NHK Submission to WAT 2025: Leveraging Preference Optimization for Japanese–English Article-Level News Translation Tasks

Hideya Mino, Rei Endo, and Yoshihiko Kawai
NHK Science and Technology Research Laboratories
1-10-11, Kinuta, Setagaya-ku, Tokyo, Japan
{mino.h-gq, endou.r-mm, kawai.y-lk}@nhk.or.jp

Abstract

This paper describes our submission to the Japanese → English: Article-level News Translation shared task as part of WAT 2025. In this shared task, participants were provided with a small but high-quality parallel corpus along with two intermediate English translations: a literal translation and a style-adapted translation. To effectively exploit these limited training data, our system employs a large language model trained via supervised fine-tuning followed by direct preference optimization (DPO), a preference learning technique for aligning model outputs with professional-quality references. By leveraging literal and style-adapted intermediate translations as negative (rejected) samples and human-edited English articles as positive (chosen) samples in DPO training, our model achieved notable improvements in translation quality. We evaluated our approach using BLEU scores and human assessments.

1 Introduction

We describe the system submitted by Team NHK as part of the the Japanese → English Article-level News Translation shared task at WAT 2025 (Shirai et al., 2025). The three shared tasks that were part of this shared task were as follows: Task 1 involved literal English translation of the Japanese articles, Task 2 involved style-adopted translation of the Japanese articles, and Task 3 involved translation into the actually published English articles from the Japanese articles. We participated in Task 3, which focused on article-level translation and required maintenance of coherence, consistency, and stylistic appropriateness beyond individual sentence-level translation. In addition to a limited amount of high-quality parallel data, two supplementary English translations for each Japanese article were provided: a literal translation, and a news-style translation, which contained edits of the literal version adapted for readability and stylistic naturalness. These two versions can be viewed as intermediate drafts that reflect different stages of the editorial translation process. Our approach leveraged these intermediate translations to improve model alignment and translation quality. We adopted a two-stage training process:

1. Supervised fine-tuning (SFT) of a large language model (LLM) on the article-level parallel corpus.

2. Direct preference optimization (DPO) (Rafailov et al., 2023) using preference pairs constructed from the provided translation variants. In DPO training, the reference English articles served as the "chosen" responses, while the literal and news-style translations acted as "rejected" responses.

We report both automatic and human evaluations showing the effectiveness of this approach.

2 System Overview

A unique aspect of this task was the availability of intermediate translation drafts alongside the official reference translations. Given the limited parallel data, we explore methods to leverage this auxiliary information to enhance translation accuracy.

Since the training corpus was too small to build a conventional neural machine translation system, we adopted an LLM fine-tuning approach. We also explored preference-based optimization methods such as reinforcement learning from human feedback (RLHF) (Ouyang et al., 2022) and DPO. These alignment methods adjust LLM behavior to better reflect human preferences and have been shown to improve performance across various natural language generation tasks (Ziegler et al., 2019).

Because the literal and news-style translations often contained lexical or syntactic deviations from the final references, they served as ideal "negative examples" for preference-based learning. We em-

Proceedings of the Twelfth Workshop on Asian Translation (WAT 2025), pages 98–102
December 24, 2025 ©2025 Association for Computational Linguistics

	Article	Token (English)		
		Original	Literal	News-style
Train	227	78,064	82,199	86,970
Development	50	15,713	17,135	17,788
Test	100	35,776	37,987	39,453

Table 1: Corpus statistics.

Learning rate schedule	cosine
Learning rate warmup	50
Sequence length	2048
Optimizer	adamW
Learning rate	5e-5
Weight decay	0.05
Micro batch size	1
Gradient accumulation steps	1
Precision	bfloat 16
Gradient clipping	1.0

Table 2: Hyperparameters for SFT with Qwen3-8B.

Learning rate schedule	cosine
Learning rate warmup	20
Sequence length	2048
Optimizer	adamW
Learning rate	1e-5
Weight decay	0.05
Micro batch size	1
Gradient accumulation steps	1
Precision	bfloat16
Gradient clipping	1.0
LoRA rank	2
LoRA alpha	4
Attention modules	q, v

Table 3: Hyperparameters for DPO with Qwen3-8B.

ployed DPO for its simplicity and training stability, constructing preference data from these preliminary translations.

3 Experimental Setup

3.1 Dataset

We used the article-level corpus provided by the WAT 2025 organizers. The dataset models the editorial workflow for translating Japanese news into English and contains 377 pairs of Japanese and English articles from Jiji Press[1], each accompanied by literal and news-style English translations. We define the following abbreviations:

- **ja_orig.**: Original Japanese article published by Jiji Press.
- **en_orig.**: Original English article also published by the same Jiji Press. This is an English version of the original Japanese article and is intended for an international audience. The content of this article may differ from the Japanese version.
- **en_literal**: Literal English translation of the Japanese article.
- **en_news-style**: A translation of the original English article edited to match the content of the original Japanese article. The order in which information is presented, vocabulary, and number of lines may differ from those of the original Japanese article.

The literal and news-style translations were newly created for this shared task. The literal translations prioritized fidelity, while the news-style versions prioritized fluency and natural English expression. Of these 377 articles, 227 belonged to the training set, 50 belonged to the development set, and 100 belonged to the test set. Table 1 shows the statistics of the corpus.

For SFT, we used (ja_orig., en_orig.) pairs. For DPO, we constructed preference tuples (x, y_r, y_c) defined as follows:

$$(x, y_r, y_c) = \begin{cases} (\text{ja\_orig.}, \text{en\_literal}, \text{en\_orig.}) \\ (\text{ja\_orig.}, \text{en\_news-style}, \text{en\_orig.}), \end{cases}$$

where x is the source article, y_r is the rejected translation, and y_c is the chosen translation.

3.2 Model and Training

We employed Qwen3-8B (Yang et al., 2025) as the base LLM. The training process consisted of two stages. First, we performed full-parameter SFT using the 227 article-level parallel pairs in the

[1] https://lotus.kuee.kyoto-u.ac.jp/WAT/jiji-corpus/2025/

Model	BLEU
GPT-4o	13.33
Zero-shot LLM	14.09
Fine-tuned LLM (SFT only)	19.54
Proposed (SFT + DPO)	**22.72**

Table 4: Official automatic evaluation results.

training set. Second, we applied DPO with *low-rank adaptation* (LoRA) (Hu et al., 2022) using 454 preference pairs constructed as described in Section 3.1.

The hyperparameters used in both stages are summarized in Tables 2 and 3. These configurations were determined based on hyperparameter tuning conducted on the development set. All experiments were carried out on a single NVIDIA A100 GPU.

3.3 Evaluation

Our system was evaluated using both automatic and human evaluations. Based on the official evaluation framework, we compared our system against three baseline systems: GPT-4o[2], zero-shot LLM, and fine-tuned LLM with (ja_orig., en_orig.) parallel data (i.e. SFT Qwen3-8B model).

For the automatic evaluation, the task organizers calculated case-sensitive BLEU (Papineni et al., 2002) scores using SacreBLEU (Post, 2018).

For the human evaluation, the task organizers employed two bilingual evaluators to assess the translation outputs of our system and the three baselines. Evaluation was conducted on 100 test articles through pairwise comparisons, separately measuring *adequacy* (semantic faithfulness) and *fluency* (linguistic naturalness). Each pair of system outputs was judged as a win, tie, or loss for our proposed model.

4 Results

4.1 Automatic Evaluation

Table 4 presents the official BLEU scores for all systems. Our proposed method (SFT + DPO) achieved the highest BLEU score, outperforming both the zero-shot and SFT-only models, thereby demonstrating the effectiveness of preference optimization in improving translation quality.

vs Baseline	Win	Tie	Lose
vs GPT-4o	5.5 / 13.5	27 / 38.5	67.5 / 48
vs Zero-shot LLM	14.5 / 19	43 / 42	42.5 / 39
vs Fine-tuned LLM (SFT only)	47 / 22	40 / 51.5	13 / 26.5

Table 5: Official human evaluation results for adequacy/fluency. Win, tie, and loss indicate the number of evaluations our proposed method won against, tied with, or lost against the baseline method.

4.2 Human Evaluation

Table 5 summarizes the official human evaluation results for adequacy and fluency. The values indicate the number of cases (out of 100) that our proposed model *won* against, *tied* with, or *lost* against each baseline, averaged across two evaluators.

Our proposed model demonstrated mixed performance in human evaluation. While it outperformed the SFT-only baseline in adequacy (47 wins vs 13 losses), indicating that DPO training improved semantic faithfulness, it underperformed in all other assessments. The model was particularly weak in fluency compared to GPT-4o (13.5 wins vs 48 losses) and zero-shot LLM (19 wins vs 39 losses), suggesting that maintaining stylistic naturalness remains a significant challenge with the current approach.

This discrepancy between BLEU and human evaluations aligns with prior observations (Sulem et al., 2018; Mathur et al., 2020) that automatic metrics often poorly capture human-perceived quality, particularly in article-level translation tasks where coherence and stylistic appropriateness play important roles.

5 Related Work

DPO (Rafailov et al., 2023) simplifies RLHF by eliminating reward modeling and directly training on preference pairs. Because of its efficiency and stability, this approach has been widely adopted in various NLP domains (Grattafiori et al., 2024; Wu et al., 2024; Sun et al., 2025).

LLMs can perform many zero- or few-shot tasks (Brown et al., 2020), but instruction or preference fine-tuning further enhances task alignment (Ouyang et al., 2022). Since collecting preference data is easier than implementing fully supervised learning, DPO offers a practical approach for adapting LLMs to domain-specific objectives. DPO (Rafailov et al., 2023) directly optimizes LLMs with preference data by removing an ex-

[2] gpt-4o-2024-11-20 version provided by Azure OpenAI.

tra reward model. We utilized DPO in this work since it is both easy to use and highly effective.

6 Conclusion

We have presented our WAT 2025 submission for Japanese→English article-level news translation. Our system leverages DPO to align LLMs using intermediate translation data as preference signals. Experimental results suggest that incorporating editorial-stage translations as negative examples allows model to achieve higher BLEU scores. Future work includes scaling this approach to handle larger datasets and exploring finer-grained document-level alignment.

Acknowledgments

These research results were obtained from the commissioned research (No. 225) by National Institute of Information and Communications Technology (NICT), Japan.

References

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, and 12 others. 2020. Language models are few-shot learners. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA. Curran Associates Inc.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, and 1 others. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. Lora: Low-rank adaptation of large language models. In *ICLR*. OpenReview.net.

Nitika Mathur, Timothy Baldwin, and Trevor Cohn. 2020. Tangled up in BLEU: Reevaluating the evaluation of automatic machine translation evaluation metrics. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4984–4997, Online. Association for Computational Linguistics.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*, volume 35, pages 27730–27744. Curran Associates, Inc.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Matt Post. 2018. A call for clarity in reporting BLEU scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Brussels, Belgium. Association for Computational Linguistics.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. In *Advances in Neural Information Processing Systems*, volume 36, pages 53728–53741. Curran Associates, Inc.

Naoto Shirai, Kazutaka Kinugawa, Hitoshi Ito, Hideya Mino, and Yoshihiko Kawai. 2025. Findings of the wat 2025 shared task on japanese-english article-level news translation. In *Proceedings of the 12th Workshop on Asian Translation*, Mumbai, India.

Elior Sulem, Omri Abend, and Ari Rappoport. 2018. BLEU is not suitable for the evaluation of text simplification. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 738–744, Brussels, Belgium. Association for Computational Linguistics.

Haoxiang Sun, Ruize Gao, Pei Zhang, Baosong Yang, and Rui Wang. 2025. Enhancing machine translation with self-supervised preference data. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 23916–23934, Vienna, Austria. Association for Computational Linguistics.

Qiyu Wu, Masaaki Nagata, Zhongtao Miao, and Yoshimasa Tsuruoka. 2024. Word alignment as preference for machine translation. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 3223–3239, Miami, Florida, USA. Association for Computational Linguistics.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, and 41 others. 2025. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*.

Daniel M Ziegler, Nisan Stiennon, Jeffrey Wu, Tom B Brown, Alec Radford, Dario Amodei, Paul Christiano, and Geoffrey Irving. 2019. Fine-tuning language models from human preferences. *arXiv preprint arXiv:1909.08593*.