# Study on Phrase Processing of English Texts by Neural Machine Translation

Danqiangyu Zhou

*Civil Aviation Flight University of China, Guanghan, Sichuan 618307, China*
*E-mail: zhoudanqiangy@outlook.com*

Received 29 October 2025; Accepted 19 March 2026

## Abstract

The current research on neural machine translation (NMT) rarely involves phrase processing, which leads to poor translation quality. This paper first gives a brief introduction to NMT and the Transformer model. Then, a statistical machine translation (SMT)-based phrase processing method that adds phrases in different suffix forms to the source-end sentences was proposed to improve translation quality. Experiments were conducted on the China Workshop on Machine Translation 2018 (CWMT2018) dataset (Chinese-English) and the WMT2014 dataset (English-German). The results showed that, among the three suffix forms, only adding the target phrase sequence in the suffix form was conducive to improving the translation quality of the Transformer model: the mean bilingual evaluation understudy (BLEU) value increased by 0.0254 on the Chinese-English dataset and by 0.0105 on the English-German dataset compared with the baseline model. Compared with NMT models such as seq2seq, the Transformer model combined with phrase processing obtained the best BLEU value, and the resulting translation was more in line with the reference translation. The results verify that the proposed method is reliable and can be applied in practice.

**Keywords:** Neural machine translation, English text, phrase, bilingual evaluation understudy.

*Journal of ICT Standardization, Vol. 14_3,* 295–310.
doi: 10.13052/jicts2245-800X.1431

## 1 Introduction

The increasing need for communication worldwide has led to the frequent transmission of multilingual information in various fields such as news and entertainment, which puts forward higher requirements for natural language processing (NLP) [1]. Machine translation (MT) is the most widely used NLP technology [2]. From statistical machine translation (SMT) to neural machine translation (NMT) [3], more and more methods have been applied. However, there are still problems with translation quality in various aspects, such as difficult translation of rare words and low translation loyalty. As NMT is the mainstream method [4], how to further improve the translation quality of NMT has become the focus of researchers' attention [5]. Yirmibesoglu et al. [6] analyzed low-resource Turkish-English NMT using Transformer architecture, compared the effects of different data enhancement methods, and discovered the effectiveness of back-translation and increasing parallel data in improving translation quality. Su et al. [7] combined the generator-discriminator framework with NMT to enhance translation quality and found through experiments that this method performed better in generating the translation that is closest to the true translation. Mi and Xie [8] studied multilingual NMT, introduced linguistic relevance assessment, and found through experiments that this method can obtain better translation quality. Wang et al. [9] designed a training framework for low-relevance languages on Mongolian-Chinese NMT and introduced hot-start transfer learning and approximate distillation to strengthen the language representation ability of the NMT encoder. The experiment found that this method achieved 3.2 improvement on bilingual evaluation understudy (BLEU). As a basic unit of natural language, phrase has a larger granularity than words, contains fixed collocations, grammatical constraints, and other features, and can reflect semantics more accurately than words. Therefore, phrase processing can effectively reduce the ambiguity of translation. Phrase processing is a very important content in SMT [10]. NMT is a word-based translation model. Few NMT studies focus on phrases. Therefore, this paper combined NMT with phrase processing of English texts to improve the translation quality of NMT and verified the performance of phrase processing combined with NMT in Chinese and English text translation through experiments on datasets. This article verifies the importance of phrase processing and provides theoretical support for the further development and optimization of NMT.

## 2 Neural Machine Translation Combined with Phrase Processing

### 2.1 NMT

NMT is a model based on neural networks, and its effectiveness has been proven in a large number of scenarios [11]. For a given source-end sentence, the target end sentence can be obtained through word-by-word translation. Convolutional neural network (CNN), gated cycle unit (GRU), and Transformer model can be used as coder decoders [12]. In NMT, for source-end sentence $x = \{x_1, x_2, \ldots, x_n\}$, an encoder is used to get intermediate representation $h_{context}$, and then the intermediate representation and context information are combined to get target sentence $\{y = y_1, y_2, \ldots, y_n\}$ in the decoder. The process can be written as:

$$h_{context} = \textit{Encoder}(x), \tag{1}$$

$$y_i \mid h_{context} \sim \textit{Decoder}(h_{context}; y_{<i}). \tag{2}$$

The optimization goal of NMT is to maximize the conditional probability of the target sequence in the dataset, and the Equation is:

$$P(y \mid x) = \prod_{i=1}^{m} P(y_i \mid y_{<i}, x; \theta), \tag{3}$$

where $P(y \mid x)$ represents the probability of source-end sentence $x$ generating target sentence $y$, $P(y_i \mid y_{<i}, x; \theta)$ is the probability of current time step $i$ under given source-end sentence $x$ and historical translation $y_{<i}$, and $\theta$ is the model parameter.

The training goal of NMT is to minimize the negative log-likelihood function of the conditional probability, and the Equation is:

$$L(\theta) = -\frac{1}{m} \sum_{i=1}^{m} \log P(y \mid x). \tag{4}$$

### 2.2 Transformer Model

Using only the attention network, the Transformer model has been proven to have better performance than CNN and recurrent neural network (RNN). With good flexibility and scalability, it shows good results in various fields such as machine vision [13], speech processing [14], and data prediction [15].

In NMT, the Transformer model has the ability to encode and model a large amount of information, so it has become one of the focuses of current NMT research. The Transformer model also uses a coder-decoder structure, which mainly includes the following content.

(1) Word embedding

It is supposed that there is source-end sentence $x = \{x_1, x_2, \ldots, x_l\}$, and its sentence length is $l$. It is transformed into word vector $E_x \in R^{l \times d_{model}}$ in the word embedding layer ($d_{model}$ is the word vector dimension), and then location coding is performed on the vector:

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{2i/d_{model}}}\right), \tag{5}$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{2i/d_{model}}}\right), \tag{6}$$

where *pos* refers to the position index and $i$ is the dimensional index.

(2) Multi-head self-attention (MHA) mechanism

Under the action of weight matrix $W^Q, W^K, W^V \in R^{d \times d_{model}}$, the word vector is transformed into query vector $Q \in R^{l \times d}$, key vector $Q \in R^{l \times d}$, and value vector $V \in R^{l \times d}$. The calculation process of attention is:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d}}\right)V, \tag{7}$$

$$head_i(Q, K, V) = Attention(QW_i^Q, KW_i^K, VW_i^V), \tag{8}$$

$$MHA(Q, K, V) = Concat(head_1, head_2, \ldots, head_h)W^O, \tag{9}$$

where $\sqrt{d}$ is a scaling factor, which is used to avoid gradient explosion or disappearance. The specific weight value acting on $V$ is obtained through the SoftMax function. After splicing the output of $h$ attention heads, the final output of the MHA layer can be obtained.

(3) Feedforward neural network (FFN)

A FFN is used for the linear transformation and nonlinear activation of the output of the MHA layer, and the Equation is:

$$FFN(x) = Relu(xW_1 + b_1)W_2 + b_2, \tag{10}$$

where $W_1, W_2, b_1$, and $b_2$ are trainable parameters.

(4) Layer normalization

The Transformer model also uses layer normalization to speed up model convergence, and the Equation is:

$$LN(x) = \alpha \times \frac{x - \mu}{\sqrt{\sigma^2}} + \beta, \tag{11}$$

where $\alpha$ and $\beta$ are trainable parameters, $\sigma$ refers to a variance, and $\mu$ is a mean value.

## 2.3 Phrase Processing

Phrase modeling is the main advantage of SMT, so this paper combines SMT with the Transformer model to implement phrase processing. Firstly, the text is preprocessed. The Berkeley Parser tool [16] is used to obtain the syntactic analysis results, and the Moses tool is used to achieve word alignment [17]. Bilingual phrases are learned and extracted based on word alignment. Then, the probability of phrase translation is calculated to obtain the phrase translation table. Suppose there is translation model $P(f \mid e)$. If a source sentence is $f = \{f_1, f_2, \ldots, f_J\}$, with a sentence length of $J$, and the target sentence is $e = \{e_1, e_2, \ldots, e_I\}$, with a sentence length of $I$. Then, the word alignment translation model can be written as:

$$P(f_1^J \mid e_1^I) = P(f_1^J, a_1^J \mid e_1^I), \tag{12}$$

where $a$ describes the correspondence between the position of the $j$-th word in $f$ and the $j$-th word in $e$. An example is shown in Table 1.

For the extracted phrases, translation probabilities are calculated.

(1) Forward phrase translation probability: refers to the phrase translation probability from the source-end sentence to the target sentence, and the Equation is:

$$P_r(\bar{e} \mid \bar{f}) = \frac{count(\bar{f}, \bar{e})}{\sum_{\bar{e}_i} count(\bar{f}, \bar{e}_i)}, \tag{13}$$

**Table 1** An example of word alignment

| Parameter | Example |
|---|---|
| e | Students study hard in the classroom every day. |
| f | 教室里的学生每天努力学习。 |
| I | 9 |
| J | 8 |
| a | {4, 7, 6, 2, 3, 1, 5, 5, 8} |

where $(\bar{f}, \bar{e})$ is the phrase pair of the source-end sentence and the target sentence, *count*$(\bar{f}, \bar{e})$ is the occurrence number of $(\bar{f}, \bar{e})$, and $\sum_{\bar{e}_i}$ *count*$(\bar{f}, \bar{e}_i)$ is the needed times for extracting all phase pairs with $\bar{f}$ as the source-end sentence.

(2) Forward lexicalization translation probability: refers to the lexicalization weighted probability from the source-end sentence to the target sentence, and the Equation is:

$$P_r(\bar{e} \mid \bar{f}, a) = \prod_{i=1}^{I} \frac{1}{\{j \mid (i,j) \in a\}} \sum_{\forall(i,j)\in a} w(e_i \mid f_j), \tag{14}$$

$$w(e_i \mid f_j) = \frac{\textit{count}(f_j, e_i)}{\sum_{e_i} \textit{count}(f_j, e_i)}, \tag{15}$$

where $a$ is the word alignment relationship between the source-end sentence and the target sentence, $(f_j, e_i)$ is the phrase pair formed by matching the source-end sentence with the target sentence, *count*$(f_j, e_i)$ is the number of occurrences of the phrase pair in the bilingual parallel corpus, and $\sum_{e_i}$ *count*$(f_j, e_i)$ is the number of occurrences of the word pair that takes $f_j$ as the lexicon of the source end sentence in parallel corpus.

(3) Reverse phrase translation probability: refers to the phrase translation probability from the target sentence to the source sentence, and the Equation is:

$$P_r(\bar{f} \mid \bar{e}) = \frac{\textit{count}(\bar{e}, \bar{f})}{\sum_{\bar{f}_j} \textit{count}(\bar{e}, \bar{f}_j)}. \tag{16}$$

(4) Reverse lexicalization translation probability: refers to the lexicalization weighted probability of the target sentence to the source sentence, and the Equation is:

$$P_r(\bar{f} \mid \bar{e}, a) = \prod_{j=1}^{J} \frac{1}{\{i \mid (i,j) \in a\}} \sum_{\forall(i,j)\in a} w(f_j \mid e_i), \tag{17}$$

$$w(f_j \mid e_i) = \frac{\textit{count}(e_i, f_j)}{\sum_{f_j} \textit{count}(e_i, f_j)}. \tag{18}$$

The resulting phrase translation table is as follows:

2012 ||| 年 the year 2012 ||| −1.5267 −0.2514 −8.1254 −6.2587

The above format corresponds to the source-end sentence ||| the target sentence ||| four phrase translation probabilities.

After the phrase translation table is obtained, the target phrase with a high probability is added to the end of the source sentence in the form of a suffix. The following three suffixes are designed in this paper.

(1) Only the target phrase sequence is added: means adding the optimal translation of the phrase extracted from the source-end sentence at the end in suffix form. An example is:

中国将为亚太地区稳定繁荣作出更大贡献。#Asia-Pacific region

(2) Only the source-end phrase sequence is added: means adding the phrase extracted from the source-end sentence to the end in suffix form. An example is:

中国将为亚太地区稳定繁荣作出更大贡献。#亚太地区

(3) Only the bilingual phrase sequence is added: means adding both the source phrase and the corresponding best translation to the end of the sentence. An example is:

中国将为亚太地区稳定繁荣作出更大贡献。#亚太地区 #Asia-Pacific region

The source-end sentence combining with phrase processing is used as input to the Transformer encoder, thereby increasing the Transformer model's phrase translation performance.

## 3 Results and Analysis

### 3.1 Experimental Setup

The experimental datasets include:

Chinese-English dataset CWMT2018 [18]: The test corpora provided by CWMT2017, CWMT2018, and WMT2018 were used, and the number of sentence pairs was 2000, 2481, and 3981, respectively.

English-German dataset WMT2014 [19]: Contains a total of 4.5 million sentence pairs. The newstest2013 was used as the validation set, and the testing was conducted using newstest2014.

NMT tasks usually use BLEU to evaluate the translation effect of the model [20]. First, the number of identical words in the model translation

result and the reference translation was compared, which is expressed by n-gram accuracy:

$$p_n = \frac{\sum_i \sum_k \min(\delta_{n_k}(T_i), \max\limits_{j \in m}(\delta_{n_k}(R_i^j)))}{\sum_i \sum_k (\delta_{n_k}(T_i))}, \tag{19}$$

where $n_k$ is the $k$-th n-gram, $T_i$ and $R_i$ are the model translation result and reference translation of the $i$-th sentence, $R_i^j$ is the $j$-th reference translation of the $i$-th sentence, and $m$ is the total number of reference translation. $\delta_{n_k}(T_i)$ and $\delta_{n_k}(R_i^j)$ are the occurrence number of $n_k$ in the model translation result and reference translation. In BLEU, length penalty factor ***BP*** was introduced to alleviate the influence of model translation length on n-gram accuracy. When the model translation length was short, the ***BP*** penalty value was assigned, and the Equation is:

$$\boldsymbol{BP} = \begin{cases} 1 & \textit{if}\, t > r \\ e^{1-r/t} & \textit{if}\, t \leq r \end{cases}, \tag{20}$$

where $t$ and $r$ are the length of model translation and reference translation.

The final BLEU is:

$$\boldsymbol{BLEU} = \boldsymbol{BP} \cdot \exp\left(\sum_{n=1}^{N} w_n \log p_n\right), \tag{21}$$

where $w_n$ is the weight parameter.

### 3.2 Result Analysis

The translation effects after adding different phrase suffix forms are shown in Table 2.

From Table 2, it can be found that adding only the suffix form of the target phrase sequence maximized the translation effect of the Transformer model, with a BLEU value increase of 0.0274 for CWMT2017, 0.0267 for CWMT2018, and 0.0219 for WMT2018 compared with the baseline model. The mean BLEU value increased by 0.0254, and the BLEU value increased by 0.0105 for the newstest2014 dataset, which verified the improvement in translation effect of the Transformer model combined with phrase processing. In contrast, the BLEU value obtained by adding only the source phrase sequence decreased for different test sets. This result indicated that adding only the source phrase sequence cannot provide useful phrase knowledge for

**Table 2** Comparison of translation effects after adding different phrase suffixes

| | Chinese-English | | | | English-German |
|---|---|---|---|---|---|
| | CWMT2017 | CWMT2018 | WMT2018 | Mean value | newstest2014 |
| Baseline (source-end sentence only) | 0.2512 | 0.2497 | 0.2533 | 0.2514 | 0.2721 |
| Adding the target phrase sequence only | 0.2786 | 0.2765 | 0.2752 | 0.2768 | 0.2826 |
| Adding the source-end phrase sequence only | 0.2321 | 0.2312 | 0.2331 | 0.2321 | 0.2545 |
| Adding the bilingual phrase sequence | 0.2469 | 0.2408 | 0.2511 | 0.2463 | 0.2612 |

Transformer translation, leading to poor translation effect. While the BLEU value obtained by adding the suffix form of bilingual phrase sequence for different test sets increased compared with the previous suffix form, it was still lower than the baseline model. This may be because the input sequence of the encoder increased after adding the bilingual phrase sequence, leading to difficulty in dynamic alignment and repeated coding problems. The training effect of the Transformer model was affected. Therefore, in the subsequent experiments, the phrase suffix form that only adds the target phrase sequence was adopted.

The translation performance under different numbers of phrase suffixes for the CWMT2017 dataset was tested (Table 3).

From Table 3, it can be seen that there was a correlation between translation performance and the number of phrase suffixes. After the number of suffixes reached four, the translation performance began to improve. This

**Table 3** Different numbers of phrase suffixes

| | BLEU value | Inference time (increased compared to baseline model) |
|---|---|---|
| Baseline | 0.2512 | – |
| 2 | 0.2214 | +3.16% |
| 4 | 0.2676 | +6.55% |
| 6 | 0.2733 | +9.82% |
| 8 | 0.2758 | +13.21% |
| 10 | 0.2791 | +16.77% |

result indicates that integrating an appropriate amount of phrase information is effective in guiding the generation of translations and can improve translation performance to a certain extent. However, in terms of inference time, as the number of suffixes increased, the inference time also rose. Therefore, in actual deployment, it is necessary to weigh the improvement of translation quality against the computational cost. If higher efficiency is desired, adding 6–8 phrase suffixes can be chosen. When 6–8 phrase suffixes are used, the BLEU value significantly improves, and the time cost is controlled at around 10–13%.

Then, the proposed method was compared with other translation models, including:

(1) sequence-to-sequence (seq2seq) [21],
(2) NMT based on a convolutional neural network (CNN),
(3) NMT based on a gated recurrent unit (GRU).

From Table 4, compared with the seq2seq model and the NMT models based on a CNN and a GRU, the Transformer model combined with phrase processing had better performance for different datasets. The mean BLEU value obtained by the proposed method was 0.2768 for the Chinese-English dataset, which was 0.0439 higher than that of the seq2seq model, 0.0378 higher than that of the NMT model based on a CNN, and 0.0276 higher than that of the NMT model based on a GRU. The BLEU value obtained for the English-German dataset was 0.2622, which was also higher than that of the other models. The results verified that the proposed method can improve the translation effect.

**Table 4** Comparison with other translation models

| | Chinese-English | | | | English-German |
|---|---|---|---|---|---|
| | CWMT2017 | CWMT2018 | WMT2018 | Mean value | newstest2014 |
| seq2seq | 0.2345 | 0.2312 | 0.2331 | 0.2329 | 0.2622 |
| NMT based on a CNN | 0.2441 | 0.2384 | 0.2345 | 0.2390 | 0.2677 |
| NMT based on a GRU | 0.2507 | 0.2472 | 0.2498 | 0.2492 | 0.2715 |
| Transformer model combined with phrase processing | 0.2786 | 0.2765 | 0.2752 | 0.2768 | 0.2826 |

**Table 5** Case analysis

| | |
|---|---|
| Source-end sentence | 请重点加强地道桥的排水工作。 |
| Adding the suffix form of the target phrase sequence | 请重点加强地道桥的排水工作。#underpass bridges |
| Reference translation | Please focus on strengthening the drainage work of underpass bridges. |
| NMT model based on a GRU | Please focus on strengthening the drainage work of the tunnel bridge. |
| Transformer model combined with phrase processing | Please focus on strengthening the drainage work of underpass bridges. |

Finally, the performance of the proposed method was further verified by a case analysis using the Chinese-English dataset.

As can be seen from Table 5, in the translation of the word "地道桥", the NMT model based on a GRU translated it as "tunnel bridge", which was inconsistent with the semantic meaning of the source-end sentence. However, the translation result obtained by using the Transformer model combined with phrase processing was "underpass bridges", which was consistent with the reference translation. The result verified the reliability of the proposed method in translation. In addition, through manual analysis of several error cases, when there are words with multiple meanings or a strong context dependence, if a bilingual phrase sequence is used as a suffix, it may introduce noise and interfere with the model's overall understanding of the source sentence. For example, in the sentence "He attended the last meeting in Washington", "last" can be translated as "上次" or "最后". If the extracted phrase alignment gives "last meeting → 最后会议", and this phrase pair is forcibly added to the suffix, the model may ignore the context (here, "Washington" implies the meeting location, and it should be "上次会议"), resulting in translation deviations. This indicates that phrase addition should be based on high-quality alignment and, when there is strong ambiguity between the source phrase and the target phrase, adding bilingual information may actually mislead the model.

## 4 Discussion

NMT generally treats a sentence as a sequence of words. After directly vectorizing each word in a sentence, it inputs them into the encoder for encoding. Based on the Transformer and the advantages of SMT in phrase processing, this paper added a target phrase sequence after the source-end

sentence. In this way, in addition to encoding words and information in the context, the encoder could also encode the phrase translation knowledge obtained from the phrase translation table to improve the translation effect.

Among the three studied suffix forms, translation performance was optimal when only the target phrase sequence was added. For Transformer, explicit target phrase information can serve as supplementary short-term memory, helping the model maintain the internal consistency of phrases and improve fluency and translation quality. Moreover, during the decoding phase, the model needs to forecast the target words from a large vocabulary. The addition of the target phrase sequence provides constraints for the decoder, narrowing down the search space and enabling the generation of more idiomatic expressions at the phrase level. Adding only a source-end phrase sequence or adding a bilingual phrase sequence may cause dynamic alignment interference, forcing the model to process two sets of signals, namely the original text and the phrase markers, simultaneously during the encoding phase. Additionally, since the source-end phrase is already encoded, adding it additionally will result in redundant encoding and information redundancy, making the encoder over-focus on these markers, thereby affecting its ability to represent other parts of the sentence and being detrimental to semantic integrity. Therefore, adding the target phrase sequence is the best choice.

In recent years, research on combining linguistic knowledge with NMT has gradually increased. However, most of the research focuses on syntactic structures or external knowledge bases. For example, Ishiwatari et al. [22] improved the decoder of NMT into a chunk-level decoder and a word-level decoder, and verified the translation effect of this method in the Workshop on Asian Translation (WAT) English-to-Japanese translation task. Hasler et al. [23] proposed an NMT method with terminology constraints and verified through experiments that using a term-phrase corpus is beneficial for improving translation quality. In current research, there are few studies on the direct integration with explicit phrases. The research in this paper did not change the model structure. It extracted relevant content from the phrase translation table and input it into the decoder, effectively improving the translation performance. It clearly demonstrated the effectiveness of only target phrases and analyzed the negative impacts of source-end phrases or bilingual phrases, providing some new empirical conclusions for the integration of phrases in NMT.

However, for long or structurally complex sentences, such as those containing multiple clauses, parenthetical expressions, or long-distance

dependencies, the boundaries and semantics of phrases may change dynamically with the context. This results in the phrase sequence in the suffix not fully matching the actual semantics expressed by the source sentence. In addition, the self-attention mechanism of the Transformer model already faces the problem of attention dilution when dealing with long sentences. Adding additional phrases may further disperse attention resources, leading to a decline in the model's ability to model the core structure. This suggests that, in future work, more refined strategies for phrase extraction and addition need to be designed. For example, syntactic analysis or context-sensitive phrase selection can be combined to meet the translation needs of complex sentences. Additionally, more phrase integration methods, such as phrase insertion and context-aware integration, can be incorporated into the research, and phrase processing can be combined with data augmentation or external knowledge sources to better address issues such as the translation of rare words.

## 5 Conclusion

Based on NMT, this paper designed a Transformer model combined with phrase processing and tested its translation effect through experiments on Chinese and English datasets. Results showed that adding the suffix form of the target phrase sequence performed best in improving the translation effect of the Transformer model, it was superior to other models such as seq2seq, and the obtained translation was more in line with the reference translation. The proposed method can be further applied in practice.

## References

[1] A. Mishra, D. Ganesh, A. Sharma, R. Vignesh, "Applying Natural Language Processing for Detecting Cybersecurity Threats Using Sentimental Analysis Techniques," *International Conference on Data Science, Machine Learning and Applications*, 594–600, 2025.

[2] H. Wang, H. Wang, "An Application System for Evaluating and Optimizing the Quality of Neural Machine Translation Corpus," in *2022 IEEE International Conference on e-Business Engineering (ICEBE)*, 178–183, 2022.

[3] S. J. Hwang, C. S. Jeong, "Integrating Pre-trained Language Model into Neural Machine Translation," *2023 2nd International Conference*

*on Frontiers of Communications, Information System and Data Science (CISDS)*, 59–66, 2023.

[4] X. Liu, J. Zeng, Z. S. J. Wang, "Exploring iterative dual domain adaptation for neural machine translation," *Knowl-based Syst.*, 283(Jan.11), 1.1–1.12, 2024.

[5] Y. Xiao, L. Wu, J. Guo, J. Li, M. Zhang, T. Qin, T. Y. Liu, "A Survey on Non-Autoregressive Generation for Neural Machine Translation and Beyond," *IEEE T. Pattern Anal.*, (10), p. 45, 2023.

[6] Z. Yirmibesoglu, T. Gungor, "Morphologically Motivated Input Variations and Data Augmentation in Turkish-English Neural Machine Translation," *ACM T. Asian Low-reso.*, 22(3), 1–31, 2023.

[7] C. Su, H. Huang, S. Shi, P. Jian, "Improving Neural Machine Translation by Transferring Knowledge from Syntactic Constituent Alignment Learning," *ACM T. Asian Low-reso.*, 21(5), 1–15, 2022.

[8] C. Mi, S. Xie, "Language relatedness evaluation for multilingual neural machine translation," *Neurocomputing*, 570(Feb.14), 127115.1–127115.15, 2024.

[9] P. Wang, H. Hou, S. Sun, N. Wu, W. Jian, Z. Yang, Y. Wang, "Hot-Start Transfer Learning Combined with Approximate Distillation for Mongolian-Chinese Neural Machine Translation," *China Conference on Machine Translation*, 12–23, 2022.

[10] M. P. Sebastian, G. S. Kumar, "Malayalam Natural Language Processing: Challenges in Building a Phrase-Based Statistical Machine Translation System," *ACM T. Asian Low-reso.*, 22(4), 1–51, 2023.

[11] B. Ahmadnia, B. J. Dorr, R. Aranovich, "Impact of Filtering Generated Pseudo Bilingual Texts in Low-Resource Neural Machine Translation Enhancement: The Case of Persian-Spanish," *Proc. Comput. Sci.*, 189, 136–141, 2021.

[12] Q. D. E. J. Ren, Y. Su, N. Wu, "Research on Mongolian-Chinese machine translation based on the end-to-end neural network," *Int. J. Wavelets Multi*, 18(01), 46–59, 2020.

[13] Q. Li, W. Xie, Y. Wang, K. Qin, M. Huang, T. Liu, Z. Chen, L. Chen, L. Teng, Y. Fang, L. Ye, Z. Chen, J. Zhang, A. Li, W. Yang, S. Liu, "A Deep Learning Application of Capsule Endoscopic Gastric Structure Recognition Based on a Transformer Model," *J. Clin. Gastroenterol*, 58(9), 937–943, 2024.

[14] P. Warule, S. Chandratre, S. P. Mishra, S. Deb, "Detection of the common cold from speech signals using transformer model and spectral features," *Biomed Signal Proces*, 93(July), 1–9, 2024.

[15] R. Rosado, O. G. Toledano-López, H. R. González, A. J. Abreu, Y. Hernandez, "Cuban Consumer Price Index Forecasting Through Transformer with Attention," *J. Autom. Mob. Robot. Intell. Syst.*, 17(2), 12–17, 2023.

[16] J. K. Kummerfeld, D. Hall, J. R. Curran, D. Klein, "Parser Showdown at the Wall Street Corral: An Empirical Investigation of Error Types in Parser Output," in *Proceeding of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 1048–1059, 2012.

[17] K. Grashchenkov, A. Grabovoy, I. Khabutdinov, "A Method of Multilingual Summarization For Scientific Documents," *2022 Ivannikov Ispras Open Conference (ISPRAS)*, 24–30, 2022.

[18] H. Yang, Y. Qin, Y. Deng, M. Wang, "NMT Enhancement based on Knowledge Graph Mining with Pre-trained Language Model," in *2020 22nd International Conference on Advanced Communication Technology (ICACT)*, 185–189, 2020.

[19] Y. Li, J. Li, M. Zhang, "Improving neural machine translation with latent features feedback," *Neurocomputing*, 463, 368–378, 2021.

[20] N. J. Suha, M. A. R. Khan, M. S. Hossain, "A Neural Machine Translation Approach for Translating Different Languages in English," *Trends Appl Sci Res*, 18(1), 169–182, 2023.

[21] K. Sun, T. Qian, X. Chen, M. Zhong, "Context-aware seq2seq translation model for sequential recommendation," *Inform Sciences*, 581, 60–72, 2021.

[22] S. Ishiwatari, J. Yao, S. Liu, M. Li, M. Zhou, N. Yoshinaga, M. Kitsuregawa, W. Jia, "Chunk-based Decoder for Neural Machine Translation," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, 1901–1912, 2017.

[23] E. Hasler, A. de Gispert, G. Iglesias, B. Byrne, "Neural Machine Translation Decoding with Terminology Constraints," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 506–512, 2018.

## Biography

**Danqiangyu Zhou**, born in September 1986, graduated from Sichuan Normal University, China, with a master's degree in June 2011. She is working at Civil Aviation Flight University of China as an associate professor. She is interested in college English teaching and literary culture.