Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

A Study on the Philosophical Connotation and English Translation Strategies of the Word “Body” in *The Analects of Confucius* Based on Computer Semantic Modeling from the Perspective of Embodied-Cognitive Translatology

Zijuan Su[1,] ✉

[1] *School of Foreign Languages and Culture, Geely University of China, Chengdu, Sichuan, 641423, China*

ABSTRACT

With the need of international dissemination of Chinese culture, the problem of translating traditional Chinese texts gradually emerges. The study embeds a computer semantic model into the English translation of *The Analects of Confucius*, and constructs a natural language understanding model based on S-LSTM network through semantic representation of natural language processing. In order to explore the performance of the S-LSTM model, it is compared with RNN, LSTM, I-LSTM and other models in terms of training time and accuracy, so as to validate the superiority of the S-LSTM model in this paper. This paper deeply explores the philosophical connotation of the character "body" in *The Analects*, and studies the structural complexity of the translation of the character "body" through the S-LSTM model. Finally, the English translation strategy of *The Analects* and other classics is proposed. Among all the comparison models, the S-LSTM model has the fastest training speed and the highest accuracy. The translation of the word "body" in *The Analects* and the local complexity of the ministry are characterized by complication. The local complexity of the noun and the subject in the source English language, and the overall complexity of the "be-passive" structure have obvious effects on the structure of the translated Chinese character "body".

Keywords: computer semantic modeling, natural language processing, S-LSTM model, English translation strategies

1. Introduction

As one of the four books, *The Analects of Confucius* has been highly honored and respected since ancient times. In the 16th century, with the arrival of European missionaries, there was a wide range of contacts and exchanges between Chinese and Western cultures, and *The Analects of Confucius*, one

✉ Corresponding author.
E-mail address: suzijuan@126.com (Z. Su).
Received 10 January 2024; Revised 25 May 2024; Accepted 20 October 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-218

of the Chinese classics, was favored by many missionaries who came to China. Some of them, out of their interest in traditional Chinese culture, and some out of the need for missionary work, coincidentally studied *The Analects of Confucius* intensively, and attempted to translate it into a foreign language and spread it overseas [1-3]. In 1593, the missionary Matteo Ricci translated the Four Books, and he was the first one to translate the Four Books into Latin and spread the Four Books to the West. He was the first to translate the Four Books into Latin and spread them to the West. He opened the prelude to the spread of Chinese classics to the West and led the trend of translating Chinese classics. Since then, missionaries and even scholars have devoted themselves to the study and translation of Chinese classics, and many translations have appeared in German, French, Spanish, Italian and so on [4-6].

As the power of the British nation flourished and the country developed rapidly, the number of missionaries coming to China from Britain also increased greatly, and the English translation of *The Analects of Confucius* began in such a historical background. The earliest English translation of *The Analects of Confucius* was published in the 30th year of Kangxi, i.e., 1691 A.D. According to the existing data, the translation only contains 80 short aphorisms, and the translator regarded Confucius as a moral philosopher and understood *The Analects of Confucius* as a moral sermon [7-9]. Although this is an extremely immature English translation of *The Analects of Confucius*, it has set a precedent for subsequent English translations [10]. From 1691 to the present, more than 30 English translations (including both abridged and full translations) have been published. Regardless of whether it is an abridged translation or a full translation, the English translation of *The Analects of Confucius* is a mixture of good and bad because of the different degrees of understanding of Confucian culture by the translators [11-12]. In addition to the different translation purposes, different translation strategies, and different character temperament of the translators themselves, the English translation of *The Analects of Confucius* is even more blossoming, each with its own unique characteristics. For example, in the first English translation of *The Analects* mentioned above, the translator treated it as a moral aphorism, and then Su Huilian's translation has obvious religious tendency, Arthur Wylie's translation has strong literary color, and Pound's translation has more creative translation points [13-15]. After entering the 20th century, many sinologists and Chinese scholars began to translate *The Analects* into English. With the production of a large number and variety of English translations of *The Analects of Confucius*, studies on English translations of *The Analects of Confucius* have also mushroomed. Especially in recent years, traditional Confucian thought and culture have been emphasized by scholars at home and abroad, and Confucian texts such as *The Analects of Confucius* have been revitalized, and the study of the English translation of *The Analects of Confucius* has become one of the academic hot spots [16-18].

The work of English translation of Chinese canonical books is an important way for Chinese culture to go out and a major initiative to show China's cultural self-confidence [19]. Since the beginning of the 21st century, China has attached great importance to the publication of English translations of classics, including them in the national strategic project, and the government has invested a lot of money to launch projects such as the "Great China Library", "Classics of China Overseas Publishing Project", "Chinese Books Overseas Promotion Program", and "Chinese Cultural Works Translation and Publishing Project", which have vigorously promoted the vigorous development of the English translation of Chinese classics [20-21]. The number of foreign translations and publications of Chinese canonical books has been increasing year by year, but there are still some gaps in the competitiveness and influence of “going out” compared with that of Europe and the United States and other cultural exporting powers, and the dissemination ability in the international community is not strong enough. The reason for this is that while there are constraints in the discourse environment of the whole international community, technically speaking, it has a lot to do with the fact that the non-modernization of Chinese canonical translations has led to the failure of the translated works to attract the foreign readers [22-23].

Based on the processing of semantic information by the natural language processing model, a natural language understanding model based on the S-LSTM network is constructed, and the vocabulary preprocessing is carried out in the original S-LSTM network, and the attention mechanism and conditional random field are added. The Semantic Analysis performance of the S-LSTM model in this paper is verified by analyzing the time cost and accuracy of the model by comparing it with the RNN, LSTM, I-LSTM and other models. This paper takes the character "body" in *The Analects* as the research object to analyze the semantic meaning, and deeply explores the philosophical connotation of the character "body" in *The Analects*. Then, the complexity of the translation structure of the character "body" is analyzed. Finally, the English translation strategy of traditional Chinese classics such as *The Analects* is proposed.

2. Computer semantic modeling

2.1. Semantic information

Semantic information refers to a type of information related to language or symbols that involves lexical, syntactic and semantic aspects [24]. It describes the meanings, relationships, and linguistic rules of linguistic units (e.g., words, phrases, sentences). The semantic information of a text is the meaning and implications conveyed in the text. It includes not only the surface meanings of words and sentences, but also the deeper meanings hidden in the text such as context, linguistic conventions, and cultural factors. In natural language processing, the understanding of semantic information is very important, because it can help computers understand the real meaning of the text, can eliminate the uncertainty of the text, and then realize a variety of natural language processing tasks.

Semantic information generally includes weakly supervised semantic information, self-supervised semantic information and external semantic information. Weakly supervised semantic information generally comes from metadata. Metadata is descriptive information used to describe the data itself or information resources. It provides detailed information about the data and information resources. In text, it mainly describes the attributes of the text itself, and common text metadata are labels, covariates, timestamps and named entities. These are some of the attributes that come with the text, which is composed of vocabulary with semantic information that allows the model to better determine the text information. In this paper, the tags of the text are mainly selected as auxiliary weakly supervised semantic information to guide the model to recognize ambiguous words, which is used to alleviate the ambiguity of the vocabulary in the text.

Self-supervised semantic information comes from expanded data, which refers to some operations on the original text based on metadata to expand the metadata into meaningful data. Some of the commonly used operations are data enhancement methods, comparison learning methods, etc. Data enhancement methods are generally methods that create new synthetic data from existing metadata to increase the amount of data. There are three general types: paraphrase-based methods, noise-based methods, and sampling-based methods. Contrastive learning methods are a type of self-supervised learning method, self-supervised learning is a method of training with unlabeled data that uses an autogenerated or self-reconstructed approach to obtain a feature representation of the data. This is achieved by training a model to perform an alternative task, such as a prediction, generation, or comparison task, and using the objective function of that task as the loss function of the trained model.Self-supervised learning is capable of extracting meaningful semantic information, such as word meanings, entities, topics, and so on, from input data. In this paper, the alternative task is a contrastive learning approach that divides the text into positive and negative samples and uses a form of self-supervised model learning, which is used to mitigate the sparsity of short texts.

External semantic information comes from external data, which refers to a set of textual data collections collected or acquired from outside, usually containing a large amount of natural language textual data. These data collections contain rich semantic information. Commonly used external data

are Wikipedia, WordNet, etc. The general usage is to pre-train the external knowledge into word vectors using word embedding techniques, and introduce the word vectors into the model as additional knowledge. This external knowledge not only contains a large amount of prior knowledge, but also enables the model to learn the semantic information between words. In this paper, the external knowledge and the knowledge learned by the model are mapped to the same latent space, aligning the semantics of the words, which is used to mitigate the irregularities of short texts.

2.2. Semantic Representation of Natural Language Processing Models

Natural Language Processing (NLP) is one of the fields of Artificial Intelligence, which involves a number of disciplines such as Computational Linguistics, Machine Learning, and Psycholinguistics [25]. One of the tasks of natural language processing is semantic analysis, and the main purpose of semantic analysis is to obtain the semantics of words, phrases, sentences, paragraphs and chapters, and all tasks related to natural language understanding can be categorized as semantic analysis. According to the language unit of the understanding object, semantic analysis can be decomposed into lexical, sentence and chapter semantic analysis. Generally speaking, lexical semantic analysis is mainly concerned with acquiring or distinguishing the semantics of words. Sentence semantic analysis mainly analyzes the semantic information of the whole sentence. Chapter semantic analysis aims to study the internal structure of the text and understand the semantic relationships between text units (sentences or paragraphs). In short, semantic analysis is to realize the semantic analysis of each language unit (words, sentences and chapters, etc.) by building models and systems, so as to achieve the purpose of understanding the real semantics of the text. This study focuses on semantic analysis at the lexical level of brain and natural language processing models. Lexical level semantic analysis can be further divided into: semantic disambiguation, lexical representation.

Text is a series of symbols with semantics, computers cannot directly encode the meaning of these symbols as humans do, so it is necessary to convert words into numeric vectors that computers can directly recognize and compute, i.e. word vectors. Word vector is the name of a set of language modeling and a collection of techniques for learning text features in natural language processing, a method of mapping words or phrases in a vocabulary to vectors of real numbers in a processing flow.

Depending on how word vectors are characterized, semantic representations by computers can be subdivided into two types: direct one-hot representations, and learned-acquired word vectors. one-hot representations convert tokens into vectors in the most common and basic way, by associating each word with a unique integer index, and then converting this integer index i into a binary vector of length N (N is the size of the lexicon), which has only the ird element as 1 and the rest of the elements as 0. The One-hot characterization is a sparse vector (the vast majority of the elements are 0), which is convenient to compute, but the dimension of the semantic vector space is very high (with the size of the dimension equal to the number of words in the corpus). The basic idea of word vector is: through the training text, each word is mapped into a vector of fixed dimension, and all these vectors are put together to form a word vector space, and each vector can be regarded as a point in the space, and the concept of "distance" is introduced into this space, so that we can determine the (lexical, syntactic, and linguistic) distance between the words according to the distance between them. Introducing the concept of "distance" in this space, we can judge the similarity between words according to the distance between them (lexical and semantic), and obtain the semantic information characterized by word vectors and the semantic relationship between word v ectors.

Because the semantics is distributed in the contextual context, so it is called distributed representation, semantically similar words appear in the contextual similarity, and vice versa, if the contextual similarity is found, then it means that the two words are semantically similar, and the word vectors are based on this law of distributed representation for learning.

In the vector space of the natural language processing model, words are represented as vectors, and the semantic similarity of two words can be obtained by calculating the similarity or distance of two

word vectors, and the cosine similarity is generally chosen as a measure of this similarity. Because the cosine value normalizes the word vectors to unit length before calculating the correlation, the effect of vector length on the similarity can be reduced.

After building a natural language processing model, the performance of the model needs to be evaluated, mainly for the performance of word vectors in accomplishing linguistic tasks.

2.3. Natural language understanding model based on S-LSTM network

2.3.1. S-LSTM. S-LSTM model [26], which deals with an utterance without taking into account the local feature extraction at the lexical level, but also refines the overall hidden state of the utterance through the lexical hidden state and makes the two supervise each other, and its specific structure is shown in Fig. 1.

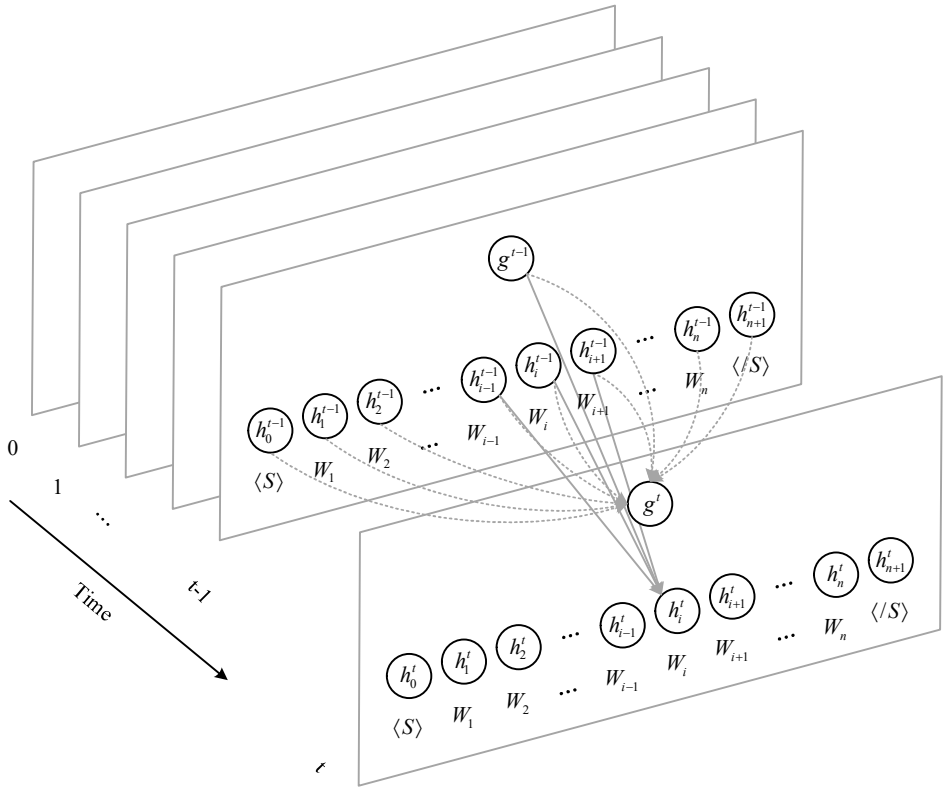


Fig. 1. S-LSTM model structure

The hidden state h of each vocabulary of the Discourse is influenced by the hidden states of neighboring vocabularies at the previous moment and the state g^{t-1} of the sentence as a whole at the previous moment. The number of overall sentence states is an adjustable hyperparameter. The vocabulary range length is also an adjustable hyperparameter, which represents the vocabulary range of the previous moment that affects the vocabulary state of the next moment during each iteration, e.g., a step size of 1 for step size s means that the hidden state h_i^t of vocabulary i at this moment is affected by the hidden state $[h_{i-1}^{t-1}, h_i^{t-1}, h_{i+1}^{t-1}]$ of the previous moment. The flow of information is controlled using a gate mechanism as shown in the following equation:

$$\xi_i^t = [h_{i-1}^{t-1}, h_i^{t-1}, h_{i+1}^{t-1}] \tag{1}$$

$$\hat{i}_i^t = \sigma(W_i \xi_i^t + U_i x_i + V_i g^{t-1} + b_i) \tag{2}$$

$$\hat{l}_i^t = \sigma(W_l \xi_i^t + U_l x_i + V_l g^{t-1} + b_l) \tag{3}$$

$$\hat{r}_i^t = \sigma(W_r \xi_i^t + U_r x_i + V_r g^{t-1} + b_r) \tag{4}$$

$$\hat{f}_i^t = \sigma(W_f \xi_i^t + U_f x_i + V_f g^{t-1} + b_f) \tag{5}$$

$$\hat{s}_i^t = \sigma(W_s \xi_i^t + U_s x_i + V_s g^{t-1} + b_s) \tag{6}$$

$$o_i^t = \sigma(W_i \xi_i^t + U_o x_i + V_o g^{t-1} + b_o) \tag{7}$$

$$u_i^t = \tanh(W_u \xi_i^t + U_u x_i + V_u g^{t-1} + b_u) \tag{8}$$

$$i_i^t, l_i^t, r_i^t, f_i^t, s_i^t = soft\max(\hat{i}_i^t, \hat{l}_i^t, \hat{r}_i^t, \hat{f}_i^t, \hat{s}_i^t) \tag{9}$$

$$c_i^t = l_i^t \cdot c_{i-1}^{t-1} + f_i^t \cdot c_i^{t-1} + r_i^t \cdot c_{i+1}^{t-1} + s_i^t \cdot c_g^{t-1} + i_i^t \cdot u_i^t \tag{10}$$

$$h_i^t = o_t^i \cdot \tanh(c_i^t) \tag{11}$$

where $\xi_{\mathrm{i}}^{'}$ represents the set of vocabulary states of the previous moment's Discourse vocabulary that have an effect on the current moment's Discourse vocabulary state, and $I_i^{'}, r_i^{'}, f_i^{'}, s_i^{'}$ as well as $i_i^{'}$ are gate mechanisms that control the flow of information contributed to $c_i^{'}$. Specifically, $i_i^{'}$ controls the information flow of the corresponding vocabulary input at this moment, $I_i^{'}$ and r_i^t control the influence of the hidden state of the left vocabulary and the hidden state of the right vocabulary in the previous moment on the hidden state of the current vocabulary, respectively, and $f_i^{'}$ controls the influence of the current vocabulary of the previous moment on the current vocabulary of the current moment, so that the S-LSTM can transfer the vocabulary information of the whole sentence to each vocabulary through the increase of the number of layers of the network. Meanwhile, $s_i^{'}$ is used to constrain the effect of the previous moment's sentence state on the current moment's word. The above parameters, which are normalized so that they sum up to 1. $o_i^{'}$ The resulting hidden states are further transformed as the final output lexical hidden states. W, U, V, b are model parameters that can be trained. σ represents the activation function. The above network structure borrows greatly from that in LSTM networks, where the tanh function is utilized to activate the current moment state and the Sigmoid activation function is used to compute the weights.

Sentence state g, which is calculated based on the hidden state of all the Discourse words in the previous moment and the sentence state in the previous moment, is calculated in the following way as shown in Eq:

$$\overline{h} = av\,g(h_0^{t-1}, h_1^{t-1}, ..., h_{n+1}^{t-1}) \tag{12}$$

$$\hat{f}_g^t = \sigma(W_g g^{t-1} + U_g \overline{h} + b_g) \tag{13}$$

$$\hat{f}_i^t = \sigma(W_f g^{t-1} + U_f h_i^{t-1} + b_f) \tag{14}$$

$$o_{\mathrm{t}} = \sigma(W_o g^{t-1} + U_o \overline{h} + b_o) \tag{15}$$

$$f_0^t, ..., f_{n+1}^t, f_g^t = \mathrm{softmax}(\hat{f}_0^t, ..., \hat{f}_{n+1}^t, \hat{f}_g^t) \tag{16}$$

$$c_g^t = f_g^t \cdot c_g^{t-1} + \sum_i f_i^t \cdot c_i^{t-1} \tag{17}$$

$$g_t = o_t \cdot \tanh(c_g^t) \tag{18}$$

$f_0^t,...,f_{n-1}^t$ and f_g^t are gate mechanisms that control the flow of information from the hidden state of all Discourse words from the previous moment and the state of sentences from the previous moment. o_t is the output gate, and w,u,b is the model parameters that can be trained.

The S-LSTM network changes the usual pattern of processing one vocabulary by one vocabulary in accordance with the chronological order used in the previous recurrent neural networks, but processes all *The Analects* vocabularies at the same time, and passes the information layer by layer according to the relationship between the proximity and distance of *The Analects* vocabularies in the continuous iteration, and at the same time refines the global state, so that the local state and the global state are supervised by each other. Such an approach speeds up the training of the network, improves the model performance, and is more compatible with the natural language understanding task, so in this paper, we choose the S-LSTM network as the base model to jointly handle the two tasks.

2.3.2. Lexical preprocessing. In the original S-LSTM network, all *The Analects* words that do not exist in the Glove word embedding dataset are directly set as "UNK", i.e., unknown words, and the label corresponds to a fixed manually designed embedding vector. However, it can be seen from the above figure that some neglected Analects words contain certain information, such as names of people, time, and other semantic slots that are crucial for accomplishing the task, and thus need to be retained. In order to retain this information as much as possible, this paper chooses to manually formulate rules for vocabulary reduction.

After the above processing, there are still some uncommon or possibly incorrect words, which often do not contain significant semantics and generally cannot be used as semantic slot values, and at the same time, ignoring them will not have a significant impact on sentence state extraction, these words are uniformly replaced by "UNK" labels.

2.3.3. Addition of attention mechanisms. As can be seen from the structure of the S-LSTM network described in the previous section, the network is supervised to generate the overall state of the sentence at the next moment from the hidden state of the vocabulary by learning a fixed weight matrix and then constraining the flow of information through a gate mechanism. This approach can only produce a weight matrix that is universal to all possible inputs because the weight matrix is fixed, and this approach has no way to further focus on the core vocabulary that has a greater impact on the sentence intent, lacks flexibility, and has a high parameter complexity. Meanwhile, it was pointed out in previous work on the ATTENTION mechanism that the result of weighted summation of lexical hidden states through the ATTENTION mechanism can better represent the overall state of the sentence as compared to the hidden states derived using the GATE mechanism, and thus the choice was made to introduce the ATTENTION mechanism [27] to refine the sentence states.

The specific method is to introduce the attention mechanism when generating the overall state of the sentence at the current moment from the hidden state of the word, take the sentence state g_{t-1} of the previous moment as the query Q, take the hidden state h_i corresponding to each word as the Key (K) and Value (V), calculate the weight corresponding to each V by finding the similarity between Q and each K, and then weight and sum V according to this weight to obtain the intermediate state c_i, and then will c_i. The sentence state g_i at the current moment is obtained by activating the constraint of the function and the output gate. Specifically:

$$e_n = a(g_{t-1}, h_i) \tag{19}$$

$$\alpha_{ti} = \frac{\exp(e_{ti})}{\sum_{k=1}^{T} \exp(e_{tk})} \tag{20}$$

$$c_t = \sum_{i=1}^{T} \alpha_{ii} h_i \tag{21}$$

$$g_t = o_t \cdot \tanh(c_t) \tag{22}$$

e_{ti} represents the weight corresponding to the i rd word at moment t, after which e_{ti} is normalized by a softmax function to produce the final practical weight. c_i is multiplied by the tanh activation function and the output gate o_t in the original model to obtain the final sentence state g_t at the current moment.

a for the attention scoring function, commonly used scoring functions are additive model, dot product model, scaled dot product model and bilinear model. The specific formulas are respectively:

$$a(g_{t-1}, h_i) = v^T \tanh(Wh_i + Ug_{t-1}) \tag{23}$$

$$a(g_{t-1}, h_i) = h_i^T g_{t-1} \tag{24}$$

$$a(g_{t-1}, h_i) = \frac{h_i^T g_{t-1}}{\sqrt{d}} \tag{25}$$

$$a(g_{t-1}, h_i) = h_i^T W g_{t-1} \tag{26}$$

where d is the dimension of the hidden state h_i^T. The additive scoring function is activated by the tanh activation function, but the computational steps are complicated and the computational efficiency is not very high. The dot product model directly multiplies the input hidden state and the overall state of the sentence in pairs to derive the corresponding weights, which is simpler to compute and highly efficient due to the use of matrix multiplication. However, in the case of large input dimensions, the variance of the output is often large, and the gradient will be small after processing by softmax, so the problem is solved by adding $\sqrt{d}$ for scaling, which is the scaled dot product model. As for the bilinear model, it can be regarded as the execution of the dot product model after the linear change of the hidden state and the overall state of the sentence respectively.

Given that the dot product model can achieve better computational efficiency, and at the same time similar to transformer, Bert and other advanced models have chosen this attention discriminant function, so this paper chooses the dot product model. The reason why the scaled dot product model is not used is because the convergence speed is slow and the model is not effective in specific experiments. After deriving the similarity and performing weighted summation, using it directly as the sentence state of the new moment can make the model more flexible, reduce the parameters, and reduce the training difficulty. The specific model effect must be verified in the experiment.

2.3.4. Addition of conditional random fields. In many previous works, all in which we input the obtained y_i^S into a Conditional Random Field (CRF) to compute the conditional probability of the corresponding slot value for each of *The Analects* vocabulary words:

$$p(y \mid x) = \frac{1}{Z(x)} \exp\left(\sum_{i,k} \lambda_k t_k(y_{i-1}, y_i, x, i) + \sum_{i,k} \mu_i s_i(y_i, x, i)\right) \tag{27}$$

$$Z(x) = \sum_{y} \exp\left(\sum_{i,k} \lambda_k t_k(y_{i-1}, y_i, x, i) + \sum_{i,l} \mu_i s_i(y_i, x, i)\right) \tag{28}$$

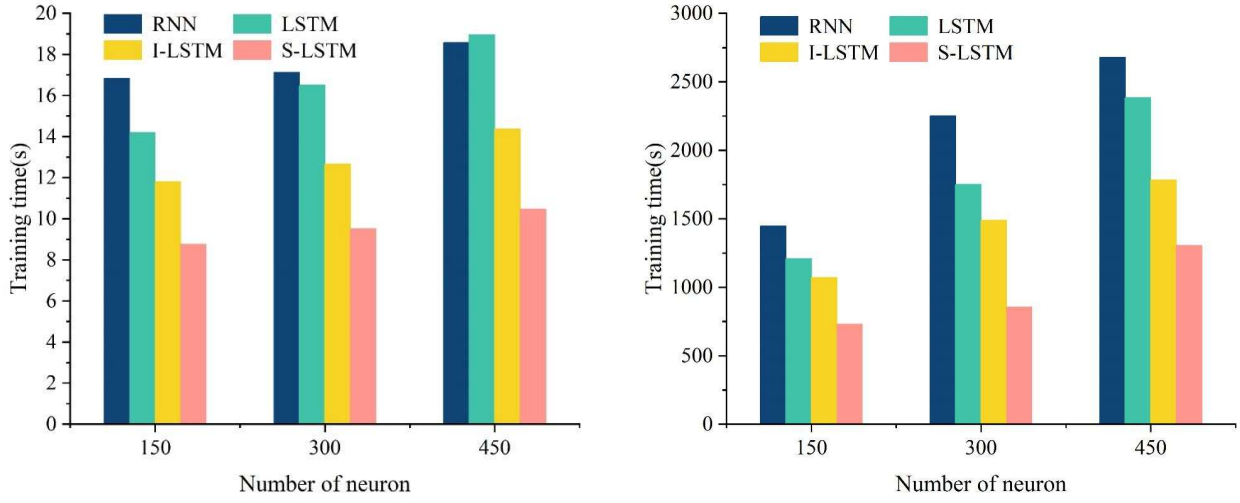
As mentioned earlier, $p(y \mid x)$ is the probability of finding y sequence labels given a sequence of sequence labels. Conditional random fields are utilized as a reinforcement of the relationships between the sequence labels of the Discourse vocabulary so that the constraints that exist between many sequence labels can be learned.

3. Semantic analysis

3.1. Semantic Modeling Analysis

3.1.1. Model time overhead validation and analysis. In order to validate the effectiveness of the proposed model in video analytics, the study conducts semantic analysis experiments with COIN dataset and Charades dataset. The study selected 500 text messages of COIN dataset and 5000 text messages of Charades dataset for model validation. Among them, 350 articles of COIN dataset and 4070 articles of Charades dataset in the training set. The test set COIN dataset 130 articles and Charades dataset 930 articles. Meanwhile, the study uses Spark as the framework of the model validation system, and uses a distributed computing system based on the remote direct data access protocol for validation, and sets the model learning efficiency to 0.1. Firstly, the model time overhead was verified in terms of the number of model neurons and the number of iterations, and RNN, LSTM, and I-LSTM were introduced to compare the results with the S-LSTM model in this paper. Among them, the training overhead of the four models with varying number of neurons is shown in Fig. 2.

A comparison of the training time overhead of the four neural network models in the COIN dataset is shown in Fig. 2(a). It can be seen that the training time overhead increases as the number of neural units increases. Among them, the S-LSTM model requires the lowest training time overall. When the number of neurons is 450, the training time required for S-LSTM is 10.46s, which is 43.67% and 44.83% lower than RNN and LSTM, respectively. Compared with I-LSTM, the training time of S-LSTM is reduced by 27.16%. Comparing the training time overheads of the four models in the Charades dataset in Fig. 2(b), it can be seen that S-LSTM models textual semantic information with a good combination of time persistence, thus avoiding extra computation. In addition, the magnitude of training time change of S-LSTM with different number of neurons can be seen that the frequency of time overhead increase at 300 neurons is significantly less than 450 neurons. This may be due to the fact that the increase in the number of neurons improves the computational efficiency of the model and reduces the training time overhead. However, the communication and memory overhead of the Spark system cluster increases, so the time overhead increases more at 450 neurons.

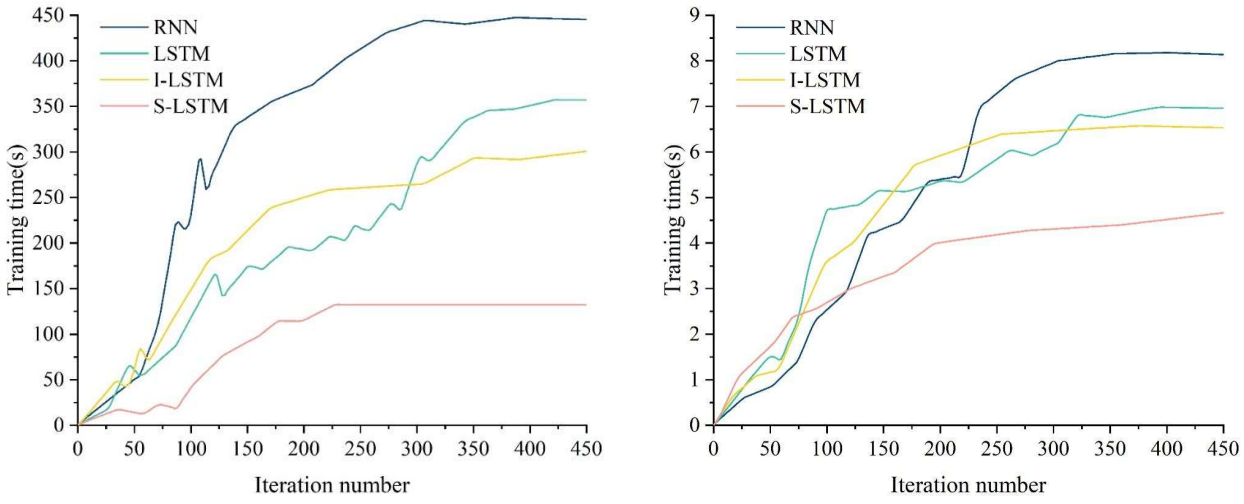


(a) Training time comparison in COIN dataset (b) Training time comparison in Charades dataset

Fig. 2. The comparison of the cost of model training in different neurons

The time overhead of 450 training iterations of the four models on the two datasets is shown in Fig. 3. From Fig. 3(a), it can be seen that the S-LSTM model requires the lowest time overhead for training on the COIN dataset and has the fastest convergence rate. After 227 iterations, the S-LSTM training time stabilizes. In contrast, RNN requires the most time overhead and LSTM converges the slowest.

Comparing the training time of the four models in the Charades dataset in Fig. 3(b), it can be seen that the I-LSTM model training iterations converge faster than S-LSTM. This may be due to the fact that the neurons of the pre-S-LSTM perform concurrent computation as well as the iterative process occupies more memory. Overall, the proposed model can utilize the time-delay feature for processing and analyzing textual information, which not only combines the advantages of neural networks, but also takes into account the characteristics of semantic analysis, and enhances the training convergence speed of the model in the two datasets while reducing the redundant computation of the model.



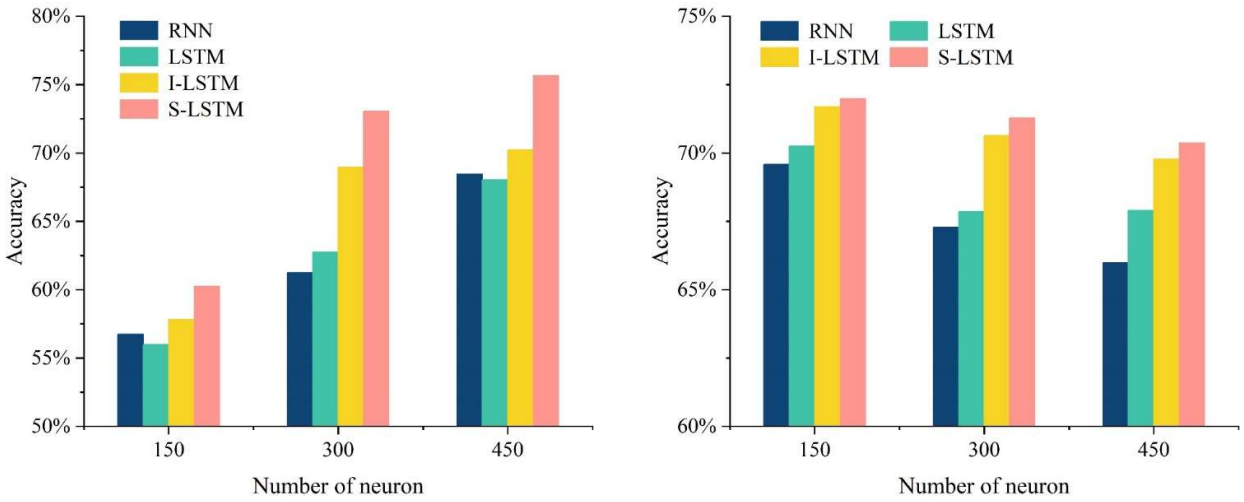
(a) Training time comparison in COIN dataset

(b) Training time comparison in Charades dataset

Fig. 3. Compare the cost of the model training time in different iterations

3.1.2. Model performance validation and analysis. Meanwhile, the study further conducted model performance validation. In order to ensure the homogeneity of the data samples used in the experiments as well as the reliability of the experiments, the study randomly divided the test set into multiple groups and arbitrarily selected a number of groups for performance testing. The accuracy validation of the four models in two data sets with 50 iterations and different numbers of neurons was first conducted. The test results are shown in Fig. 4.

As can be seen from Fig. 4(a), the accuracy of all models increases with the number of neurons. Among them, the S-LSTM model has the highest accuracy. When the number of neurons is 150, the accuracy of S-LSTM model increases by 3.51% and 4.28% over RNN and LSTM, respectively. When the number of neurons is increased to 450, the S-LSTM model accuracy is 75.68%, which is 5.45% to 7.65% more accurate than the other three models. From the test results of the four models in the Charades dataset in Fig. 4(b), it can be seen that the increase in the number of neurons did not lead to an increase in model accuracy. When the number of neurons is 450, the gap between the accuracy of S-LSTM model and I-LSTM model is not obvious. However, the difference in accuracy with RNN and LSTM models is large. Combining the above results, it can be seen that the performance of S-LSTM is more superior. This may be due to the fact that S-LSTM is able to update the gradient of backpropagation based on the text semantic analysis.



(a) Accuracy comparison in COIN dataset (b) Accuracy comparison in Charades dataset

Fig. 4. Comparison of model accuracy in different neurons

In order to compare the performance of the four models more intuitively, the study conducted model accuracy and F1 score tests with 300 neurons and 50 iterations in two datasets, and the specific results are shown in Table 1. It can be seen that in the COIN dataset, the F1 scores of the four models are above 80%. Among them, the accuracy of the RNN model is only 77.58%, and its test result is the worst compared with other models. In contrast, the S-LSTM model has an accuracy of 94.52% and an F1 score of 95.67%.The test results on the Charades dataset further confirm the superiority of S-LSTM. Compared with other models, the accuracy of S-LSTM model increased by 4.60% to 12.05%. This indicates that the improved LSTM language model based on semantic analysis proposed by the study is feasible in the field of text analysis. It also shows that the method proposed by the study can reduce the time efficiency while ensuring the accuracy, thus improving the training efficiency of the language model.

Table 1. Performance comparison of different models in COIN and Charades datasets

Model	COIN dataset		Charades dataset	
	Accuracy (%)	F1 (%)	Accuracy (%)	F1 (%)
RNN	77.58	82.39	75.68	75.38
LSTM	81.06	84.62	75.04	75.16
I-LSTM	86.47	88.96	82.49	85.34
S-LSTM	94.52	95.67	87.09	89.17

3.2. Analysis of the Connotation and Translation of the Word "Body" in The Analects of Confucius

3.2.1. Semantic analysis. There are few direct descriptions of "body" in the Confucian classics, and most of the discussions on "body" are mentioned in the discussion of "benevolence" and "propriety", and are noted as part of the practice of "benevolence" and "propriety". However, it is undeniable that "body" occupies an important position in the Confucian discourse, and self-cultivation is synonymous with Confucian moral cultivation. In Confucius's interpretation, the "body" is no longer a physical life in a purely physiological sense, but a "body symbol" with strong agency and practicality, expressing personal character and blooming the moral meaning of life.

In *The Analects*, Confucius paid some attention to the "body". According to the author's statistics, the word "body" appears 17 times in *The Analects*. Among them, it appears once as a measure word:

"There is a long body and a half." The rest of the place appears "body", such as. "My Heavens and My Body: Seeking for Others and Not Being Loyal? Making friends and not believing them? Are you used to it? Zeng Tzu took loyalty, faithfulness, and knowledge as the criteria for examining himself every day, and if there was any, he changed it, and if he did not, he encouraged it, and as the "body" that practiced loyalty, faithfulness, and knowledge, he could accept introspection. "Wanting to clean oneself and being inconely" is a desire to clean oneself but disturbs the great ethical relationship, so it can be seen that purifying oneself may also have a negative impact on one's own social ethics. "If you are close to the person who is not good, the gentleman will not enter", the gentleman does not approach the person who is personally involved in doing bad things, here, the "body" is the carrier of evil. "Do not lower your will, do not humiliate your body", and "body" can become the object of humiliation. In addition, under the moral code of conduct, courtesy, righteousness, loyalty, and faithfulness, "the king can bring his body", "kill himself to become benevolent", "body" can die or kill, and he must sacrifice his life to fulfill his faith and righteousness and show loyalty. It can be seen that the meaning of "body" is not only a physical body in a physiological sense, it can be saved, caused, forgotten, humiliated, killed, and cleaned..... Although it still refers to the body, it also derives from the meaning of self, self and life. Carrying out various moral practices through the "body" is the "body" of morality.

Zi said: "Self-denial is benevolence." Confucius advocated "benevolence", which externally emphasized the legacy of clan democracy in the ruling system of the Zhou Dynasty and opposed brutal exploitation and oppression. From the inner aspect, it is the perfection and pursuit of individual personality. Confucius emphasized the ability to learn self-discipline and thus to create benevolence. Benevolence is invisible: Ambition is in the body, it is tangible. There must be a set of guidelines to regulate the body, which is "etiquette". Confucius's "rites" are divided into two levels: ritual and essence. "Li" must be concretized through rituals, and the "body" of the body becomes the carrier of "Li". "Don't look at incivility, don't listen to incivility, don't speak incivility, don't move incivility", "benevolence" has found an external constraint scale. Under the continuous carving of self and external etiquette, "etiquette" gradually forms the inner essence of human beings, reaching the highest standard of "benevolence" respected by Confucianism.

To sum up, the "body" in *The Analects* is an external tool for Confucianism to practice personality ideals and improve self-moral cultivation, and it is a physical operation for thinking about self-cultivation. For Confucius, the "body" is not only a private physiological body, but also a functional existence in social relations. Through the use of "body", the benevolence is put into reality.

3.2.2. Translation analysis:

1) Analysis of the structural complexity of the character "body" in translated Chinese, original English and original Chinese. Table 2 shows the descriptive results of the structural complexity of the word "be" in the three text types of *The Analects*. The results of Kruskal-Wallis test showed that there were significant differences in the four indexes among the three text types ($H=374.56$, $p<0.001$ of the distance between name and subject). $H=791.24$, $p<0.001$ of the dependent distance between nouns and passive markers. $H=1726.84$, $p<0.001$. $H=92.67$, $p<0.001$).

The results of post hoc tests showed that the average dependency distances of noun-object dependencies and “body” structures in transliterated Chinese were significantly higher than those in original English ($p<0.001$), while there was no significant difference from those in original Chinese ($p>0.05$). The dependency distance between nouns and passive markers is significantly higher than that of original English and lower than that of original Chinese ($p<0.001$). Dependency distance between nouns and doers is significantly lower than that of original English and higher than that of original Chinese ($p<0.001$). This indicates that the local complexity of noun and subject and the overall complexity of the “body” structure in translated Chinese are not significantly different from those in the original Chinese, while the local complexity of noun and passive markers shows simplifying features and the local complexity of noun and subject shows complicating features.

Table 2. Descriptive results of structural complexity of 3 textual types

Index	Translated Chinese		Original English		Original Chinese	
	Mean	SD	Mean	SD	Mean	SD
Dependent distance between noun and acceptor	6.08	5.42	4.96	4.74	6.57	6.45
Dependent distance between noun and passive mark	2.18	3.41	1.38	0.68	2.26	1.97
Dependent distance between noun and casting object	1.89	2.09	5.17	5.42	1.47	0.97
Average structural dependent distance	3.52	2.75	3.14	2.38	3.63	2.83

2) Analysis of the complexity of the passive structure of English-Chinese translation. Table 3 shows the descriptive results of the passive structural complexity of the translation in English and Chinese. The results of Spearman correlation analysis showed that there was a significant positive correlation between the nouns in source English and the translated Chinese and the average dependence distance of the structure of the word "body" ($r=0.075$, $p=0.033$). The average dependence distance of the character "body" was $r=0.186$, $P<0.001$), but there was no significant correlation between the dependence distance between nouns and passive markers, and between nouns and agents ($P>0.05$). This indicates that the local complexity of the noun and the subject in the source English and the overall complexity of the "be-passive" structure have obvious effects on the structure of the translated Chinese character "body", while the effects of the noun and passive marker, and the local complexity of the noun and the agent are not obvious.

Table 3. Descriptive results of passive structural complexity in English and Chinese

Index	Original English		Translated Chinese	
	Mean	SD	Mean	SD
Dependent distance between noun and acceptor	5.36	5.16	6.17	5.44
Dependent distance between noun and passive mark	1.35	0.68	2.19	2.39
Dependent distance between noun and casting object	4.86	3.19	1.88	1.85
Average structural dependent distance	3.34	2.49	3.62	2.79

3.3. English Translation Strategies

Culture is a collection of symbols, a web of meanings woven by human beings themselves, and in order to gain a true understanding of culture, it is necessary to carefully study the various components of the symbolic system and their interrelationships by means of “ethnography”, which is inevitably based on the re-conceptualization and elaboration of the culture of another people (the researcher).

At the interpretive level, translation is the same as ethnographic description. Appiah, a professor of philosophy at Princeton University in the United States, applied it to translation studies and proposed the concept of "deep translation", the core meaning of which is "deep contextualization, the use of footnotes, annotations, commentaries and other methods in the process of translating texts, so as to place the text in a rich cultural and contextual context, and then integrate the meaning obscured by the text with the translator's intention". Based on the analysis of the semantics and translation of the morpheme "body" in *The Analects*, the author believes that "deep translation" can be used as a translation strategy for the traditional Chinese classic Analects.

“Deep translation” is a specific translation method of the translator, which is to construct a space of interaction between the reader and the cultural and historical context of the text by adding interpretive materials, i.e., to build a ‘web of meaning’ in the translated text, so as to promote the target language reader's respect and understanding of the other's culture and to express respect for the source language's culture. In other words, a “web of meaning” is constructed in the translated text, so as to promote the target language readers' respect for and understanding of other cultures, and to express their respect for the source language culture. In terms of effect, “this type of translation combines

translation with rigorous academic research, and actually belongs to the category of academic translation, which is suitable for cultural books, academic works and a few literary works that contain rich cultural information, and its recipients are foreign readers and researchers who are interested in the original text and the culture behind it".

In terms of the presentation of the translated text, "in-depth translation" "can be crystallized into footnotes, endnotes, notes, two-line notes, in-text cryptic notes, as well as prefaces, treks, dedications, postscripts, appendices, glossaries, acknowledgements, and other categories". In order to achieve deep translation, Appiah suggests that translators reconstruct the historical context in which the original text was produced - the web of meanings - through various annotations, in order to help readers of the translated language understand the culture of the source language more deeply. Deep translation is in essence incremental translation, and the added information about the source language and source culture is mainly placed in the annotation section, the concept of which itself encompasses the method of implementing deep translation.

4. Conclusion

This paper takes the character "body" in *The Analects* as the research object, constructs a natural language understanding model based on S-LSTM to analyze it, excavates its philosophical connotation, and explores the translation strategies of traditional Chinese classics such as *The Analects*.

In comparison with other semantic models, the S-LSTM model requires the lowest training time overall. When the number of neurons is 150, 300, and 450, the training time required by S-LSTM is 8.76s, 9.51s, and 10.46s on the COIN dataset, and 728.86s, 856.13s, and 1303.4s on the Charades dataset, respectively. All of which are the smallest among all the test results. At the same time, reducing the redundant computation of the model improves the training convergence speed. When the number of neurons is 150, 300, 450, the accuracy of S-LSTM model is 69.58%, 67.28%, 65.98% on COIN dataset, and 71.98%, 71.28%, 70.37% on Charades dataset, respectively.The accuracy of S-LSTM model is the highest among all models.

Through the analysis of the complexity of the structure of the character "body" in *The Analects*, it is found that the local complexity of translated Chinese nouns and deeds is characterized by complication. The local complexity of the noun and the subject in the source English language, and the overall complexity of the "be-passive" structure have obvious effects on the structure of the translated Chinese character "body".

References

[1] Yang, L., & Zhou, G. (2024). Dissecting The Analects: an NLP-based exploration of semantic similarities and differences across English translations. Humanities and Social Sciences Communications, 11(1), 1-12.

[2] Li, Y., Xiang, R., Xiu, S., & Qu, W. (2024). A Study on Translation of Numbers in the Analects of Confucius from the Perspective of Cultural Translation Theory——a Case study of the Translation Version by Gu Hongming. In SHS Web of Conferences (Vol. 185, p. 01015). EDP Sciences.

[3] Min, F. (2021). A critical study on translation of the analects: An ideological perspective. International Journal of Translation, Interpretation, and Applied Linguistics (IJTIAL), 3(1), 45-54.

[4] Hu, W., & Jia, X. (2021). Comparative Analysis of Two Famous English Versions of Analects of Confucius. Sino-US English Teaching, 18(10), 279-283.

[5] Tao, Y. (2018). FROM MONOLOGUE TO DIALOGUE: WESTERN TRANSLATORS'PERSPECTIVES ON TRANSLATING KEY CULTURAL CONCEPTS IN THE ANALECTS. Translation Review, 102(1), 46-66.

[6] He, M. (2017). A comparative multidimensional study of the English translation of Lunyu (The Analects): A corpus-based analysis. GEMA Online Journal of Language Studies, 17(3), 37-54.

[7] Fangfang, Q., & Huanhai, F. (2017). A Comparison between the English Translations of The Analects of Confucius by James Legge and Ku Hung-Ming. In Asia International Symposium on Language, Literature and Translation (p. 78).

[8] Shi, Y., & Hu, W. (2024). A Study on the Translation of Culturally Unique Chinese Terms Based on Corpora—Taking English Translations of The Analects as an Example. Journal of Linguistics and Communication Studies, 3(4), 43-47.

[9] Tsai, C. C. (2018). The Analects (pp. 45-192). Princeton University Press: Princeton, NJ, USA.

[10] Bergeton, U. (2019). Found (and lost?) in translation: Culture in The Analects. Harvard Journal of Asiatic Studies, 79(1), 49-95.

[11] Zhou, Q. (2023). A Study of Xu Yuanchong's English Translation of the Analects of Confucius from the Perspective Eco-translatology. International Journal of Education and Humanities, 6(2), 72-75.

[12] Liu, H. (2020, December). A study on the reader reception of English translations of The Analects-data analysis with Python. In 2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE) (pp. 2381-2387). IEEE.

[13] Chen, M. (2022). A Study of the French Translation Strategy of High-frequency Culture-loaded Terms in The Analects of Confucius Based on Corpus. Proceedings of the 9th ICELAIC.

[14] Tao, Y., & Li, W. (2024). Reading Confucius in translation: An empirical study of western academic readers' responses to English translations of The Analects. Translation and Interpreting Studies.

[15] Lin, Q. (2024). Visualization Analysis of the Translation and Dissemination of The Analects Based on CiteSpace (2013-2023). Journal of Humanities and Social Sciences Studies, 6(5), 126-132.

[16] Min, F. (2021). Understanding and translating Confucian philosophy in the Analect s: a sociosemiotic perspective. Semiotica, 2021(239), 287-306.

[17] Liu, Z., & Sun, C. (2024). Direction of translation and textual cohesion: A study of two Spanish translations of The Analects using Coh-Metrix-Esp. CÍRCULO de Linguística Aplicada a la Comunicación, 100, 29.

[18] Ni, P. (2017). Understanding the Analects of Confucius: A new translation of Lunyu with annotations. State University of New York Press.

[19] Wang, Y. (2024). Retranslation of The Analects of Confucius and the Causes for Retranslation. Journal of Art, Culture and Philosophical Studies, 1(1).

[20] Youlan, T. A. O., & Yiyi, H. U. (2024). A Deep-Learning-Driven Study on the Reception of Translated Works: A Case Study of Sentiment Analysis for Readers' Reviews on The Analects Translated by DC Lau. Journal of Foreign Languages, 47(6), 72-80.

[21] Qiu, J., & Zhang, J. (2018). Getting to Know Confucius: A New Translation of the Analects. Translated by Lin Wusun.(Beijing: Foreign Languages Press, 2010. v+ 23, Pp. 359. Paperback. ISBN 978-7-119-06165-8). Journal of Chinese Philosophy, 45(3-4), 261-263.

[22] Yang, D. (2022). An Investigation on English Translations of Culture-Loaded Words in The Analects of Confucius from the Eco Perspective: A Case Study of the English Translation of Lectures on China's Traditional Political Thoughts. Editorial Board, 7.

[23] Hou, Y., & Sun, Y. (2019). A Corpus-Based Comparative Analysis of Cohesive Devices in Two English Translations of The Analects of Confucius. International Journal of Languages. Literature and Linguistics, 5(4), 247-252.

[24] Brielle C. Stark,Sarah Grace Dalton & Alyssa M. Lanzi. (2025). Access to context-specific lexical-semantic information during discourse tasks differentiates speakers with latent aphasia, mild cognitive impairment, and cognitively healthy adults. Frontiers in Human Neuroscience1500735-1500735.

[25] Avan Kumar,Harshitha Chandra Jami,Bhavik R. Bakshi,Manojkumar Ramteke & Hariprasad Kodamana. (2025). An evolutionary study on technologies for polyethylene terephthalate waste recycling using natural language processing. Computers and Chemical Engineering109011-109011.

[26] Sharma Tripti & Prasad Sanjeev Kumar. (2024). Enhancing cybersecurity in IoT networks: SLSTM-WCO algorithm for anomaly detection. Peer-to-Peer Networking and Applications(4),2237-2258.

[27] Syed Afraz Hussain Shah,Ubaid Ahmed,Muhammad Bilal,Ahsan Raza Khan,Sohail Razzaq,Imran Aziz & Anzar Mahmood. (2025). Improved electric load forecasting using quantile long short-term memory network with dual attention mechanism. Energy Reports2343-2353.