International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-9

https://doi.org/10.70917/ijcisim-2026-0384
Article

The Construction of Information Flow Delivery and Reader Interpretation Mechanism of English and American Literary Works Based on Semantic Network Analysis

Dan Wang [1,*]

[1] Foreign Language school, Henan University of Animal Husbandry and Economy, Zhengzhou 450011, China
* Correspondence author: wangdan.81@163.com

Abstract: This paper demonstrates the information flow transmission process of English and American literary works through two dimensions: lexical and discourse, and explores the discourse information reception from the reader's perspective. Taking the classic Anglo-American literature A Tale of Two Cities as the research object, relying on the lexical analysis technology to count the high-frequency words in the work, and the semantic network technology to conduct co-word analysis. Graph Attention Network GAT is utilized for relevance retrieval to construct a mechanism for readers' interpretation of English and American literature. The results show that Alexander Manette has the highest frequency of occurrence, reaching 1,901 times. Its pointwise centrality and middle centrality were the largest, 1476.000 and 503.062, respectively, and its proximity centrality was the smallest, 48.000. The highest degree of connection with "Alexander Manette" is "Charles Darnay", with a total of 157 occurrences. Analyzing the information transfer characteristics of the discourse from the perspective of author's information processing and introducing semantic networks for reader interpretation analysis provide new ideas and methods for the dissemination of English and American literature.

Keywords: semantic network; information flow transmission; GAT network; co-word analysis; English and American literature

1. Introduction

In the development history of western literature, Anglo-American classic literature has great influence. British and American classic literature has high literary value and is an important channel for readers to understand the ideology and culture of British and American countries and improve their own cultural literacy [1-3]. The history and culture of Britain and the United States are rich and colorful, and the background of the times and many social customs are written into the literary works. Many famous writers in the world literary world are from Britain and the United States, and the works of the two countries are widely praised [4-6]. For example, the works of Ernest Hemingway, Emerson, Shakespeare and Mark Twain have a more distinctive linguistic style, which expresses specific humanistic ideas and a unique period of time [7-9]. Reading and appreciating British and American classic literature not only helps readers to familiarize themselves with the western cultural background, but also helps to discover the differences between Chinese and western cultures [10-11]. In particular, the appreciation of some British and American classic literary works helps readers to appreciate the unique charm of Western literary language in characterization, subject structure and linguistic features, and readers can continuously improve their ability to appreciate British and American literary works through the appreciation of British and American classic literary works [12-15].

In this paper, we provide an in-depth description of lexical analysis and semantic network techniques, and propose an improved graph attention network GAT based on graph convolutional network GCN. Explore the factors affecting readers' reading in English and American literature through rare words and illumination comparisons. Relying on word frequency display, semantic network and semantic retrieval links to build a mechanism for reader interpretation. Utilizing lexical analysis technology to carry out word frequency statistics to obtain the high-frequency words in the text of the work. Select the high-frequency words to construct the co-word matrix, and visualize the display through semantic network technology. Introduce graph attention network GAT for semantic retrieval to explore the connection between high-frequency words.

2. Description of key technologies

2.1. Lexical Analysis Techniques and Text Semantic Expansion Techniques

2.1.1. Lexical analysis

Lexical analysis techniques are techniques that recognize individual substrings of an input string as words or symbols. For natural language processing, it mainly includes word segmentation, lexical annotation, dependent syntax analysis, named entity recognition and relation extraction. For this project, since the main processing is Chinese text, and Chinese text does not have the same separator as English text, there are two tasks involved: text segmentation and lexical labeling.

For Chinese word separation, the common method is prefix dictionary tree matching algorithm. For lexical annotation task, the more classical algorithm is the algorithm based on the Hidden Markov Model, who first uses the transfer probability to calculate the state transfer matrix, and then solves the optimal solution by the Viterbi algorithm. In essence, both tasks can be regarded as sequence annotation tasks, the purpose of which is to predict a marker for each position of a given string, which can be used to determine what kind of elements are actually in the corresponding substring positions. With deep learning becoming the mainstream of research, there is now a lot of neural network-based research for the above two tasks as well.

In general, for the above two tasks, there has been a relatively deep research in industry and academia, and some toolkits have been developed to facilitate the use of others. The main common tools are: jieba, LTP and so on.

2.1.2. Semantic networks

Semantic web is a way to express human knowledge through data in the form of web. In the semantic network, human knowledge is mainly divided into entities and relations, corresponding to the graph data, i.e., nodes and edges. Take the event "John gave Mary a book" as an example, its semantic network expression is shown in Figure 1.

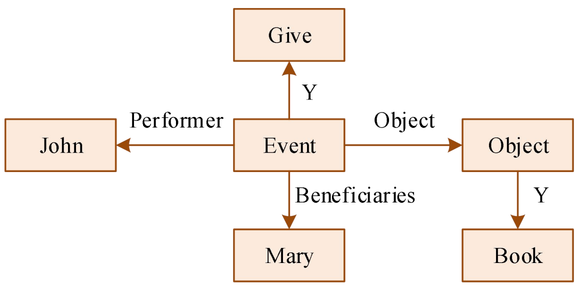

Figure 1. Sample semantic network

As can be seen from the figure, the semantic network can express the knowledge of the human world very flexibly, but its various forms of expression also bring difficulties in practical application. In order to make the semantic network can be applied to the computer to automatically analyze the data, the concept of knowledge mapping came into being, knowledge mapping provides the data form for the "entity-relationship-entity" ternary form, but its essence is still a kind of revealing the relationship between the entities of the semantic network. Semantic network. It has a wide range of applications, including intelligent search, in-depth Q&A, and graph-based search on social networks. In China, there is also a fairly mature semantic network.

2.2. Graph Convolutional Neural Networks

2.2.1. Graph Convolutional Networks GCNs

The GCN network is a structure that can handle data relations on non-Euclidean spaces. This network is derived from the improvement of previous graph convolution algorithms and is able to solve some node classification problems stably and efficiently by the propagation of graph information. This graph convolution network is obtained by a first-order approximation based on the previously studied local spectral graph convolution network. It is assumed that the Laplace matrix of the relational graph of the data is L, as in equation (1).

$$L = D - A \tag{1}$$

where D is the degree matrix of the graph vertices, i.e., the number of edges whose diagonal elements are each node, and A is the adjacency matrix of the graph, i.e., $A_{ij} = 1$ if the i rd node in the graph is connected to the j th node by an edge, and $A_{ij} = 0$ if there is no edge connection. Then the eigenfunctional decomposition of the Laplace matrix L is as in Eq. (2).

$$L = I_N - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} = U \Lambda U^{-1} = U \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} U^T \tag{2}$$

where U is the square matrix consisting of all eigenvectors, $U = \left(\overrightarrow{u_1}, \overrightarrow{u_2}, \ldots, \overrightarrow{u_n}\right)$, and since the Laplace matrix is a semi-positive definite symmetric matrix whose eigenvectors are orthogonal to each other, U is an orthogonal matrix with $UU^T = I$, and hence $U^{-1} = U^T$; Λ is a diagonal matrix consisting of eigenvalues corresponding to the eigenvectors. The conventional convolution of two functions is the inverse transformation of the product of the Fourier transforms of the two functions, and for the graphical model the convolution can be analogized to the computational process as in equation (3).

$$g_\theta * x = U g_x U^T x = U \begin{pmatrix} \theta_1 & & \\ & \ddots & \\ & & \theta_n \end{pmatrix} U^T x \tag{3}$$

where x is the characterization vector of the graph node, the Fourier transform of x can be expressed as $\hat{x} = U^T x$; g_θ is the convolution kernel designed to adapt to the needs of the scene, and $\theta_i = \hat{h}\left(\lambda_1\right)$ is the Fourier transform of the convolution kernel.

In some practical application scenarios, the feature decomposition of the Laplace matrix and the matrix multiplication of the eigenvectors are too complicated to compute, which will make the whole process very consuming of computational resources, so the truncated expansion of the Chebyshev polynomials of the $T_K(x)$ th to the K th order is used as an approximation of g_θ, and the improved graph convolution is shown in Eq. (4).

$$g_{\theta'} * x \approx \sum_{k=1}^{K} \theta'_k T_k(\tilde{L}) x \tag{4}$$

where θ' is the vector of Chebyshev coefficients; $\tilde{L} = \frac{2}{\lambda_{\max}} L - I_N$, $\lambda_{\max}$ are the maximum eigenvalues of L. The improved graph convolution does not need to do feature decomposition again, which greatly simplifies the computation.

The above graph convolution calculation combined with the specific graph structure and features of graph nodes can effectively model and process the elemental relationships in non-Euclidean space.

2.2.2. Graph Attention Network GAT

Graph Attention Network GAT is a further generalization of graph convolutional networks. The original graph convolutional network treats all neighboring nodes of the center node in the same way, which is reflected in the neighbor matrix as all neighboring relationships are marked as 1. However, in real scenarios there are important and unimportant relationships between nodes, and the GCN does not reflect this difference in the degree of importance, so instead of having only two values of 0 and 1 in the graph neighbor matrix of the GAT, the weights are different for the nonzero element multiplied by different weights so as to reflect the importance of the relationship of the central node to its neighboring nodes, which is calculated by the eigenvectors of the two nodes with the attention mechanism, so this improved network is called graph attention network.

For the input graph node feature matrix $h=\left\{\vec{h}_1,\vec{h}_2,\ldots,\vec{h}_N\right\}$, where $\vec{h}_i$ is the feature vector of a node in the graph model, the attention score of node i to node j is calculated as in equation (5).

$$e_{ij}=attention\left(W\vec{h}_i,W\vec{h}_j\right) \tag{5}$$

where W is a linear transformation matrix.

In order to make the comparison in the attention scores between different nodes more reasonable, the attention scores are normalized with softmax function as in equation (6).

$$\alpha_{ij}=soft\max\left(e_{ij}\right)=\frac{\exp\left(e_{ij}\right)}{\sum_{k\in N_i}\exp\left(e_{ik}\right)} \tag{6}$$

Where N_i is the numbered set of neighboring nodes of node i. Then the graph attention calculation process of the graph model is shown in equation (7).

$$\vec{h}_i^l=\sigma\left(\sum_{j\in \mathrm{N}_i}\alpha_{ij}^k W^k\vec{h}_j\right) \tag{7}$$

where $\sigma(\cdot)$ is the nonlinear transformation.

The computation of GAT does not rely on the complete graph, and for each node, only the influence of neighboring nodes on the center node is computed, so it has high computational efficiency, and at the same time, it has a better ability to reason about relationships than GCN. In this paper, we draw on the idea of GAT to utilize dependencies more efficiently in fine-grained sentiment analysis and improve the performance of the model.

3. Construction of reader interpretation mechanism based on semantic network analysis

3.1. Introduction to the corpus

In this paper, we take British and American literature as the object of study and select seven classic British and American literary works as the experimental corpus. The reference corpus used is the British National Corpus (BNC), a collection of 100 million words of modern English samples, consisting of 4124 texts representing a wide range of modern English, from which we extracted a total of 329 works of fiction to form the reference fiction corpus.

3.2. Discourse information flow transfer

3.2.1. Lexical dimensions

Rare words refer to word forms that appear only once in the corpus and can be used to reflect the degree of creativity of the corpus. The lower the proportion of rare words in the total classifier, the less creative the corpus is, the higher its degree of normalization, and the difficulty of readers' reading comprehension will be reduced accordingly. Therefore, this paper utilizes Wordsmith software to count rare words in the corpus, supplemented by manual checking and exclusion, and subsequently, the frequency and frequency of rare words are calculated. The statistical results are shown in Table 1.

Table 1. Comparison of rare words

	Generic symbol	Number of rare words	Frequency of rare words
BNC Fiction Library	159733	69897	43.76%
The Complete Works of Shakespeare	11572	4862	42.02%
Wuthering Heights	13869	5573	40.18%
One Hundred Years of Solitude	10783	4272	39.62%
Pride and Prejudice	15872	6672	42.04%
Tess of the d'Urbervilles	8673	3684	42.48%
A Tale of Two Cities	8819	3276	37.15%
The Great Gatsby	8676	3762	43.36%

As can be seen from Table 1, firstly, the frequency of the use of rare words and the frequency of rare words in the 7 British and American classic literature works are significantly lower than that of the BNC novel library, which indicates that they prefer to use a smaller number of rare words, which reflects the fact that the British and American classic literature works have chosen to cater to the reading habits of the English-speaking readers in order to improve the reader care. Secondly, among the seven British and American literary works, the frequency of the use of rare words in A Tale of Two Cities is only 37.15%, which is significantly lower than that of the other six works, which means that the language standard of A Tale of Two Cities is more standardized and more in line with the required standard of the English language, and the readers' reading comprehension will be less difficult with it.

3.2.2. Discourse dimensions

Whether a discourse can be understood and accepted by the target language readers, discourse articulation plays an important role. In the realization of articulation, illumination is the most commonly used grammatical means in English. Among them, illumination mainly includes personal illumination and indicative illumination; personal illumination is mainly realized by personal pronouns, possessive pronouns and noun possessive pronouns; indicative illumination is mainly realized by indicative pronouns, adverbs of time and place, definite articles and so on. In this paper, Wordsmith software is utilized to retrieve the illumination variables in the corpus, and the results are shown in Table 2.

Table 2. Discourse dimension comparison

	Generic symbol	Number of rare words	Frequency of rare words
BNC Fiction Library	159733	50872	31.85%
The Complete Works of Shakespeare	11572	4662	40.29%
Wuthering Heights	13869	5487	39.56%
A Tale of Two Cities	10783	4835	44.84%
Pride and Prejudice	15872	6753	42.55%
Tess of the d'Urbervilles	8673	3289	37.92%
One Hundred Years of Solitude	8819	3428	38.87%
The Great Gatsby	8676	3387	39.04%

As can be seen from Table 2, firstly, the frequency of the use of illuminating means in all the seven British and American classic literature works is significantly higher than that of the BNC novel library, indicating that all the seven British and American classic literature works make great use of the typical linguistic features of the English language, that is, they are more standardized and standardized in terms of the use of language, in order to reduce the reading barriers of the English-speaking readers. Secondly, in two works, A Tale of Two Cities and Pride and Prejudice, the frequency of the use of illocutionary devices is 44.84% and 42.55%, respectively, which is much more than the other five works. This indicates that the two works pay more attention to the degree of articulation of the discourse and pay more attention to the reading experience of English readers.

3.3. Mechanisms for reader interpretation

3.3.1. Word Frequency Display

Taking A Tale of Two Cities as an example, the word frequency statistics of the corpus data of A Tale of Two Cities after word separation are shown in Table 3. As can be seen from Table 3, Alexander Manette has the highest frequency of occurrence, reaching 1,901 times. According to the characteristics of the data to summarize the cultural information dimensions to which the high-frequency words belong, it can be seen that: (1) the background of the story: the French Revolution. (2) Place of occurrence: Paris, London. (3) Main characters: Alexander Manette, Charles Darnay, Lucy Manette and so on. It can be seen that through the statistics of high-frequency words in novels, the core elements in novels can be effectively extracted, which can help readers to analyze the structure of the novels, the emotional tone and its cultural background.

Table 3. Word frequency statistics

Serial number	Frequently used words	Word frequency
1	Alexandre Manette	1901
2	Charles Darnay	1583
3	Lucie Manette	1208
4	Paris	1103
5	Javert	785
6	London	763
7	Carton	664
8	Defarge	493
9	Barsad	302
10	French Revolution	186
...	...	...

3.3.2. Semantic Web Presentation

The first 10 high-frequency words in A Tale of Two Cities are selected to construct a co-occurrence matrix, and its co-occurrence knowledge graph is shown in Figure 2, with 10 word frequencies and 86 co-occurrence lines. The nodes in the graph represent the high-frequency words, the size of which is the importance of the word frequencies, and the lines represent the co-occurrence relationship between the word frequencies. In the center of the graph is “Alexander Manette”, indicating that he is the most important character in the book.

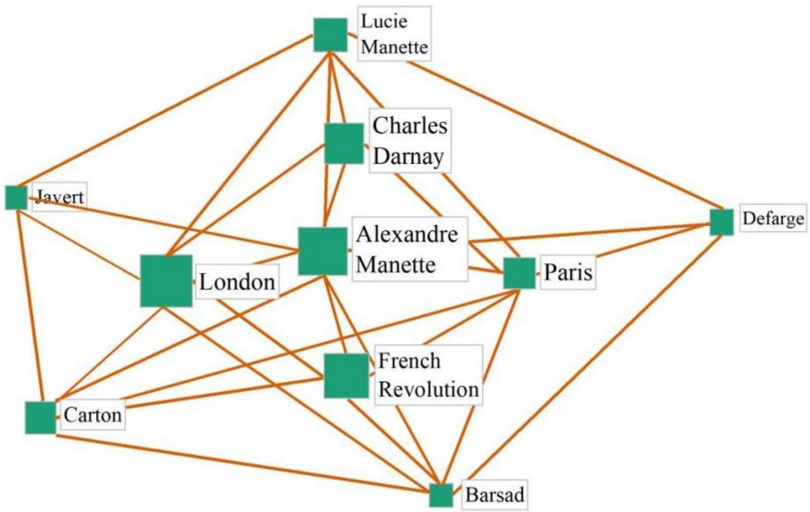

Figure 2. High frequency word meaning network

The constructed high-frequency word data co-word matrix is imported into Ucinet, and the point-degree centrality, intermediate centrality and proximity centrality of each node are obtained, and the results of point-degree centrality, intermediate centrality and near-centrality are shown in Table 4. In the semantic network nodes of high-frequency words in A Tale of Two Cities, the maximum point degree centrality is 1476.000, and the minimum point degree centrality is 258.000, among which the point degree centrality of "Alexander Manette" is 1476.000, followed by "Charles Darnay", "London", "Paris" and "French Revolution", with a point degree centrality of 984.000, 932.000, 915.000 and

901.000, respectively. It can be seen that these nodes are in a relatively central position in the semantic network and have greater "control" over the entire semantic network. The intermediate centrality of a node measures the ability of the node to control the interaction of other nodes. In the semantic network nodes of high-frequency words in A Tale of Two Cities, the maximum middle centrality is 503.062, and the minimum intermediate centrality is 30.432, among which the node with the largest intermediate centrality is "Alexander Manette", indicating that "Alexander Manette" is in the center of the semantic network and controls the connection of most nodes. There are 4 nodes with intermediate centrality below 100, namely "Javier", "Carlton", "Balsa", and "Devre", indicating that these nodes are located at the edge of the semantic network. The smaller the proximity centrality of a node, the more it is located in the center of the semantic network. The node with the smallest nearness to centrality among the high-frequency words in A Tale of Two Cities is "Alexander Manette", with a nearness of 48.000, which shows that the node is located in the center of the semantic network. To sum up, "Alexander Manet" is the protagonist of "A Tale of Two Cities".

Table 4. Centrality analysis of high-frequency words

Frequently used words	Point degree center degree	Intermediate centrality	Centrality of proximity
Alexandre Manette	1476.000	503.062	43.000
Charles Darnay	984.000	369.985	54.000
Lucie Manette	593.000	186.083	69.000
Paris	915.000	382.832	58.000
Javert	403.000	47.082	79.000
London	932.000	326.983	52.000
Carton	397.000	86.003	81.000
Defarge	258.000	30.432	90.000
Barsad	359.000	92.000	77.000
French Revolution	901.000	308.490	55.000

3.3.3. Semantic search

Keyword search based on GAT can better understand the semantic meaning of search terms and discover other conceptual entities associated with them. Taking "Alexander Manet", the main character in A Tale of Two Cities, as an example, the connection between "Lucy Manet", "Charles Darnay", "Paris", "London", "Javier" and "Carlton" is searched, and the search results are shown in Figure 3. After observation, it was found that "A Tale of Two Cities" has the highest connection with "Alexander Manette" is "Charles Darnay", with a total of 157 places; The lowest degree of connection is "Carlton", with a total of 8 locations.

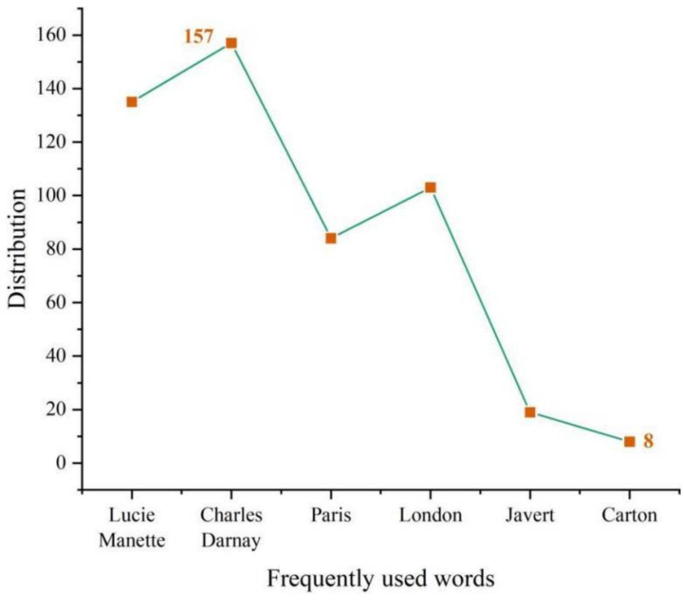

Figure 3. Search results of high-frequency words

4. Conclusion

Utilizing word frequency display, constructing semantic network and semantic retrieval to build a reader interpretation mechanism, the study and analysis are carried out on the example of the classic British and American literature A Tale of Two Cities.

Alexander Manet was the most frequent at 1,901. According to the characteristics of the data, the cultural information dimension to which the high-frequency words belong can be summarized, and the background of the story, the place of occurrence and the main characters can be seen. The semantic network shows that "Alexander Manette" has the largest point centrality and intermediate centrality of 1476.000 and 503.062, respectively, and the minimum near centrality is 48.000, which shows that the node is located in the center of the semantic network, that is, "Alexander Manette" is the protagonist of "A Tale of Two Cities". The highest degree of connection with "Alexander Manet" is "Charles Darnay", with a total of 157 places; The lowest degree of connection is "Carlton", with a total of 8 locations.

To summarize, the use of word frequency display can quickly identify the theme and key information of a literary work, semantic networks help to reveal the multi-level connection between characters and background, and semantic retrieval provides readers with a basis for identifying character relationships. To build a mechanism for readers' interpretation of English and American literature in this way can help readers capture the hidden information and deep structure in novels, and make a profound exploration of the textual understanding of literary works.

References

1. Bogue, R. (2013). On the Superiority of Anglo-American Literature. Deleuze Studies, 7(3), 302-318.
2. Reisner, P. (2017). Contemporary Anglo-American. American Communities: Between the Popular and the Political, 35, 187.
3. Okyayuz, A. Ş. (2017). Anglo-American teen literature in translation. New Review of Children's Literature and Librarianship, 23(2), 148-171.
4. Christianson, F. Q., & Thorne-Murphy, L. (Eds.). (2017). Philanthropic Discourse in Anglo-American Literature, 1850–1920. Indiana University Press.
5. Yingjun, L. (2012). An Exploration of the Model of “Stage” Teaching the Anglo-American Literature in the Network Environment. IERI Procedia, 2, 49-54.
6. Al-Omari, K. M., & Obeidat, M. M. (2022). Perspectives of the English Department Professors of the Hashemite University on Anglo-American Literature and Culture: A Pedagogic Stance. Teaching American Literature, 12(1).
7. Ambayon, E. E., & Millenes, C. (2020). Virtual Approach in Teaching Anglo-American Literature and Students’ Comprehension Skills. Available at SSRN 3723600.
8. Zhao, Y. J. (2014). A Study of the Influence of Afro-American Culture on Anglo-American Fiction. English Language and Literature Studies, 4(1), 69.
9. Marcinkiewicz, P. (2017). Polish Translations of Anglo-American Literature and the Question of Ideology: From Romanticism to Twentieth-Century Avant-Gardes. TranslatoLogica: a journal of Translation, Language and Literature, 1.
10. Kyriakidou, E. (2019). Petr Antene, Campus Novel Variations: A Comparative Study of an Anglo-American Genre. European journal of American studies.
11. Dumitrescu, I. (2018). The experience of education in Anglo-Saxon literature (Vol. 102). Cambridge University Press.
12. Stanković, A. K., & Ignjatović, S. (2021). TEACHING ANGLO-AMERICAN CULTURE AT THE ENGLISH DEPARTMENT: THE PAST, THE PRESENT AND THE FUTURE. Facta Universitatis, Series: Linguistics and Literature, 097-105.
13. Anténe, P. (2015). Campus novel variations: A comparative study of an Anglo-American genre. Palacký University Olomouc.

14. Zsámba, R. (2020). Contemporary Perspectives on Language, Culture and Identity in Anglo-American Contexts. Eger Journal of English Studies, (20), 121-125.
15. Marabe, J. V., & Petalla, M. B. (2023). Macro communication skills and Anglo-American literary competence of Grade 9 students in a public high school. Asian Journal of Language, Literature and Culture Studies, 6(3), 272-288.