JCSE Journal of Computing Science and Engineering

Regular Paper
Journal of Computing Science and Engineering,
Vol. 19, No. 3, September 2025, pp. 94-100

English Classical Translation under the Knowledge Difference Based on Transformer

Chuanxue Zhang*
School of Foreign Languages, Zhengzhou Shengda University, Xinzheng, Henan, China
zcx_xue@outlook.com

Abstract

When performing machine translation on poems in English ancient literature and classics, this paper adopted a two-stage translation method. In the first stage, the Transformer model was used to translate English poems into vernacular translations. In the second stage, an encoder and a decoder constructed with a long short-term memory (LSTM) were used to convert the vernacular translations into Chinese poems. Meanwhile, a back-translation strategy was adopted when training the encoder and decoder in the second stage. After that, simulation experiments were carried out. In the experiments, the two-stage algorithm was compared with the multilingual bidirectional and auto-regressive transformers (mBART), traditional LSTM, and traditional Transformer models. The findings suggest that the translation algorithm can accurately translate English poetry and align the translated text more closely with the stylistic characteristics of Chinese poetry.

Category: Natural language processing

Keywords: English poetry; Translation; Chinese poem; Transformer

I. INTRODUCTION

In the context of high integration of globalization and informatization, cross-linguistic and cross-cultural exchanges are becoming increasingly frequent. As a crucial bridge connecting different civilization systems, translation is gaining more prominence in its status and role [1, 2]. Among various types of translation, poetry translation is considered challenging due to its deep historical and cultural background and unique language style. Classical poetry not only differs from modern English in linguistic structure but also presents a high level of complexity in rhetorical devices, semantic expressions, and cultural connotations [3]. Traditional translation methods, such as rule-based or statistics-based translation algorithms, often struggle with poetry translation, making it difficult to fully capture the deep semantic relationships in the text and accurately convey the cultural connotations of the original text [4]. With the rapid advancements in artificial intelligence technology, the neural network machine translation algorithm represented by the Transformer model opens up new possibilities for English poetry translation. Zhang et al. [5] proposed a deep feature fusion model and integrated it into a deep learning architecture for sentence matching task. Guan et al. [6] utilized word dependencies to match similar sentences, with experimental results demonstrating that combining word and word dependencies enhances the ability to extract matching features between two sentences. Based on a deep generalized left-right (GLR) parsing model, Wang and Zhao [7] utilized various methods of parameter initialization in the embedding layer and multi-layer computation methods in the sentence decoder,

Open Access http://dx.doi.org/10.5626/JCSE.2025.19.3.94 http://jcse.kiise.org

Received 2 September 2025; **Revised** 11 October 2025; **Accepted** 10 November 2025
*Corresponding Author

 pISSN: 1976-4677 eISSN: 2093-8020

to establish a method framework for English long and short sentence translation and recognition. Experimental results indicated that the deep GLR model-based method improved the accuracy of the model parameters. This paper applied the Transformer model and to the initial stage of English poetry translation to obtain the vernacular translation. Subsequently, the encoder and decoder, constructed by the long short-term memory (LSTM) algorithm, were employed in the second stage to convert the vernacular translation into Chinese poetry. The encoder-decoder training in the second stage utilized the back-translation strategy. Moreover, simulation experiments were conducted. The novelty of this paper lies in adopting a two-stage form to translate English poems. In the process of directly converting English poems into Chinese poems, traditional single-stage translation algorithms need to take into account both the conveyance of the poem's meaning and the transformation of different poetic styles between the two languages. This often leads to an antagonism between the two goals, potentially producing results that satisfy neither the accurate conveyance of meaning nor the transformation of poetic styles. However, the two-stage translation algorithm translated English poems into vernacular Chinese, and then the vernacular Chinese was converted into Chinese poems. This approach reduced the difficulty of directly translating English poems into Chinese poems. At the same time, the back-translation strategy was used to improve the accuracy of converting vernacular Chinese into Chinese poems.

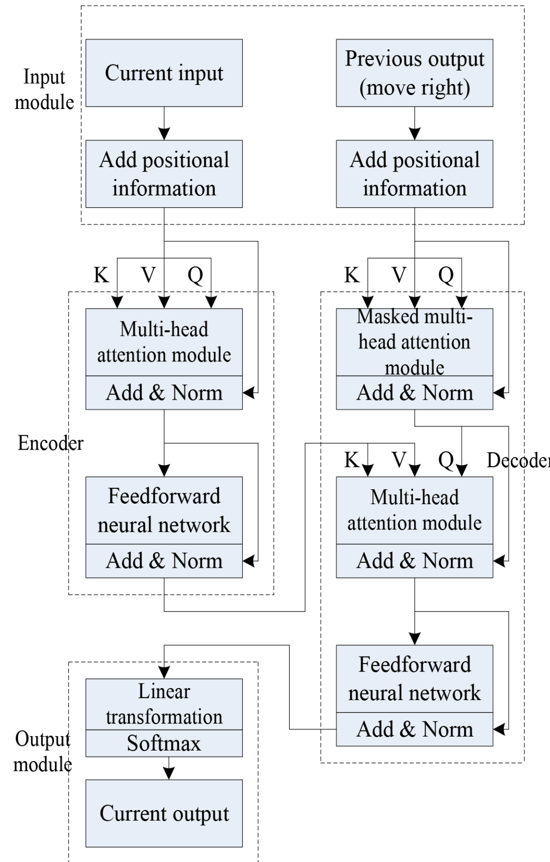

Fig. 1. Basic structure of the Transformer model.

II. ENGLISH POETRY TRANSLATION BASED ON TRANSFORMER

A. Transformer Model

In this paper, the Transformer model is employed for translating English poetry, and its fundamental structure is illustrated in Fig. 1. The input module comprises the current input of the sequence text and the output obtained from the calculation of the previous input, i.e., the previous output [8]. Both inputs require the addition of positional information before the "current input" is fed into the encoder and the "previous output" is input to the decoder. The calculation formula of the multi-head attention module in the encoder for the current input:

$$\begin{cases} Q_h = XW_Q^h \\ K_h = XW_K^h \\ V_h = XW_V^h \\ Z_h = attention(Q_h, K_h, V_h) \\ MultiHead = Linear(Concat(Z_1, Z_2, \cdots Z_h)) \end{cases} \tag{1}$$

where Q_h, K_h, V_h are the query, key, and value of the h-th attention head, respectively, X is the input text, W_Q^h, W_K^h, W_V^h are the trainable weight matrices used to compute Q_h, K_h, V_h, *attention*() is the computation function of the attention score, *Concat*() is the concatenation function of the attention score of multiple attention heads [9], *Linear*() is the linear transformation function, and *MultiHead* is the final output of the multi-head attention. The results computed by the multi-head attention module and the original input of this module are computed in the residual connection normalization module, the computed result is forward computed in the feed-forward neural network, the output is processed with the residual connection normalization module, and the processed result is input to the decoder [10].

The difference between the masked multi-head attention module in the decoder and the multi-head attention module lies only in introducing lower triangular matrix *Mask* in the formula for calculating attention score Z_h, which ensures that the attention head can only see the current and previous text. The "previous output" is input to the decoder after adding the positional information and then calculated by the masked multi-head attention module [11]. The result is processed with the residual connection normalization module. The processed result is input into the multi-head attention module together with the encoder's calculation result. The former and W_Q^h constitute Q_h, and the latter and W_K^h, W_V^h constitute K_h, V_h. Then, the calculation result is input into the output module to obtain the character probability distribution of translation through the linear transformation and softmax function. The character sequence of translation with the highest probability is selected.

B. Translation of English Poetry

Machine translation of English poems can translate English into relatively straightforward Chinese. However, poetry is a distinct literary genre, and there are format disparities between English and Chinese poems. Utilizing machine translation algorithms for the vernacular translation of English poetry cannot convey the format, rhyme, and metaphor although it can convey the content. To enhance the translation quality of English poetry, this paper introduced a two-stage model combining machine translation and poem generation for translating English poems. The operational principle of this model is converting English poems into vernacular language using a machine translation algorithm in the first stage, followed by transforming the vernacular language into the form of Chinese poetry utilizing a poetry generation algorithm in the second stage [12].

As mentioned earlier, the entire translation process is divided into two stages: using the Transformer model [13] to translate English poems into vernacular language and converting vernacular language into the Chinese poetry form. The conversion of vernacular language into Chinese poetry can be seen as a form of machine translation in principle. Hence, this paper also employed the encoder-decoder method to process the vernacular. Given that the Transformer model used in the first stage has been elaborated in the previous text, we will not describe it further and will concentrate on the second stage. During the training of the translation model in the second stage, to address the shortage of parallel corpus data between vernacular and Chinese poems, a back-translation strategy is utilized to create a pseudo-parallel corpus. This back-translation strategy converts the vernacular language into Chinese verse through an encoder and a decoder. Subsequently, noise is introduced to the verse, and it is converted back to vernacular language through another set of encoder and decoder. The parameters of the encoders and decoders are adjusted based on the disparities between the converted vernacular language and the original vernacular language. The training workflow of the algorithm for translating English poetry is depicted in Fig. 2.

1) An English poem is input and translated into Chinese vernacular text through the Transformer model.

2) The Chinese vernacular text is fed into the vernacular encoder 1 for forward computation. The vernacular encoder utilizes the LSTM algorithm, and the sequence of hidden states generated from the LSTM computation during forward computation is used as the intermediate vector.

3) The intermediate vector is input into the poem decoder 1 for forward computation. The poem decoder also utilizes the LSTM algorithm, and during forward computation, it generates the character probability distribution of the Chinese poem. Subsequently, the sequence with the highest probability is selected [14].

4) Noise is added to the converted Chinese poem by randomly deleting some words and shifting word positions.

5) The noise-processed Chinese poem is fed into poem encoder 2 for forward computation. The sequence of hidden states calculated after calculating using the LSTM algorithm is used as the intermediate vector.

6) The intermediate vector is passed to vernacular decoder 2 for forward computation. During this process, the forward computation generates a probability distribution for the vernacular text. Subsequently, the sequence with the highest probability is chosen.

7) Whether the training can be terminated is determined. The termination criteria include reaching the predefined number of training sessions or if the loss function has converged. The formula for the loss function is:

$$\begin{cases} l_{lm} = \underset{S\in S}{E}(-\log P(S|D_s(E_s(S_N))) + \underset{T\in T}{E}(-\log P(T|D_t(E_t(T_N))) \\ l_{bt} = \underset{S\in S}{E}(-\log P(S|D_s(E_t(T_S))) + \underset{T\in T}{E}(-\log P(T|D_t(E_s(S_T))) \\ l = \alpha_1 l_{lm} + \alpha_2 l_{bt} \end{cases} \tag{2}$$

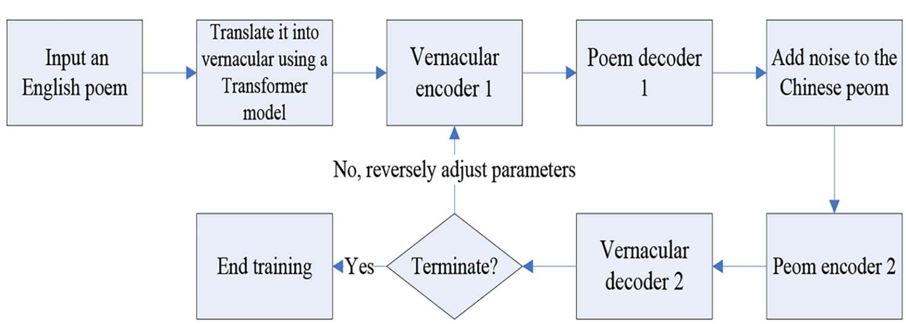

Fig. 2. Training process for translating English poems.

where l_{lm} is the language modeling loss, l_{bt} is the back-translation loss, l is the composite loss, α_1 and α_2 are scaling factors, S is the vernacular language, T is the Chinese poem, S_N and T_N are the noise-added vernacular language and Chinese poem, T_S is the Chinese poem obtained after translating S, S_T is the Chinese poem obtained after translating T, E_s and E_t are the encoder function for vernacular languages and Chinese poems, D_s and D_t are the decoder function for vernacular languages and Chinese poems.

8) Upon reaching the termination condition, the training stops; otherwise, the composite loss is calculated using Eq. (2), and the parameters of the two sets of encoders and decoders are adjusted inversely based on the calculated loss.

III. SIMULATION EXPERIMENTS

A. Experimental Data

The experimental data utilized in this paper were sourced from the Chinese English Conference Proceedings Parallel Corpus - Core (CECPC-Core), which is a research output of the major project funded by the National Social Science Foundation of China. The text has undergone processing involving sentence alignment and lexical assignment. Among the data from this corpus, 800 English poems were chosen, with 70% designated for the training set and 30% for the test set. Since the two-stage machine translation algorithm adopted in this paper first translated English poems into vernacular Chinese and then converted them into the form of Chinese poems, and the translation models of the two stages were trained independently, in addition to the English poems and their parallel translations in the corresponding format in the sample set, a vernacular translation of the English poems also needs. This paper used the form of manual translation to obtain the vernacular translations of English poems.

B. Experimental Setup

The English poem translation algorithm proposed in this paper was divided into two phases. The first phase involved translating an English poem into Chinese vernacular using the Transformer model. This phase included the following relevant parameters: a six-layer encoder, a six-layer decoder, 24 attention heads in both the encoder and decoder, and a rectified linear unit (ReLU) activation function for the feed-forward neural network. In the second phase, the Chinese vernacular text was further converted into Chinese poetic text. This process employed the LSTM algorithm to construct the encoder and decoder. Through orthogonal experiments, the parameters were 512 nodes in the LSTM input layer, 1,024 nodes in the hidden layer, and the ReLU activation function. The output layer of the decoder utilized beam search [15] to output the character sequence with the highest probability, and the specification of the beamwidth was 10. The translation algorithm was trained with a learning rate of 0.02 and a maximum of 400 training iterations. The other two translation algorithms were introduced to compare with the two-stage algorithm. They did not follow the two-stage format but directly translated English poems into Chinese poems. One of them was the traditional Transformer model, and the other one used the traditional LSTM algorithm to construct the encoder and decoder. The two-stage algorithm was also compared with the multilingual bidirectional and auto-regressive transformers (mBART) model [16].

C. Evaluation Criteria

Both bilingual evaluation understudy (BLEU) indicator and manual evaluation were used to evaluate the performance of the English poem translation algorithm. The BLEU indicator can be calculated using:

$$BLEU = B \cdot \exp\left(\sum_{n=1}^{N} \omega_n \log p_n\right) \tag{3}$$

where ω_n is the weight of a word, p_n is the word percentage, and B is the penalty factor.

For the human evaluation, 30 native Chinese-speaking experts with a Bachelor of Arts degree were invited to assess the translated Chinese poems. The evaluation covered five perspectives: fluency, semantic coherence, translation consistency, poetic quality, and translation score. Each perspective was rated on a scale of 1 to 5.

D. Experimental results

The partial results of the four translation algorithms for an English poem are presented in Table 1. It can be seen that the mBART, traditional LSTM, and traditional Transformer models tended to favor direct translation and output text formats that closely resemble the original text. In contrast, the two-stage algorithm proposed in this paper not only captured the essence of the original text but also maintained a more poetic Chinese format in the translation.

The BLEU indicator was employed to objectively score the four translation algorithms, and the BLEU scores of the algorithms under different N-gram precisions are displayed in Table 2. As the N-gram size increased, the BLEU scores of all four translation algorithms decreased. Moreover, under the same N-gram precision, the two-stage algorithm achieved the highest BLEU score, followed by the traditional Transformer model and the mBART model, and the traditional LSTM yielded the lowest BLEU score.

Finally, the four translation algorithms were manually evaluated by experts, and the results are depicted in Fig. 3. The figure illustrates that in terms of fluency, semantic

Table 1. Translations of an English poem under different translation algorithms

	Example 1	Example 2
Original text	You say that you love rain, but you open your umbrella when it rains... You say that you love the sun, but you find a shadow spot when the sun shines... You say that you love the wind, But you close your windows when wind blows... This is why I am afraid; You say that you love me too...	When you are old and grey and full of sleep, And nodding by the fire, take down this book, And slowly read, and dream of the soft look. Your eyes had once, and of their shadows deep; How many loved your moments of glad grace, And loved your beauty with love false or true, But one man loved the pilgrim Soul in you, And loved the sorrows of your changing face; And bending down beside the glowing bars, Murmur, a little sadly, how Love fled. And paced upon the mountains overhead, And hid his face amid a crowd of stars.
Reference translation	君言爱雨雨初降，伞底犹开影两行。 暖日倾城光潋滟，却寻幽处避辉芒。 风来入户帘先闭，语到情深意自藏。 闻道于君亦怜我，怎堪心事落斜阳？	美人迟暮云鬓白，盹寐炉畔孤难耐； 取下此卷细读来，梦萦柔瞳伴绿黛。 多少公子馋玉貌，真情假意恋媚娇； 唯有某君不同调，爱汝心灵及衰老。 炽热炉边身趋前，戚戚诉说悲百感； 逝去之爱步高山，繁星之中隐其颜。
The translation of the traditional LSTM algorithm	你说你喜欢雨， 但下雨的时候你却撑开伞 …… 你说你爱太阳， 但当阳光普照时，你会找到一个阴暗的地方 …… 你说你爱风， 但当风吹起的时候，你关上了窗户 …… 这就是我害怕的原因； 你说你也爱我 ……	当你老了、灰了、睡着了， 在火边打盹时，取下这本书， 然后慢慢阅读，梦见轻柔的眼光。 你的眼睛过去有深深的阴影； 多少爱在你年轻的岁月， 爱你的美无论这爱是真是假， 但一个男人爱你的灵魂， 也爱你变化容貌中的皱纹； 在升起的炉子旁垂头， 轻述，一点点悲伤，为何爱会消逝。 在山顶缓慢踱步， 在群星中隐藏他的脸庞。
The translation of the traditional Transformer algorithm	你说你爱雨， 可细雨落下的那一刻，你却撑伞而行，将它隔于咫尺之外 …… 你说你爱阳光， 可在暖阳倾洒的午后，你却寻一处阴凉，避开那片温柔的明亮 …… 你说你爱风， 可当清风拂面时，你却悄然合上窗棂，将它的轻语拒之门外 …… 这便是我心底的惶然 —— 你也说，你爱我 …… 可我会想，在你靠近的时候，是否也在悄悄设限？	当你老了，头白了，睡意昏沉， 炉火旁打盹，请取下这部诗歌， 慢慢读，回想你过去眼神的柔和， 回想它们昔日浓重的阴影； 多少人爱你青春欢畅的时辰， 爱慕你的美丽，假意或真心， 只有一个人爱你那朝圣者的灵魂， 爱你衰老了的脸上痛苦的皱纹； 垂下头来，在红光闪耀的炉子旁， 凄然地轻轻诉说那爱情的消逝， 在头顶的山上它缓缓踱着步子， 在一群星星中间隐藏着脸庞。
mBART	你爱雨，你这么说 但你在雨落时撑开伞 你爱阳光，你这么说 但你在光照时躲进阴影 你爱风，你这么说 但你在风起时关上了窗户 这便是我的害怕之处 因你也说了你爱我 ……	当你年老头白昏昏欲睡， 在炉边打盹前取下诗歌， 一边阅读一边回忆过往温柔的眼神， 回想其过往浓厚的阴影； 有多少人或真心或假意地爱慕你的年轻时光、你的美丽。 唯有一人喜爱你那灵魂，以及你那衰老面容上的皱纹； 在发光的炉子畔垂下头来缓缓诉说爱情的消失， 在山顶上缓缓踱着步子，将脸庞藏于群星之中。
The translation of the two-stage algorithm	君言爱雨雨初降，伞底犹开影两行。 暖日倾城光潋滟，却寻幽处避辉芒。 风来入户帘先闭，语到情深意自藏。 闻道于君亦怜我，怎堪心事落斜阳？	美人迟暮云鬓白，盹寐炉畔孤难耐； 取下此卷细读来，梦萦柔瞳伴绿黛。 多少公子馋玉貌，真情假意恋媚娇； 唯有某君不同调，爱汝心灵及衰老。 炽热炉边身趋前，戚戚诉说悲百感； 逝去之爱步高山，繁星之中隐其颜。

Table 2. BLEU of different translation algorithms with different N-gram precisions

	BLEU score			
	1-gram	**2-gram**	**3-gram**	**4-gram**
The translation of the traditional LSTM algorithm	6.95	4.58	3.12	2.15
The translation of the traditional Transformer algorithm	10.31	8.95	6.87	4.21
mBART	10.36	8.89	6.97	4.25
The translation of the two-stage algorithm	18.54	17.11	15.87	13.24

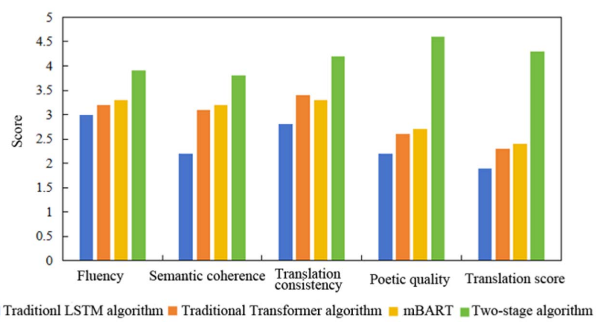

Fig. 3. Manual scoring.

coherence, translation consistency, poetic quality, and translation score, the two-stage algorithm received the highest scores, followed by the traditional Transformer model and the mBART model, and the traditional LSTM received the lowest scores.

IV. CONCLUSION

This paper applied the Transformer model in the initial stage of English poem translation to generate vernacular translations. Subsequently, the encoder and decoder built using the LSTM algorithm were utilized in the second stage to convert the vernacular translations into Chinese poems. The training of the encoder and decoder in the second stage used the back-translation strategy. Simulation experiments were conducted to compare the outcomes with those achieved by the mBART, traditional LSTM, traditional Transformer algorithms. It was found that the two-stage algorithm not only captured the essence of the original text but also output a translation that adheres more closely to the Chinese poetic format. The BLEU score of the two-stage algorithm was the highest, followed by the mBART and traditional Transformer algorithm, while the traditional LSTM algorithm yielded the lowest BLEU score. The scores obtained by the two-stage algorithm ranked the highest, followed by the traditional Transformer and mBART algorithms, and the traditional LSTM algorithm obtained the lowest scores.

CONFLICT OF INTEREST

The author has declared that no competing interests exist.

REFERENCES

1. H. Huang, “The Cooperative principle based study on passive sentence translation in science and technology English—to improve the quality and efficiency of non literature translation,” *Open Journal of Modern Linguistics*, *12*(2), 179-187, 2022. https://doi.org/10.4236/ojml.2022.122014
2. Y. Ababneh and R. M. Al Rousan, “The influence of translation on the Arabic sentence structure: electronic journalism as a model,” *Sino-US English Teaching*, vol. 17, no. 1, pp. 18-27, 2020.
3. M. Araujo, A. Pereira, and F. Benevenuto, “A comparative study of machine translation for multilingual sentence-level sentiment analysis,” *Information Sciences*, vol. 512, pp. 1078-1102, 2020. https://doi.org/10.1016/j.ins.2019.10.031
4. T. Zhang and J. Wu, “Study on error correction method of English long sentence translation based on support vector machine,” *International Journal of Reasoning-Based Intelligent Systems*, vol. 15, no. 3-4, pp. 284-289, 2023. https://doi.org/10.1504/IJRIS.2023.136395
5. X. Zhang, W. Lu, F. Li, X. Peng, and R. Zhang, “Deep feature fusion model for sentence semantic matching,” *Computers, Materials and Continua*, vol. 61, no. 2, pp. 601-616, 2019. https://doi.org/10.32604/cmc.2019.06045
6. X. Guan, J. Han, Z. Liu, and M. Zhang, “Sentence similarity algorithm based on fused bi-channel dependency matching feature,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 34, no. 7, article no. 2050019, 2020. https://doi.org/10.1142/S0218001420500196
7. H. Wang and C. Zhao, “English long and short sentence translation and recognition method based on deep GLR model,” *Computational Intelligence and Neuroscience*, vol. 2022, article no. 3119477, 2022. https://doi.org/10.1155/2022/3119477
8. Q. Yang, L. Yu, S. Tian, and J. Song, “Collaborative semantic representation network for metaphor detection,” *Applied Soft Computing*, vol. 113, article no. 107911, 2021. https://doi.org/10.1016/j.asoc.2021.107911
9. X. Yang, T. Zhang, and C. Xu, “Semantic feature mining for video event understanding,” *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol.

12, no. 4, article no. 55, 2016. https://doi.org/10.1145/2962719
10. H. Xi, "The design of complex semantic machine translation model for foreign linguistics," *Boletin Tecnico,* vol. 55, no. 15, pp. 473-481, 2017.
11. M. Liu, L. Zhang, H. Hu, L. Nie, and J. Dai, "A classification model for semantic entailment recognition with feature combination," *Neurocomputing*, vol. 208, pp. 127-135, 2016. https://doi.org/10.1016/j.neucom.2016.01.096
12. Z. Peng, "The approaches of internet public opinion research," *Library Journal*, vol. 35, no. 12, pp. 63-68, 2016. https://doi.org/10.13663/j.cnki.lj.2016.12.009
13. A. E. Alpysbayeva, S. A. Ashymkhanova, and M. A. Buribayeva, "Translator's laboratory or translation history of m. zhumabayev's poetry into English," *Keruen*, vol. 85, no. 4, pp. 240-252, 2024. https://doi.org/10.53871/2078-8134.2024.4-18
14. Y. Xia, "Reproduction, variation and reconstruction of the Bashu regional image in the English translation of tang poetry," *International Journal of Language and Linguistics*, vol. 13, no. 1, pp. 1-6, 2025. https://doi.org/10.11648/j.ijll.20251301.11
15. R. K. Chakrawarti, H. Mishra, and P. Bansal, "Review of machine translation techniques for idea of Hindi to English idiom translation," *International Journal of Computational Intelligence Research*, vol. 13, no. 5, pp. 1059-1071, 2017.
16. Y. Liu, J. Gu, N. Goyal, X. Li, S. Edunov, M. Ghazvininejad, M. Lewis, and L. Zettlemoyer, "Multilingual denoising pre-training for neural machine translation," *Transactions of the Association for Computational Linguistics*, vol. 8, pp. 726-742, 2020. https://doi.org/10.1162/tacl_a_00343

Chuanxue Zhang https://orcid.org/0009-0009-9983-9036

Chuanxue Zhang received her Master's degree in English language and literature from Faculty of Foreign Languages of Ningbo University in 2012; received her bachelor's degree in English from Faculty of Foreign Languages of Huarui College of Xinyang Normal University in 2009. She is currently an English teacher in School of Foreign Languages of Zhengzhou Shengda University. Her research interests include the theory and practice of English teaching, the literature of English.