (2021). A method of English test knowledge graph construction. Journal of Computer and Communications, 9(9), 99-107.
[5] Zhou, L.-Y., & Wang, Y.-Y. (2025). Simulation of personalized English learning path recommendation system based on knowledge graph and deep reinforcement learning. Scientific Reports, 15(1), Article 34554.
[6] Hou, B., Lin, Y., Li, Y., et al. (2025). KG-PLPPM: A knowledge graph-based personal learning path planning method used in online learning. Electronics, 14(2), Article 255.
[7] Zhang, S., Wang, X., Ma, Y., et al. (2023). An adaptive learning method based on knowledge graph. Frontiers in Educational Research, 6(6).

DOI: 10.54254/2755-2721/2026.GU33743

[8] Wang, F., Zhang, L., Chen, X., et al. (2019). Research on personalized learning path discovery based on differential evolution algorithm and knowledge graph. In International Conference on Data Service (pp. 285-295). Springer Singapore.
[9] Li, L. (2025). Personalized college English vocabulary learning path design incorporating adaptive algorithms. International Journal for Housing Science and its Applications. https: //doi.org/10.70517/ijhsa463712
[10] Chen, J.-Y., Saeedvand, S., & Lai, I.-W. (2023). Adaptive learning path navigation based on knowledge tracing and reinforcement learning. arXiv, arXiv: 2305.04475.