ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

The Role of Lexical Semantic Network Construction in Chinese College Students' English Reading

Kaifang Fan[1,*]

[1] Suqian University Suqian Jiangsu 223800

SUMMARY: *In order to explore the effect of lexical semantic network construction on Chinese college students' English reading comprehension, this paper selects 40 reading texts from 286 non-English major college students to construct a corpus. After preprocessing, 2316 valid words are preserved, and 4876 valid semantic relationship edges are formed. This paper constructs a lexical semantic network by means of natural language processing, semantic similarity calculation and graph model, and uses a reading test to test its influence on reading performance, word sense inference and discourse integration. The results show that the average clustering coefficient of the network is 0.312, the average path length is 5.27, and the proportion of the largest connected subgraph is 88.51%, which indicates that the lexical relations have obvious aggregation and connectivity. With the improvement of semantic network structure, the total score of reading increased from 66.3 to 82.1, the score of word meaning inference increased from 62.7 to 78.9, and the score of discourse integration increased from 64.1 to 81.4. The results show that lexical semantic network can effectively improve reading comprehension performance by enhancing semantic activation, inter-sentence cohesion and discourse construction, and provide a new method basis for college English reading teaching and intelligent reading support.*

KEYWORDS: *Lexical semantic network; English reading comprehension; Chinese college students; Natural Language Processing*

1 Introduction

In the English reading process of Chinese college students, vocabulary is not only the basic unit for literal recognition, but also plays an important role in activating semantic connections, organizing textual information and supporting reasoning and judgment. A solid foundation has been established on the relationship between vocabulary knowledge and reading comprehension. Tong et al. (2023) found that there was a significant correlation between second language vocabulary knowledge and reading level, and vocabulary fluency would have a moderating effect on this relationship [1]. Chen and Zhang (2023) proposed that lexical depth knowledge is not a single dimension, and the role of depth knowledge at different levels in reading comprehension is obviously different [2]. Dagnaw (2023) further pointed out that vocabulary breadth and vocabulary depth together constitute an important lexical basis for reading comprehension, and they are complementary in explaining reading performance [3]. Rabadi (2023) also believes that the reading comprehension level of English learners is not only affected by the number of words, but also continuously restricted by the degree of semantic mastery of words [4]. These studies show that the key problem in reading

* fiona0060@163.com
https://doi.org/10.65102/is2026714 Published: 30 April, 2026

comprehension is no longer limited to "how many words to know", but gradually turns to "how words are organized, how they are activated, and how they participate in discourse meaning construction". In this context, a re-examination of Chinese college students' English reading comprehension from the perspective of the construction of lexical semantic network can not only deepen the explanation of reading processing mechanism, but also help to respond to the current practical problem of the disconnect between vocabulary learning and comprehension in college English reading.

From the research status at home and abroad, on the one hand, the existing results continue to deepen the understanding of vocabulary factors, on the other hand, they also begin to introduce digital technology and intelligent tools to improve the effect of reading support. Wang and Zhang (2025) proposed through cross-lag analysis that vocabulary breadth, vocabulary depth and reading comprehension do not have a one-way influence, but a dynamic interaction relationship [10]. Pei et al. (2025) found that morphological knowledge can influence the English reading performance of Chinese college students through the mediating effect of vocabulary knowledge and grammar knowledge, which indicates that vocabulary processing in reading has a more complex hierarchical structure [11]. In terms of technology-supported reading, Zhang and Perez-Paredes (2024) studied Chinese EFL learners' mobile dictionary usage behavior in reading tasks and pointed out that digital tools would change the path of word meaning acquisition and the way of reading processing [5]. Jin and Liu (2024) showed that hybrid computer dynamic assessment can effectively diagnose and promote the development of reasoning reading in Chinese EFL classroom [6]. Zhang (2024) proposed that computer-aided interaction design such as highlighting and annotation would have a significant impact on EFL learners' reading comprehension [7]. Wang and Kabilan (2025) found that task-based reading activities supported by wechat can improve Chinese EFL learners' reading comprehension and learning engagement [9]. These studies provide an important basis for technology-enabled English reading. However, they still mainly focus on external tools, scaffolding forms and intervention effects, and lack a deeper structural explanation of how the semantic relationships between words are systematically modeled and further affected by reading comprehension.

In recent years, the development of artificial intelligence and semantic computing provides new research conditions for the visualization, structuralization and computability of lexical relations. Wei and Hu (2025) studied incidental vocabulary learning under the condition of "reading plus listening" in the context of Chinese EFL classroom, indicating that multimodal input helps to promote vocabulary accumulation and comprehension development [12]. Lin et al. (2025) pointed out that AI-supported pre-reading scaffold can not only improve English reading comprehension level, but also improve learning motivation and learning attitude [13]. Liu and Ma (2025) conducted a case study on vocabulary learning of low-level second language learners, and proposed that the teaching method based on customized GPT and corpus scaffold has strong pertinency [14]. Compared with general digital tools, semantic networks and knowledge graphs put more emphasis on the association organization between words and their cognitive representation, which provides a new entry point for explaining semantic processing in reading. Feng and Liu (2024), by comparing the lexical semantic network structure of L1 and L2, pointed out that there are identifiable differences in learners' vocabulary organization at the global and local levels [8]. Yang et al. (2025) extracted semantic network from students' thinking aloud data and found significant correlation between it and learning performance, indicating that semantic network can become an important representation tool for learning analysis [15]. From the perspective of instructional design, Liu (2025) proposed that the semantic association organization of English words based on knowledge graph is helpful to enhance the structured relationship between words [16]. Although these studies have provided methodological inspiration for the application of semantic networks and knowledge graphs in language learning, there are still few studies that directly apply lexical semantic network construction to the test of English reading comprehension of Chinese college students, especially the lack of a complete research framework linking linguistic interpretation, computer modeling and empirical analysis of reading.

Based on the above research background and current situation, this paper intends to focus on the English reading situation of Chinese college students, explore the role of lexical semantic network construction in reading comprehension, and embed computer technologies such as natural language processing, semantic similarity calculation and graph structure modeling into the research process. The aim of this paper is to reveal the influence mechanism of lexical semantic network on reading comprehension, explain how the semantic connection between words plays a role in word meaning inference, intersentence integration and discourse comprehension, and further investigate the applicability of this mechanism in Chinese college students' English reading. Aiming at this goal, this paper constructs a lexical semantic network analysis framework from the perspective of the relationship between vocabulary knowledge and reading comprehension, examines the relationship between network characteristics and reading performance combined with reading tasks, and discusses the application value of semantic network in college English reading instruction and intelligent reading support based on the empirical results. Compared with the existing research, the innovation of this paper is mainly reflected in three aspects. First, the research perspective is further advanced from the single index such as word breadth and depth to the relationship network level between words, so as to explain the problem of semantic organization in reading comprehension more systematically. Second, natural language processing, graph model and knowledge graph are introduced into English reading research to enhance the visualization and computability of lexical semantic relations. Thirdly, based on the Chinese college students, the role of lexical semantic network in the localization context is studied, which provides new research basis for the reform of college English reading teaching and the design of intelligent reading support tools.

2 Theoretical basis and analytical framework

2.1 The Inner Relation between Lexical Semantic Relations and English Reading Comprehension

English reading comprehension is not a mechanical recognition of the meaning of an isolated word, but a continuous process of information activation, sentential meaning integration, discourse construction and reasoning judgment based on the semantic relationship of words. Synonymy, near-synonymy, hypernymy, collocation, category affiliation and context connection among words form the basis of semantic network in reading processing. When reading, learners first activate relevant semantic nodes through word form recognition and existing word representation, and then complete local meaning discrimination by virtue of the association strength between words, and continuously adjust the scope of semantic interpretation in the process of inter-sentence cohesion and discourse advancement. If the semantic relationship of words is clear and tightly organized, learners are more likely to grasp the sentence logic, identify the anaphora relationship, infer the implicit information, and form a more complete representation of the text meaning. If there is no stable relationship between words, the reading process is easy to stay at the word level, which leads to fragmentation of understanding, obstacle of reasoning and deviation of grasping the main idea. It can be seen that lexical semantic relations not only affect the efficiency of word sense extraction, but also affect the depth of sentence comprehension and the quality of discourse integration. They are the core intermediary link connecting lexical knowledge and reading comprehension performance, and their action path is shown in Figure 1.

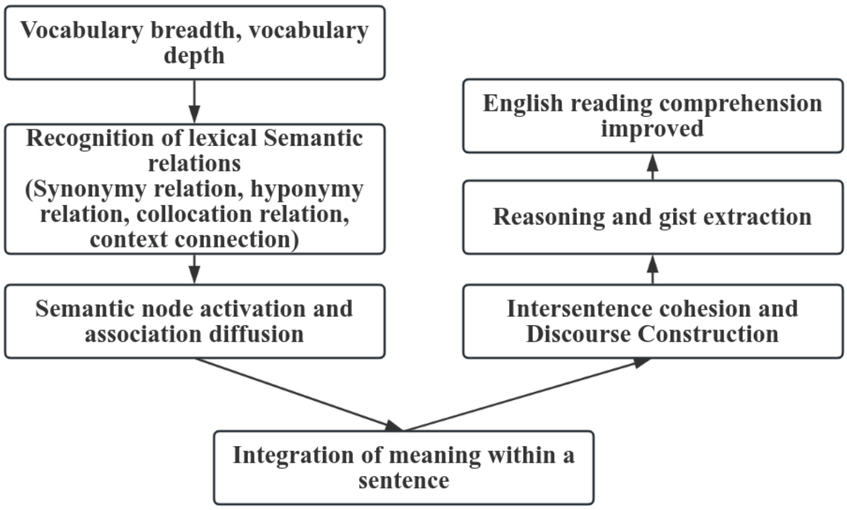

Figure 1: Schematic diagram of the intrinsic relationship between lexical semantic relations and English reading comprehension

2.2 Structural Features of Lexical Semantic Networks and Their cognitive Interpretation

The lexical semantic network is not a simple accumulation of lexical sets, but an organized structure formed by lexical nodes relying on semantic similarity, hyponymy subordination, collocation co-occurrence and situation correlation. Its structural characteristics are mainly shown as centrality, clustering, hierarchy and path connectivity. Centrality reflects the hub position of some high-frequency core words in the network, which usually assumes the function of semantic aggregation and information distribution, and is an important fulpoint for quickly locking the topic in reading. Clustering is reflected in the fact that words with similar meaning or function related tend to form local semantic clusters, which is helpful for learners to complete semantic classification and local integration in the reading process. The hierarchy is manifested by the expansion of lexical meaning along the direction of "upper concept-lower members-concrete context", which makes the discourse meaning construction have progressive characteristics from abstract to concrete and from core to edge. Path connectivity reflects the degree of reachability between different semantic clusters, which directly affects the efficiency of cross-sentence information linking and implicit relation inference.

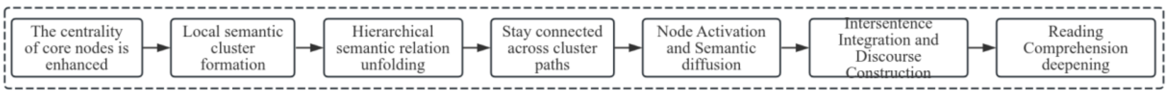

Figure 2: Schematic representation of lexical semantic network structural features and their cognitive interpretation

As shown in Figure 2, when the network structure presents high center aggregation, clear hierarchy and stable connectivity, learners can quickly complete node activation, relationship diffusion and discourse integration on the basis of word sense recognition, and then form a more complete discourse representation. On the contrary, if the network structure is loose, the connection is broken or the hierarchy is chaotic, the semantic processing is easy to stagnate at the local meaning level, resulting in insufficient information integration and deep understanding. Thus, the lexical semantic network is both a structural mapping of the organizational state of lexical knowledge and an important cognitive representation of the quality of semantic processing during reading comprehension.

2.3 Natural Language Processing and the Applicability of Graph Models for Semantic Network Construction

The construction of lexical semantic network is essentially to transform the lexical units, semantic clues and association patterns scattered in the text into computable and analyzable structured representations. In this process, natural language processing undertakes the basic functions of text cleaning, word recognition, relation extraction and semantic representation, which can transform the inflection, collocation co-occurrence, hyponymy cues and context correlation in the original reading materials into stable semantic relationship data. The graph model further organizes these relationships in the form of nodes and edges, so that the connection strength, cluster distribution, center position and path structure between words can be visualized. The combination of the two can not only improve the systematicity and operability of semantic network construction, but also transform the implicit semantic organization mode into measurable network characteristics.

Table 1: Applicability of NLP and graph models in semantic network construction

Technical link	Main functions	Role in the construction of semantic networks
Text preprocessing	Word segmentation, lemmatization, stop word processing	Unified vocabulary form to reduce noise interference
Vocabulary recognition	Keywords, part-of-speech and semantic units are extracted	Determine the basic composition of the network nodes
Relation extraction	Identify co-occurrence, collocation, hypernymy and contextual connections	Semantic connections between words are formed
Semantic computation	Calculate the similarity, association strength and weight	Quantify the strength of the relationship between nodes
Graph structure modeling	Construct node, edge and weight network	Present the overall semantic organization form
Network analysis	Centrality, clustering, connectivity and other metrics are calculated	Explain the structural features and reading meaning of lexical semantic network

As shown in Table 1, natural language processing focuses on solving the question "how semantic relations are recognized and computed", while graph models focus on answering the question "how semantic relations are organized and interpreted". For English reading research, this combination of techniques can not only preserve the linguistic properties of lexical semantic relations, but also reveal the potential influence of lexical network structure on reading comprehension, so it has strong method adaptability and analysis value.

2.4 Analysis framework and research hypothesis of lexical semantic network affecting reading comprehension

Based on the previous discussion on lexical semantic relations, semantic network structure and its computational representation, this paper regards lexical semantic network as an intermediary structure connecting lexical knowledge and reading comprehension performance, and constructs an analytical framework of "network structure feature-semantic processing-reading comprehension results". Specifically, the centrality, clustering, hierarchy and connectivity of the lexical semantic network reflect the closeness of the organization of words in the text, the way of meaning aggregation and the ability to reach across nodes. These structural properties are not directly equivalent to reading performance, but further affect reading comprehension performance by affecting the efficiency of word sense activation, intra-sentence meaning integration, inter-sentence relationship cohesion and discourse theme construction. The more prominent the core nodes in the network, the easier it is for learners to grasp the topic of the text. The clearer the local semantic cluster is, the more conducive it is to complete the relevant information classification and local inference. The more stable the hierarchical relationship is, the more it can support the progressive understanding from the surface meaning to the deep semantic. The stronger the cross-cluster connectivity is, the more helpful it is to complete the cross-sentence information integration and implicit relation recognition. Therefore, the lexical semantic network is not an external additional factor, but a structural reflection of the semantic organization ability in the process of reading comprehension.

On this basis, this paper proposes the following research hypothesis: H1, the structural characteristics of lexical semantic network have a significant impact on Chinese college students' English reading comprehension. The clearer the network organization and the more stable the connection, the better the reading comprehension performance. H2, the influence of lexical semantic network on reading comprehension is mainly realized through semantic processing such as word meaning inference, intersentence integration and discourse construction, among which semantic integration ability plays a key mediating role. H3, There are differences in the lexical semantic network structure of students with different English levels. The network of high-level learners is more centered aggregation, hierarchical clarity and path connectivity, so they show stronger overall grasp and deep inference ability in reading comprehension tasks. The above analytical framework and research hypothesis together constitute the theoretical basis for the subsequent network construction, variable measurement and empirical test of this paper.

3 Construction and research design of Lexical semantic Network

3.1 Research object, corpus source and data preprocessing

This study takes non-English major college students in Chinese universities as the research object. A total of 312 undergraduates from two comprehensive universities are selected as the initial sample by stratified cluster sampling method, covering three grades from freshman to junior. In order to ensure the homogeneity of the sample and the validity of the data, after eliminating the samples with missing examinations, abnormal answering time and incomplete questionnaire information, 286 valid samples were finally obtained, with an effective rate of 91.67%. Among them, 128 were boys and 158 were girls; 174 have passed CET-4 and 112 have failed. The corpus source is composed of two parts: teaching reading materials +

standardized test texts. One part is selected from college English textbooks and extended reading materials, with a total of 24 articles. The other part is selected from CET-4 simulation reading and Academic English training texts, a total of 16 texts, a total of 40 texts. The total number of words in the original corpus is 21864, and the total number of word forms is 4872. After lemmatization and function word cleaning, 2316 content words are retained to form the basic corpus for subsequent semantic network construction.

Data preprocessing mainly includes five steps: text cleaning, word segmentation and tagging, lemmatization, stop word filtering and low-frequency word elimination. Firstly, the title number, special symbols, repeated marks and characters without semantic payload were removed to reduce the noise interference. Then, the text is segmented and part-of-speech tagged, and only core semantic units such as nouns, verbs, adjectives and adverbs are retained. Considering the influence of different text lengths on word frequency statistics, the word frequency is standardized. Suppose the original occurrence frequency of a word wi in the whole corpus is fi, and the total word frequency is N, then its standardized word frequency is defined as follows.

$$F_i=\frac{f_i}{N}\times 10^4 \tag{1}$$

Here, Fi represents the normalized occurrence frequency per ten thousand words. In order to avoid the interference of extremely low frequency words on network sparsity, suppose that the total number of texts is D, and the word wi appears in di texts, then the document distribution rate is denoted as:

$$P_i=\frac{d_i}{D} \tag{2}$$

In this study, the retention conditions are Fi≥2 and Pi≥0.05, which means that the term appears at least twice in every ten thousand words and is distributed in at least 5% of the texts. After this process, the corpus is compressed from the original 4872 word forms to 2316 effective word units, and the data redundancy rate is 52.46%. On this basis, the corresponding structure of "students' reading performance data-text corpus data-word semantic relationship data" is further constructed, which lays a data foundation for the subsequent generation of word semantic network, network index extraction and reading comprehension test.

3.2 Lexical Relation Extraction and Semantic Similarity Computation

After the corpus cleaning and the word element selection, we further extract the relationship between effective words, and calculate the semantic similarity according to this, so as to form the edge weight basis of the lexical semantic network. Considering that the semantic connection in English reading is reflected in both the local co-occurrence of words and the overall context distribution, this paper uses the method of "co-occurrence relation extraction + distributed semantic computation + weighted fusion" to construct the inter-word relationship. Based on the corpus processed in Section 3.1, 2316 effective lemmas are finally entered into relation calculation, covering 40 reading texts, and the total number of sentences is 1684. In order to balance the local context correlation and computational stability, the sentence is used as the basic segmentation unit, and the sliding window length is set to 5. After preliminary scanning, 48271 candidate word pairs were obtained. After eliminating the accidental co-occurrence relationships that only appear once, 12684 high-confidence word pairs are retained, accounting for 26.28% of the initial candidate relationships.

Lexical relation extraction relies first on co-occurrence statistics. Suppose that the terms wi and wj appear together in window k, then a co-occurrence event is recorded. In the whole window, its co-occurrence frequency is denoted as cij, then:

$$c_{ij}=\sum_{k=1}^{K} I_{ij}^{(k)} \tag{3}$$

where $I_{ij}^{(k)}=1$ means that wi and wj co-occur in the KTH window, otherwise it is 0. K is the total number of Windows. The total number of Windows generated in this study is 5248. If only raw co-occurrence frequency is used, it is easy to overestimate the surface connection between high-frequency words. Therefore, conditional probability is further introduced to describe the directional association strength. Let the occurrence frequency of term wi be fi, then the conditional probability of wj given the occurrence of wi is defined as follows.

$$P(w_j \mid w_i)=\frac{c_{ij}}{f_i} \tag{4}$$

This index is able to reflect the local tendency to activate w_j by w I, but is still affected by the high-frequency background distribution. In order to weaken the frequency bias of common terms, we further use point mutual information to measure the significant co-occurrence degree of word pairs:

$$PMI(w_i,w_j)=\log\frac{P(w_i,w_j)}{P(w_i)P(w_j)} \tag{5}$$

Among them,

$$P(w_i,w_j)=\frac{c_{ij}}{\sum_i \sum_j c_{ij}}, \quad P(w_i)=\frac{f_i}{N}, \quad P(w_j)=\frac{f_j}{N} \tag{6}$$

N is the total number of words in the corpus, which is 21864. Considering that low frequency word pairs are easy to produce inflated PMI values, this paper uses positive point mutual information to correct them:

$$PPMI(w_i,w_j)=\max\{PMI(w_i,w_j),0\} \tag{7}$$

The results show that 9347 (73.69%) of the 12684 valid word pairs have PPMI greater than 0, which indicates that stable local semantic relations can be formed in the corpus.

Only relying on co-occurrence statistics is still not enough to fully reflect the semantic proximity, so we further calculate the word vector similarity from the perspective of distributed semantics. Firstly, the vector representation of each term is extracted based on the pre-trained English word embedding model, and the embedding vector of term w_i is denoted as $v_i \in R^d$ with dimension d=300. The cosine similarity between any two terms is defined as:

$$Sim_{cos}(w_i,w_j)=\frac{v_i \cdot v_j}{\|v_i\|\|v_j\|} \tag{8}$$

This index mainly reflects the proximity of terms in the overall context distribution. After calculating all valid word pairs, the mean, median and standard deviation of cosine similarity are 0.284, 0.251 and 0.117, respectively, indicating that the semantic proximity between different word pairs is significantly differentiated. In order to unify the co-occurrence strength and distributed semantics into the same dimension, PPMI and cosine similarity are respectively standardized by range:

$$\widetilde{\mathrm{PPMI}}_{ij}=\frac{\mathrm{PPMI}_{ij}\text{-}\mathrm{PMI}_{min}}{\mathrm{PMI}_{max}\text{-}\mathrm{PMI}_{min}} \tag{9}$$

$$\widetilde{\mathrm{Sim}}_{ij}=\frac{\mathrm{Sim}_{cos}(w_i,w_j)\text{-}\mathrm{Sim}_{min}}{\mathrm{Sim}_{max}\text{-}\mathrm{Sim}_{min}} \tag{10}$$

On this basis, the comprehensive semantic relation weights are constructed as follows.

$$W_{ij}=\alpha\,\widetilde{\mathrm{PPMI}}_{ij}+(1\text{-}\alpha)\,\widetilde{\mathrm{Sim}}_{ij} \tag{11}$$

Here, α is the weight coefficient of local co-occurrence information. Considering that reading semantic processing depends on both local context and relatively stable semantic neighborhood relations, this paper sets α=0.6, which moderately increases the contribution of co-occurrence relations in the comprehensive weight. The calculation results show that the mean value of the comprehensive weight of all valid word pairs is 0.347, and the standard deviation is 0.146.

In order to avoid the network too thin or too dense caused by too many weak links, it is necessary to filter the comprehensive weight. In this paper, the comprehensive weight threshold τ is used to determine the edge retention:

$$A_{ij}=\begin{cases}1, & W_{ij}\geq\tau\\ 0, & W_{ij}<\tau\end{cases} \tag{12}$$

Here, A_{ij} indicates whether a connection is established between the terms w_i and w_j. Through multiple rounds of trial calculation, it is found that when τ=0.35, the network density and interpretability are relatively balanced. At this time, 4876 relationship edges are retained, and each node connects 4.21 edges on average, and the network density is 0.00182. Furthermore, in order to reflect the relative strength difference between nodes, this paper does not adopt pure binary network, but forms a weighted adjacency matrix G=[gij] on the basis of retaining connections, where:

$$g_{ij}=A_{ij}\cdot W_{ij} \tag{13}$$

The obtained matrix not only retains the structural information of "whether connected", but also contains the degree information of "how strongly connected", which can provide the basis for the subsequent calculation of network indicators such as centrality, clustering coefficient and path length.

In general, lexical relation extraction and semantic similarity calculation are not simply technical splicing, but the local collocation connections, intra-sentence co-occurrence patterns and global semantic distribution in the reading text are integrated into the analysis framework. The former ensures that the relation extraction can be close to the specific reading context, and the latter makes the lexical connection have more stable semantic interpretation power beyond a single text fragment. The weighted relation network formed by the integration of the two can more truly reflect the structural characteristics of the lexical semantic organization in the reading materials. It also lays a methodological foundation for the subsequent construction of lexical semantic network and its correlation analysis with reading comprehension performance.

3.3 Construction method of word semantic network based on computer technology

After the lexical relation extraction and semantic similarity calculation, we further construct a lexical semantic network by using the graph computing method to realize the structured expression of lexical nodes, relation edges and their weights. The process is not a simple superposition of word-pair relations, but integrates word co-occurrence information, distributed semantic information and network topology constraints into the same modeling framework, so that the semantic relations are transformed from discrete statistical results into computable and interpretable network structures. Specifically, let the set of valid terms obtained after filtering be:

$$V=\{v_1,v_2,\ldots,v_n\} \tag{14}$$

Here, n=2316 represents the total number of network nodes. If there is a stable semantic connection between any two lexical nodes vi and v_j, an edge e_{ij} is established, and the lexical semantic network can be expressed as follows.

$$G=(V,E,W) \tag{15}$$

Here, E is the set of edges and W is the set of edge weights. If there is no effective relation between words, then $e_{ij}=0$; If there is a connection, the edge weight is determined by the comprehensive semantic relation weight W_{ij} calculated above.

To ensure that semantic information from different sources can be represented collaboratively in the same network, the original relation matrix is weighted and fused. Suppose that the relationship matrix obtained based on co-occurrence statistics is $C=[c_{ij}]$, and the relationship matrix obtained based on semantic similarity is S=[sij], then the initial semantic weight matrix after fusion can be written as follows.

$$M=\lambda C+(1-\lambda)S \tag{16}$$

Here, $\lambda\in[0,1]$ is the fusion parameter. Combined with the previous setting, this paper takes $\lambda=0.6$ to highlight the fundamental role of local context relations in reading texts in network construction. Since the value interval of different elements in matrix M is not completely consistent, it needs to be further normalized. Let the maximum and minimum values of elements in the matrix be Mmax and Mmin respectively, then the normalized weights are defined as follows.

$$m_{ij}'=\frac{m_{ij}-M_{min}}{M_{max}-M_{min}} \tag{17}$$

After normalization, all edge weights are mapped to the interval [0,1], which is conducive to the subsequent threshold screening and network sparsity control.

Considering that it is easy to cause structural redundancy and interpretation deviation if the lexical semantic network retains all weak links, this paper adopts the double constraint method to realize network pruning. The first step is to filter the global threshold, that is, keep only those values that satisfy:

$$m_{ij}'\geq\theta \tag{18}$$

The relation edge of, where θ=0.35. The second step is the local nearest neighbor constraint, which only keeps the top k neighbor nodes with the highest weight for each node to prevent a few high-frequency words from forming excessive dense connections. Let the set of candidate neighbors of node vi be N(vi), then its reserved adjacency set is denoted as follows.

$$N_k(v_i)=\mathrm{TopK}\{m_{ij}' \mid v_j \in N(v_i)\} \tag{19}$$

Here, k=8 is taken in this paper. Under the double constraints, 4876 valid edges were finally retained to form a more stable sparse weighted network. The whole construction process is shown in Figure 3.

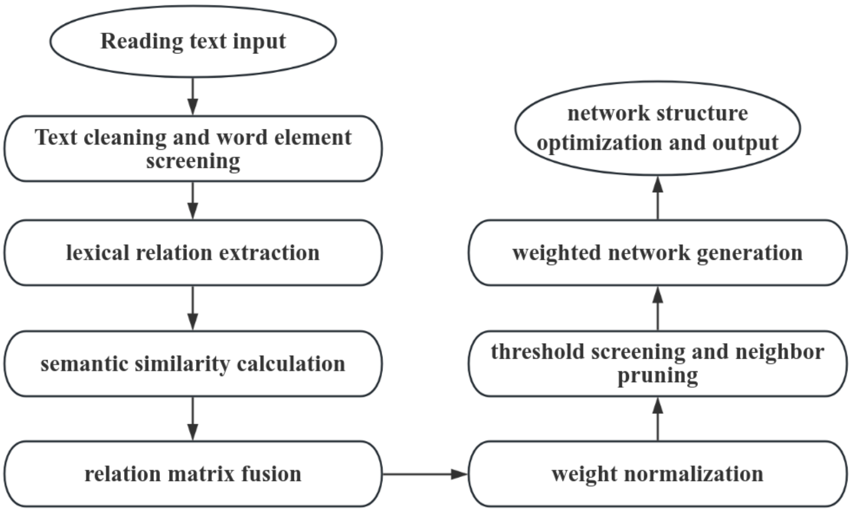

Figure 3: Flowchart of lexical semantic network construction based on computer technology

As shown in Figure 3, the generation of lexical semantic network is not a single algorithm step, but is completed by four levels of "text processing - relation calculation - structural constraints - network output". In order to eliminate the bias caused by directional statistics, the network matrix is further symmetrized. Let the initial adjacency matrix be A=[aij]. If there is a connection between node vi and vj, then:

$$a_{ij}=\begin{cases} m_{ij}', & m_{ij}' \geq \theta \text{ 且 } v_j \in N_k(v_i) \\ 0, & \text{Others} \end{cases} \tag{20}$$

Since the lexical semantic relations are more suitable to be characterized as undirected relations on the whole, the final weighted adjacency matrix is obtained by using the average symmetrization method:

$$a_{ij}^{*}=\frac{a_{ij}+a_{ji}}{2} \tag{21}$$

The final network is thus constructed as follows.

$$A^{*}=[a_{ij}^{*}]_{n\times n} \tag{22}$$

Here, A* is a symmetric weighting matrix, which can reflect whether the nodes are connected and the connection strength.

After the network is generated, in order to enhance the structure interpretability, we further introduce node weight and local density constraints to optimize the network topology. Firstly, the weighted strength of each node is calculated to describe the semantic connection ability of a word in the whole network. Let the weighted strength of node v_i be s_i, then:

$$s_i=\sum_{j=1}^{n} a_{ij}^{*} \tag{23}$$

The larger the index is, the richer the connections of the word in the semantic network and the stronger the semantic radiation ability. Second, calculate the node degree value:

$$d_i=\sum_{j=1}^{n} I\,(a_{ij}^{*}>0) \tag{24}$$

where $I(\cdot)$ is the indicative function. This value is 1 if $a_{ij}^{*}>0$, and 0 otherwise. The node degree value is used to measure the number of direct connections a certain term has established with other terms. In order to avoid topological dominance of the whole network caused by a small number of high-frequency nodes, this paper performs logarithmic smoothing of node strength:

$$\hat{s}_i=\log\,(1+s_i) \tag{25}$$

After smoothing, the weight differences of high-strength nodes are moderately compressed, and the network structure is more stable.

In addition, in order to preserve the local semantic clusters without losing the overall connectivity of the network, we introduce a weighted density constraint in the topology optimization stage. Let the weighted density of the network be ρw, then:

$$\rho_w=\frac{\sum_{i\neq j} a_{ij}^{*}}{n(n-1)} \tag{26}$$

This metric reflects the weight level retained on average between unit node pairs. According to the calculation, the weighted density of the word semantic network constructed in this study is 0.00182, indicating that the overall network is relatively sparse, but it still retains effective connections sufficient to support semantic diffusion and structure identification. In order to ensure that the network does not have too many isolated nodes, this paper further statistics the proportion of isolated nodes. If a node satisfies di=0, it is regarded as an isolated node. The proportion of isolated nodes is defined as follows.

$$R_{iso}=\frac{N_{iso}}{n} \tag{27}$$

where, N_{iso} is the number of isolated nodes. After processing, the proportion of isolated nodes is controlled within 4.79%, which indicates that the network has good overall coverage while maintaining sparsity.

At the network expression level, this paper finally represents each word as a two-layer structure of "node attribute + edge weight relationship". The node attribute vector can be written as follows.

$$x_i=(f_i,\hat{s}_i,d_i,p_i) \tag{28}$$

where f_i is the normalized term frequency, $\hat{s}_i$ is the smoothed weighted strength, d_i is the node degree value, and p_i is the document distribution rate of the term in the text. The vector is used to preserve the basic statistical characteristics of words in the corpus, and the edge weight matrix A* is responsible for characterizing the relationship structure between words. With the combination of node attributes and network topology, the lexical semantic network is no longer just a static relationship graph, but a composite representation system with corpus information, semantic information and structural information.

In order to facilitate the subsequent network analysis and visualization output, this paper uses the weighted undirected graph form for the final storage of the network, and uses G*=(V,E*,W*,X) to represent its complete structure, where E* is the effective edge set after screening,W* is the final edge weight set,X ={x1,x2,... ,xn} is the node attribute set. The lexical semantic network constructed by this method can not only intuitively display the overall pattern of lexical semantic organization in the reading text, but also provide a unified data basis for the subsequent calculation of indicators such as centrality, clustering coefficient, path length and community division.

In general, the construction method of word semantic network based on computer technology is essentially a process of transforming the word relationship in reading text from "linear distribution" to "graph structure organization". In the technical path, this method comprehensively utilizes the relationship recognition ability of natural language processing, the weight modeling ability of semantic computing and the structural expression ability of graph model, so as to make the implicit semantic links between words explicit and convert them into measurable, comparable and interpretable network representations. This construction method can not only reflect the organization of words in reading materials more truly, but also provide a reliable method support for the subsequent investigation of the relationship between word semantic network and reading comprehension.

3.4 Web Metrics Design and Reading Comprehension Measurement

After completing the construction of lexical semantic network, the research focus is no longer just to describe whether the network exists, but to further identify the differences in network structure and their correspondence with reading comprehension performance. Based on this, this paper designs from two levels: network structure representation and reading comprehension measurement. The former is used to describe the organization state of word semantic network, and the latter is used to quantify students 'comprehension level in reading tasks, thus providing a unified data basis for subsequent correlation analysis, regression test and difference comparison.

In terms of network index design, this paper mainly selects core indicators that can reflect the importance of nodes, the degree of local aggregation and the overall level of connectivity. Firstly, the node degree is used to measure the number of direct connections a certain term has established with other terms. If a node is connected with more words, it means that the word has stronger connection ability in the semantic network. The degree value of node vi is defined as follows.

$$d_i = \sum_{j=1}^{n} I\,(a_{ij}>0) \tag{29}$$

Here, $I(a_{ij}>0)$ means 1 if there is a connection between node vi and v_j, and 0 otherwise. Corresponding to the node degree, the node strength further considers the edge weight difference to reflect the overall strength of the semantic connection of a word, which is defined as:

$$s_i = \sum_{j=1}^{n} a_{ij} \tag{30}$$

Node degree emphasizes "connection quantity", node strength emphasizes "connection quality", and the combination of the two can more accurately identify the core words in the network.

Secondly, in order to investigate whether words tend to form local semantic clusters, this paper introduces the clustering coefficient index. If there are also strong connections between adjacent nodes of a word, it means that the region of the word has high semantic aggregation, which is more conducive to local semantic integration. The clustering coefficient of node vi can be expressed as follows.

$$C_i = \frac{2e_i}{d_i(d_i-1)} \tag{31}$$

Here, e_i represents the number of connections that actually exist between the neighboring nodes of node v_i. Furthermore, the clustering coefficients of all nodes were averaged to obtain the overall clustering level of the network. The higher the clustering coefficient, the easier it is for words to form stable semantic clusters around a certain topic, which plays an important role in intra-sentence meaning integration and local inference.

Thirdly, in order to characterize the overall semantic transfer efficiency of the network, this paper uses the average path length index. This metric reflects the average number of steps required to connect any two lexical nodes by the shortest semantic path, which is defined as follows.

$$L = \frac{1}{n(n-1)} \sum_{i \neq j} l_{ij} \tag{32}$$

Here, lij denotes the shortest path length between node vi and vj. The shorter the average path length, the stronger the reachability between different semantic clusters, the higher the efficiency of information diffusion in the network, and the easier it is for learners to complete cross-sentence integration and implicit relation inference in reading.

In addition to the above indicators, network density is used to describe the overall connectivity tightness in this paper. The network density can be expressed as follows.

$$D = \frac{2m}{n(n-1)} \tag{33}$$

Here, m is the actual number of edges and n is the number of nodes. This index can reflect the overall distribution density of lexical semantic relations, but too high density may mask the real structural differences, so this paper uses it in combination with clustering coefficient and path length, rather than as a judgment basis alone.

For the measurement of reading comprehension, this paper sets up four categories of items according to the research objectives: word meaning comprehension, intersentence integration, discourse gist grasp and reasoning judgment, so as to ensure that the measurement results can correspond to the main links that the lexical semantic network may play. The reading test consists of 4 texts and 20 questions, with a total score of 100, including 20 points for word meaning comprehension, 25 points for intersentence integration, 25 points for discourse theme, and 30 points for reasoning and judgment. The overall reading comprehension score of each student is recorded as:

$$R_i=\sum_{k=1}^{4} r_{ik} \tag{34}$$

Here, r_{ik} represents the score of the ith student in the KTH category reading task. In order to facilitate subsequent comparison of scores of different dimensions and model estimation, this paper further standardizes the original scores:

$$Z_i=\frac{R_i-\bar{R}}{SD_R} \tag{35}$$

Among them, $\bar{R}$ is the average score of the sample and SDR is the standard deviation of the sample. After standardization, the reading performance of different students can be unified under the same dimension, so as to facilitate correspondence analysis with network indicators.

In order to improve the stability of the measurement results, this paper pre-tested the items before the formal test, and adjusted the individual items according to the discrimination and difficulty coefficient. The final reading comprehension measurement tool can not only reflect students' mastery of explicit word meaning, but also reflect their integration and inference ability at the discourse level, which is theoretically consistent with the lexical semantic network constructed in the previous paper. Therefore, the network index is responsible for revealing the semantic organization of words, and the reading comprehension measurement is responsible for presenting learners 'actual comprehension results. The combination of the two measures can systematically examine the mechanism of lexical semantic network on Chinese college students 'English reading comprehension.

3.5 Experimental process and data analysis method

The experimental process of this study focuses on "corpus processing-network construction-reading measurement-statistical test", and matches the lexical semantic network features with students 'reading comprehension performance one by one. Before the formal experiment, 40 reading texts were screened, cleaned and normalized according to the above standards, and then the lexical relations were extracted and a weighted semantic network was generated. Then 286 effective subjects were selected to participate in the reading test. The test material consisted of 4 reading texts and 20 questions, covering four dimensions of word meaning understanding, intersentence integration, text theme grasp and reasoning judgment. In order to reduce the interference of order effect and fatigue effect, the test was implemented according to a unified length of time, a unified instruction language and a unified answering environment. After the completion of the reading task, the basic information and English level data of the students were collected synchronously, and the individual reading scores were correlated with the semantic network features of the corresponding texts to form a paired database of "subject-text-network indexes-reading scores". The whole experimental process is as described in the previous section, which not only ensures the correspondence between corpus processing and reading measurement, but also provides a structured data basis for subsequent effect tests.

For data analysis, this paper uses a combination of descriptive statistics, reliability tests, correlation analysis, difference tests and regression analysis. Descriptive statistics is mainly used to present the mean, standard deviation, maximum and minimum values of reading scores and network indicators, so as to grasp the overall distribution characteristics of the sample. The reliability test mainly examines the internal consistency of the reading test to ensure the stability of the measurement results. To test the linear relationship between lexical semantic network and reading comprehension, Pearson correlation coefficient was used for preliminary analysis, which was expressed as follows.

$$r=\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2\sum_{i=1}^{n}(Y_i-\bar{Y})^2}} \quad (36)$$

Here, X_i denotes the observed values of network metrics and Y_i denotes the reading comprehension score. If the correlation reached a significant level, a regression model was further constructed to investigate the predictive effect of network indicators on reading comprehension. In this paper, the total score of reading comprehension is taken as the dependent variable, the average node degree, the average node strength, the clustering coefficient, the average path length and the network density are taken as the independent variables, and the English level, grade and CET-4 passing are taken as the control variables. The model form is as follows:

$$R_i=\beta_0+\beta_1Deg_i+\beta_2Str_i+\beta_3Clu_i+\beta_4Path_i+\beta_5Den_i+\beta_6C_i+\varepsilon_i \quad (37)$$

Here, R_i is the reading comprehension score of the ith student, C_i is the set of control variables, and ε_i is the random error term. The model was used to identify the independent explanatory effects of different network features on reading performance.

In order to further test the differences in the role of lexical semantic network among students with different English proficiency, this paper divided the sample into a high level group and a low level group according to the quartile of the total score of reading comprehension, and used the independent sample ttt test to compare the differences in the core network indicators between the two groups. When multiple dimension scores need to be compared simultaneously, a supplementary test combined with one-way ANOVA is performed. Considering that the research hypothesis of this paper involves the process chain of "word meaning inference, intersentence integration and discourse construction", hierarchical regression will be used to investigate the mediating effect of semantic processing dimension between network features and reading comprehension. After controlling the basic variables, network indicators and procedural reading dimensions will be gradually added to compare the changes in the interpretation rate of the model. To determine whether the influence of lexical semantic network on reading comprehension is realized through specific processing links.

In terms of technical implementation, text preprocessing, lexical relation extraction, semantic similarity calculation and network index extraction are mainly completed in Python environment, visual expression of semantic network and local structure recognition are realized by graph analysis tools, and statistical analysis is carried out in SPSS and regression analysis environment. By integrating natural language processing, graph modeling and statistical testing methods into the same research flow, this paper can not only systematically present the structural characteristics of lexical semantic network, but also examine its actual effect on Chinese college students 'English reading comprehension at the data level, thus forming a complete analysis chain from network construction to effect verification.

4 Testing the effect of lexical semantic network on college students' English reading comprehension

4.1 Overall structural features of lexical semantic networks

The word semantic network as a whole shows obvious characteristics of "sparse connection, local aggregation, and core prominence". The final network included in the analysis contains 2316 nodes and 4876 edges, and the network density is 0.00182, indicating that words are not generally directly connected with each other, but are mainly connected around words with strong semantic correlation. The average degree is 4.21, and the average weighted strength is 1.46, indicating that most words have some semantic connection ability. The average clustering coefficient is 0.312, indicating that a relatively stable local semantic cluster has been formed in the network, which is conducive to the integration of local meaning in reading. The average path length is 5.27, and the proportion of the largest connected subgraph is 88.51%, which indicates that most words can be indirectly connected through short paths, and the overall semantic reachability is good. At the same time, the modularity is 0.421, which indicates that the network has certain characteristics of community differentiation, and can clearly reflect the organization state of word semantic relations in reading texts. On the whole, the network has both structural stability and semantic discrimination, which lays a foundation for the subsequent test of its role in reading comprehension. The overall structural characteristics of the lexical semantic network are shown in Table 2.

Table 2: Overall structural characteristics of the lexical semantic network

Index Name	Numerical value
Number of nodes	2316
Number of edges	4876
Network density	0.00182
Average degree	4.21
Average weighted intensity	1.46
Average clustering coefficient	0.312
Average path length	5.27
Maximum proportion of connected subgraphs	88.51%
modularity	0.421

4.2 The Effect of Lexical Semantic Network on reading comprehension performance

After the overall structural features have been confirmed, this paper further examines the influence of lexical semantic network on college students' English reading comprehension performance. In order to improve the interpretability of the results, according to the comprehensive level of the three core indicators of node degree, clustering coefficient and connectivity, the corresponding texts of the sample are divided into three levels: low, medium and high, and the changes of students 'reading comprehension scores under different levels are compared. The results show that with the improvement of the lexical semantic network structure, the total score of students 'reading comprehension shows a continuous upward trend. Under the condition of low-level semantic network, the average reading scores of students concentrated between 63.4-68.7. In the medium level condition, the score increased to between 67.2 and 74.8. In the high level condition, the score further rises to between 72.6 and 82.3. On the whole, the clearer the network structure, the more stable the local aggregation, and the more smooth the cross-node connectivity, the better the performance of students in the reading task.

From the perspective of different dimensions, connectivity had the most obvious improvement on reading performance. The average score of the high level connectivity condition was 82.3, which was significantly higher than that of the low level condition (68.7), with a difference of 13.6 points. The score corresponding to clustering coefficient improved from 66.1 to 77.9. The influence of node degree is relatively stable, but also shows a significant positive change, from 63.4 points to 72.6 points. The results show that the lexical semantic network not only affects local word sense recognition, but also further affects the overall reading comprehension performance by enhancing the ability of cross-sentence information cohesion and discourse semantic integration. The results of regression analysis also showed that node degree, clustering coefficient and connectivity had positive predictive effects on the total score of reading, and connectivity had the highest explanatory power. This indicates that whether a more stable and accessible semantic relationship can be formed between words is an important factor affecting the reading comprehension performance of college students in the process of English reading. The relevant results are shown in Figure 4.

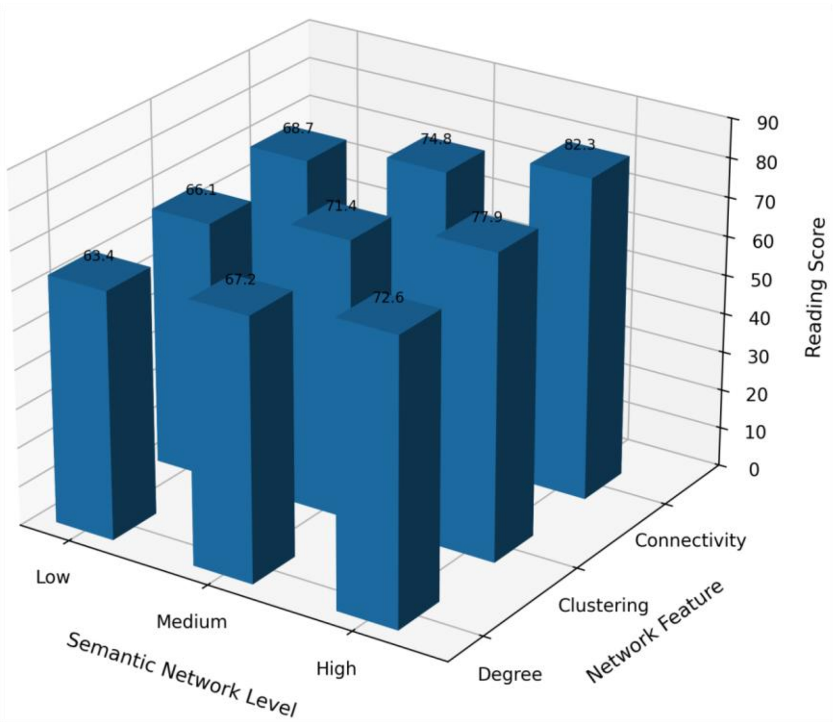

Figure 4: Effects of lexical semantic network features on reading comprehension performance

4.3 The Influence of Lexical Semantic Network on Word Sense Inference and discourse Integration

In the specific processing of reading comprehension, word sense inference and discourse integration are the two key dimensions that best reflect the role of lexical semantic network.

The former mainly reflects learners' ability to activate and filter word meanings according to context cues, while the latter reflects learners' integration level of cross-sentence information, paragraph logic and overall meaning relations. Based on the lexical semantic network constructed in the previous paper, this paper further compares the performance of students in these two dimensions under the conditions of different network levels. The results show that with the progress of the lexical semantic network from low level to high level, the scores of students in word sense inference and discourse integration show a continuous upward trend, and the improvement of discourse integration is slightly higher than that of word sense inference. As shown in Figure 5, under the condition of low-level semantic network, the average scores of word sense inference and discourse integration are 61.8 and 64.2 respectively. In the medium level condition, the scores of the two items increased to 69.5 and 72.8 respectively. In the high level condition, the score further increased to 78.4 and 81.6. It can be seen that the lexical semantic network can not only promote local word sense discrimination, but also improve learners 'ability to grasp the context relationship and discourse logic by enhancing the connection strength and cross-node accessibility between words.

From the perspective of the change trend, the total increase of word sense inference is 16.6 points, the total increase of discourse integration is 17.4 points, and the latter is slightly higher by 0.8 points, indicating that the lexical semantic network has a more obvious role in promoting high-level discourse processing. This result shows that when there is a clear semantic clustering, a stable hierarchical relationship and a smooth cross-cluster path between words, learners can complete the transition from local word meaning to overall discourse meaning faster, and then improve the efficiency of implicit information recognition and topic integration. On the contrary, if the network structure is loose and the node connection is weak, although learners can complete part of word meaning judgment with the help of local clues, they are more likely to break in cross-sentence cohesion and overall construction. In general, the lexical semantic network has a significant effect on word sense inference and discourse integration, and the effect on discourse integration is stronger, which further indicates that the lexical semantic network is not only involved in the lexical level, but deeply involved in the multi-level semantic processing in reading comprehension.

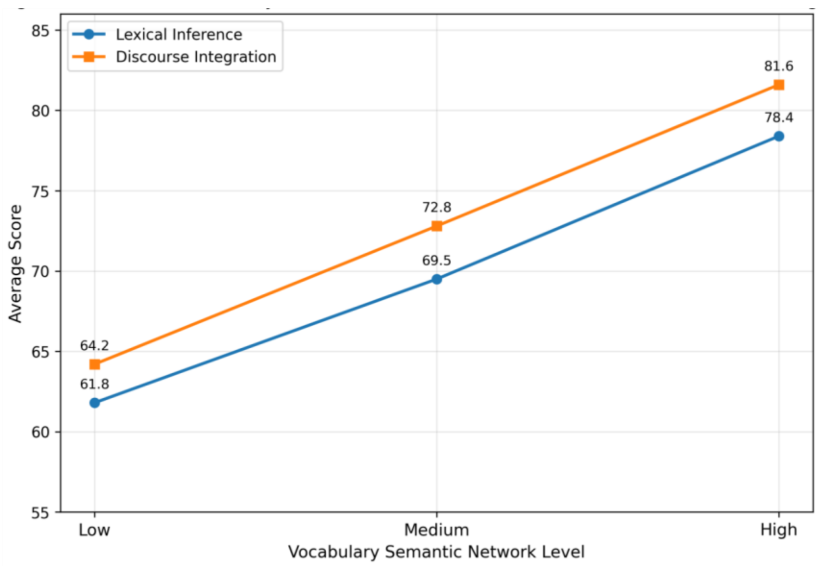

Figure 5: Influence of lexical semantic network on word sense inference and discourse integration

4.4 The differential performance of students with different English proficiency

In order to further test the applicability of lexical semantic network in different groups of students, this paper divided the sample into low, medium and high English proficiency groups according to the total score of reading comprehension and the English proficiency test results, and compared their performance differences in the total score of reading comprehension, word sense inference and discourse integration. The results show that students with different English levels show a relatively stable gradient distribution in all indicators. The high level group is always higher than the middle level group, and the middle level group is significantly higher than the low level group. As shown in Figure 6, the mean total score of reading comprehension in the low-level group was 66.3, word sense inference was 62.7, and discourse integration was 64.1. The three scores of the middle level group increased to 74.2, 70.6 and 72.5 respectively; The high level group further increased to 82.1, 78.9 and 81.4. It can be seen that with the improvement of English level, students 'performance in local word sense discrimination and overall discourse construction is enhanced synchronously.

In terms of specific differences, the difference between the low level group and the high level group in the total score of reading comprehension was 15.8 points, the difference in word meaning inference was 16.2 points, and the difference in discourse integration was 17.3 points, indicating that the difference in English proficiency was not only reflected in the change of vocabulary recognition ability, but also reflected in the continuous expansion of high-level discourse processing ability. Among them, the gap between the groups is the largest in the discourse integration dimension, indicating that high-level learners can better play the supporting role of lexical semantic network in cross-sentence information cohesion, topic extraction and implicit relation recognition. In contrast, although low-level learners can complete part of word sense judgment with the help of local semantic connections, they still face great difficulties in transforming discrete lexical information into coherent discourse meaning. In general, lexical semantic network has a promotion effect on students with different English levels, but its effect will be more obvious with the improvement of English level, which indicates that the function of lexical semantic network in reading comprehension has certain hierarchical adaptability and difference.

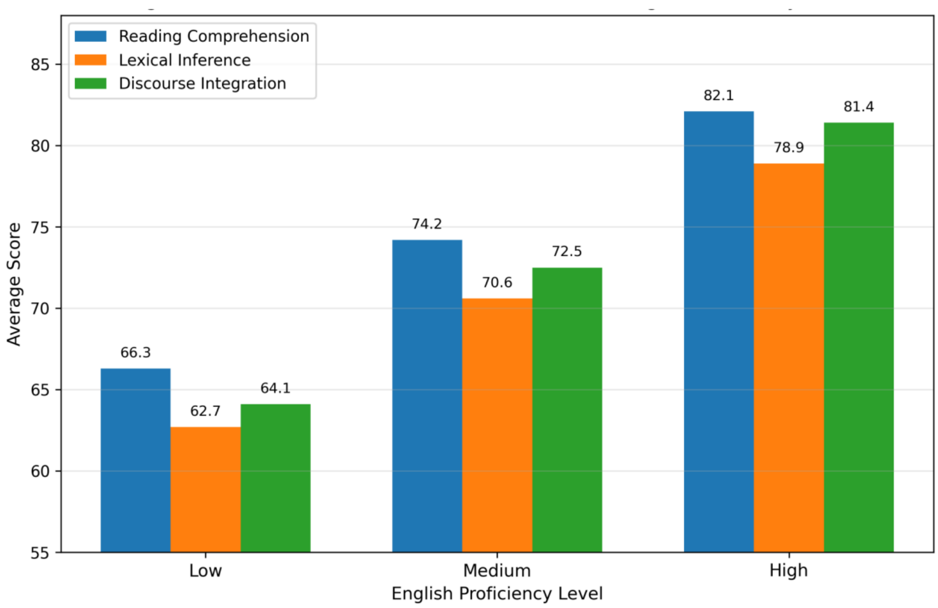

Figure 6: Differential performance of students with different English levels

4.5 Discussion of results and explanation of action mechanism

These results indicate that lexical semantic network plays a stable role in promoting college students' English reading comprehension, and this role does not stop at the level of word recognition, but further extends to higher-level processing such as word meaning inference, inter-sentence cohesion and discourse integration. The overall structure of the network shows that the lexical relationships in the reading corpus have obvious local clustering and cross-node connectivity, which indicates that words participate in the construction of meaning in a related and structured way in the discourse. Therefore, when the central node in the semantic network is more prominent, the local semantic cluster is clearer, and the cross-cluster path is more smooth, it is easier for learners to quickly enter the semantic field from the keywords, and complete the cross-sentence integration and topic extraction on the basis of local understanding.

From the perspective of mechanism, the influence of lexical semantic network on reading comprehension is mainly realized through two paths. One path is reflected in the improvement of semantic activation efficiency, that is, the strong association between words can shorten the time of word sense retrieval and matching, and help learners to complete the word sense confirmation in the context faster. The other path is reflected in the enhancement of discourse construction ability, that is, stable semantic connections help integrate scattered information into coherent meaning chains, thereby improving reasoning judgment and overall understanding quality. The difference results of students with different English levels further show that the facilitation of lexical semantic network has a clear hierarchical effect. High-level learners can more effectively use the semantic clues and structural relationships in the network to complete deep processing, while low-level learners rely more on local word meanings and surface clues, and have not yet fully utilized the supporting role of semantic network in discourse understanding. It can be seen that the lexical semantic network is not only an explicit representation of the organizational state of lexical knowledge, but also an important internal mechanism affecting the quality of English reading comprehension.

5 Conclusions and Implications

The results show that lexical semantic network can significantly promote Chinese college students' English reading comprehension. The semantic network constructed based on 286 college students, 40 reading texts and 2316 effective lemma shows that the network contains 4876 effective edges, the average clustering coefficient is 0.312, the average path length is 5.27, and the maximum connected subgraph accounts for 88.51%, which indicates that the lexical relationship in the reading corpus has obvious aggregation and connectivity. Further examination shows that with the improvement of semantic network structure, the total score of students' reading increased from 66.3 to 82.1, word sense inference from 62.7 to 78.9, and discourse integration from 64.1 to 81.4, indicating that lexical semantic network not only affects local word sense recognition, but also improves the overall score of students' reading. More deeply involved in the process of inter-sentence cohesion, discourse construction and reasoning and judgment. The theoretical contribution of this paper is to promote the research of vocabulary from single indicators such as breadth and depth to the relationship network level, and systematically reveal the mechanism of lexical semantic organization on reading comprehension. The value of this method lies in the combination of natural language processing, semantic similarity calculation and graph model analysis, which forms a computable and interpretable reading research path. For college English reading teaching, we should pay more attention to the training of semantic connections between words, and improve students' comprehension ability by means of synonymous chains, hypernymy chains, collocation groups and thesaurus networks. There are still some limitations in this paper. For example, the sample size mainly focuses on 286 students, the research object is mainly non-English major college students, and the scope of corpus and dynamic tracking are still limited. In the future, we can further expand the sample, introduce longitudinal data, and combine knowledge graph, intelligent reading support system and generative artificial intelligence tools to carry out more in-depth research on the role of lexical semantic network in different reading tasks.

Funding

(Supported by Social Science Foundation of Jiangsu Province of China (Grant No.22YYD003)

References

[1] Tong Y, Hasim Z, Abdul Halim H. The relationship between L2 vocabulary knowledge and reading proficiency: The moderating effects of vocabulary fluency[J]. Humanities and Social Sciences Communications, 2023, 10(1): 555.

[2] Chen T, Zhang D. Different aspects of vocabulary depth knowledge in L2 Chinese reading comprehension: Comparing higher-and lower-proficiency readers[J]. Foreign Language Annals, 2023, 56(3): 786-806.

[3] Dagnaw A T. Revisiting the role of breadth and depth of vocabulary knowledge in reading comprehension[J]. Cogent Education, 2023, 10(1): 2217345.

[4] Rabadi R I. Examining the role of breadth and depth of vocabulary knowledge in reading comprehension of English language learners[J]. Jordan Journal of Modern Languages & Literatures, 2023, 15(1): 327-345.

[5] Zhang D, Pérez-Paredes P. Chinese EFL learners' use of mobile dictionaries in reading comprehension tasks[J]. System, 2024, 121: 103221.

[6] Jin C, Liu Y. Diagnosing and promoting learners' L2 inferential reading development through hybrid computerised dynamic assessment in the Chinese EFL classroom[J]. Computer Assisted Language Learning, 2024: 1-28.

[7] Zhang P. Effects of highlights and annotations on EFL learners' Reading comprehension: an application of computer-assisted interactive reading model[J]. Computer Assisted Language Learning, 2024: 1-33.

[8] Feng X, Liu J. The Structure of Lexical-Semantic Networks at Global and Local Levels: A Comparison between L1 and L2[J]. Complexity, 2024, 2024(1): 8644384.

[9] Wang Y, Kabilan M K. Investigating the impact of WeChat-mediated TBLT on Chinese EFL learners' reading comprehension and engagement in higher education: a mixed-methods study[J]. Computer Assisted Language Learning, 2025: 1-33.

[10] Wang T, Zhang H. Reciprocal effects of vocabulary breadth, vocabulary depth, and reading comprehension: a cross-lagged panel analysis in Chinese-speaking EFL learners[J]. Humanities and Social Sciences Communications, 2025, 12(1): 1-10.

[11] Pei Z, Zhang H, Xu H, et al. A Finer-Grained Exploration of Morphological Knowledge in EFL Reading Among Chinese College Learners: Mediation Through Vocabulary and Grammar Knowledge[J]. International Journal of Applied Linguistics, 2025.

[12] Wei R, Hu D. Incidental Vocabulary Learning Through Reading While Listening in the Chinese EFL Classroom Context: A Longitudinal Study[J]. SAGE Open, 2025, 15(2): 21582440251334243.

[13] Lin C C, Lin T H, Tang C K. Enhancing English reading comprehension, learning motivation and attitude through AI-Supported Pre-Reading scaffolding[J]. Journal of Computer Assisted Learning, 2025, 41(6): e70150.

[14] Liu J, Ma Q. Supporting low-proficiency L2 learners’ vocabulary learning with custom GPT-scaffolded corpus-based language pedagogy: a case study[J]. Computer Assisted Language Learning, 2025: 1-37.

[15] Yang P, Jeoung S, Cromley J, et al. Semantic Networks Extracted from Students’ Think-Aloud Data are Correlated with Students’ Learning Performance[C]//Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing. 2025: 25802-25815.

[16] Liu Z. Teaching design of English vocabulary semantic association based on knowledge graph[C]//Proceedings of the 2025 2nd International Symposium on Artificial Intelligence for Education. 2025: 70-76.