

---

AI translation · View original & related papers at  
[chinarxiv.org/items/chinaxiv-202507.00031](https://chinarxiv.org/items/chinaxiv-202507.00031)

---

# Research and Development of an Intelligent Manchu-Chinese Bidirectional Translation System

**Authors:** Minfei Wu, Sun Xiaoxin, Yin Minghao, Suhua Wang, Wang Suhua

**Date:** 2025-06-29T19:09:35+00:00

## Abstract

**[Purpose/Significance]** The protection and utilization of Manchu archives face severe challenges. Developing a Manchu-Chinese mutual translation system based on the Transformer artificial intelligence model holds significant importance for the inheritance of endangered linguistic and cultural heritage. **[Method/Process]** The system employs a Transformer-based machine translation architecture, with model optimization tailored to the characteristics of the Manchu language. The system possesses self-learning and self-memory capabilities, continuously improving translation accuracy through ongoing optimization. **[Results/Conclusion]** The Transformer-based Manchu-Chinese mutual translation system achieves efficient conversion between Manchu and Chinese, filling the technological gap in Manchu information processing. This system demonstrates significant application value in the digitization of Manchu archives and Manchu-Chinese cultural exchange, providing an innovative technical solution for the protection and inheritance of minority languages.

## Full Text

### Research and Construction of an Intelligent Manchu-Chinese Machine Translation System

**Wu Minfei<sup>1</sup>, Sun Xiaoxin<sup>1</sup>, Yin Minghao<sup>1</sup>, Wang Suhua<sup>2</sup>**

<sup>1</sup>School of Computer Science and Technology, Northeast Normal University, Changchun 130117

<sup>2</sup>Changchun Humanities and Sciences College, Changchun 130117

## Abstract

**[Purpose/Significance]** The preservation and utilization of Manchu archives confront significant challenges. Developing an intelligent Manchu-Chinese machine translation system based on the Transformer artificial intelligence model holds profound importance for the inheritance of endangered linguistic and cultural heritage. **[Method/Process]** This study employs a Transformer-based neural machine translation architecture, optimized according to the linguistic characteristics of Manchu. The system incorporates self-learning and memory mechanisms, continuously improving translation accuracy through iterative optimization. **[Result/Conclusion]** The Transformer-based Manchu-Chinese machine translation system achieves efficient cross-lingual conversion, addressing a critical technological gap in Manchu language information processing. This system demonstrates substantial application value in the digitization of Manchu archives and Manchu-Han cultural exchange, providing an innovative technological solution for the preservation and transmission of minority languages.

**Keywords:** Manchu-Chinese Intelligent Machine Translation; Transformer Model; Data Augmentation; Transfer Learning

**Classification Number:** H211

## 1. Introduction

Manchu, as a crucial language within the Altaic family's Manchu-Tungusic branch, carries rich historical and cultural information. However, with its speaker population dwindling, UNESCO designated it as a critically endangered language in 2009 [1]. Consequently, the rescue and preservation of Manchu linguistic and cultural heritage has become an urgent priority, garnering widespread recognition and high-level attention from both national authorities and society at large [2]. Manchu-Chinese translation not only satisfies the practical need for conversion between Chinese and Manchu scripts but also facilitates the inheritance of Manchu culture and promotes interethnic cultural exchange. Traditional neural machine translation methods, such as rule-based systems and statistical neural machine translation, exhibit inherent limitations when handling the substantial typological differences between Manchu and Chinese, making it difficult to achieve high-quality translation results.

With the rapid advancement of artificial intelligence technology, natural language processing has witnessed revolutionary breakthroughs, particularly in machine translation. The Transformer model, as one of the most dominant deep learning architectures today, has become the preferred approach for machine translation tasks due to its exceptional parallel processing capabilities and powerful modeling of long-distance dependencies. However, despite these impressive technological advances, many minority languages—especially low-resource languages like Manchu—still face enormous challenges in modern information technology applications.

Constructing a Manchu-Chinese translation system carries multiple layers of

significance. On one hand, it can facilitate the preservation and transmission of Manchu by providing convenient linguistic tools for Manchu speakers. On the other hand, it can enhance cultural exchange and mutual understanding between Han and Manchu ethnic groups, thereby promoting national unity. As relevant research has noted, Manchu language and script held a special status during the Qing Dynasty, being designated as the “national language” and “national script,” and played an instrumental role in the socio-economic and cultural development of the Manchu people during the early Qing period.

The Transformer architecture, with its self-attention mechanism and parallel computing capabilities, demonstrates superior performance in neural machine translation, addressing the temporal constraints of recurrent neural networks for sequence modeling and effectively capturing long-distance dependencies. Applying Transformer to Manchu-Chinese translation can overcome the limitations of traditional methods and enhance both translation accuracy and fluency.

This paper introduces, for the first time, a Manchu-Chinese parallel corpus; sample contents are shown in Table 1. In the table, the “manchu” field represents Manchu transliteration, while the “chinese” field represents the corresponding Chinese text. Based on the generated Manchu-Chinese sentence pairs and combined with the Transformer machine translation model, we construct a Manchu-Chinese bidirectional translation system, conduct in-depth research on its application in Manchu-Chinese translation tasks, explore optimal translation strategies, and provide technical support for Manchu resource preservation and cross-lingual cultural dissemination.

This study aims to develop an efficient and accurate Manchu-Chinese translation system utilizing the advanced Transformer model. Although machine translation technology has achieved remarkable progress, research applying these technologies to the specific language pair of Manchu and Chinese remains relatively scarce. Our system is designed to provide accurate and efficient translation services, thereby promoting the preservation and active use of the Manchu language.

The contributions of this research are manifested in three primary aspects. First, we developed a Transformer model specifically tailored for Manchu-Chinese translation tasks. Second, we leveraged high-quality bilingual parallel corpora for model training and, addressing the characteristics of Manchu as a low-resource language, explored effective training strategies. Third, we implemented a user-friendly interface to demonstrate the practical application effects of the translation system.

Through this work, we hope to contribute to the development of translation technologies for minority languages and provide support for the preservation and transmission of Manchu. In the following sections, we will elaborate on the system architecture, training data, evaluation methods, and experimental results to showcase our research findings and future development directions.

## 2. Literature Review

Kalchbrenner and Blunsom [3] proposed a continuous translation model for translating sentences from source to target languages. Meng et al. [4] introduced the application of convolutional neural networks for machine translation, organically integrating them with statistical machine translation models. Gehring et al. [5] opted to apply convolutional neural networks to encoder-decoder architectures as encoders. Meanwhile, Kalchbrenner et al. [3] proposed the ByteNet machine translation model, which achieved state-of-the-art performance at the character level but yielded unsatisfactory results at the word level.

Gehring et al. [6] proposed ConvS2S, a fully CNN-based machine translation model. Sutskever et al. [7] introduced an encoder-decoder system employing recurrent neural networks (RNN). In the same year, Bahdanau et al. [8] incorporated attention mechanisms into RNN-based machine translation models, thereby avoiding the limitations imposed by fixed-length source sentence representations and enhancing translation accuracy and quality.

Subsequently, the Transformer [9] model emerged, demonstrating exceptional performance in machine translation tasks. It not only generates high-quality translations but also significantly reduces training time compared to previous models, substantially improving training efficiency. Since its inception, the Transformer model has rapidly become the dominant architecture in machine translation due to its superior performance and advantages, exerting profound influence on subsequent research and applications.

Machine translation technology relies on large-scale bilingual parallel corpora for model training. However, like Manchu translation research, the vast majority of languages face the challenge of scarce large-scale parallel datasets. To address the difficulties arising from the shortage of bilingual parallel corpora, the academic community has developed a series of machine translation strategies for low-resource environments.

Liang et al. [10] investigated multi-teacher distillation from subnet space simulation and permutation variants of single-teacher models. Its unique advantage lies in the ability of a single teacher model to output multiple variables without adding extra parameters or significantly increasing training costs. Singh et al. [11] proposed a semi-supervised method to improve translation quality for extremely low-resource language pairs such as English-Manipuri. Kumar et al. [12] employed reinforcement learning to address domain adaptation, training language models to select out-of-domain sentences semantically similar to in-domain data, thereby improving machine translation performance.

Transfer learning enables the migration of models trained on high-resource languages to similar low-resource language training tasks, yielding effective results. Jiang et al. [13] proposed applying transfer learning to lexical constraint models to effectively solve domain mismatch problems in machine translation tasks. Liu et al. [14] introduced a k-nearest-neighbor transfer learning method comprising

parent-child representation alignment and child-aware datastore construction, improving inference efficiency and ensuring consistency between model output representations by selectively extracting parent datastores based on relevance to child models. Huang et al. [15] used English as a pivot language to initialize parameters for Chinese-Vietnamese translation models via Chinese-English and English-Vietnamese translation models. Xue et al. [16] transferred pretrained BERT network parameters to Transformer models to address issues such as poor long-sentence translation and word sense disambiguation, achieving model performance optimization.

Data augmentation techniques are commonly employed to alleviate training data scarcity for low-resource languages. Gao et al. [17] proposed soft contextual data augmentation, which overcomes the limitations of traditional random word replacement, deletion, or swapping by randomly selecting words in sentences and performing soft replacement based on probability distributions over the vocabulary, thereby achieving fine-grained semantic adjustment. Nguyen et al. [18] adopted a multi-model training strategy, training multiple models in both target-to-source and source-to-target translation directions, using these models to translate training data and generate two sets of synthetic training data, significantly expanding the original dataset. Duan et al. [19] based their approach on dependency parse trees, considering the grammatical role of words in sentences and selecting and modifying words with specific probabilities to make the data augmentation process more targeted. Wu et al. [20] proposed mixSeq, a simple yet effective data augmentation method that constructs longer input-output pairs by randomly concatenating two input sequences and their corresponding target sequences, enabling rapid dataset expansion. Ko et al. [21] utilized monolingual data from low-resource languages and parallel data from similar high-resource languages for training, employing unsupervised adaptive methods to facilitate translation between low-resource languages. Xia et al. [22] proposed an effective two-step rotation data augmentation method that uses high-resource languages as pivots to generate pseudo-parallel corpora through unsupervised machine translation. Kondo et al. [23] generated longer sentences by concatenating two sentences from given parallel corpora, improving machine translation model performance on long-sentence translation. Jin et al. [24] proposed a novel data augmentation method called AdMix, which generates new training samples through linear interpolation of original sentences and sentences with slight discrete noise. Zhang [25] first extracted dictionaries from parallel corpora and then performed word substitution based on these dictionaries to improve machine translation performance for different language pairs in low-resource scenarios. Yang et al. [26] employed data augmentation methods to expand training corpora and utilized the cross-lingual pretrained model mRASP to improve the quality of Tibetan-Chinese bidirectional machine translation.

Additionally, numerous scholars have conducted research on the architecture and training methods of multilingual pretrained models. Li et al. [27] proposed the multilingual conditional masked language pretraining model CeMAT, which achieved significant performance improvements in both autoregressive

and non-autoregressive machine translation. Fan et al. [28] introduced a new many-to-many multilingual translation model M2M100 capable of direct translation between 100 languages, outperforming English-centric multilingual models. Raffel et al. [29] proposed the T5 model, achieving leading results in numerous English-centric natural language processing (NLP) tasks. Subsequently, Xue et al. [30] proposed the multilingual variant mT5, which inherits all advantages of the T5 model and demonstrates excellent performance on multiple multilingual datasets. Meta AI developed the NLLB model [31] based on sparsely-gated mixture-of-experts conditional computation, trained on data obtained through innovative and effective data mining techniques tailored for low-resource languages.

### 3. Construction of Manchu-Chinese Bidirectional Translation Model Based on Transformer

#### 3.1 Transfer Learning and Model Fine-tuning

Transfer learning is an important machine learning methodology whose core principle involves acquiring knowledge from source tasks and migrating it to target tasks with certain similarities. It is particularly suitable for scenarios with insufficient training data and can significantly reduce the data volume required for target tasks.

In traditional natural language processing approaches, models typically rely on large-scale parallel corpora and are only applicable to specific languages. Training and test data must be independently and identically distributed and sufficiently large to achieve satisfactory results. In contrast, transfer learning overcomes these limitations, with the key being the ability to share model parameters between source and target tasks. The specific operational process includes first training a well-performing model on high-resource tasks, then transferring its parameters to the target task, and fine-tuning on the new task to achieve knowledge transfer.

Transfer learning is particularly suitable for machine translation tasks with scarce data. In the current neural machine translation (NMT) field, except for high-resource language pairs like English-Chinese, the vast majority of language pairs lack sufficient bilingual parallel corpora. Through transfer learning, universal linguistic knowledge acquired from high-resource language training can be utilized to significantly improve translation performance for low-resource language pairs.

In neural machine translation, “parameter isolation” is an important technique for improving model training efficiency and performance. This method selectively freezes parameters of certain model layers to avoid overfitting and catastrophic forgetting, thereby enhancing model generalization capabilities. Different layers in neural networks are responsible for extracting semantic information at different levels, with lower layers typically learning basic grammar and se-

mantic information. Continuing to update these layers during fine-tuning may disrupt existing knowledge. Freezing lower-layer parameters while focusing on training higher-layer parameters relevant to the new task helps the model adapt to target tasks more quickly and efficiently.

This method proves particularly effective in multilingual translation tasks. Significant differences exist in grammar and vocabulary across languages, and parameter isolation enables models to reduce interference by freezing parameters of irrelevant languages when handling multiple language pairs. For example, when simultaneously processing English-French and Chinese-Japanese translation, partial network layers can be frozen separately according to typological differences to improve model stability across different language pairs.

Furthermore, parameter isolation also demonstrates significant advantages in cross-domain translation. For instance, when migrating a general-domain English-Chinese translation model to the medical domain, lower-layer general language parameters can be frozen while only higher-layer parameters responsible for terminology recognition and expression are fine-tuned. This approach preserves the model's general translation capabilities while enabling rapid adaptation to specialized domains, avoiding overfitting and improving translation quality.

### 3.2.1 Dataset Construction

The data in this study primarily originates from modern academic works, including contemporary monographs on Manchu language, literature, folklore, and arts. These materials contain not only rigorous academic analysis but also research findings on contemporary Manchu usage, neologisms, and grammatical variations. The Manchu-Chinese bilingual texts reflect translation needs in modern academic contexts, enabling the model to adapt to knowledge dissemination and academic exchange scenarios. Data sources also include Manchu-Chinese daily life dialogues to facilitate translation in conversational contexts.

Through extensive efforts, we have collected over 4,000 parallel sentence pairs (Manchu-Chinese). Data sources include *Ji Yonghai's 800 Modern Manchu Sentences*, *365 Manchu Sentences* (compiled by He Rongwei), *Jin Biao's 150 Manchu Sentences*, *2230 Dialogue Sentences*, *Daily Phrase Recording Companion Sentences*, *174 Daily Simple Oral Sentences*, among others. After data augmentation, the total dataset size exceeds 11,800 parallel sentence pairs, as shown in Table 3 .1.

**Table 2** Datasets

### 3.3.2 Data Preprocessing

The first step involves data cleaning. Specialized text cleaning scripts are employed to remove garbled text, duplicate entries, and non-Manchu/Chinese characters (such as interspersed English abbreviations and numeric symbols that

constitute interference), ensuring each text entry is clean and suitable for subsequent processing.

The second step is data annotation. Manual annotation is adopted to process cleaned texts into Manchu-Chinese parallel sentence pairs, for example: {"manju": "si saiyvn?", "chinese": "你好! "}.

The final step is data partitioning. The corpus is divided into training, validation, and test sets at a ratio of 8:1:1. The training set is used for initial model learning and parameter adjustment, the validation set periodically evaluates model performance during training to prevent overfitting, and the test set provides final objective performance evaluation after model training completion.

Through comprehensive and meticulous dataset construction and management, we establish a solid data foundation for the Transformer-based Manchu-Chinese translation system, with the potential to train high-performance, scenario-adaptive intelligent translation models.

### 3.3.3 Experimental Model Parameter Settings

The following provides a detailed introduction to model parameter settings:

**Training Parameters:** Number of epochs: 50. Batch size: 3074. Maximum sentence length: 128.

**Input and Output Parameters:** Maximum input sequence length ( $\max_{\{input\}}\{\{length\}\}$ ): defaults to 128. Maximum target sequence length ( $\max_{\{target\}}\{\{length\}\}$ ): defaults to 128.

**Training Configuration:** Batch size (batch\_size): defaults to 32. Learning rate (learning\_rate): defaults to 1e-5. Number of training epochs (epoch\_num): defaults to 50.

**Model Parameters:** Pretrained model checkpoint (model\_checkpoint): defaults to "Helsinki-NLP/opus-mt-zh-en".

**Optimizer Parameters:** Optimizer used: defaults to AdamW. Learning rate scheduler (lr\_scheduler): defaults to linear scheduler.

**Hardware Configuration:** Device used: selects "cuda" or "cpu" based on availability. Random seed (seed): ensures result reproducibility, defaults to 42.

These parameters can be adjusted in the code to accommodate different training requirements and dataset characteristics.

### 3.3.4 Evaluation Metrics

Translation quality metrics primarily include human evaluation and automatic evaluation. Automatic evaluation is time-saving and labor-efficient, and highly correlated with human evaluation results, making it popular among machine

translation researchers. Below, we introduce the automatic evaluation methods involved in our experimental section.

BLEU (Bilingual Evaluation Understudy) is a classic, widely used, and long-standing automatic evaluation metric in machine translation, proposed by Papineni et al. [32] in 2002. The BLEU metric provides quantitative measurement of translation accuracy.

COMET (Cross-lingual Optimized Metric for Evaluation of Translation) is a neural framework proposed by Unbabel AI that constructs a multilingual machine translation evaluation model based on ranking and regression. As an automatic evaluation metric, this model can accurately predict machine translation quality by leveraging information from source and target language sentences, thereby demonstrating correlation with state-of-the-art human evaluation.

During inference, the model generates a triplet for machine translation hypotheses. The quality score assigned to machine translations is determined based on the harmonic mean of distances between source sentence embeddings and reference sentence embeddings.

COMET fully utilizes recent advances in cross-lingual modeling. It can capture semantic similarity beyond the lexical level and more effectively distinguish high-performance neural machine translation (NMT) systems. In this paper, we use the unbabel-comet toolkit to calculate COMET scores, which range between [0,1]. When the score reaches

### 3.3.5 Comparative Experiments on Transfer Learning

**Table 3** Comparative experiments on transfer learning

Model	BLEU Score
Transformer	4.99
Transformer + Transfer Learning	34.57

In the Manchu-Chinese translation system, the Transformer-based model's BLEU score improved from 4.99 to 34.57 after incorporating transfer learning. This significant improvement can be attributed to transfer learning's enhancement of model performance in several aspects:

**Data Augmentation and Knowledge Transfer:** Transfer learning allows models to leverage abundant resources from source languages (such as Chinese) to enhance target language (such as Manchu) learning. Since Manchu data may be relatively limited, direct training could lead to model overfitting or insufficient learning of complex linguistic features. The pretrained model learns knowledge from the source language, helping improve target language translation accuracy.

**Enhanced Adaptability:** Transfer learning helps models enhance adaptability. Manchu and Chinese may have specific expressions in certain domains. Through transfer learning, models can better adapt to linguistic features across different domains.

**Improved Generalization:** Transfer learning enhances model generalization capabilities. With only a small amount of Manchu data, models may struggle to generalize to unseen sentences or vocabulary. Through transfer learning, models can leverage extensive source language data to learn more universal linguistic patterns, demonstrating stronger generalization in target language translation.

**Parameter Sharing:** In transfer learning, models can share partial parameters between source and target languages, improving training efficiency. Simultaneously, parameter sharing enables the model to transfer useful information between source and target languages, further optimizing translation performance.

**Attention Mechanism and Contextual Understanding:** The core of the Transformer model is the self-attention mechanism, which captures long-distance dependencies in input sequences. With transfer learning assistance, models can better understand contextual relationships between source and target languages, generating more accurate and fluent translation results.

In summary, transfer learning significantly improves BLEU scores in the Manchu-Chinese translation system, primarily due to data augmentation, enhanced domain adaptability, improved generalization, parameter sharing, and better attention mechanisms with contextual understanding. These factors collectively enable the model to more accurately capture and translate complex linguistic features between Manchu and Chinese.

### 3.3.6 Comparative Experiments on Model Fine-tuning

**Table 4** Comparative experiments of model fine-tuning

Strategy	BLEU Score
Freeze encoder and decoder FC1 layers	28.12
Freeze encoder and decoder FC2 layers	28.23
Freeze encoder and decoder output layers	24.45
Freeze encoder layers	78.25
Freeze decoder layers	20.01
Freeze first layer of encoder and decoder	85.12
Freeze first two layers of encoder and decoder	27.45

As shown in Table 3.3, the baseline experiment achieved a BLEU score of 81.65, indicating the model possesses certain transfer capabilities without specific freezing strategies. When freezing the FC1 modules of both encoder and decoder layers, the BLEU score dropped to 28.12, demonstrating that FC1 modules play

a crucial role in transfer learning, and freezing them significantly impacts the model's adaptability to language pairs.

When freezing the FC2 modules of encoder and decoder layers, the BLEU score decreased to 28.23, indicating that FC2 modules, similar to FC1 modules, are important fully connected layers, and freezing them also causes substantial performance degradation.

Freezing the output layers of encoder and decoder caused the BLEU score to drop significantly to 24.45, showing that this has a major direct impact on output layers, with freezing severely limiting the model's output capabilities.

Freezing the encoder layers resulted in a BLEU score decrease to 78.25, but the performance remained relatively high. This may indicate the importance of the encoder while also suggesting the decoder's capacity to compensate to some extent for the frozen encoder.

Freezing the decoder layers caused the BLEU score to plummet to 20.01, demonstrating that decoder layers are highly specific for generating target language, and freezing them severely hinders the model's adaptation to new target languages.

Freezing the first layers of both encoder and decoder significantly improved the BLEU score to 85.12, indicating that freezing shallow networks helps preserve the model's general features while allowing deep networks to adapt to new language pairs. This may also suggest that first-layer learning could be influenced by other layers, or that first-layer weights were already well-optimized during pretraining and require no further adjustment.

Freezing the first two layers of encoder and decoder caused the BLEU score to drop dramatically to 27.45, approaching the level observed when freezing FC1 and FC2. This indicates that the first two layers are crucial for model performance.

This study reveals the roles of different modules in the Helsinki-NLP/opus-mt-zh-en model for Chinese-to-Manchu machine translation through experiments with various transfer learning strategies. The results demonstrate that freezing shallow networks enhances model transferability, while freezing deep networks or specific modules may limit model adaptability. These findings provide valuable references for future transfer learning strategies in cross-lingual machine translation tasks.

### 3.3.7 Comparative Experiment on Changing Evaluation Metrics for Manchu-to-Chinese

**Table 5** Comparative experiments of full-to-medium replacement of evaluation indicators

Metric	Score
BLEU (maximum 100)	12.70 → 0
COMET (maximum 1)	0.3316 → 0.6952

In Manchu-Chinese translation training experiments, we observed that BLEU scores for Manchu-to-Chinese tasks were significantly lower than those for Chinese-to-Manchu tasks. Generally, translation quality in both directions is positively correlated—better Chinese-to-Manchu training results should correspond to better Manchu-to-Chinese results. Upon investigation, we determined that BLEU may be inappropriate for evaluating Manchu-to-Chinese translation. This is because a single Manchu sentence may have multiple reasonable Chinese translations, and Chinese expression is highly flexible. The diversity of reference translations directly reduces BLEU scores. Additionally, Chinese translations may trigger length penalties if they are shorter than reference translations. For example, if the reference translation is “*我认为这是一个重要的问题*” (I think this is an important issue) while the system outputs “*这是重要问题*” (This is important issue), the translation is semantically correct but penalized for brevity.

This study decided to address these issues by using COMET in conjunction with BLEU. Since COMET scoring is based on semantic similarity, it reduces dependence on surface forms. As shown in Table 4.4, when using BLEU as the evaluation metric, the score was 12.70 at epoch=2, then suddenly dropped to 0 and remained at 0 for all subsequent training epochs. However, after switching to COMET as the evaluation metric, the COMET score continuously improved from an initial 0.3316 to 0.6952 after 50 epochs. COMET scores range between [0,1], and the final score of 0.6952 demonstrates the effectiveness of Manchu-to-Chinese training.

## 4. Design and Implementation of Manchu-Chinese Bidirectional Translation System Based on Transformer

### 4.1 System Overview

As a language carrying rich historical and cultural information, Manchu faces significant challenges in preservation and transmission. Constructing an efficient and accurate Manchu-Chinese translation system is of great importance. This system aims to achieve automatic translation from Manchu text to Chinese text, employing a statistical neural machine translation architecture combined with rule-based and deep learning technologies. It encompasses key technologies including text digitization, tokenization, part-of-speech tagging, bilingual corpus construction, and translation model training. The system is expected to achieve high translation quality, broad coverage, strong robustness, and ease of use, providing powerful tools for Manchu cultural research and transmission.

## 4.2 System Prototype Diagram

**Figure 1** Prototype of the Manchu-Chinese neural machine translation system [FIGURE:1]

The prototype design of the Manchu-Chinese neural machine translation system is shown in Figure 4 [FIGURE:4].1. In the system's frontend interface, users can input source language text in the upper input box. After completing the input, clicking the “Translate” button automatically sends the text to the model for translation processing. Upon completion, the system displays the generated target language translation in the lower text box, concluding the translation process.

## 4.3 System Effect Diagrams

**Figure 2** Renderings of Chinese to Manchu translations

### Manchu-Chinese Translation of Qing Dynasty Documents

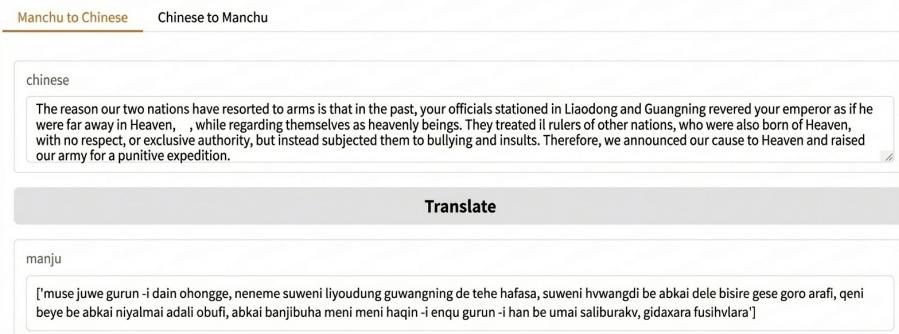


Figure 1: Figure 2

**Figure 3** Manchu to Chinese translation rendering

The final effects of the Manchu-Chinese bidirectional translation system are shown in Figure 4.2 (Chinese-to-Manchu translation rendering) and Figure 4.3 (Manchu-to-Chinese translation rendering). As demonstrated, when users select “Chinese-to-Manchu” translation and input the Chinese example “大爷，我们该走了” (Uncle, we should go), the system outputs the Manchu sentence “sakada amji,be yabume oho.” Conversely, when users select “Manchu-to-Chinese” translation and input the Manchu sentence “sakada amji,be yabume oho,” the system outputs the Chinese sentence “大爷，我们该走了” (Uncle, we should go). This demonstrates that the system can effectively complete Manchu-Chinese bidirectional translation tasks.

**Qing Dynasty Document Manchu-Chinese Translation**[Manchu to English](#) [English to Manchu](#)

Manchu

juwan ninggun de, bi ginjeu be tuwaname tumsi, lousa de juwe tanggy niyalma adabufi.

**Translate**

English Translation

On the 16th, when I went to inspect Jinzhou, I ordered Turlush and Lousa to lead two hundred soldiers,

Figure 2: Figure 3

## 5. Conclusions and Limitations

This paper addresses the challenge of Manchu-Chinese translation. Given the endangered status of Manchu and the limitations of traditional translation methods, we propose a Manchu-Chinese translation system based on the Transformer architecture. Our research adopts transfer learning principles, introducing a pretrained model originally trained for English-Chinese translation and significantly improving low-resource translation performance through fine-tuning on Manchu-Chinese parallel corpora.

Methodologically, we conducted deep preprocessing of collected bilingual corpora tailored to Manchu-Chinese linguistic characteristics. Simultaneously, we employed parameter isolation strategies to differentially freeze and fine-tune various modules of the pretrained model, preserving core linguistic representation capabilities while adapting the model to Manchu-Chinese translation task requirements.

Experimental results demonstrate that the proposed translation system significantly outperforms both traditional statistical neural machine translation models and pure Transformer-based models in terms of accuracy and fluency. Experimental results under parameter isolation strategies far exceed the baseline performance of the pretrained model, proving the effectiveness of our research methodology.

However, no system design can be perfect, and this system is no exception. Compared to Chinese-to-Manchu translation results, Manchu-to-Chinese translation results are noticeably inferior, possibly due to limited model adaptability.

This research not only provides an efficient and reliable technical solution for Manchu-Chinese translation, promoting digital preservation of Manchu resources and cross-lingual cultural dissemination, but also offers reusable technical pathways and practical experience for neural machine translation research

on endangered and low-resource languages, holding important theoretical significance and application value.

In summary, while this design still shows certain deficiencies, these limitations also provide possible directions for future improvement and expansion. We believe that through continuous technical upgrades and refinements, we can certainly construct a more efficient Manchu-Chinese bidirectional translation system.

## References

- [1] Wang Se, Lu Zhong. 20 People Vow to Awaken 238 Volumes of Manchu Archives[N]. Guangming Daily, 2014-11-29(04).
- [2] Dong Shaohua. Representative Fu Chunli: Awakening Sleeping Manchu Archives[N]. Xinjiang Daily, 2015-03-07.
- [3] Kalchbrenner N, Blunsom P. Recurrent continuous translation models[C]. Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. 2013: 1700-
- [4] Meng F, Lu Z, Wang M, et al. Encoding source language with convolutional neural network for machine translation[C]. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. 2015: 20-30.
- [5] Gehring J, Auli M, Grangier D, et al. A convolutional encoder model for neural machine the Association for the 55th Annual Meeting of translation[C]. Proceedings of Computational Linguistics. 2016: 123-135.
- [6] Gehring J, Auli M, Grangier D, et al. Convolutional sequence to sequence learning[C]. Proceedings of the 34th International Conference on Machine Learning. 2017: 1243-1252.
- [7] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks[J]. Advances in Neural Information Processing Systems, 2014, 27.
- [8] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[C]. Proceedings of the 3rd International Conference on Learning Representations. 2015: 1-15.
- [9] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances In Neural Information Processing Systems, 2017, 30.
- [10] Liang X, Wu L, Li J, et al. Multi-teacher distillation with single model for neural machine translation[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2022, 30: 992-1002.
- [11] Singh S M, Singh T D. Low resource machine translation of english-manipuri: A semisupervised approach[J]. Expert Systems with Applications, 2022, 209: 118187.
- [12] Kumar A, Pratap A, Singh A K, et al. Addressing domain shift in neural machine translation via reinforcement learning[J]. Expert Systems with Applications, 2022, 201: 117039.
- [13] Jiang H, Zhang C, Xin Z, et al. Transfer learning based on lexical constraint mechanism in low-resource machine translation[J]. Computers and Electrical

Engineering, 2022, 100:

- [14] Liu S, Liu X, Wong D F, et al. kNN-TL: k-nearest-neighbor transfer learning for lowresource neural machine translation[C]. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. 2023: 1878-1891.
- [15] Huang Jihao, Yu Zhengtao, Yu Zhiqiang, et al. Chinese-Vietnamese Machine Translation Based on Transfer Learning[J]. Journal of Xiamen University, 2021, 60(01): 104-108.
- [16] Xue Junjie. Research on Machine Translation Optimization Model Based on Transfer Learning Technology[J]. Automation & Instrumentation, 2023, 10:
- [17] Gao F, Zhu J, Wu L, et al. Soft contextual data augmentation for neural machine the Association for the 57th Annual Meeting of translation[C]. Proceedings of Computational Linguistics. 2019: 5539-5544.
- [18] Nguyen X P, Joty S, Wu K, et al. Data diversification: A simple strategy for neural machine translation[J]. Advances in Neural Information Processing Systems, 2020, 33: 10018-10029.
- [19] Duan S, Zhao H, Zhang D. Syntax-aware data augmentation for neural machine translation[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2023, 31: 2988-2999.
- [20] Wu X, Xia Y, Zhu J, et al. mixSeq: A simple data augmentation methodfor neural machine translation[C]. Proceedings of the 18th International Conference on Spoken Language Translation. 2021: 192-197.
- [21] Ko W J, El-Kishky A, Renduchintala A, et al. Adapting high-resource NMT models to translate low-resource related languages without parallel data[C]. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. 2021: 802-812.
- [22] Xia M, Kong X, Anastasopoulos A, et al. Generalized data augmentation for lowresource the Association for the 57th Annual Meeting of translation[C]. Proceedings of Computational Linguistics. 2019: 5786-5796.
- [23] Kondo S, Hotate K, Kaneko M, et al. Sentence concatenation approach to data augmentation for neural machine translation[C]. Proceedings of the 2021 Conference of the the Association for Computational Linguistics: Student North American Chapter of Research Workshop. 2021: 143-149.
- [24] Jin C, Qiu S, Xiao N, et al. AdMix: A mixed sample data augmentation method for neural machine translation[C]. Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence. 2022: 4171-4177.
- [25] Zhang Baoxing. Data Augmentation Method for Low-Resource Language Machine Translation Based on Dictionary[J]. Intelligent Computer and Applications, 2024, 14(03): 67-75.
- [26] Yang Dan, Yongcuo, Renqing Zhuoma, et al. Research on Tibetan-Chinese Bidirectional Machine Translation Based on mRASP[J]. Computer Technology and Development, 2023, 33(12): 200-206.
- [27] Li P, Li L, Zhang M, et al. Universal conditional masked language pre-training for neural machine translation[C]. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. 2022: 6379-6391.

- [28] Fan A, Bhosale S, Schwenk H, et al. Beyond english-centric multilingual machine translation[J]. *Journal of Machine Learning Research*, 2021, 22(107): 1-48.
- [29] Raffel C, Shazeer N, Roberts A, et al. Exploring the limits of transfer learning with a unified text-to-text transformer[J]. *Journal of Machine Learning Research*, 2020, 21(140): 1-
- [30] Xue L, Constant N, Roberts A, et al. mT5: A massively multilingual pre-trained text-to-text transformer[C]. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2021: 483-498.
- [31] Costa-jussà M R, Cross J, Çelebi O, et al. No language left behind: Scaling humancentered <https://doi.org/10.> (2022-08-25)[2023-05-26]. *translation*[EB/OL]. machine 48550/arXiv.2207.04672.
- [32] Papineni K, Roukos S, Ward T, et al. Bleu: a method for automatic evaluation of machine the Association for the 40th Annual Meeting of translation[C]. *Proceedings of Computational Linguistics*. 2002: 311-318.

*Source: ChinaXiv — Machine translation. Verify with original.*