構文類似度報酬を用いた GRPO を適用した Reasoning モデルによる特許請求項の日英翻訳

辻本祥吾 [1]　帖佐克己 [2]　永田昌明 [2]　笹野遼平 [1]
[1] 名古屋大学 大学院情報学研究科　[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

概要

特許請求項は，独特な書式と複雑な構文構造を持つ法的文書である．そのため，その翻訳には，原文の複雑な構文構造を正確に理解するだけでなく，翻訳先言語における特許固有の構文構造に従うことが求められる．本研究では，原文の構文構造をより正確に捉えるために Reasoning モデルを活用し，翻訳先言語における特許固有の構文構造を翻訳に反映させるために，構文構造の類似度を測る FastKASSIM を従来の BLEU や COMET といった評価尺度とともに強化学習の報酬として組み込む．実験の結果，Qwen3-8B をベースに学習したモデルが，約 30 倍の規模の Qwen3-235B ベースのモデルに匹敵する，あるいは上回る性能を達成した．また，FastKASSIM を報酬に用いることが，翻訳先言語での構文構造を反映した翻訳に寄与することを定性的に示した．

1　はじめに

特許請求項は発明の技術的範囲を法的に規定するものであり，複雑な修飾関係を持つ名詞句といった独特な書式で記述される．そのため，その翻訳においては，原文である特許請求項の複雑な構文構造を正確に理解することと，翻訳先言語における特許固有の構文構造を翻訳に忠実に反映させることが重要な課題となっている．

近年，大規模言語モデル（LLM）は，推論過程（Reasoning）を明示的に生成させることで，複雑なタスクでの性能が向上することが示されている．こうした推論能力の活用は，特許請求項の翻訳において，原文のより正確な理解に寄与すると期待されるが，その有効性は十分には検証されていない．

また，LLM を特定のタスクに適応させる手法として Group Relative Policy Optimization（GRPO） [1] が注目されている．先行研究 [2, 3] では，BLEU [4] や COMET [5] といった一般的な翻訳の評価尺度を報酬として用いた GRPO を機械翻訳タスクに適用しており，翻訳精度の顕著な向上が報告されている．しかし，これらの尺度のみでは，特許請求項の翻訳において，翻訳先言語での構文構造を十分に反映させることはできない．例えば，特許請求項は一般に全体で一つの名詞句を構成し，英語では発明の構成要素が文脈に応じて「;」（セミコロン）や「,」（カンマ）で区切られるのに対し，日本語では「、」（読点）で区切られるといった違いがある．これらの尺度は，主に表層的な一致や翻訳品質に焦点を当てており，必ずしも特許固有の構文構造が翻訳に反映されるとは限らない．

本研究では，第一に，原文の複雑な構文構造を理解する能力を向上させるため，特許請求項翻訳に Reasoning モデルを適用する．第二に，翻訳先言語での構文構造を翻訳に反映させるため，構文構造の類似度を測る尺度である FastKASSIM [6] を GRPO の報酬に組み込むことを提案する．実験の結果，日英間の特許請求項翻訳において，Qwen3-8B をベースに学習したモデルが約 30 倍のパラメータ数を持つ Qwen3-235B ベースのモデルに匹敵する，あるいは上回る性能を達成することを確認した．

2　関連研究

LLM は最終出力の前に推論過程を明示的に生成することで，複雑なタスクにおける性能が向上することが報告されている [7, 8]．また，Reasoning モデルは直喩や隠喩といった直訳が難しい表現の翻訳や，特に文学作品において，長い文脈や文化的背景の理解を要する翻訳に有効であることが示唆されている [9, 10]．一方で，一般的な分野での機械翻訳タスクにおいては，こうした推論能力の有効性は限定

的であるとも報告されている [2, 11]．これは，一般的な分野のテキストは多くの場合単純な構文構造をしており，深い推論を必要としないためだと考えられる．それに対し，特許請求項は複雑に入り組んだ構文構造を特徴としている．その複雑さを踏まえると，Reasoning モデルは，標準的な翻訳モデルでは扱いが難しい構文構造を正確に捉えることができ，特許請求項の翻訳において有効であると考えられる．

GRPO [1] は，タスクに合わせた報酬を用いて LLM を特定のタスクに効率的に最適化する強化学習手法である．MT-R1-Zero [2] は，教師ありファインチューニングやコールドスタートを必要とせず，自動評価尺度を報酬に組み込んだ GRPO を機械翻訳に適用することで，翻訳品質が向上することを報告している．この手法では，推論過程を<think>タグ，最終的な翻訳を<translate>タグで囲んで出力するようにプロンプトで LLM に指示を与える．学習時の報酬 r は，タグ構造が正しい場合は 1，誤りを含む場合は −1 となる S_{format} と，翻訳品質を測る S_{metric} の二つの要素から式 (1) のように定義される．S_{metric} には，表層的な一致を評価する BLEU [4]，原文と参照訳に基づいて翻訳品質を評価する COMET [5]，そして参照訳を用いない品質推定を行う COMETKiwi [12] といった尺度の組み合わせが用いられている．

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

一方で，これらの報酬は一般的な翻訳品質を向上させるが，構文構造の逸脱に対して明示的なペナルティを与えるものではない．そのため，翻訳先言語における特許固有の構文構造に従うことが求められる特許請求項翻訳のような分野においては，こうした標準的な尺度のみでは不十分である．

FastKASSIM [6] は tree kernel を用いて二つの構文木間の部分木の重複を定量化することで構文類似度を測定する尺度である．計算コストが高い木編集距離に対し，FastKASSIM は効率的に語彙の重複に依存しない構造を捉えることができるうえ，構文的な類似・非類似の識別タスクにおける人手評価との相関が高いことが報告されている．特許請求項翻訳において，原文の構文構造を正確に捉えることと翻訳先言語での構文構造を翻訳に反映させることを両立させるために，FastKASSIM のような構文構造を捉える尺度を報酬に組み込むことは不可欠である．

3 提案手法

特許請求項の翻訳では，原文の複雑な構造を理解し，翻訳先言語における特許固有の構文構造に従う翻訳を行うことが求められる．そこで本研究では，これらの要件を満たすため，Reasoning モデルに対し，構文構造の類似度を報酬に用いた GRPO を適用する手法を提案する．第一に，特許請求項の複雑な構文構造を正確に捉えるために Reasoning モデルを採用する．モデルは，推論能力を活用することで構文構造の深い解析を行うことができ，その推論に基づいて翻訳を生成することが可能になる．第二に，翻訳先言語での構文構造を反映させるため，MT-R1-Zero の枠組みに基づき，GRPO に構文類似度を報酬として組み込む．具体的には，従来の評価尺度に加え，構文構造の類似度を測る尺度である FastKASSIM を S_{metric} に導入する．なお，FastKASSIM は非対称な尺度であるため，参照訳と生成文の間の類似度を双方向で計算し，その平均値を用いる．この構文類似度報酬によって参照訳の構文構造に従った翻訳を行うようにモデルを促す．

4 実験設定

4.1 タスクとデータセット

本研究では，公開されている大規模日英特許対訳コーパスの JaParaPat [13] を利用して，日英および英日方向の特許請求項翻訳における提案手法の評価を行った．学習データとしては 2016 年の特許から日英方向・英日方向の翻訳データにそれぞれ 6565 文対，評価データとしては 2020 年の特許から同様にそれぞれ 1000 文対，コーパスから特許請求項に該当する部分を抽出して使用した[1)]．また，日英方向のデータは日本に出願した特許とその特許を優先権主張の対象とする米国の特許との対から，英日方向のデータはその逆の特許対から抽出した．

4.2 使用モデル

GRPO の学習対象モデルとして Qwen3-8B [14] を thinking mode で使用し，日英および英日の両方向で単一のモデルの学習を行った[2)]．また，ベースラインとして，Qwen3-8B の thinking mode と non-thinking mode，より大規模なモデルとして，bitsandbytes

1) 各データセットの文長に関する統計情報については付録 A を参照．
2) 実装の詳細は付録 B を参照．

表 1　日英・英日翻訳の自動評価結果．Model 欄では，FastKASSIM・BLEU・COMET・COMETKiwi をそれぞれ F・B・C・K と略記している．Avg. 欄は 5 つの尺度の平均を示し，Fail 欄はタグ構造の誤りや同じ出力の繰り返しのために翻訳の抽出に失敗した用例数を示す．

Model				JA-EN							EN-JA						
F	B	C	K	Fast KASSIM	BLEU	COMET	COMET Kiwi	XCOMET XXL	Avg.	Fail	Fast KASSIM	BLEU	COMET	COMET Kiwi	XCOMET XXL	Avg.	Fail
8B w/o Thinking				57.87	30.84	78.33	77.80	81.16	65.20	0/1000	47.29	19.16	81.41	82.07	67.83	59.55	0/1000
8B w/ Thinking				56.82	27.63	77.90	78.46	82.66	64.69	0/1000	47.91	19.18	81.14	82.34	68.56	59.83	0/1000
○				**65.13**	32.42	78.93	76.51	80.92	66.78	0/1000	**57.06**	25.62	81.06	77.86	57.19	59.76	3/1000
	○			59.61	**42.80**	81.16	76.87	81.10	68.31	1/1000	52.16	**38.07**	83.80	80.20	69.95	64.84	2/1000
○	○			64.38	41.65	81.15	76.83	80.40	68.88	0/1000	54.10	36.09	83.60	81.11	68.96	64.77	4/1000
		○		60.10	36.85	80.78	78.90	83.78	68.08	0/1000	51.60	24.81	**85.78**	83.10	73.17	63.69	0/1000
○		○		63.44	36.51	80.10	78.37	83.08	68.30	2/1000	55.75	28.98	84.86	81.91	68.47	63.99	1/1000
	○	○		60.74	42.44	**81.45**	77.01	81.06	68.54	1/1000	50.17	36.96	84.61	82.83	71.85	65.28	1/1000
○	○	○		63.17	39.75	80.30	78.09	83.09	68.88	0/1000	54.33	34.87	84.35	81.77	70.65	65.19	1/1000
			○	53.82	22.63	78.81	**80.60**	83.92	63.96	0/1000	41.62	16.92	82.31	**84.27**	72.08	59.44	0/1000
○			○	63.03	30.19	80.17	79.73	84.43	67.51	0/1000	55.96	26.77	83.40	83.08	71.22	64.09	1/1000
	○		○	60.08	40.70	80.57	78.41	82.93	68.54	0/1000	49.20	36.12	84.55	83.43	72.16	65.09	0/1000
○	○		○	64.37	40.81	81.12	77.73	80.54	**68.91**	0/1000	54.07	35.62	84.24	82.47	70.74	**65.43**	3/1000
235B-Instruct				54.26	29.33	79.15	78.75	83.60	65.02	0/1000	49.06	27.74	82.51	82.93	71.50	62.75	0/1000
235B-Thinking				56.40	27.86	78.84	79.10	84.43	65.33	0/1000	49.64	24.87	82.16	83.10	69.45	61.84	2/1000
GPT-5				57.01	36.84	80.48	78.92	**84.63**	67.58	0/1000	50.20	26.84	83.91	83.58	**74.17**	63.74	0/1000

による 4bit 量子化を行った Qwen3-235B-A22B-Thinking-2507 や Qwen3-235B-A22B-Instruct-2507 と，GPT-5（`gpt-5-2025-08-07`）を使用した．

4.3　報酬と評価尺度

報酬設計は，MT-R1-Zero [2] と同様，式 (1) に基づくものとし，S_{metric} には FastKASSIM・BLEU・COMET・COMETKiwi の 4 つの尺度の組み合わせのうち，COMET と COMETKiwi を併用する場合を除いた 11 通りについて検証を行った．複数の報酬を組み合わせる場合はそれらのスコアの和を S_{metric} として用いた[3]．翻訳性能の評価には，これら 4 つの尺度に加え，報酬計算に使用していない中立的な尺度として XCOMET-XXL [15] を用いた．なお，FastKASSIM の計算に必要な構文解析には Stanza [16] を用いた．

5　結果と分析

5.1　自動評価結果

実験結果を表 1 に示す．まず，ベースモデルにおける Reasoning の影響について確認する．人手評価と高い相関を持つ翻訳品質を評価する尺度である XCOMET-XXL に注目すると，235B の英日翻訳を除く全ての場合で，Reasoning モデルが，Reasoning を用いないモデルを上回った．これは，推論によって原文の複雑な構文構造への理解が深まり，意味内容をより正確に把握できていることを示唆している．

次に，GRPO と構文類似度報酬の有効性について述べる．GRPO で学習したモデルは，報酬に対応した評価尺度において顕著なスコアの向上を示した．提案手法で注目した FastKASSIM のスコアにおいては，FastKASSIM を単独で報酬に用いた場合に，日英方向において+8.31pts，英日方向において+9.15pts の向上を示し，翻訳先言語における特許固有の構文構造がより反映された翻訳が促されることが確認された．さらに，FastKASSIM を他の尺度と組み合わせて報酬に用いた場合においても，MT-R1-Zero での知見と同様に，それぞれの報酬に対応する評価尺度でのスコアの向上が見られた．特に，FastKASSIM と COMETKiwi を組み合わせた構成では，両方の尺度で高いスコアを示しただけでなく，日英翻訳の XCOMET-XXL において Qwen3-235B の Thinking モデルと並ぶ 84.43 のスコアを達成した．

最後に，提案手法と約 30 倍のパラメータ数を持つ Qwen3-235B や，GPT-5 を比較する．本結果では，Qwen3-8B に提案手法を適用したモデルが，構文類似度および翻訳品質の観点で，Qwen3-235B を上回る，あるいは匹敵する性能を達成した[4]．また，ベースラインにおいて，Reasoning の導入やモデルサイズの大規模化のいずれも FastKASSIM のスコアを大幅には向上させていないことから，翻訳先言語での構文構造を反映した翻訳を行う能力は一般的

3）　各尺度の感度の違いを考慮するために min-max 正規化の適用も検討したが，FastKASSIM による構文構造における改善が低減したため，本研究では採用しなかった．

4）　付録 C にて文長ごとの性能比較を行った．

表 2　特許請求項の日英翻訳の定性比較

原文	**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。**
参照訳	**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)

な言語能力とは独立しており，FastKASSIM を用いた明示的な最適化が必要であることが示唆される．GPT-5 に対しても，GRPO の報酬に用いた尺度においては上回る性能を示しており，XCOMET-XXL においては日英翻訳で同等の性能を達成した一方で英日翻訳では及ばない結果となった．これは日本語の構文解析の不正確さに起因すると考えられ，解析エラーがノイズを含んだ報酬となることで，翻訳品質に関する最適化に悪影響を及ぼしたと考えられる．

5.2　翻訳先の文構造を反映した翻訳の例

提案手法の有効性を検証するため，特許請求項の日英翻訳について定性分析を行った．表 2 に示す例において，参照訳の，“The fastening method according to claim 1,” という名詞句の構造と，“from the one whose ... to the one whose ...” という原文の「締結する順番が早いものから遅いもの」にも対応する「もの」(the one）を用いた構造の 2 点が翻訳文に反映されているかに注目する．

表より，GRPO を用いていないモデルは，意味内容を保持した翻訳を生成できているものの，翻訳先言語における特許固有の構文構造を十分に捉えきれていないことがわかる．Qwen3-8B の出力は名詞句ではなく文の形式をとっており，また，Qwen3-235B-Thinking の出力は名詞句となっているものの，“from the earliest to the latest” という部分が名詞的要素の「もの」を用いた構造を明示的に表現できていない．同様に，COMETKiwi 報酬モデルの出力は高い XCOMET-XXL スコアを示し，意味内容を保持しているにもかかわらず，必要な構文構造を反映できていない．これは，COMETKiwi が翻訳品質を評価するための指標であり，構文構造からの逸脱に対して明示的なペナルティを与えないためだと考えられる．対照的に，FastKASSIM & COMETKiwi 報酬モデルの出力は，名詞句の形式をとり，かつ「もの」に対応する構造を正しく反映している．この翻訳は XCOMET-XXL と FastKASSIM の両方で高いスコアを示しており，意味内容と構文構造の両面において高品質な翻訳となっていることが分かる．この事例は FastKASSIM を報酬に用いることが翻訳先言語での構文構造の反映に寄与することを示唆する．

6　おわりに

本研究では，原文の複雑な構文構造の正確な理解と，翻訳先言語における特許固有の構文構造を反映した生成が求められる特許請求項の翻訳タスクに対し，Reasoning モデルに構文構造の類似度尺度である FastKASSIM を報酬に用いた GRPO を適用する手法を提案し，その有効性を検証した．実験の結果，Reasoning モデルの利用によって翻訳品質が向上する傾向がみられ，推論過程が複雑な構文構造のより正確な理解を促すことが確認された．さらに，構文類似度尺度を GRPO の報酬に組み込むことで，高い翻訳品質を維持しつつ，翻訳先言語での構文構造をより反映した翻訳が実現された．特に，提案手法で学習した 8B モデルは，約 30 倍のパラメータ数を持つ 235B モデルと比較して，同等以上の性能を達成した．以上の結果は，特許翻訳のようなドメイン制約の強いタスクにおいて FastKASSIM のような構文類似度尺度を報酬に導入することが，比較的小規模なモデルで高品質な翻訳を実現する上で有効なアプローチであることを示す．

参考文献

[1] DeepSeek-AI. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning, 2025.

[2] Zhaopeng Feng, Shaosheng Cao, Jiahan Ren, Jiayuan Su, Ruizhe Chen, Yan Zhang, Jian Wu, and Zuozhu Liu. MT-r1-zero: Advancing LLM-based machine translation via r1-zero-like reinforcement learning. In Christos Christodoulopoulos, Tanmoy Chakraborty, Carolyn Rose, and Violet Peng, editors, **Findings of the Association for Computational Linguistics: EMNLP 2025**, pp. 18685–18702, Suzhou, China, November 2025. Association for Computational Linguistics.

[3] Jiaan Wang, Fandong Meng, and Jie Zhou. Deeptrans: Deep reasoning translation via reinforcement learning, 2025.

[4] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In Pierre Isabelle, Eugene Charniak, and Dekang Lin, editors, **Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics**, pp. 311–318, Philadelphia, Pennsylvania, USA, July 2002. Association for Computational Linguistics.

[5] Ricardo Rei, Craig Stewart, Ana C Farinha, and Alon Lavie. COMET: A neural framework for MT evaluation. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu, editors, **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, pp. 2685–2702, Online, November 2020. Association for Computational Linguistics.

[6] Maximillian Chen, Caitlyn Chen, Xiao Yu, and Zhou Yu. FastKASSIM: A fast tree kernel-based syntactic similarity metric. In Andreas Vlachos and Isabelle Augenstein, editors, **Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics**, pp. 211–231, Dubrovnik, Croatia, May 2023. Association for Computational Linguistics.

[7] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, brian ichter, Fei Xia, Ed Chi, Quoc V Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, **Advances in Neural Information Processing Systems**, Vol. 35, pp. 24824–24837. Curran Associates, Inc., 2022.

[8] Takeshi Kojima, Shixiang (Shane) Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large language models are zero-shot reasoners. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, **Advances in Neural Information Processing Systems**, Vol. 35, pp. 22199–22213. Curran Associates, Inc., 2022.

[9] Sinuo Liu, Chenyang Lyu, Minghao Wu, Longyue Wang, Weihua Luo, Kaifu Zhang, and Zifu Shang. New trends for modern machine translation with large reasoning models, 2025.

[10] Jiaan Wang, Fandong Meng, Yunlong Liang, and Jie Zhou. DRT: Deep reasoning translation via long chain-of-thought. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, editors, **Findings of the Association for Computational Linguistics: ACL 2025**, pp. 6770–6782, Vienna, Austria, July 2025. Association for Computational Linguistics.

[11] Mao Zheng, Zheng Li, Bingxin Qu, Mingyang Song, Yang Du, Mingrui Sun, and Di Wang. Hunyuan-mt technical report, 2025.

[12] 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. CometKiwi: IST-unbabel 2022 submission for the quality estimation shared task. In Philipp Koehn, Loïc Barrault, Ondřej Bojar, Fethi Bougares, Rajen Chatterjee, Marta R. Costa-jussà, Christian Federmann, Mark Fishel, Alexander Fraser, Markus Freitag, Yvette Graham, Roman Grundkiewicz, Paco Guzman, Barry Haddow, Matthias Huck, Antonio Jimeno Yepes, Tom Kocmi, André Martins, Makoto Morishita, Christof Monz, Masaaki Nagata, Toshiaki Nakazawa, Matteo Negri, Aurélie Névéol, Mariana Neves, Martin Popel, Marco Turchi, and Marcos Zampieri, editors, **Proceedings of the Seventh Conference on Machine Translation (WMT)**, pp. 634–645, Abu Dhabi, United Arab Emirates (Hybrid), December 2022. Association for Computational Linguistics.

[13] Masaaki Nagata, Makoto Morishita, Katsuki Chousa, and Norihito Yasuda. JaParaPat: A large-scale Japanese-English parallel patent application corpus. In Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue, editors, **Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)**, pp. 9452–9462, Torino, Italia, May 2024. ELRA and ICCL.

[14] 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, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jing Zhou, Jingren Zhou, Junyang Lin, Kai Dang, Keqin Bao, Kexin Yang, Le Yu, Lianghao Deng, Mei Li, Mingfeng Xue, Mingze Li, Pei Zhang, Peng Wang, Qin Zhu, Rui Men, Ruize Gao, Shixuan Liu, Shuang Luo, Tianhao Li, Tianyi Tang, Wenbiao Yin, Xingzhang Ren, Xinyu Wang, Xinyu Zhang, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yinger Zhang, Yu Wan, Yuqiong Liu, Zekun Wang, Zeyu Cui, Zhenru Zhang, Zhipeng Zhou, and Zihan Qiu. Qwen3 technical report, 2025.

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

[16] Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. Stanza: A python natural language processing toolkit for many human languages. In Asli Celikyilmaz and Tsung-Hsien Wen, editors, **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations**, pp. 101–108, Online, July 2020. Association for Computational Linguistics.

A　学習データおよび評価データ

本実験で使用したデータセットの統計情報を表 3 に示す．表中の平均，標準偏差，最大値は日本語・英語ともに文字数に基づいて算出している．

表 3　学習・評価データの文字列長

データセット			平均	標準偏差	最大値
学習データ	日 → 英	原文	119.85	104.30	1724
		参照訳	317.52	245.05	1797
	英 → 日	原文	245.11	179.45	1407
		参照訳	117.56	102.26	1411
評価データ	日 → 英	原文	109.14	88.45	638
		参照訳	316.92	256.04	1967
	英 → 日	原文	275.16	208.71	1289
		参照訳	127.76	143.32	1729

B　実装詳細

本研究では，学習フレームワークの実装に PyTorch を用い，強化学習のパイプラインには verl ライブラリ，ロールアウトの生成には vLLM を採用した．全ての実験は NVIDIA RTX 6000 Ada Generation GPU を 4 基搭載した単一ノード上で実施した．Qwen3-8B モデルの学習には，1 エポックあたり約 35 時間を要した．また，メモリ使用効率を最適化するため，Fully Sharded Data Parallel（FSDP）を利用し，テンソル並列サイズを 2 に設定した．具体的なハイパーパラメータは表 4 に示す．

表 4　実験設定の詳細．

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

C　文字列長と翻訳性能の分析

日英翻訳における参照訳の長さと FastKASSIM スコアや XCOMET スコアの関係を分析する．FastKASSIM と XCOMET のどちらの尺度においても，参照訳が長いほどスコアが低下しており，求められる翻訳が長いほど難しいタスクとなっていることが示されている．FastKASSIM スコアにおいて，どの文字列長についても FastKASSIM を GRPO の報酬に含めて学習を行ったモデルが他のモデルと比較して高いスコアを示していることや，XCOMET スコアにおいて，FastKASSIM と COMETKiwi を強化学習の報酬に用いたモデルが，GPT-5 や Qwen3-235B と共に，他のモデルと比較して高いスコアを示していることから，提案手法は文章の長さによらず有効であることが確認された．

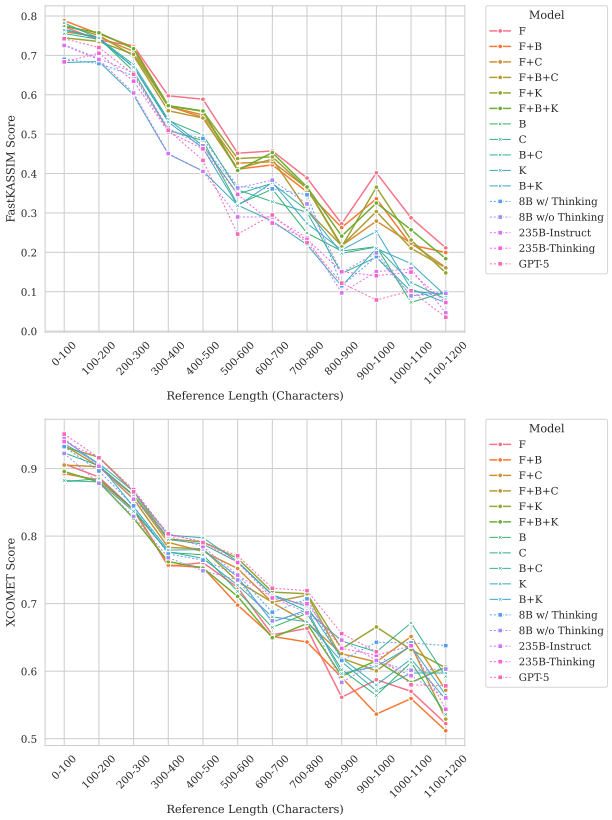

図 1　日英翻訳における参照訳の長さごとの，FastKASSIM スコアの平均（上）と XCOMET スコアの平均（下）．凡例では，FastKASSIM・BLEU・COMET・COMETKiwi をそれぞれ F・B・C・K と略記しており，GRPO による学習を行った各モデルを，学習において報酬として用いた尺度の組み合わせで表している．