Improving Japanese-English Patent Claim Translation via Reasoning and GRPO with Syntactic Reward

Shogo Tsujimoto[1] **Katsuki Chousa**[2] **Masaaki Nagata**[2] **Ryohei Sasano**[1]
[1]Graduate School of Informatics, Nagoya University [2]NTT, Inc.
tsujimoto.shogo.i7@s.mail.nagoya-u.ac.jp katsuki.chousa@ntt.com
masaaki.nagata@ntt.com sasano@i.nagoya-u.ac.jp

Abstract

Patent claims are legal documents with distinctive formats and complex syntactic dependencies. Translating these claims requires not only accurately understanding the complex syntactic structure of the source claims but also ensuring the translation follows the structural conventions of claims in the target language. In this study, we leverage reasoning models to disentangle the complex syntactic structures of source claims. Furthermore, to ensure the translation reflects the structural requirements in the target language, we incorporate FastKASSIM—a syntactic similarity metric—into the reward of reinforcement learning, alongside conventional metrics such as BLEU and COMET. Experimental results confirm that our model, trained on Qwen3-8B, outperforms Qwen3-235B in syntactic similarity while simultaneously achieving comparable translation quality, despite the latter having approximately 30 times more parameters. Automatic and human evaluations show that models trained with rewards combining FastKASSIM and translation quality metrics such as COMET or COMETKiwi achieve strong performance, suggesting that syntactic rewards help improve patent-specific structural conformity while maintaining translation quality.

1 Introduction

Patent claims legally define the technical scope of an invention using distinctive formats, such as noun phrases with complex syntactic dependencies. Translating these claims presents two challenges: accurately understanding the complex syntactic structure of the source claims and ensuring that the translation faithfully reflects the structural conventions of the target language.

Recent advances in Large Language Models (LLMs) demonstrate that generating a reasoning process before a final answer can improve performance on complex tasks. While leveraging such reasoning capability in patent claim translation may facilitate a more precise understanding of the source text, the full extent of this benefit has yet to be fully established.

Additionally, to effectively align LLMs with such specific tasks, Group Relative Policy Optimization (GRPO) (DeepSeek-AI, 2025) has attracted attention. Previous studies (Feng et al., 2025; Wang et al., 2025b) have attempted to apply GRPO to machine translation by using standard translation quality metrics, such as BLEU (Papineni et al., 2002) and COMET (Rei et al., 2020), as rewards, demonstrating significant improvements in translation accuracy. However, in patent claim translation, relying solely on these metrics is insufficient to reflect the target-side structural conventions of patent claims.[1] These metrics focus primarily on surface-level agreement or semantic similarity and do not necessarily ensure adherence to the structural constraints specific to patents.

In this study, we first apply reasoning models to patent claim translation to improve the ability to understand complex source structures. Evaluating such translations requires assessing two orthogonal dimensions: semantic quality and structural adherence in the target language. Second, to explicitly enforce the latter, we propose incorporating FastKASSIM (Chen et al., 2023), a syntactic similarity metric, into the GRPO reward. Through experiments on Japanese-English patent claim translation, we confirm that our model, trained on Qwen3-8B, outperforms Qwen3-235B in syntactic similarity while achieving comparable translation quality, despite the latter having approximately 30 times more parameters.

[1]For example, an English claim must generally form a single noun phrase where constituent elements of the invention are usually separated by ; (semicolons) and , (commas) depending on the context, whereas in Japanese, by 、(Japanese commas).

Proceedings of the 15th International Joint Conference on Natural Language Processing and the 5th Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics, pages 95–105
November 6-10, 2026 ©2026 Association for Computational Linguistics

2 Related Work

Explicitly generating a reasoning process before a final answer has been reported to improve performance on complex tasks (Wei et al., 2022; Kojima et al., 2022). Initial studies suggest that reasoning models are beneficial for non-literal translations, such as metaphors and similes, as well as for translations that require long contexts and cultural understanding, particularly in literary works (Liu et al., 2025; Wang et al., 2025a). On the other hand, in general-domain machine translation tasks, it is reported that the benefits of such reasoning capabilities are limited (Feng et al., 2025; Zheng et al., 2025). This limitation is likely because general-domain texts typically feature relatively simple syntax, which does not necessitate deep reasoning. In contrast, patent claims are characterized by intricate source structures. Given this high complexity, reasoning capabilities would be effective for translating patent claims, as they can facilitate the disentanglement of complex dependencies that pose difficulties for standard translation models.

GRPO (DeepSeek-AI, 2025) is a reinforcement learning approach that efficiently tailors LLMs for specialized tasks by optimizing their outputs according to task-specific reward signals. MT-R1-Zero (Feng et al., 2025) applied GRPO to machine translation without supervised fine-tuning or cold start data, by employing automatic evaluation metrics as reward functions. This method prompts the LLMs to enclose the thinking process within `<think>` tags and the final translation in `<translate>` tags. They demonstrated that using translation quality metrics, such as BLEU (Papineni et al., 2002), COMET (Rei et al., 2020), and COMETKiwi (Rei et al., 2022), as rewards directly improves translation quality. While these rewards improve general translation quality, they do not explicitly penalize syntactic deviations. Consequently, relying solely on standard metrics is insufficient for domains such as patent translation, where adherence to target-side structural conventions is essential.

FastKASSIM (Chen et al., 2023) is a metric that measures syntactic similarity by quantifying the overlap of subtrees between two syntax trees using tree kernels. In contrast to computationally expensive tree edit distance, FastKASSIM efficiently captures structural isomorphism regardless of vocabulary overlap and demonstrates a high correlation with human judgment. Unlike standard metrics, FastKASSIM directly quantifies structural isomorphism. Therefore, incorporating such a syntactic metric into the reward function is essential to bridge the gap between semantic correctness and structural conformity in patent claim translation.

3 Proposed Method

Translating patent claims requires understanding the complex structure of the source claims and generating translations that conform to the sentence structure of the target claims. In this study, we propose a method that combines reasoning models with a syntactic reward in GRPO to address these requirements. We first adopt reasoning models to accurately capture the complex syntactic structures of patent claims. By utilizing reasoning capability, the model can perform a deep analysis of dependency structures, generating translations based on this reasoning. Second, to reflect the target-side structural conventions, we incorporate a syntactic reward into GRPO based on the MT-R1-Zero framework.

The reward r during training consists of two elements: S_{format}, which is 1 if the format constraints are satisfied and -1 otherwise, and S_{metric}, which measures translation quality and structural conformity. Specifically, r is defined as follows:

$$r = \begin{cases} S_{format} - 2, & \text{if } S_{format} = -1 \\ S_{format} + S_{metric}, & \text{if } S_{format} = 1 \end{cases}. \quad (1)$$

To explicitly enforce structural adherence alongside semantic quality, we define S_{metric} as the sum of the scores from a selected set of evaluation metrics $\mathcal{M}$:

$$S_{metric} = \sum_{m \in \mathcal{M}} S_m. \quad (2)$$

Here, $\mathcal{M} \subseteq \{\text{FastKASSIM}, \text{BLEU}, \text{COMET}, \text{COMETKiwi}\}$ denotes the set of metrics selected for the reward, and S_m represents the score calculated by metric m. Note that we exclude the simultaneous use of COMET and COMETKiwi (i.e., $\{\text{COMET}, \text{COMETKiwi}\} \not\subseteq \mathcal{M}$). For FastKASSIM, which is an asymmetric metric, we define its score $S_{\text{FastKASSIM}}$ as the average of the bidirectional similarities between the reference ref and the generated translation $trans$:

$$S_{\text{FastKASSIM}} = \frac{1}{2}(\text{FastKASSIM}(ref, trans) + \text{FastKASSIM}(trans, ref)). \quad (3)$$

This syntactic reward encourages the translation to be faithful to the syntactic structure of the reference. Crucially, because this evaluation is based on the similarity between the model's output and the reference translation, our approach is independent of the source language and focused solely on the target side. In the patent domain, syntactic and structural conventions are determined by the filing country's practices and legal frameworks. Therefore, our approach is designed to adapt to these country-specific requirements, such as the structural differences between Japanese and English claims.

4 Experimental Setup

4.1 Task and Dataset

We evaluated our method on patent claim translation using the claim subset of the publicly available patent parallel corpus, JaParaPat (Nagata et al., 2024). In this paper, we primarily focus our evaluation and analysis on the Japanese-to-English (JA-EN) direction, as it holds greater practical significance and higher demand in intellectual property operations, such as foreign patent filings. The experimental results and analysis for the English-to-Japanese (EN-JA) direction are provided in Appendix D.

4.2 Models Used

We employed Qwen3-8B in thinking mode as the base model for GRPO training, jointly training a model on both Japanese↔English directions.[2] As baselines, we used Qwen3-8B and Qwen3-235B-A22B in both thinking and non-thinking modes, alongside GPT-5.[3]

4.3 Rewards and Evaluation Metrics

As defined in Section 3, we used combinations of FastKASSIM, BLEU, COMET, and COMETKiwi as the metric reward.[4]

To evaluate translation performance, in addition to these four metrics, we used XCOMET-XXL (Guerreiro et al., 2024) as a neutral metric not used in the reward calculation. Because it was excluded from the GRPO reward calculation and trained with a different objective from COMET and COMETKiwi, it serves as a reliable benchmark free from reward-hacking bias. Similarly, to evaluate structural fidelity independently of the FastKASSIM reward, we employed Tree Edit Distance (TED) for configurations that achieved a balance between syntactic and semantic metrics. We used Stanza (Qi et al., 2020) for parsing required for both FastKASSIM and TED calculations. Specifically, we applied TED to constituent phrase structure trees with leaf nodes removed, normalized by the number of nodes:

$$\mathrm{TED}_{\mathrm{norm}}(T_1, T_2) = 1 - \frac{\mathrm{TED}(T_1, T_2)}{|T_1| + |T_2| - 2} \quad (4)$$

where T_1 and T_2 denote the constituent phrase structure trees of the two sentences being compared, and $|T|$ represents the number of nodes in tree T.

To assess the statistical significance of the key comparisons discussed in Section 5, we performed paired bootstrap resampling (10,000 iterations) across the full test set.

Additionally, we conducted human evaluation using two complementary protocols. First, following Direct Assessment with a continuous rating scale (Graham et al., 2013), evaluators assigned a direct assessment score on a 0–100 scale. Second, inspired by the question-based evaluation framework of LiTransProQA (Zhang et al., 2025), we employed a checklist-style evaluation in which evaluators rated each criterion on a five-point scale from 1 to 5. Our human evaluation was conducted by a linguistic data service provider specializing in syntactic treebank construction and reference translation development. Each sample was assessed by an expert in syntactic analysis and translation, then double-checked by another professional. The detailed human evaluation protocol is provided in Appendix C.

5 Results and Analysis

5.1 Automatic Evaluation

The left side of Table 1 shows the automatic evaluation results. First, we examine the impact of reasoning on the base models. Focusing on XCOMET-XXL, a semantic similarity metric highly correlated with human evaluation, models using reasoning outperformed those without it across both the 8B and 235B model scales in the JA-EN direction.[5] This

[2] See Appendix A for detailed settings and Appendix B for the prompt template.

[3] We used `gpt-5-2025-08-07` via the OpenAI API.

[4] We explored applying min-max normalization; however, our experiments did not confirm its effectiveness, as detailed in Appendix E.

[5] Reasoning models did not improve BLEU and FastKASSIM. However, these metrics are insufficient for capturing

Model				Automatic evaluation					Human evaluation					
F	B	C	K	Fast KASSIM	BLEU	COMET	COMET Kiwi	XCOMET XXL	Direct Assessment	A Fidelity	B Coverage	C Scope	D Syntax	E Fluency
8B w/o Thinking				57.87	30.84	78.33	77.80	81.16	67.00	4.79	4.63	4.89	4.65	4.81
8B w/ Thinking				56.82	27.63	77.90	78.46	82.66	65.20	4.65	4.69	4.92	4.65	4.76
○				**65.13**	32.42	78.93	76.51	80.92	59.20	4.75	4.57	4.88	4.84	3.96
	○			59.61	**42.80**	81.16	76.87	81.10	83.80	4.92	4.73	4.88	4.91	4.95
○	○			64.38	41.65	81.15	76.83	80.40	81.00	4.84	4.75	4.95	4.88	4.77
		○		60.10	36.85	80.78	78.90	83.78	83.20	4.89	4.75	4.91	4.93	4.85
○		○		63.44	36.51	80.10	78.37	83.08	**90.80**	**4.96**	4.77	**4.97**	**4.98**	4.92
	○	○		60.74	42.44	**81.45**	77.01	81.06	87.00	4.89	4.75	4.95	4.90	**5.00**
○	○	○		63.17	39.75	80.30	78.09	83.09	85.80	4.93	4.87	4.95	4.87	4.89
			○	53.82	22.63	78.81	**80.60**	83.92	61.20	4.80	4.65	4.77	4.47	4.92
○			○	63.03	30.19	80.17	79.73	84.43	79.20	4.88	4.75	4.89	4.90	4.80
	○		○	60.08	40.70	80.57	78.41	82.93	86.20	4.91	4.83	4.92	4.94	4.83
○	○		○	64.37	40.81	81.12	77.73	80.54	72.40	4.93	4.72	4.93	4.94	4.32
235B-Instruct				56.57	34.70	79.88	78.81	84.43	72.00	4.64	4.81	4.93	4.75	4.83
235B-Thinking				57.32	32.97	79.57	79.03	**85.05**	87.20	4.92	4.85	4.93	4.90	4.91
GPT-5				57.01	36.84	80.48	78.92	84.63	87.60	4.93	**4.95**	**4.97**	4.86	4.85

Table 1: Automatic and human evaluation results for JA-EN patent claim translation. The left block shows automatic evaluation scores, and the right block shows human evaluation scores. In the "Model" columns, F, B, C, and K denote FastKASSIM, BLEU, COMET, and COMETKiwi used as rewards, respectively. "Direct Assessment" denotes the average independent direct assessment score on a 0–100 scale, while A–E are average item-wise checklist scores on a five-point scale from 1 to 5. The detailed human evaluation criteria are provided in Appendix C.

suggests that reasoning enhances the understanding of the complex syntactic structures of the source claims, improving comprehension of the semantic content. Based on this evidence, we adopted the reasoning model as the base for GRPO training.

Regarding the effectiveness of the syntactic reward, GRPO-trained models showed improvements in the scores corresponding to the rewards. While optimizing directly on evaluation metrics raises concerns about reward hacking, our independent evaluation using XCOMET-XXL—which was not used as a reward—suggests that the improvements generalize beyond the reward signal. Specifically, using FastKASSIM as the sole reward improved the FastKASSIM score substantially by +8.31pts in JA-EN, confirming enhanced adherence of the target structure. A similar trend was also observed in EN-JA, where FastKASSIM-based rewards consistently improved syntactic similarity, as detailed in Appendix D. Furthermore, when FastKASSIM was combined with other metrics, improvements were observed in the evaluation metrics corresponding to each reward, consistent with the findings in MT-R1-Zero. Notably, the configuration combining FastKASSIM and COMETKiwi not only maintained high scores in both metrics but also achieved the highest XCOMET-XXL of 84.43 among the 8B models in JA-EN. This indicates that combining FastKASSIM with a semantic metric can balance target-side structural conformity and semantic quality in automatic evaluation. To confirm this statistically, we compared the models trained with FastKASSIM & COMET against BLEU & COMET, and FastKASSIM & COMETKiwi against BLEU & COMETKiwi. We observed statistically significant improvements in both FastKASSIM and XCOMET-XXL at the $p < 0.05$ level, confirming that integrating a syntactic reward enhances structural fidelity while maintaining high translation quality. As shown in the qualitative analysis in Appendix F, generated outputs reflect target-side structural conventions, and we observed reward-dependent differences in the reasoning traces, particularly in references to fluency and syntactic structure, suggesting that FastKASSIM gains represent structural alignment rather than metric hacking.

Our Qwen3-8B-based model also outperforms Qwen3-235B in syntactic similarity while achieving comparable translation quality, despite the latter having approximately 30 times more parameters. Since neither reasoning nor scaling from 8B to 235B substantially improved FastKASSIM for the baselines, patent-specific structural conformity appears to require explicit optimization with FastKASSIM.

To ensure that the structural improvements

source comprehension, as FastKASSIM focuses on target-side structure while BLEU penalizes the paraphrases. To better evaluate deep understanding, we focus on XCOMET-XXL.

Model	JA-EN	EN-JA
8B w/o Thinking	0.794	0.749
8B w/ Thinking	0.796	0.747
FastKASSIM & COMET	**0.817**	0.752
FastKASSIM & COMETKiwi	0.805	**0.756**

Table 2: Syntactic similarity evaluated by normalized Tree Edit Distance (TED).

are not merely artifacts of optimizing directly for FastKASSIM, we conducted an additional evaluation using Tree Edit Distance (TED). We specifically evaluated the models trained with FastKASSIM & COMET and FastKASSIM & COMETKiwi, as these configurations demonstrated a strong balance of high semantic and syntactic similarities. As shown in Table 2, these configurations consistently improved syntactic similarity in both translation directions compared to the 8B baselines, confirming that the improvement in syntactic similarity is also observed with a metric not used during training.

5.2 Human Evaluation

The human evaluation results (the right side of Table 1) further clarify the effect of each reward combination. The model trained with FastKASSIM and COMET as rewards achieved the highest direct assessment score among all models (90.8), with consistently high item-wise scores and a particularly strong Claim Form & Syntax score (D=4.98). This aligns with the automatic evaluation results, where FastKASSIM-based rewards substantially improved syntactic similarity, and indicates that jointly optimizing syntactic similarity and semantic translation quality is well suited to patent claim translation. Notably, while FastKASSIM alone improved structural similarity, it degraded fluency in human evaluation (E=3.96); adding COMET recovered fluency (E=4.92), showing that COMET complements FastKASSIM by preserving translation quality.

Comparing reward combinations further reveals that FastKASSIM is most effective when paired with a semantic metric. Combining BLEU with COMET improved over single-metric rewards but did not surpass combining FastKASSIM with COMET, suggesting that explicit syntactic similarity is more complementary to COMET than surface-level n-gram overlap for this task. The combination of FastKASSIM with COMETKiwi also improved over either metric alone, although the gains were smaller than those of the other two-reward combinations—possibly because neither reward directly uses the reference translation as a semantic target. In contrast, adding FastKASSIM to BLEU-based rewards did not consistently improve human scores, suggesting potential reward conflicts between surface-level n-gram matching and syntactic-structure optimization. Overall, these results indicate that FastKASSIM contributes to claim-specific structural aspects such as Element Coverage, Scope & Logical Constraints, and Claim Form & Syntax, but should be combined with an appropriate semantic reward to maintain fidelity and fluency.

6 Conclusion

We propose a novel approach that combines reasoning models and GRPO with a syntactic reward to address two challenges in patent claim translation: understanding complex source structures and adhering to the structural conventions in the target language. Experimental results confirmed that the use of reasoning models improves semantic similarity, indicating that the reasoning process contributes to a better understanding of patent claims characterized by complex sentence structures. Furthermore, by explicitly incorporating a syntactic similarity metric into the GRPO reward function, we achieved translations that faithfully reproduce the structural conventions unique to the target language while maintaining semantic accuracy. This demonstrates that for tasks like patent translation, combining syntactic metrics such as FastKASSIM with semantic quality metrics such as COMET is effective for achieving high-quality translations with efficient model sizes.

Limitations

In this study, we verified the effectiveness of the proposed method on the Japanese-English language pair. Given that these two languages exhibit vastly different syntactic structures, and the tools utilized (e.g., Stanza and XCOMET) support numerous languages, the positive results suggest the potential applicability and robustness of our method across other languages and structure-constrained domains (e.g., legal and contract documents). However, extending verification to a broader range of languages remains a subject for future work.

Furthermore, the reliability of FastKASSIM depends on the accuracy of the underlying syntactic

parser. Syntactic parsers may produce inaccurate parse trees for the complex and lengthy structures characteristic of patent claims, potentially introducing noise into both the training reward and automatic evaluation. While GRPO optimization on large datasets appears to smooth out these individual errors, developing training methods robust to parsing failures is an important future direction.

Additionally, our approach assumes access to high-quality reference translations. While such data is readily available in the patent domain thanks to multinational filings (Nagata et al., 2024), adapting this method to settings with noisier or less standardized references presents an intriguing avenue for future research.

Finally, a key limitation lies in the lack of evaluation metrics specifically designed for formatting and structural constraints. Strict writing conventions are required not only in patents but also in broader legal documents; however, established automatic or human evaluation scales for assessing adherence to such specific styles do not currently exist. Moreover, patent claim translation lies at the intersection of translation expertise and patent-domain expertise. While translation experts may not necessarily be familiar with patent-specific claim structures, patent-domain experts may not necessarily be trained to assess translation quality. This makes it difficult to secure suitable evaluators for large-scale expert human evaluation. Therefore, developing standardized evaluation frameworks for these structural constraints is an essential direction for future research.

Ethical Considerations

Patent claims are legally binding documents where precise terminology and structure are critical. Although our proposed method improves translation quality and structural conformity, the generated translations may still contain errors or hallucinations. Consequently, the generated outputs require expert verification before being used officially due to the sensitive legal nature of patent claims. Regarding the human evaluation, the assessment was outsourced to a professional external company, ensuring evaluators were fairly compensated at standard commercial rates. Furthermore, the evaluation dataset consists of publicly available patent claims containing no personal, sensitive, or harmful information, posing no psychological risk to the evaluators.

References

Maximillian Chen, Caitlyn Chen, Xiao Yu, and Zhou Yu. 2023. FastKASSIM: A fast tree kernel-based syntactic similarity metric. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 211–231, Dubrovnik, Croatia. Association for Computational Linguistics.

DeepSeek-AI. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *Preprint*, arXiv:2501.12948.

Zhaopeng Feng, Shaosheng Cao, Jiahan Ren, Jiayuan Su, Ruizhe Chen, Yan Zhang, Jian Wu, and Zuozhu Liu. 2025. MT-r1-zero: Advancing LLM-based machine translation via r1-zero-like reinforcement learning. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 18685–18702, Suzhou, China. Association for Computational Linguistics.

Yvette Graham, Timothy Baldwin, Alistair Moffat, and Justin Zobel. 2013. Continuous measurement scales in human evaluation of machine translation. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 33–41, Sofia, Bulgaria. Association for Computational Linguistics.

Nuno M. Guerreiro, Ricardo Rei, Daan van Stigt, Luisa Coheur, Pierre Colombo, and André F. T. Martins. 2024. xCOMET: Transparent machine translation evaluation through fine-grained error detection. *Transactions of the Association for Computational Linguistics*, 12:979–995.

Takeshi Kojima, Shixiang (Shane) Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2022. Large language models are zero-shot reasoners. In *Advances in Neural Information Processing Systems*, volume 35, pages 22199–22213. Curran Associates, Inc.

Sinuo Liu, Chenyang Lyu, Minghao Wu, Longyue Wang, Weihua Luo, Kaifu Zhang, and Zifu Shang. 2025. New trends for modern machine translation with large reasoning models. *Preprint*, arXiv:2503.10351.

Masaaki Nagata, Makoto Morishita, Katsuki Chousa, and Norihito Yasuda. 2024. JaParaPat: A large-scale Japanese-English parallel patent application corpus. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 9452–9462, Torino, Italia. ELRA and ICCL.

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.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 101–108, Online. Association for Computational Linguistics.

Ricardo Rei, Craig Stewart, Ana C Farinha, and Alon Lavie. 2020. COMET: A neural framework for MT evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2685–2702, Online. Association for Computational Linguistics.

Ricardo Rei, Marcos Treviso, Nuno M. Guerreiro, Chrysoula Zerva, Ana C Farinha, Christine Maroti, José G. C. de Souza, Taisiya Glushkova, Duarte Alves, Luisa Coheur, Alon Lavie, and André F. T. Martins. 2022. CometKiwi: IST-unbabel 2022 submission for the quality estimation shared task. In *Proceedings of the Seventh Conference on Machine Translation (WMT)*, pages 634–645, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Jiaan Wang, Fandong Meng, Yunlong Liang, and Jie Zhou. 2025a. DRT: Deep reasoning translation via long chain-of-thought. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 6770–6782, Vienna, Austria. Association for Computational Linguistics.

Jiaan Wang, Fandong Meng, and Jie Zhou. 2025b. Deeptrans: Deep reasoning translation via reinforcement learning. *Preprint*, arXiv:2504.10187.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, brian ichter, Fei Xia, Ed Chi, Quoc V Le, and Denny Zhou. 2022. Chain-of-thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems*, volume 35, pages 24824–24837. Curran Associates, Inc.

Ran Zhang, Wei Zhao, Lieve Macken, and Steffen Eger. 2025. LiTransProQA: An LLM-based literary translation evaluation metric with professional question answering. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 29087–29109, Suzhou, China. Association for Computational Linguistics.

Mao Zheng, Zheng Li, Bingxin Qu, Mingyang Song, Yang Du, Mingrui Sun, and Di Wang. 2025. Hunyuan-mt technical report. *Preprint*, arXiv:2509.05209.

A Detailed Settings

We used JaParaPat (Nagata et al., 2024) for the evaluation. For each direction, we used 6,565 sentence pairs published in 2016 for training and 1,000 pairs published in 2020 for evaluation. The JA-EN data were extracted from patent pairs filed in Japan and their U.S. counterparts claiming priority, while the EN-JA data were extracted from the reverse patent pairs.

We implemented our training framework using PyTorch, utilizing the `verl` library for the Reinforcement Learning (RL) pipeline and `vLLM` for efficient rollout generation. All experiments were conducted on a single node equipped with four NVIDIA RTX 6000 Ada Generation GPUs. The training of the Qwen3-8B model took approximately 35 hours for 1 epoch. To optimize memory usage, we utilized Fully Sharded Data Parallel (FSDP) and set the Tensor Parallelism (TP) size to 2. The specific hyperparameters are summarized in Table 3.

Hyperparameter	Value
Training	
Base Model	Qwen3-8B
Learning Rate	5×10^{-7}
Global Batch Size	8
Max Sequence Length	2048
Total Epochs	1
GRPO Rollout Size	8
Inference	
RL Framework	verl
Inference Engine	vLLM
Tensor Parallel Size	2

Table 3: Hyperparameters and settings used for training.

B Prompt Template Used for Training

In this appendix, we detail the exact prompt template utilized during the training phase of our model (Table 4). To elicit a structured reasoning process, the model is instructed to generate its internal thought process within `<think>` tags prior to outputting the final translation within `<translate>` tags. The prompt does not contain any rules specific to patent translation.

The placeholders enclosed in curly brackets denote variables dynamically replaced by the corresponding instances from the training dataset: `{src_language}` and `{tgt_language}` represent the source and target languages (e.g., English and Japanese), respectively, and `{src_text}` is the actual input sentence to be translated.

C Human Evaluation Protocol

We conducted human evaluation on 25 samples randomly extracted from the JA-EN evaluation set.

User Prompt
A conversation between User and Assistant. The User asks for a translation from {src_language} to {tgt_language}, and the Assistant solves it. The Assistant first thinks about the reasoning process in the mind and then provides the user with the final translation. The reasoning process and final translation are enclosed within <think> </think> and <translate> </translate> tags, respectively, i.e., <think> reasoning process here </think><translate> final translation here </translate>. User:{src_text} Assistant:

Table 4: The prompt template used for training the model on the translation task. The template incorporates instructions for Chain-of-Thought reasoning using `<think>` and `<translate>` tags.

The evaluation was performed by bilingual language experts through a professional external annotation company. Each sample was assessed by an expert in syntactic analysis and translation, then double-checked by another professional. To ensure that the evaluation captured patent-specific requirements, we used patent-translation-aware criteria covering both general translation quality and claim-specific structural constraints.

We used two complementary protocols. First, following Direct Assessment with a continuous rating scale (Graham et al., 2013), evaluators assigned a direct assessment score on a 0–100 scale. Second, inspired by the question-based evaluation framework of LiTransProQA (Zhang et al., 2025), we conducted a checklist-style evaluation for five patent-specific criteria. For each criterion, evaluators assigned a score on a five-point scale from 1 to 5, where higher scores indicate better quality. The item-wise criteria used in the checklist-style evaluation are shown in Table 5.

D Results on English-to-Japanese Translation

As shown in Table 6, we also evaluated our method on the English-to-Japanese (EN-JA) direction. The results show a trend similar to JA-EN in terms of syntactic similarity: using FastKASSIM as a reward substantially improved the FastKASSIM score, and models with FastKASSIM-based combined rewards generally outperformed the baselines on this metric. This suggests that syntactic rewards can encourage target-side structural conformity in both translation directions.

However, unlike in JA-EN, FastKASSIM did not clearly contribute to translation quality in EN-JA. Among the 8B models, the highest XCOMET-XXL score was achieved by the model trained with COMET alone, rather than by FastKASSIM-based combined rewards. This difference may stem from the lower accuracy of Japanese syntactic parsing compared with English parsing. Because FastKASSIM relies on syntactic parses, parsing errors in Japanese patent claims may introduce noisy reward signals during training, making it harder to improve semantic quality while optimizing structural conformity.

E Min-max normalization

Metrics such as BLEU and FastKASSIM are highly sensitive and tend to exhibit large variances compared to semantic metrics like COMET and COMETKiwi. Consequently, when the reward is defined as a simple sum of these metrics, the training process tends to be disproportionately influenced by BLEU and FastKASSIM due to their larger score fluctuations. To address this imbalance, we experimented with applying min-max normalization to each metric to ensure they contribute equally to the total reward. As shown in Table 6, applying normalization resulted in a decrease in FastKASSIM and BLEU scores compared to the non-normalized counterparts, while COMET and COMETKiwi scores improved. Therefore, since FastKASSIM and XCOMET are both important for this task, we used the simple sum as the reward function.

F Qualitative Analysis

F.1 Translation Outputs

We conducted a qualitative analysis (Table 7) to verify whether the generated translations successfully capture target-side structural conventions. Specifically, we examine two conventions: (1) the global noun phrase structure (e.g., “The fastening method...”); and (2) the syntactic structure for the nominalizer “もの” (thing/one) in the source phrase “... ものから... ものへ”, which should be reflected as the reference phrase “from the one whose ... to the one whose ...” to match the reference conventions.

Human Evaluation Criterion and Description
Direct Assessment An independent direct quality assessment assigned by the evaluator on a 0–100 scale.
A. Content Fidelity (1) Is the core inventive concept (i.e., what the invention is about) accurately preserved? (2) Are the technical effects and functions explicitly stated in the source text free from distortion? (3) Is the translation completely free from any added technical content not present in the source text (hallucination)?
B. Element Coverage (1) Are all essential constituent elements (apparatus components, steps, or ingredients) in the source text fully translated? (2) Are the relationships between elements (e.g., "comprising," "connected to," "disposed on") accurately preserved? (3) Are quantities, ranges, and conditions (e.g., "one or more," "at least," "consisting of") faithfully maintained?
C. Scope & Logical Constraints (1) Is the claiming scope (e.g., "comprising," "consisting of," "characterized in that") left unaltered? (2) Are the modifier antecedents (the targets modified by "wherein" or "wherein the...") correctly maintained? (3) Does the interpretation of parallel vs. modifying relationships match the source text?
D. Claim Form & Syntax (1) Is the hypothesis formatted as a single-sentence claim? (2) Is the statutory category of the claim (e.g., apparatus, method, composition) accurately preserved? (3) Is the consistency of reference terms and definite/indefinite articles (e.g., "a," "the," "said," "the said") strictly maintained? (4) Does the hypothesis preserve the noun-phrase-like structure of patent claims, rather than becoming a narrative or explanatory sentence?
E. Fluency & Grammar (1) Is the translation grammatically correct in the target language? (2) Does the translation utilize natural and standard phrasing appropriate for patent documents? (3) Is the text completely free from ambiguity, redundancy, or phrasing that compromises readability?

Table 5: Human evaluation criteria. Direct Assessment denotes a direct quality assessment on a 0–100 scale, following prior work on Direct Assessment with continuous rating scales (Graham et al., 2013). For the checklist-style evaluation inspired by LiTransProQA (Zhang et al., 2025), each sub-question under A–E was rated on a five-point scale from 1 to 5, and the reported score for each criterion is the average of its sub-question scores.

Norm.	Model F	Model B	Model C	Model K	JA-EN Fast KASSIM	JA-EN BLEU	JA-EN COMET	JA-EN COMET Kiwi	JA-EN XCOMET XXL	EN-JA Fast KASSIM	EN-JA BLEU	EN-JA COMET	EN-JA COMET Kiwi	EN-JA XCOMET XXL
8B w/o Thinking					57.87	30.84	78.33	77.80	81.16	47.29	19.16	81.41	82.07	67.83
8B w/ Thinking					56.82	27.63	77.90	78.46	82.66	47.91	19.18	81.14	82.34	68.56
	○				**65.13**	32.42	78.93	76.51	80.92	**57.06**	25.62	81.06	77.86	57.19
		○			59.61	**42.80**	81.16	76.87	81.10	52.16	**38.07**	83.80	80.20	69.95
	○	○			64.38	41.65	81.15	76.83	80.40	54.10	36.09	83.60	81.11	68.96
○	○	○			62.93	41.55	81.32	77.15	80.99	54.39	35.52	84.10	82.05	69.59
			○		60.10	36.85	80.78	78.90	83.78	51.60	24.81	**85.78**	83.10	73.17
	○		○		63.44	36.51	80.10	78.37	83.08	55.75	28.98	84.86	81.91	68.47
○	○		○		62.67	36.70	81.15	78.12	81.17	55.41	28.73	85.10	81.92	68.60
		○	○		60.74	42.44	81.45	77.01	81.06	50.17	36.96	84.61	82.83	71.85
○		○	○		60.09	41.03	80.66	78.21	83.50	49.39	36.43	84.84	82.94	72.53
	○	○	○		63.17	39.75	80.30	78.09	83.09	54.33	34.87	84.35	81.77	70.65
○	○	○	○		61.61	40.57	**81.48**	77.56	82.02	54.02	34.70	84.57	82.53	70.82
				○	53.82	22.63	78.81	**80.60**	83.92	41.62	16.92	82.31	**84.27**	72.08
	○			○	63.03	30.19	80.17	79.73	84.43	55.96	26.77	83.40	83.08	71.22
○	○			○	60.62	25.91	79.69	80.16	84.41	52.25	20.37	82.69	83.84	71.34
		○		○	60.08	40.70	80.57	78.41	82.93	49.20	36.12	84.55	83.43	72.16
○		○		○	59.41	37.43	80.55	79.64	83.73	48.52	35.10	84.51	83.80	73.35
	○	○		○	64.37	40.81	81.12	77.73	80.54	54.07	35.62	84.24	82.47	70.74
○	○	○		○	61.34	37.47	80.40	79.25	84.06	52.06	33.87	84.55	83.58	72.73
235B-Instruct					56.57	34.70	79.88	78.81	84.43	48.90	27.84	83.38	83.39	73.39
235B-Thinking					57.32	32.97	79.57	79.03	**85.05**	48.36	25.62	82.67	82.51	71.57
GPT-5					57.01	36.84	80.48	78.92	84.63	50.20	26.84	83.91	83.58	**74.17**

Table 6: Comprehensive automatic evaluation results for JA-EN and EN-JA patent claim translation with and without reward normalization. "Norm." indicates the application of min-max normalization to rewards. In the "Model" column, F, B, C, and K denote FastKASSIM, BLEU, COMET, and COMETKiwi, respectively.

Regarding the baseline models, they fail to fully capture the required structures despite preserving semantic content. The output of the Qwen3-8B w/ Thinking model fails to maintain the global noun phrase structure, generating a complete sentence instead. While the larger Qwen3-235B-Thinking model successfully forms a global noun phrase, it fails to preserve the specific internal structure for "もの" (outputting "from the earliest to the latest"). These failures suggest that the general fluency and natural generation capabilities optimized for standard domains become counterproductive in the patent domain, where strict adherence to distinctive formatting is required.

As evident in Table 7, the output of the COMETKiwi reward model, despite having a high XCOMET-XXL score and preserving semantic content, does not take the form of a noun phrase, failing to reflect the necessary sentence structure. In contrast, the models using FastKASSIM & COMETKiwi as rewards successfully form global noun phrases and correctly reflect the "もの" structure. These translations achieve high scores in both XCOMET-XXL and FastKASSIM, confirming that the syntactic reward explicitly enforces adherence to target-side structural conventions without sacrificing translation quality.

These observations confirm that incorporating FastKASSIM into the reward function significantly contributes to generating translations that adhere to the target-side structural conventions of patent claims.

F.2 Reasoning Traces

To further examine how the syntactic reward guides the model to disentangle complex syntax, we conducted a qualitative analysis of the generated reasoning traces. We observed distinctive behaviors driven by the different reward formulations:

- **Base Thinking Model (Qwen3-8B w/ Thinking):** The model recognizes the patent domain (e.g., noting "accuracy is crucial" for patent documents) but struggles to translate this awareness into the strict structural constraints of patent claims, and we observed instances where it failed to construct the output as a single overarching noun phrase.
- **Semantic Reward Only (COMETKiwi):** Even when training is performed, if the syntactic structure of the target language is not included in the reward metric (as with COMETKiwi), it is not prioritized during the reasoning process. In fact, we observed that the COMETKiwi-trained model tends to focus on fluency, often producing traces that include phrases like "putting it all together smoothly." This emphasis on fluency can make it difficult for the model to reflect the required single noun phrase structure, which in some cases results in translations consisting of multiple sentences, failing to adhere to the claim structure.
- **Syntactic + Semantic Reward (FastKASSIM & COMETKiwi):** In contrast, for the model trained with both rewards, our qualitative analysis suggests that this combination more explicitly guides the model to construct the correct syntax. In many cases, we observed that the traces demonstrate a focus on identifying the core structure, with the model often anchoring on phrases such as "The main part is..." (e.g., to identify "A method of joining..."). This process appears to allow the model to correctly establish the primary noun phrase and appropriately integrate subsequent clausal modifiers, effectively preserving the hierarchical structure typical of complex patent claims.

Source	**前記カートリッジは円柱状であり、前記カートリッジにおいて、前記複数の種類の締結部材は、同一円周上において前記締結する順番が早いものから遅いものへと順に配置される、請求項１に記載の締結方法。**
Reference	**The fastening method according to claim 1, wherein the cartridge has a columnar shape and the plurality of types of fastening members are arranged on the same circumference in an order from the one whose order of fastening is early to the one whose order of fastening is late in the cartridge.**
8B w/ Thinking	The aforementioned cartridge is cylindrical, and in the cartridge, the multiple types of fastening members are arranged in order from those with an earlier fastening sequence to those with a later one on the same circumference, as claimed in claim 1. (FastKASSIM = 0.793, XCOMET = 0.946)
235B-Thinking	The fastening method according to claim 1, wherein the cartridge is cylindrical and the plurality of types of fastening members are arranged on the same circumference within the cartridge in the order of their fastening sequence from the earliest to the latest. (FastKASSIM = 0.789, XCOMET = 0.952)
COMETKiwi	The aforementioned cartridge is cylindrical. In this cartridge, multiple types of fastening components are arranged sequentially around the same circumference, from those fastened earlier to those fastened later. This is a fastening method described in claim 1. (FastKASSIM = 0.246, XCOMET = 0.960)
FastKASSIM & COMETKiwi	A method of joining as described in claim 1, wherein the cartridge is cylindrical and the plurality of types of joining members are arranged in sequence from those that are joined earlier to those that are joined later on the same circumference within the cartridge. (FastKASSIM = 0.891, XCOMET = 0.955)

Table 7: Qualitative comparison of Japanese-English patent claim translation generated by baseline models and our proposed models. Green text indicates preservation of structure or content, while red indicates failure to preserve structure.