

# ***Progress in Literary Translation Based on Reinforcement Learning***

**Weizhe Peng**

*School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing, China  
pengweizhe@bupt.edu.cn*

**Abstract.** Apart from semantic equivalence, literary translation prioritizes the transmission of style, cultural imagery and rhetorical devices, which poses considerable difficulties for automatic translation. Traditional neural machine translation is often plagued by a relative lack of fluency in expression, a lack of cultural meaning, and underutilized creativity in literary works. Unlike previous areas of machine translation, these areas can analyze in depth and improve their literary quality through continuous interaction between reinforcement learning (RL) and the environment. This article will comprehensively assess the current state of education and related aspects in the field of reinforcement learning. First, it provides the fundamental principles and benefits of applying reinforcement learning in the translation process. He then explores these applications by encountering three problem areas: how to solve data loss problems in translation of low-resource languages; How to manage latency and quality with real-time translation; How to maintain consistency of terminology and style in theme-centered translation. In addition, the micro-consultation examined commonly used data sets, measurement standards and obstacles faced by researchers. Research has shown that the reward function has elements of writing, fluidity, fidelity and styling, allowing the model to produce translations that attract the viewer's attention. This challenge still needs to be addressed by establishing reliable reward models, stabilizing training and overcoming costly computing costs. The way forward could include reward mechanisms based on LLM, reinforcement learning algorithms to capture non-political boundaries, and a unified assessment system to integrate culture and language into a real reference framework.

**Keywords:** reinforcement learning, language translation, low-resource language translation, real-time language translation, domain-specific

## **1. Introduction**

The reality of globalization and the promotion of humanist and cultural movements have strengthened the existence of literary publishing. Literary translation is not only a language conversion, but also a strategically important act of cultural, artistic and aesthetic recreation. The heart of the problem lies in the fact that the original text in the target language is constructed symmetrically in terms of semantics, style, rhetorical devices (such as metaphors and symbols), and subtle emotional differences. Although neural modeling machine translation systems have achieved

considerable success in everyday areas, their training objectives are primarily focused on appropriate maximum probability, which makes the algorithms ignore the real mind hidden in the original text. These systems are widely used in literary texts, so their engines often generate grammatically correct but simple translations, avoiding capturing the literary quality and artistic style of the original text. In other words, what poetry lacks most is the transfer of style, idiomatic expression and proper rhythm. In addition, the capacity to adapt to infrastructure involving low-resource languages, real-time interactive scenarios, and legal and medical fields is limited.

Structure  
To address the bottleneck in sentence translation, correctors use reinforcement learning, an advanced training paradigm emphasising learning from interactive feedback [1]. Reinforcement learning does not require such rigid input-output pairs as supervised learning, but efficient implementation alone does not solve these issues. In contrast, the reward system in reinforcement learning algorithms guides the model to develop translations towards higher quality and alignment with specific end goals, such as higher BLEU scores, human preference evaluations, or strict literary style [1]. In the case of literary translation, reward signals can have multidimensional and highly intense characteristics, capable of assessing fluency, fidelity, deviation from the original style, and the success of conveying cultural imagery [2]. This enables translation models to explore translation spaces beyond simple probability maximisation, resulting in more adventurous and artistic translations.

However, portraying meaning and presenting the appeal of language imagery rooted in cultural and stylistic features—significant challenges in automatic translation—receive greater attention in literary translation. Traditional neural machine translation models for literary texts face issues such as rigid style, loss of cultural nuances, and lack of creativity. Reinforcement learning (RL) is a method of development involving continuous interaction with the environment (similar to human evaluators judging translation quality), thus providing a new perspective for improving the quality of machine-translated literature. This paper attempts to track this reinforcement learning process through analysis, supporting literary translation and key related scenarios. First, the article outlines the basic principles and advantages of using reinforcement learning in translation tasks. Then, it provides an in-depth analysis from three specific application angles: first, reinforcement learning can address data sparsity in low-resource language translation; another is that RL can be used to find an appropriate balance between cost and quality in real-time translation tasks; finally, reinforcement learning can maintain terminology consistency to achieve the proper style in domain-specific translation. Secondly, this review briefly points out the most important datasets, evaluation methods, and challenges in these areas. Research shows that designers of reward functions (such as prioritising fluency and accuracy in translation) help train models to output literature more effectively. Constructing and training reward functions is a current challenge. Future considerations may include reward mechanisms based on large language model applications, efficient non-strategic reinforcement learning algorithms, and coherent cross-language and cross-cultural evaluation frameworks. This paper analyses the effectiveness of reinforcement learning methods in terms of word expression, control over language paradigms, and mutual understanding in cultural contexts. This study provides a systematic perspective and valuable reference for deepening the application of reinforcement learning in the field of literary translation.

## 2. Application field

Reinforcement learning proposes a method to optimise the decision-making process of translation models, aiming to achieve the best results. The basic function of the agent (translation model) is to study translation behaviours (lexical choices and sentence adjustments) and receive expected reward

signals from the environment (i.e., the evaluation system). The agent then selects strategies to maximise the expected cumulative future rewards and ultimately learns to adopt strategies. When designing reward signals for literary translation, it is also necessary to go beyond automatic programme metrics such as BLEU, incorporating artistic considerations such as style, originality, and cultural sensitivity [2]. This will be illustrated in detail through three specific cases.

## 2.1. Low-resource language translation

The main problem with low resource language translation is the severe shortage of high-quality parallel corpora, which makes it difficult to train standard neural translation models and even leads to poor model performance. Reinforcement learning aims to introduce a reward mechanism that can effectively utilize various supervised signals (even non parallel corpora) to guide model training. Currently, low resource translation based on reinforcement learning follows two strategies: Q-learning based class interferometry and dual neural networks. The first type is reinforcement learning and quality assessment called upon by researchers. For example, in Korean Chinese neural translation, the theory of combining reinforcement learning and automatic quality assessment has been applied to this problem [3]. This method initially evaluates candidate translations by constructing a quality assessment model, using these scores as reward signals for reinforcement learning. The translation model itself is regarded as an intelligent agent, and the behavior generated by translation is optimized through policy gradient methods to maximize the expected benefits provided by the quality assessment model. In this way, it reduces the reliance on highly available parallel corpora, allowing the model to start with very little sample data and feedback [3]. The next ones are multi task reinforcement learning and curriculum reinforcement learning. In low resource research on language translation, creating multi task reward functions that collaborate with similar tasks (such as language modeling tasks and anti translation tasks) can promote the sharing of knowledge representations between different tasks, and the performance of tasks depends on the degree of generalization under low resource conditions [4]. Meanwhile, researchers have proposed RLDA-NMT, a method for domain adaptation in the form of reinforcement learning. Here, practice utilizing monolingual data in the target domain, which lacks parallel data, and implement a reward model to select generated translations during the reinforcement process to reduce the problem of over alignment in low resource domains [5]. Next, one of the research directions is to use human feedback for reinforcement learning to further improve low resource translation performance [6]. The core of implementing such schemes lies in the ability of reinforcement learning to extract more useful models from a small amount of data, by providing tokens suitable for completing tasks (reward tokens based on vector space of downward and upward translation or concatenation words). Therefore, the problem of data scarcity can only be partially alleviated. For example, when calculating translations containing lower resource languages, a mixed method (trialfa) is used. This is a combination of preference optimization methods and hypergeometric gamma rewards, showing a significant increase in the execution of translation tasks involving languages such as Tigris and Kachary [7]. In addition, this framework can also be seen as a possible way to translate literary works that are difficult to read.

## 2.2. Real-time translation

To achieve this goal, real-time translation systems should generate translations starting from the first input part of the source language content. The key point is to find the optimal balance between translation quality and output delay. Reinforcement learning is highly suitable for modeling such

sequential decision-making processes. In this case, the system must always choose between either being in a "waiting" state or starting to translate the received content. Research has shown that reinforcement learning is significantly effective in solving these tasks. Usually, this method defines real-time translation as a sequential decision problem. The process is as follows: at any given time step, the agent (translation system) decides whether to continue waiting for more input or generate translation output based on the accumulated bilingual data so far, targeting a certain language prefix [8]. The reward system includes facial values, which are the similarity between the final translation result and the reference translation, as well as penalties for translation delays (such as the time period during which the translated subject is triggered). No matter what content is received at each time step, the agent will learn a set of strategies to coordinate the correct waiting time to maintain it, so that the overall translation system can achieve a good balance between two conflicting environmental characteristics faced simultaneously: translation delay and translation/overall quality of daily life [8]. Other studies have further improved the reward design, such as immediate rewards for translation fluency and penalties for untranslated content if the risk exceeds a certain percentage. In fact, these advancements enable models to process more complex sentences and structures at the same speed as subtitles, while presenting excellent results. In recent years, research on simultaneous translation systems using large-scale language models has brought new progress. One approach to solving such problems is based on Divergent Adaptive Policy (APBD), which decouples the adaptation of translation models from the translation models. This method is based on the evaluation of possible biases in the relationships that may arise from future information, and makes read and write decisions. Tests have also shown that this technology has reduced the trade-off between translation accuracy and latency through multiple benchmark tests [9]. Machine learning methods applied to real-time translation of literature (such as simultaneous translation during literary festivals), including making decisions based on the unique properties of literary texts. For example, the strategy of a well-trained reinforcement learning agent can determine whether to wait for further semantic unit construction or choose to wait before word concatenation in the case of incomplete metaphors or parallel structures. Therefore, literary rhetoric will be disseminated in a complete manner rather than hastily.

### 2.3. Specialized translation

Translation in specialized areas such as law, medicine and technology places extremely high demands in terms of terminological accuracy, formal stylistic suitability and consistency. Traditional methods generally use vocabulary lists and align parallel bodies, although they are feasible in these areas, there are difficulties in flexibly using structure-rich vocabulary and phrases in context. This method can also be found in reinforcement learning, as long as it is implemented and applied to tasks that improve style by setting rules. The characteristics of professional translation methods are fixed methods and vocabulary limitations. In machine translation, existing methods combine the power of neural networks with two approaches: the vocabulary constraints imposed during the training phase and the decoding phase, to gently influence the output. Furthermore, reinforcement learning frameworks seem to accept this concept in the most natural way possible. Thanks to specialization, the availability of appropriate words can be incorporated into the reward function, that is, if the correct words that match the target word are used. Temporary learning benefits, and the model masters and begins to follow the rules that define the available words. In addition to terminology, domain-specific style patterns are also very important. Therefore, legal texts must be firm and non-personalized, while medical texts must be clear and concise. By integrating a style workbook into reinforcement learning, consistency with translation characteristics can be achieved.

When implementing reinforcement learning with translation, this deep reward model is often accompanied by the detailed reward model mentioned above. For example, the reward function may consist of some rewards for word order correctness, formal sentence structure, and semantic fidelity. By interacting with the reward environment, the model creates a translation strategy that not only focuses on semantically equivalent propositions, but also pays attention to domain-specific terminology and style boundaries and responds accordingly. A framework called "Translation Revision" has been introduced for a more effective translation-oriented revision of the corpus. This framework assigns two tasks to major linguistic models: finding and tracking the trajectories of unsatisfied constraints, and forming the model through observation and revision. The proportion of forced translation in various vocabulary and structural tensions is 15% higher than that of the large standard universal language model [10]. This method allows machine translation systems to adapt more flexibly to specialized and standardized subdomains of the translated literature. For example, science fiction literature often creates specific professional terms, while historical novels have specific languages at specific times.

#### 2.4. specific data sets and evaluation metrics

For research on the application of reinforcement learning in literary translation, specific data sets and evaluation metrics are equally crucial. Therefore, this article summarizes the data sets and evaluation methods that can be used in the field of reinforcement learning for literary translation. Concerning datasets: First, there are versatile translation datasets (such as the daily domain dataset of the WMT common task) that are often used as a reference for performance tests. Secondly, there are specialized data sets for literary translation; among them, the corpus MMTE (Metrics for Evaluating Machine Translation Quality of Metaphoric Language) is an important resource [2]. This corpus focuses on metaphorical language translation, includes several language pairs, and provides human annotations for metaphors. With regard to translation quality assessment, this approach is particularly suitable for training and evaluating reinforcement reward models that focus on literary rhetoric [2]. Third, there are low-resource, domain-specific data sets. For low-resource research, benchmarks such as FLoRes are available; For specialized areas, parallel bodies such as JRC-Acquis in the legal field and EMEA in the medical field provide valuable resources. These data sets can be used to build and validate RL effectiveness in specific scenarios. Regarding evaluation metrics, first, there are automated metrics: in addition to versatile metrics such as BLUE, RED and METEOR, to evaluate literary quality, context-based semantic similarity metrics such as BERTScore can be used, as well as metrics specifically designed to evaluate style transfer. A systematic review of methods for evaluating the quality of machine translation provides important references for the design of RL reward functions [11]. Second, there is human evaluation: given the highly subjective nature of the quality of literary translation, human evaluation remains the gold standard. Typically, a Likert scale is used to rate translated texts on dimensions such as fluidity, fidelity, style appropriateness, and overall quality. Third, there is LLM-based evaluation: with the advancement of large language models, the use of advanced models such as GPT-4 as judges to assign multidimensional scores to translations (such as fluency and fidelity mentioned above) has become an effective and reliable alternative evaluation method, often used in building reward models for reinforcement training [11].

Although reinforcement learning has brought new approaches to optimizing literary translation, it still faces many challenges. Therefore, reward design is a key bottleneck. It is essential to understand the challenges faced by reinforcement models that can reliably and faithfully assess the quality of translated literary works. Currently, most academic work relies on proxy platforms for human validation or large-scale linguistic models, which is an expensive task and can also lead to

bias [2]. Secondly, it is about maintaining driving stability and mixing several working conditions. However, this method also has some drawbacks: first, they are often unstable in terms of training (PPO) and have low sampling efficiency, as they can result in considerable training costs in situations requiring repeated experiments. Finally, the conciliation of multiple objectives is also part of the subtitle of the work. In addition to serving primarily as a means of precise semantic translation, literary translation also reflects stylistic correctness and cultural arrangement. There are challenges to building a reward function that aligns all target frameworks, which also means the complexity of the task.

## 2.5. Future research

Future research is expected to lead to success in these areas: first, using frameworks of large linguistic models to advance more complex reward systems that can deeply understand semantics and aesthetics without the need to extend algorithms under development. An example here is a subtle evaluation metric used in MMT research—the concept of metaphorical equivalence, which can ideally serve as a basis for building more complex literary translation reward functions [2]. second, by integrating more reliable and stable non-strategic techniques with Wide Language Models (MLL) adjustment, RL algorithms can be improved to maximize sample efficiency and training robustness. By designing professional data sets covering various genres and languages, a broader evaluation framework is built, then cutting-edge developments are made to further improve the field [12]. third, consider using reinforcement learning techniques for peripheral and advanced translation, which will process the interactive comments of human editors to get a fast, high-end personalized translation, producing very pleasant effects.

## 3. Conclusion

This study reviews the latest developments and applications of reinforcement learning in the field of literary translation. In this regard, in the face of the main obstacles of literary translation - finding a balance between style uniqueness, cultural transmission and originality - given the sequential mode of decision-making of reinforcement learning and the process of reward-based optimization thinking, its performance far exceeds traditional supervised learning modes and seems superior. This article provides in-depth analysis from three key application dimensions: In resource scarcity scenarios, reinforcement learning can use learning techniques such as high-quality assessment and curriculum design to derive effective translation models from limited data. In real-time translation tasks, reinforcement learning tends to model the decision process of "waiting for translation" to find a balance between translation delay and quality, that is, the actual translation will only be formed after the target data has been fully processed. The main conclusion of this study is that the selection of complex reward methods in areas with model-based evaluators has led to greater success in these applications. However, this area faces some challenges, including the low quality of the reward, the lack of robustness and effectiveness in the reinforcement training offering, and the multi-objective optimization objective. In the future, the overall improvement of learning quality through reinforcement in deeper level work and machine 2D art translation will have more humanistic importance, realizing a triple combination of superior core systems, effective training strategies, and extensive evaluation frameworks.

## References

- [1] Cao, Y., Sheng, Q. Z., McAuley, J., & Yao, L. (2023). Reinforcement learning for generative AI: A survey. arXiv preprint arXiv: 2308.14328.
- [2] Wang, S., Zhang, G., Wu, H., Loakman, T., Huang, W., & Lin, C. (2024, November). MMTE: Corpus and metrics for evaluating machine translation quality of metaphorical language. In *Proceedings of the 2024 conference on empirical methods in natural language processing* (pp. 11343-11358).
- [3] Liu, F., Zhao, Y., Jin, G., Lu, X., Jin, Z., & Cui, R. (2023, October). Korean-Chinese machine translation method based on independent language features. In *China Conference on Machine Translation* (pp. 37-49). Singapore: Springer Nature Singapore.
- [4] Wang, Y. (2025). Research on low-resource language machine translation based on reinforcement learning. *Information & Computer*, 37(7), 92–95.
- [5] Zhang, H., Liu, M., Li, C., Chen, Y., Xu, J., & Zhou, M. (2024, May). A reinforcement learning approach to improve low-resource machine translation leveraging domain monolingual data. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* (pp. 1486-1497).
- [6] Wang, L., & Xiao, Y. (2024). Improving Low-Resource Machine Translation Using Reinforcement Learning from Human Feedback. *Intelligent Automation & Soft Computing*, 39(4).
- [7] Nair, A. R., Gupta, D., Paul, B., & Bhavani, J. S. (2025). Enhancing Low-Resource Indian Language Machine Translation Using Large Language Models With Preference Optimization and Hypergeometric-Gamma Reward. *IEEE Access*, 14, 1641-1665.
- [8] Arivazhagan, N., Cherry, C., Te, I., Macherey, W., Baljekar, P., & Foster, G. (2020, May). Re-translation strategies for long form, simultaneous, spoken language translation. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 7919-7923). IEEE.
- [9] Zhao, L., Fan, K., Luo, W., Jing, W., Wang, S., Zeng, Z., & Huang, Z. (2023, December). Adaptive policy with wait-k model for simultaneous translation. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 4816-4832).
- [10] Huang, P., Mu, Y., Wu, Y., Li, B., Xiao, C., Xiao, T., & Jingbo, Z. (2024, July). Translate-and-revise: Boosting large language models for constrained translation. In *Proceedings of the 23rd Chinese National Conference on Computational Linguistics (Volume 1: Main Conference)* (pp. 1059-1074).
- [11] Wang, J., Zhuang, C., & Wei, Y. (2024). Machine translation quality assessment: Methods, applications, and prospects. *Foreign Language and Literature*, 40(3), 1–12.
- [12] Liang, J., Tang, H., Ma, Y., Liu, J., Zheng, Y., Hu, S., ... & Hao, J. (2025). Squeeze the soaked sponge: Efficient off-policy reinforcement finetuning for large language model. arXiv preprint arXiv: 2507.06892.