Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Reinforcement Learning for the Computational Interpretation of Classical Medical Heritage Texts

Si Xie
Hunan University of Traditional Chinese Medicine
Wei Liu
weiliu@hnucm.edu.cn
Hunan University of Traditional Chinese Medicine
Jueling Luo
Changsha Aotu Pigai Intelligent Technology Co., Ltd.
Xiyue Song
Hunan University of Traditional Chinese Medicine
Mei Ouyang
Hunan University of Traditional Chinese Medicine
Wanjin Song
Hunan University of Traditional Chinese Medicine
Wei Hu
Hunan University of Traditional Chinese Medicine

Article

Keywords: Traditional Chinese Medicine classics, Intangible cultural heritage, Heritage text translation, Reinforcement learning, Epistemic fidelity, Digital humanities

Posted Date: March 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9067668/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Reinforcement Learning for the Computational Interpretation of Classical Medical Heritage Texts

Si Xie[a], Wei Liu[a,*], Jueling Luo[c], Xiyue Song[a], Mei Ouyang[a], Wanjin Song[b], Wei Hu[a]

[a]*School of Informatics, Hunan University of Chinese Medicine, Changsha, Hunan, 410208, People's Republic of China*
[b]*School of Traditional Chinese Medicine, Hunan University of Chinese Medicine, Changsha 410208, China*
[c]*Changsha Aotu Pigai Intelligent Technology Co., Ltd., Changsha, Hunan, China*

Abstract

Traditional Chinese Medicine (TCM) classics are a major form of intangible heritage, preserving historically layered medical knowledge, diagnostic logic, and therapeutic epistemologies. For heritage-text digitisation, interpretive fidelity and epistemological continuity are as critical as linguistic fluency. We present R1-TCM-Translator, a heritage-oriented framework for ancient-to-modern Chinese medical translation that combines multi-objective reinforcement learning (GRPO) with a structured six-step reasoning process to make cultural-epistemic reconstruction explicit and auditable. Experiments on a philologically curated parallel corpus of 15,387 sentence pairs from eight representative TCM classics show consistent improvements over supervised fine-tuning baselines and strong general-purpose large models. R1-TCM-Translator-8B demonstrates consistent gains on lexical-alignment and semantic-consistency metrics, specifically BLEU and COMET, indicating improved cross-text interpretive stability rather than metric-specific optimisation alone. Fine-grained analyses further show an approximately 23-percentage-point increase in specialised terminology accuracy and improved consistency in interpreting complex pathogenesis-related semantics. These findings suggest that reward-guided structured reasoning can improve epistemic fidelity in digitised medical heritage archives beyond surface-form

*Corresponding author
Email address: weiliu@hnucm.edu.cn (Wei Liu)

translation quality. By embedding semantic-fidelity objectives directly into optimisation, the framework operationalises heritage-aware translation as an auditable alignment process rather than a purely generative task. While doctrine-dense passages and composite interpretive risks remain challenging, the framework provides a reproducible computational pathway for digital preservation, knowledge modelling, and digital-humanities research on medical heritage texts.

Keywords: Traditional Chinese Medicine classics, Intangible cultural heritage, Heritage text translation, Reinforcement learning, Epistemic fidelity, Digital humanities

---

1. Introduction

Traditional Chinese Medicine (TCM) classics are foundational archives of medical civilisation and core carriers of intangible cultural heritage, preserving historically layered diagnostic logic, therapeutic epistemologies, and doctrinal reasoning across millennia [1, 2, 3]. Beyond their historical value, these texts function as knowledge-bearing artefacts whose philological integrity directly influences the fidelity of intergenerational knowledge transmission. In this sense, the interpretation of classical TCM is not only a linguistic exercise, but also a heritage-science challenge concerning epistemic fidelity and long-term preservation of medical knowledge. The global impact of knowledge derived from classical Chinese medicine, exemplified by artemisinin-related discoveries rooted in historical materia medica traditions [4, 5], also highlights the contemporary relevance of preserving and accurately interpreting such texts.

In the context of large-scale heritage digitisation, transforming classical medical texts into machine-readable and computationally interpretable forms remains a major challenge [6]. Recent heritage-science studies show that knowledge-graph infrastructure and digital curation workflows are central to managing heterogeneous cultural-heritage resources at scale [7, 8, 9, 10]. This direction is increasingly integrated with broader digital-heritage infrastructures, including 3D preservation workflows and GIS-enabled conservation analytics [11, 12]. Digitisation is not equivalent to image acquisition or optical character recognition (OCR) conversion: for heritage texts with compressed syntax and historically layered terminology, interpretive transformation directly determines whether core epistemic structures are preserved

or distorted. Classical medical texts are also embedded within long-standing interpretive traditions, where meaning is shaped by commentary lineages and evolving semantic conventions. For TCM classics, semantic drift during translation may alter causal reasoning in pathogenesis, diagnostic categorisation, and treatment principles, thereby weakening transmission fidelity in digital archives and downstream scholarly use.

Existing systems are typically optimised for general linguistic quality rather than heritage-specific epistemic fidelity. General-purpose large language models are effective across a broad range of linguistic tasks, yet they may fail to adequately model historically situated terminology, implicit epistemic relations, and context-dependent interpretive rules in classical medical corpora. As a result, such systems may produce linguistically fluent outputs that fail to maintain epistemic fidelity of the source tradition. Prior work in ancient Chinese natural language processing (NLP) and TCM informatics has provided useful resources and downstream applications, including corpus digitisation of classical TCM texts [13], clinical knowledge warehousing [14], and TCM knowledge-graph construction and applications [15, 16]. However, a dedicated optimisation paradigm for heritage-oriented epistemic fidelity is still lacking.

To address this gap, we introduce R1-TCM-Translator, a reinforcement learning framework tailored for heritage text interpretation. The framework combines a structured six-step reasoning process to make cultural-epistemic reconstruction explicit and auditable with heritage-grounded multi-objective reward alignment under GRPO, enabling optimisation guided by heritage-specific semantic and epistemic requirements. Rather than optimising only surface-form similarity, our design explicitly targets semantic adequacy and interpretability of reasoning traces as operational dimensions of epistemic fidelity during ancient-to-modern language transformation.

This study contributes to digital humanities and computational heritage science in four principal aspects; the following paragraphs detail dataset construction, framework design, empirical evidence, and broader methodological relevance. We established the TCM-Ancient-Modern corpus, a philologically curated parallel dataset containing 15,387 sentence pairs from eight representative TCM classics. We further proposed a heritage-aware reinforcement learning framework for classical medical text interpretation, designed to integrate structured reasoning with reward alignment. Empirically, the proposed framework improved BLEU and COMET over supervised fine-tuning baselines and strong general-purpose models, with consistent trends across

multiple test subsets. Fine-grained evaluation further showed improved consistency in pathogenesis-related semantics. To align with heritage-science objectives, we evaluated not only lexical and semantic metrics but also causal-semantic preservation and epistemic fidelity under expert-informed criteria. These results support a reusable computational pipeline for digital preservation and knowledge modelling of medical heritage texts. The framework is designed around domain-agnostic components, namely structured reasoning and reward alignment, that support adaptation to other classical medical traditions beyond Chinese corpora.

1.1. Positioning within Digital Humanities and Heritage Science

Unlike general-purpose translation systems, the present work is situated within heritage science, where knowledge-preservation and knowledge-organisation methods are increasingly examined through data-driven computational workflows [7, 10, 17]. From a digital humanities perspective, translating classical TCM texts is a form of computational interpretation that mediates access to historical knowledge systems [18, 19]. This positioning is consistent with recent text-centric heritage analytics, where NLP and corpus-quality assessment are used to evaluate conservation discourse and reliability of historical textual evidence [20, 21, 22]. In addition, recent work on epistemic translation and cross-boundary knowledge translation supports treating epistemic fidelity as a methodological requirement in heritage-text interpretation [23, 17]. In this sense, computational translation becomes an epistemic mediation process rather than a purely linguistic transformation. Concretely, this perspective is operationalised through translation decisions that are auditable at the levels of term disambiguation, causal semantics, and reasoning trace consistency. By embedding epistemic fidelity into optimisation objectives, this work demonstrates a practical pathway for methodologically integrating AI into heritage science.

2. Materials

2.1. Heritage corpus design and governance principles

To align dataset construction with heritage-science requirements, we treated TCM classics as knowledge-bearing heritage artefacts rather than generic linguistic material. Corpus curation followed three governance principles: philological reliability, epistemic fidelity, and auditability. In practice, we used authoritative editions as base texts, preserved clinically and

doctrinally meaningful propositions, and maintained traceable alignment and correction records.

All source texts and modern renderings were collected from authoritative published volumes, rather than web-crawled or crowd-authored sources. This choice prioritises textual reliability and reduces uncontrolled variation in heritage-sensitive semantic analysis, in line with evidence that document quality and textual noise materially affect downstream heritage-text analytics [20].

2.2. Source editions, representativeness, and doctrinal-bias control

The working corpus contains eight core TCM classics used in model development and evaluation: *Huangdi Neijing*, *Maijing*, *Jingui Yaolue*, *Nanjing*, *Shanghan Lun*, *Bianque Xinshu*, *Wenbing Tiaobian*, and *Sisheng Xinyuan*. Edition screening and authority validation were conducted by TCM faculty members, and selected volumes are listed in Appendix Appendix A, Table A.13.

Corpus selection aimed to preserve intra-traditional diversity while maintaining curricular authority, covering:

- **Historical layers**: early canonical foundations to later doctrinal developments;
- **Knowledge domains**: diagnostics, pathogenesis, therapeutics, and formula reasoning;
- **Textual forms**: theoretical exposition, aphoristic formulations, and clinically oriented prose.

Where multiple modern interpretive traditions exist, we prioritised pedagogically standard editions widely used in national TCM curricula and institutional teaching to reduce sectarian doctrinal bias.

Variant handling and editorial normalisation. All selected texts are authoritative editions. Variant readings across publication lines were documented during curation and harmonised to the designated base edition for each classic. Readings affecting core medical meaning were explicitly checked to maintain semantic stability across books. Original punctuation in the selected printed editions was retained. Marginal commentary and explanatory notes were excluded from source-side extraction unless they were part of the translated sentence unit.

2.3. Alignment workflow and quality control

The final TCM-Ancient-Modern dataset contains 15,387 ancient-modern sentence pairs. Sentence-level alignment was performed by TCM postgraduate students, followed by teacher sampling review. The workflow included extraction, preliminary alignment, sampled teacher audit, and final correction.

Specifically:

- **Student alignment**: TCM postgraduate annotators constructed ancient-modern pairs from authoritative printed editions;
- **Teacher audit**: TCM faculty sampled 100 aligned pairs and reviewed alignment boundaries and semantic correspondence;
- **Final correction**: ambiguous or unstable pairs were revised or removed.

In a sampled teacher audit of 100 aligned pairs, 1 case required correction, corresponding to a correction rate of 1.0%. The issue was mainly an alignment-boundary inconsistency. The affected pair was revised and logged, indicating stable alignment quality under the defined protocol and reinforcing the importance of explicit correction traceability in heritage-text workflows [20].

2.4. Dataset characteristics and heritage-specific complexity

Beyond corpus size, the dataset exhibits heritage-specific epistemic complexity, including semantic shift between classical and modern usage, doctrinally loaded terminology, zero-anaphora, and paratactic logic. These properties motivate structured reasoning and reward-aligned optimisation rather than purely surface-form translation. Figure 1 shows the sample distribution across the eight source classics.

Table 1 presents representative translation examples from different types of TCM classics in the dataset.

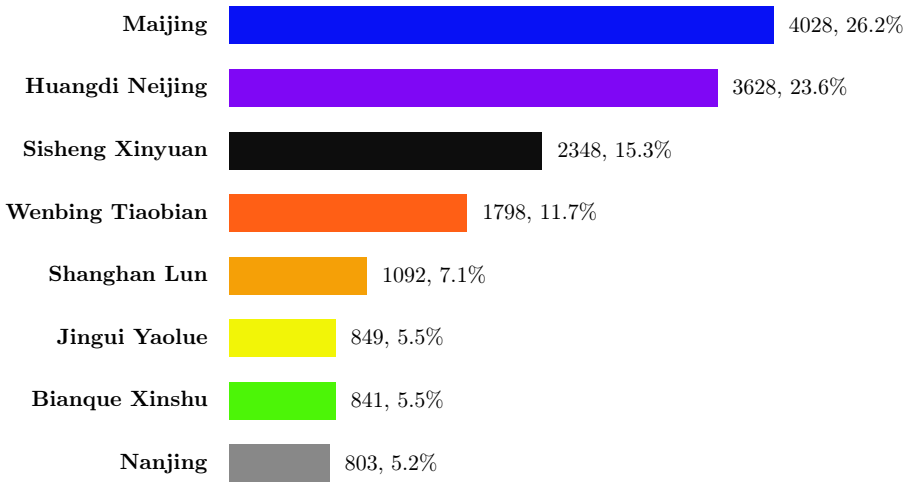


Figure 1. Distribution of samples across classics in the TCM-Ancient-Modern dataset. The dataset covers eight representative TCM classics and contains 15,387 parallel sentence pairs in total. *Maijing* contributes 4,028 pairs, accounting for 26.2% and forming the largest share, followed by *Huangdi Neijing* with 3,628 pairs and a share of 23.6%.

Table 1. Representative translation examples from the TCM-Ancient-Modern dataset

Classic	Original, Classical Chinese	Reference, Modern Chinese
Huangdi Neijing	秋脉如浮，何如而浮?	黄帝问：秋天的脉象如浮，那么怎样才算浮呢?
Shanghan Lun	脉浮数者，法当汗出而愈。	脉象浮数，为病在表，照理应当用发汗法治疗，汗出后就能痊愈。
Maijing	脉有阳盛阴虚，阴盛阳虚，何谓也?	脉象有阳盛阴虚、阴盛阳虚的情况，是怎么回事?
Wenbing Tiaobian	以花易豆者，恶其呆滞也。	这里用扁豆花代替扁豆，是因为扁豆性质呆滞。

2.5. General-domain open-source SFT corpus

We additionally use the public Hugging Face instruction-tuning source `SylvanL/Traditional-Chinese-Medicine-Dataset-SFT` to provide broad domain grounding, with paper-level accounting exceeding 1.2 million samples across structured knowledge text, terminology interpretation resources, and unstructured classical materials. For reproducibility, we release the project-aligned processed training files (`sft_medical_knowledge_source1/2/3.json`); therefore, scale statistics in this section follow source-dataset accounting, whereas released files correspond to processed splits used in our pipeline.

2.6. Licensing and reproducibility statement

The TCM-Ancient-Modern dataset will be open-sourced on GitHub. To support full reproducibility, we release the aligned sentence pairs, annotation procedures, processing scripts, and training, validation, and test split metadata. For anonymous review, the repository link is provided in supplementary materials and will be made public upon publication.

2.7. Scale rationale

While modest relative to web-scale corpora, the dataset size reflects the labour-intensive nature of expert-audited heritage alignment. In this study, corpus governance prioritises epistemic traceability and philological reliability over raw scale.

3. Methods

3.1. Heritage-oriented design principle

Method design is guided by **epistemic fidelity**, i.e., preserving diagnostically and historically grounded meaning in classical medical texts rather than maximising surface fluency alone. Concretely, optimisation targets three heritage-relevant dimensions: **semantic fidelity**, **causal-semantic preservation**, and **reasoning auditability**. This text-centric formulation is aligned with recent heritage-science applications of AI to manuscript analysis and condition-sensitive interpretation workflows [24]. Operationally, epistemic fidelity is assessed through semantic correctness, causal relation preservation, and trace-structure consistency.

Figure 2 provides an overview of the proposed method pipeline.

3.2. Task formulation

Let x denote a classical TCM source sentence, r a structured reasoning trace, and y the modern Chinese translation. Conventional machine translation typically models $p_\theta(y \mid x)$. In contrast, we model joint generation with explicit reasoning:

$$p_\theta(y, r \mid x) = p_\theta(r \mid x)\, p_\theta(y \mid x, r), \tag{1}$$

where r serves as an interpretable intermediate representation for epistemic mediation. This factorisation operationalises a "first interpret, then translate" strategy.

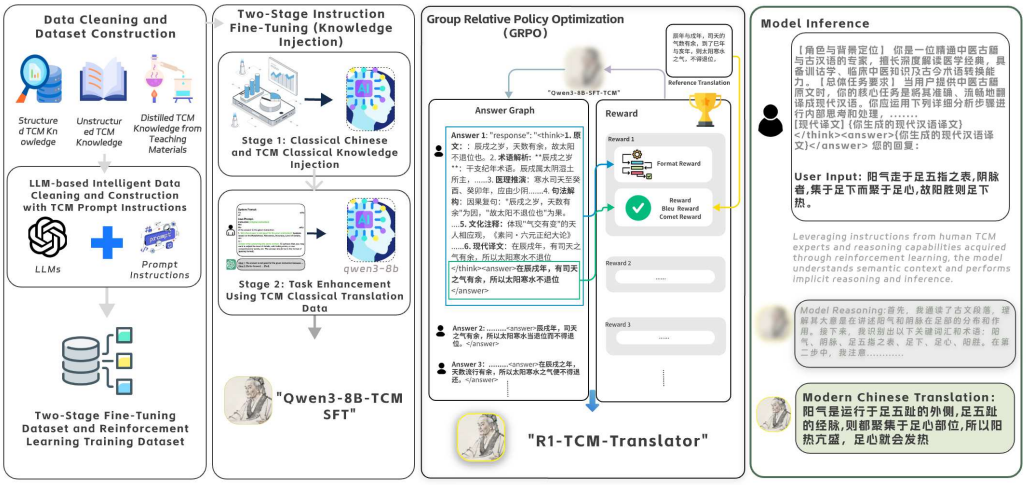


Figure 2. Overview of the R1-TCM-Translator method pipeline.

3.3. Two-stage training framework

We adopt a curriculum-style two-stage framework to separate domain grounding from reasoning optimisation.

Stage I: Domain grounding via SFT. Base models are first supervised fine-tuned on large-scale open TCM resources to acquire background knowledge of terminology, concepts, and basic doctrinal relations.

Stage II: Structured reasoning adaptation. Models are then finetuned on six-step reasoning-trace data distilled from a teacher model, with the objective of learning both interpretable traces and final translations under the joint target $p_\theta(y, r \mid x)$.

3.4. Six-step reasoning scaffold and audit criteria

The reasoning scaffold contains six steps: **source text**, **term interpretation**, **medical causal inference**, **syntactic reconstruction**, **cultural and epistemic note**, and **modern translation**. A worked six-step example and the full prompting template are provided in Appendix Appendix B and Appendix Appendix C. To improve auditability, each step is linked to a minimal criterion:

- **Term interpretation**: key medical terms are explicitly mapped to context-appropriate meanings;

- **Medical causal inference**: core causal relations in pathogenesis or treatment logic are stated;
- **Syntactic reconstruction**: major ellipsis and parataxis structures are recoverable in modern syntax;
- **Cultural and epistemic note**: doctrinal context is not contradicted by the generated interpretation;
- **Trace coherence**: step order and internal consistency are preserved across the chain.

3.5. Reward design

Let G denote the model output with reasoning trace r and translation y. The total reward is:

$$R_{\text{total}}(G) = R_{\text{format}}(G) + R_{\text{quality}}(G), \tag{2}$$

with equal unit weighting for the two reward modules.

3.5.1. Format reward R_{format}

This term checks whether the six-step structure is present and ordered, implemented via rule-based pattern matching:

$$R_{\text{format}}(G) = R_{\text{struct}}(G) + R_{\text{trace}}(G). \tag{3}$$

R_{struct} rewards valid tag/section structure; R_{trace} rewards complete and ordered reasoning steps. Coefficients are selected by pilot runs and rewards are normalised to avoid domination by any single component.

3.5.2. Quality reward $R_{quality}$

$$R_{\text{quality}}(G) = R_{\text{bleu}} + R_{\text{comet}}, \tag{4}$$

where R_{bleu} denotes the normalised BLEU term and R_{comet} measures semantic adequacy. BLEU is used as a lexical-overlap proxy during optimisation.

3.6. GRPO-based policy optimisation

We use GRPO to optimise multi-objective rewards without a separate learned critic. For each prompt q, a group of outputs $\{o_i\}_{i=1}^{G}$ is sampled from the old policy. The objective is:

$$\mathcal{J}_{\mathrm{GRPO}}(\theta) = \mathbb{E}\Bigg[\frac{1}{G}\sum_{i=1}^{G}\min\Big(\rho_i(\theta)A_i,\ \mathrm{clip}(\rho_i(\theta),1-\epsilon,1+\epsilon)A_i\Big) - \eta\, D_{\mathrm{KL}}(\pi_\theta\|\pi_{\mathrm{ref}})\Bigg], \tag{5}$$

with policy ratio

$$\rho_i(\theta) = \frac{\pi_\theta(o_i \mid q)}{\pi_{\theta_{\mathrm{old}}}(o_i \mid q)}. \tag{6}$$

Let $s_i = R_{\mathrm{total}}(o_i)$ be the sample reward. Group-relative standardised advantage is:

$$A_i = \frac{s_i - \mathrm{mean}(\{s_j\}_{j=1}^{G})}{\mathrm{std}(\{s_j\}_{j=1}^{G})}. \tag{7}$$

GRPO is suitable here because group-relative ranking reduces single-output reward noise and stabilises optimisation for tasks requiring both interpretable reasoning traces and semantically faithful outputs. In practice, the KL term is computed with a token-level approximation under sampled trajectories.

3.7. Evaluation protocol

To align with heritage-science objectives, we reported both automatic and heritage-oriented metrics. BLEU and COMET are treated as reproducible proxies for lexical and semantic adequacy, not as direct measurements of epistemological integrity [25, 26]:

- **Automatic metrics**: BLEU and COMET;
- **Heritage-oriented automatic analyses**: causal-semantic preservation and reasoning-trace structural consistency under predefined rule-based criteria.

Challenge-wise evaluation covered semantic shift, cultural and epistemic context preservation, zero-anaphora, and parataxis. This setup mitigates edition-specific bias and reveals robustness beyond aggregate scores.

3.8. Implementation and robustness controls

We reported full implementation details for reproducibility, including base model versions, tokenizer, maximum sequence length, learning rate, batch size, group size G, training epochs, sampling temperature, and random seed settings.

To reduce overfitting risk, we used fixed training, validation, and test splits, checkpoint selection on validation objectives, and subset-consistency reporting across books and challenge categories. Repeated runs with different seeds were used to verify trend consistency.

4. Experiments

4.1. Experimental Setup

To comprehensively evaluate R1-TCM-Translator on classical TCM translation, we conducted experiments on eight representative TCM classics. We used BLEU and COMET as primary metrics [25, 26], where BLEU captures lexical overlap and COMET serves as an automatic proxy for semantic fidelity.

We compared three model groups: **Base** models, namely Qwen3-0.6B and Qwen3-8B; **SFT** models; and **SFT+GRPO final** models in the R1-TCM-Translator series. In this paper, `Qwen3-0.6B-SFT+GRPO` and `R1-TCM-Translator-0.6B` refer to the same final-weight checkpoint, and similarly for the 8B version; we therefore report the final checkpoints under the unified name `R1-TCM-Translator`. We also report `Qwen3-0.6B-GRPO` as a single-stage reference model trained with direct GRPO from the base checkpoint, without an intermediate SFT stage.

All compared models at inference were decoded three times per query under fixed decoding settings, and reported scores were averaged across the three outputs. This protocol estimates decoding-time variability rather than training-time variance.

The same chain-of-thought prompting template was used for all compared models, with identical instruction structure and output requirements. To ensure fairness, we kept prompt content, prompt order, and decoding protocol fixed across models.

The test set contains **3,241** aligned pairs in total, distributed as follows: *Huangdi Neijing* with 804 pairs, *Jingui Yaolue* with 198 pairs, *Nanjing* with 159 pairs, *Bianque Xinshu* with 185 pairs, *Maijing* with 816 pairs, *Shanghan*

Lun with 215 pairs, *Sisheng Xinyuan* with 489 pairs, and *Wenbing Tiaobian* with 375 pairs.

4.1.1. Baselines

Closed-source general models. We queried OpenAI GPT-4o and GPT-4.1. GPT-4o is a widely used multimodal foundation model. GPT-4.1 supports long-context inference.

Open-source general model. We evaluated DeepSeek-V3.1, a Mixture-of-Experts model with 671B total parameters and 37B active parameters.

Open-source reasoning model. We evaluated DeepSeek-R1, a reinforcement-learning-trained model for reasoning-intensive tasks.

In-domain base models. We used Qwen3-0.6B and Qwen3-8B for domain adaptation and staged optimisation. This setup enabled controlled analysis of model-capacity effects and post-training gains.

Model-selection rationale. The baseline suite covers both closed-source and open-source models, both general-purpose and reasoning-oriented paradigms, and different parameter scales. This allows a balanced evaluation of frontier general models and the effectiveness of our in-domain optimisation pipeline under the same prompting protocol.

4.1.2. Implementation Details for SFT

SFT settings extracted from the training log are reported in Appendix Appendix D, Table D.15.

4.1.3. Implementation Details for GRPO

GRPO training was implemented with `unsloth` and run on NVIDIA **H200** hardware. Full hyperparameter settings are reported in Appendix Appendix E, Table E.16.

4.2. Main Results

Table 2 reports detailed performance across all eight classics. Overall, R1-TCM-Translator-8B substantially outperforms strong general-purpose models on BLEU with a score of 41.5 while remaining competitive on COMET-based semantic fidelity with a score of 79.6. On BLEU, it exceeds GPT-4o by 29.2 points, GPT-4.1 by 27.7 points, and DeepSeek-R1 by 20.7 points. On

COMET, it stays in a high-performance range comparable to DeepSeek-V3.1 at 78.9 and DeepSeek-R1 at 78.4.

R1-TCM-Translator-0.6B also shows strong parameter efficiency, with BLEU at 29.5, higher than DeepSeek-R1 and GPT-4.1, and COMET at 77.4, close to larger general-purpose models. As a no-SFT single-stage reference, Qwen3-0.6B-GRPO, trained with direct GRPO from the base model, reaches 24.8 BLEU and 77.3 COMET in macro average across the eight classics; compared with this no-SFT setting, the two-stage SFT+GRPO 0.6B model yields a further 19.0% improvement in BLEU and 0.1% improvement in COMET, indicating that direct GRPO without prior SFT is less effective than applying GRPO after SFT.

From a stage-wise perspective, SFT and GRPO are complementary. SFT primarily improves domain lexical grounding, whereas GRPO contributes larger additional gains in both lexical and semantic-fidelity proxy metrics. These results are consistent with improved interpretive stability in heritage-text translation, but do not by themselves directly establish improved heritage knowledge transmission quality.

Table 2. Translation performance on TCM classic texts using BLEU and COMET. Reported scores are means over three decoding runs per query under fixed decoding settings. The Avg. column reports macro averages across the eight classics. Best value in each column is boldfaced.

	Huangdi Neijing		Jingui Yaolue		Nanjing		Bianque Xinshu		Maijing		Shanghan Lun		Sisheng Xinyuan		Wenbing Tiaobian		Avg.	
Model	BLEU	COMET	BLEU	COMET	BLEU	COMET	BLEU	COMET	BLEU	COMET	BLEU	COMET	BLEU	COMET	BLEU	COMET	BLEU	COMET
Closed																		
GPT-4o	16.4	73.2	7.7	72.9	11.4	72.9	13.3	79.5	18.4	70.1	8.6	73.7	7.0	74.3	15.2	76.2	12.3	74.1
GPT-4.1	18.6	74.2	10.3	73.3	7.5	73.9	9.4	78.9	12.9	70.9	8.6	73.0	10.1	73.4	33.0	76.8	13.8	74.3
Open																		
General Purpose LLMs																		
DeepSeek-V3.1	22.2	78.0	13.0	**79.4**	5.2	78.2	36.2	84.4	14.2	75.2	19.5	**77.3**	9.1	77.9	23.6	80.4	17.9	78.9
DeepSeek-R1	37.1	77.2	10.3	78.4	8.3	78.2	37.5	84.3	14.4	75.2	15.5	76.1	9.1	78.0	33.8	80.2	20.8	78.4
Ours																		
Qwen3-0.6B	3.2	66.3	1.7	64.3	0.6	65.1	3.7	54.9	7.7	64.5	1.1	65.2	4.1	67.9	5.2	69.8	3.4	64.7
Qwen3-0.6B-SFT	15.3	67.9	9.8	68.6	6.6	67.7	28.4	76.0	18.1	67.9	13.3	66.9	14.3	69.7	21.8	70.8	16.0	69.4
Qwen3-0.6B-GRPO (direct GRPO, no SFT)	32.1	76.1	14.3	76.4	21.2	76.7	55.0	83.6	29.5	74.9	20.1	74.0	9.2	77.9	16.9	78.8	24.8	77.3
Qwen3-8B	13.1	75.0	9.2	75.2	7.6	73.0	28.6	72.6	19.4	72.9	11.1	72.9	12.7	75.9	17.4	78.2	14.9	74.5
Qwen3-8B-SFT	27.1	74.2	9.5	74.5	8.7	74.7	38.1	81.8	15.3	73.1	22.9	71.4	8.7	74.9	16.4	76.5	18.3	75.1
R1-TCM-Translator-0.6B (SFT+GRPO)	**59.9**	76.0	8.8	76.0	17.4	77.6	48.2	83.5	13.7	75.6	21.1	73.9	12.7	77.9	**54.5**	78.5	29.5	77.4
R1-TCM-Translator-8B (SFT+GRPO)	38.4	**78.8**	**14.2**	78.7	**18.8**	**79.5**	**58.9**	**85.6**	**67.4**	**77.8**	**53.0**	75.5	**36.6**	**80.1**	45.1	**80.5**	**41.5**	**79.6**

4.3. Ablation Study

To quantify stage-wise contributions, we conducted an ablation analysis across three training configurations, namely Base, SFT, and SFT+GRPO, as shown in Table 3.

Ablation results show complementary effects of the two training stages. SFT primarily improves lexical and domain grounding, while GRPO contributes larger additional gains in both BLEU and COMET-based semantic fidelity. We also observe limited or slightly negative COMET changes in

Table 3. Ablation study: impact of training stages on translation performance

Configuration	Qwen3-0.6B			Qwen3-8B		
	BLEU	COMET	**ΔBLEU / ΔCOMET**	BLEU	COMET	**ΔBLEU / ΔCOMET**
Base	3.4	64.7	–	14.9	74.5	–
+ SFT	16.0	69.4	+12.6 / +4.7	18.3	75.1	+3.4 / +0.6
+ SFT + GRPO	29.5	77.4	+13.5 / +8.0	41.5	79.6	+23.2 / +4.5

some SFT settings, indicating a possible objective mismatch between SFT data characteristics and semantic-fidelity evaluation.

5. Empirical Findings on Heritage-Specific Semantic Preservation

5.1. Cross-text epistemic consistency

Domain mismatch remains a core challenge in machine translation, and this challenge is amplified in heritage texts where terminology, rhetorical style, and epistemic assumptions vary across sources. In classical TCM translation, different canonical texts encode distinct terminological systems and interpretive traditions. We therefore frame this analysis as **cross-text epistemic consistency** rather than a generic MT domain-transfer problem.

While BLEU and COMET do not directly measure epistemological integrity, we used them as reproducible proxy indicators and complemented them with phenomenon-level qualitative audits. This design reduces reliance on surface-form metrics alone and links automatic evaluation to heritage-oriented interpretive risks.

SFT improves baseline quality but shows limited cross-text interpretive stability. We fine-tune Qwen3 models on parallel data spanning all eight classics and evaluate cross-text performance, as shown in Table 2. Although SFT improves overall scores over base models, limitations remain evident: Qwen3-0.6B-SFT reaches BLEU 16.0 and COMET 69.4, whereas Qwen3-8B-SFT reaches BLEU 18.3 and COMET 75.1, with substantial variation across books.

Qualitative inspection identifies three recurring issues, as shown in Table 4:

- **Terminology confusion**: key normative expressions are misread. A representative case is that "法当" is interpreted as descriptive status rather than therapeutic instruction.

- **Register instability**: book-specific rhetorical styles are not consistently preserved.
- **Inference-related misinterpretation**: conditional relations are sometimes rendered as process descriptions, as in "与水则哕".

Table 4. Examples of cross-text interpretive mismatch in classical TCM translation. Red: terminology mismatch; blue: mismatch in register and style; green: hallucination-like misinterpretation.

Source	Original	Reference	SFT Output
伤寒论	脉浮数者，法当汗出而愈。	脉浮数的，应当汗出后就能痊愈。	脉浮数而出现的，是病已解表。
伤寒论	欲饮水者，与水则哕。	想要饮水的人，如果给予喝水，就会引起呕吐症状。	欲饮水、饮水中，就会产生气逆的症状。
黄帝内经	东方生风，风生木。	东方产生风，风产生木。	东方的风会生长树木。
脉经	寸口脉沉而迟，沉则为寒，迟则为虚。	寸口脉沉而迟，沉脉表示寒证，迟脉表示虚证。	寸口脉沉迟，沉是因为寒冷，迟是因为虚弱。
温病条辨	太阴温病，脉浮而缓，身不甚热，微渴者。	太阴温病，脉象浮而缓，身体不太热，稍微口渴的。	太阴温病，脉浮缓，身体发热不明显，有点口渴。
金匮要略	师曰：夫病痰饮者，当以温药和之。	老师说：患痰饮病的人，应当用温性药物调和治疗。	医生说：痰饮病人，要用温热的药。
难经	脉有三部，寸关尺也。	脉有三个部位，分别是寸、关、尺。	脉分为三部分，即寸脉、关脉和尺脉三个位置。
扁鹊心书	保命之法，灼艾第一。	保全生命的方法，艾灸是第一位的。	保命的方法，用艾条灸是最好的。

These observations suggest that SFT captures useful lexical mappings but remains limited in preserving cross-text epistemic consistency.

GRPO yielded stronger interpretive stability. Compared with SFT, GRPO-trained models show more stable behavior on phenomena requiring structured

inference and doctrinally consistent interpretation as shown in Table 5. We deliberately describe this as an *association* with improved interpretive performance, rather than making a direct causal claim about "deep reasoning activation."

Table 5. Comparison of SFT vs. SFT+GRPO outputs. Green highlights improvements in SFT+GRPO outputs.

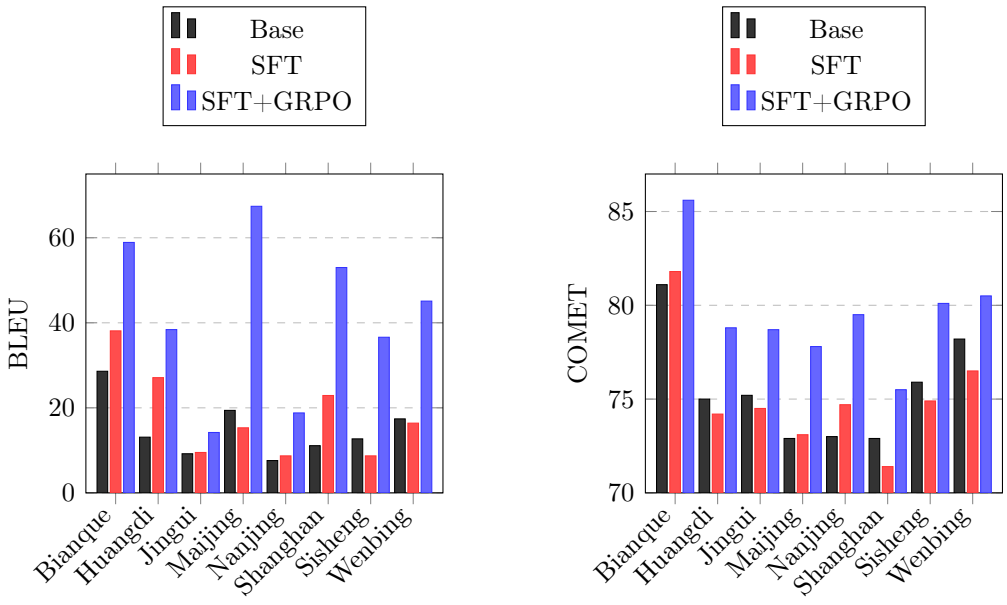
Source	**Original**	**SFT Output**	**SFT+GRPO Output**	**Interpretive Note**
伤寒论	脉浮数者，法当汗出而愈。	脉象浮数的，说明邪未入里，应当用汗法，调和营卫之气，待汗出后病势就解除。	脉浮数的，应当发汗，汗出病就会痊愈。	Better preservation of the normative meaning in "法当"
伤寒论	欲饮水者，与水则哕。	口渴饮水的，如果给予饮水，就会发生呕吐。	想喝水的，给予水就会呕吐。	Correctly preserves the conditional relation in "与水"
伤寒论	发热而厥，七日，下利者，为难治。	发高热而四肢厥冷，七天来又有腹泻的，是病势已深，较难治愈。	发热而四肢厥冷，七天后又出现腹泻的，这是难治之证。	More accurate temporal relation handling
黄帝内经	东方生风，风生木。	东方的风会生长树木。	东方产生风，风产生木。	Better doctrinal consistency under Five-Phase semantics

Quantitatively, the 0.6B GRPO model reaches BLEU 29.5 and COMET 77.4, improving over SFT by 84.4% BLEU and 11.5% COMET. The 8B GRPO model reaches BLEU 41.5 and COMET 79.6, improving over SFT by 126.8% BLEU and 6.0% COMET. These gains are consistent with stronger cross-text interpretive stability.

Model capacity and post-training interaction. After SFT+GRPO, the 8B model, with BLEU at 41.5 and COMET at 79.6, outperforms the 0.6B model, with BLEU at 29.5 and COMET at 77.4, indicating that larger capacity better accommodates cross-text epistemic variation. At the same time, 0.6B-SFT+GRPO surpasses 8B-SFT by 11.2 BLEU and 2.3 COMET, suggesting that heritage-aligned post-training can substantially compensate for scale limitations.

Overall, these findings support a heritage-oriented interpretation: structured reward alignment improves computational handling of cross-text epistemic risk, rather than merely increasing benchmark scores.

Figure 3 compares Base, SFT, and SFT+GRPO across all eight classics.



(a) BLEU **(b)** COMET

Figure 3. Performance of Qwen3-8B across eight TCM classics. Panel a shows BLEU as lexical alignment, and panel b shows COMET as a proxy for semantic fidelity.

5.2. Terminology preservation under low-frequency conditions

In heritage-text translation, low-frequency terms are not only a statistical sparsity issue but also a preservation risk, because rare terms often carry specialised historical meaning. We therefore analyse this component as **terminology preservation under low-frequency conditions**, rather than as a generic rare-word benchmark.

Following prior rare-word analyses in neural machine translation [27, 28], term frequencies are estimated from the GRPO training corpus, and test-set term alignment accuracy is reported by frequency bins. Figures 4 and 5 present cross-model and key-model comparisons.

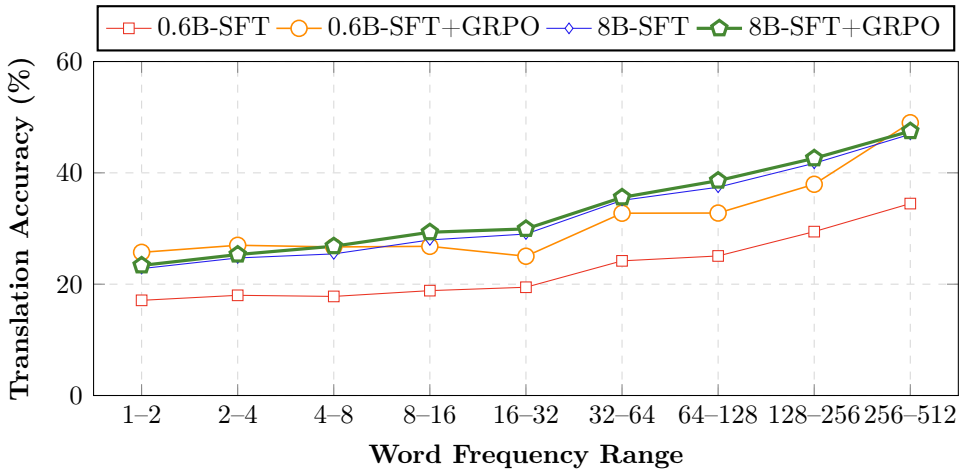


Figure 4. Term-alignment accuracy across frequency bins for four reproducible open-source model settings.

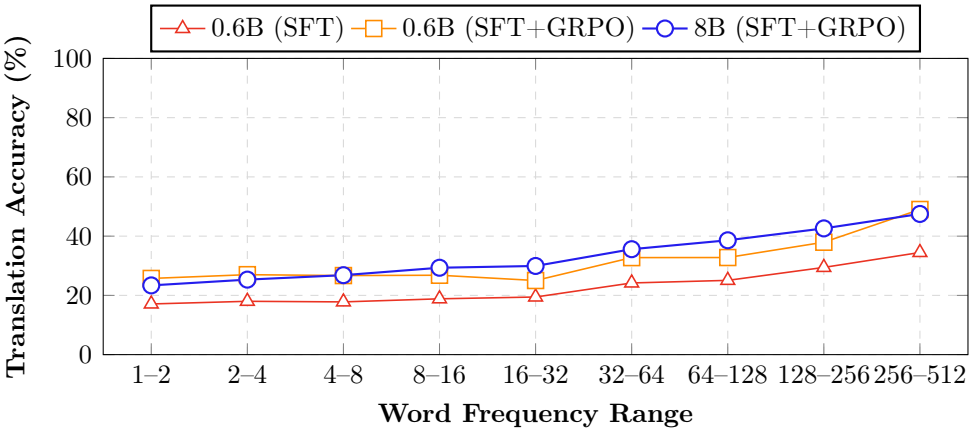


Figure 5. GRPO-related gains in term-alignment accuracy for 0.6B-SFT, 0.6B-SFT+GRPO, and 8B-SFT+GRPO.

Key findings.

- **Frequency-sensitive behavior is universal**: all models show higher accuracy in higher-frequency bins, indicating persistent sparsity effects

in terminology preservation.

- **GRPO improved low-frequency term handling**: for the 0.6B model, overall term-alignment accuracy increases from 19.78% under SFT to 28.09% under SFT+GRPO; for 8B, it increases from 25.98% to 27.50%.

- **Low-frequency bins remain challenging**: in the <16-frequency region, 0.6B improves from 17.93% to 26.55%, while 8B changes from 25.23% to 26.21%.

- **Model capacity remains important**: larger models maintain stronger performance across bins, but GRPO narrows part of the gap between small and large models.

These results should be interpreted at the component level: the contribution of our method is not universal dominance on every isolated submetric, but a heritage-aligned optimisation framework combining structured reasoning and reward alignment for broader interpretive stability. Detailed statistics are provided in Appendix Appendix F, Table F.17.

5.3. Causal-chain preservation in long and paratactic structures

We analyse sentence-length effects as supporting evidence for **causal-chain preservation** in compressed classical syntax. In classical medical texts, longer spans often contain nested conditions, implicit causality, and paratactic linkage; preserving these relations is central to epistemic fidelity.

Following prior MT length analyses [29], the test set is partitioned into ten source-length bins, from 1–10 to 91–100 characters, and BLEU and COMET trends are examined by bin, as shown in Fig. 6.

Key findings.

- **GRPO showed stronger medium- and long-range preservation**: in the 21–30 bin, 0.6B BLEU increases from 20.6 to 41.3 and COMET from 0.712 to 0.785; 8B reaches BLEU 57.4 and COMET 0.805.

- **Capacity effect remains clear**: 8B consistently outperforms 0.6B across bins, especially in the 41–50 range, indicating better handling of long-range dependencies.

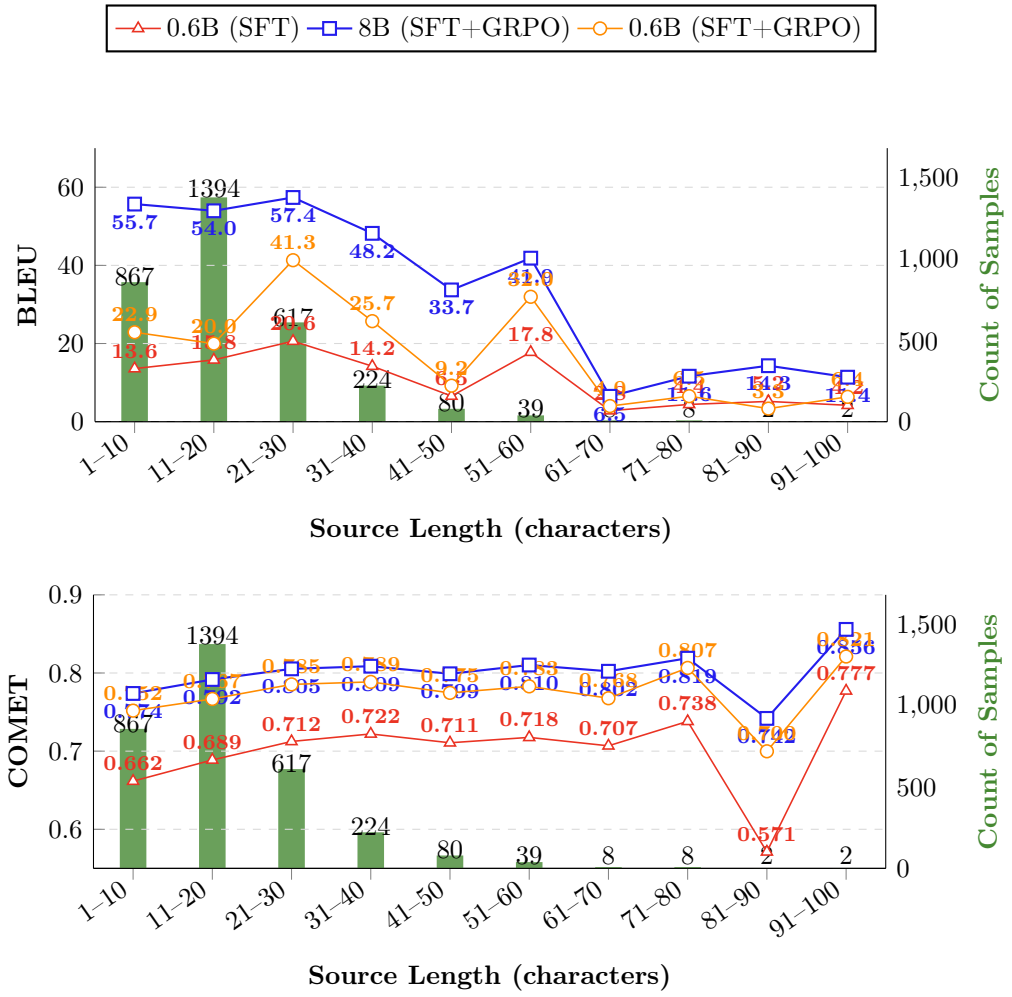


Figure 6. BLEU and COMET by source-length bins. Bars on the right y-axis show sample counts, while lines on the left y-axis show model performance across Base, SFT, and SFT+GRPO settings.

- **Very long inputs remain difficult**: performance declines in the 81–90 range; for example, 8B COMET falls to 0.742, suggesting unresolved challenges in preserving complex multi-clause medical reasoning.
- **Distribution imbalance constrains certainty at long lengths**: short bins dominate the test set, with 1,394 samples in the 11–20 range, while >60-character bins are sparse, which limits robustness of extreme-length conclusions.

To provide direct case-level evidence in the main text, Table 6 reports representative sentences with implicit conditional/causal chains and compares Qwen3-8B-SFT with Qwen3-8B-SFT+GRPO outputs.

Table 6. Case study: long and paratactic conditional-causal chain preservation in *Shanghanlun*.

Len.	Original	Reference	SFT	SFT+GRPO	Interpretive note
21	恶水者，若下之，则里冷不嗜食，大便完谷出；	病人不愿喝水的，如果误用攻下，致阴寒内感，就会出现不思饮食、大便完谷不化；	如果呕吐清水，用攻下的方法，就会使病人出现腹中冷，厌恶进食，大便有未消化的食物排出。	怕喝水的病人，如果误用攻下法治疗，就会引起里寒、食欲不振，大便中含有未消化的食物；	SFT misreads "恶水" as "vomiting water"; GRPO restores the condition-cause chain more faithfully.
43	但头汗出，身无汗，剂颈而还，小便不利，渴引水浆者，此为瘀热在里，自必发黄，茵陈汤主之。	如果仅见头部出汗，到颈部为止，身上无汗，小便不通畅，口渴想喝汤水，这是湿热郁滞在里，势必出现肌肤发黄，用茵陈蒿汤主治。	但头汗出而身无汗齐颈而还，小便不利，口渴饮水的，这是瘀热在里，不久必发黄，以茵陈汤主治。	只是头部出汗，身体无汗，汗出至颈部即停止，小便不通畅，口渴想喝水的人，这是瘀热在里，必然会出现发黄，用茵陈汤治疗。	Both preserve the chain; GRPO is clearer on clause boundaries and causal progression.

Overall, this analysis supports the same heritage-oriented interpretation as above: GRPO improves structure-sensitive translation behavior, but full preservation of long- range causal semantics in highly compressed classical prose remains an open problem. Additional examples are provided in Appendix Appendix G, Table G.18.

5.4. Four heritage-relevant interpretive risk dimensions

To avoid a fragmented benchmark-style presentation, we organise challenge analysis into four heritage-relevant interpretive risk dimensions. These are semantic shift across historical language layers, cultural-epistemic context preservation, zero-anaphora recovery, and paratactic logic preservation. This framing links empirical evaluation directly to semantic preservation and epistemic continuity in digitised heritage texts.

While BLEU and COMET do not directly measure epistemological integrity, we used them as reproducible proxy indicators and complemented them with phenomenon-level qualitative audits. The challenge proportions in Table 7 follow the predefined phenomenon annotation protocol described in Materials. Challenge BLEU/COMET results are reported as macro averages over books under strict index alignment.

Table 7. Distribution of four heritage-relevant interpretive risk dimensions in the evaluation set.

Risk dimension	Total samples	Samples containing the phenomenon	Ratio, %
Semantic shift, classical-to-modern	3241	2927	90.31
Cultural-epistemic context	3241	2965	91.48
Parataxis and implicit logic	3241	2406	74.24
Zero anaphora	3241	1893	58.41

5.4.1. Semantic shift across historical language layers

Classical TCM texts contain substantial diachronic semantic variation, where the same lexical form may encode different meanings in modern Chinese. This dimension evaluates whether models preserve historically situated meaning rather than producing superficially fluent paraphrases.

Figure 7 summarises model performance on this subset.

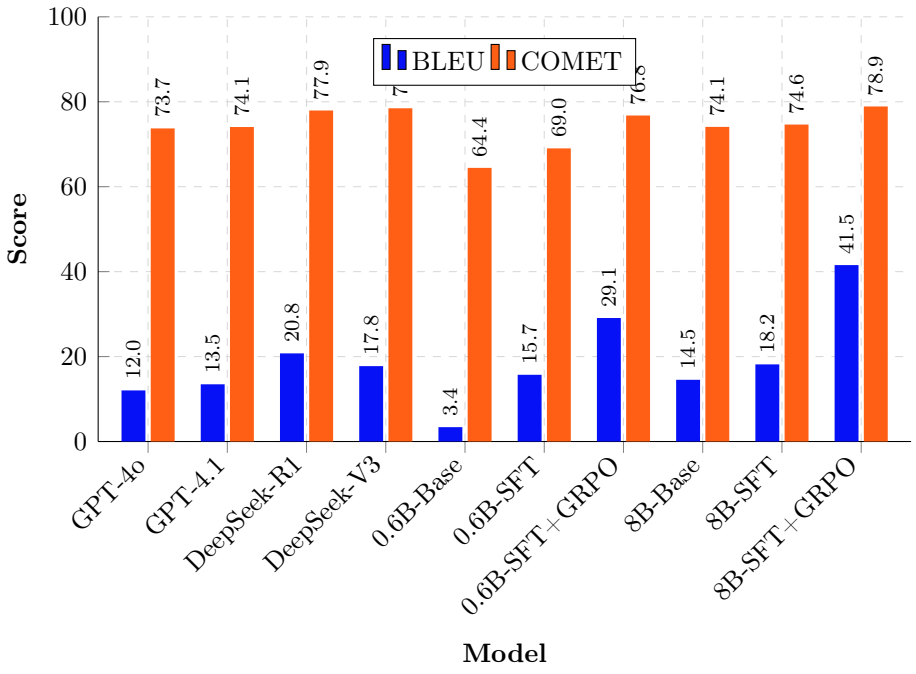


Figure 7. Results on the semantic-shift subset, with BLEU and COMET used as proxy metrics.

Key observations are as follows:

- **The SFT+GRPO 8B model performs strongest overall** on this subset, with BLEU at 41.5 and COMET at 78.9.

- **GRPO delivered consistent gains** over SFT on semantic-shift cases, particularly where explicit interpretive reconstruction is required.
- **Frontier baselines from both closed-source and open-source families remain competitive on COMET**, but often show lower BLEU on terminologically dense instances.
- **Model capacity matters**: 8B variants generally outperform 0.6B variants on historically shifted semantics.

Table 8 provides case-level comparisons between SFT and SFT+GRPO outputs.

Table 8. Case study: semantic-shift translation. Red marks lexemes with historical semantic shift.

Source	Original	Reference	SFT Output	SFT+GRPO Output	Interpretive Note
温病条辨	若再加面黄肢逆，则非前汤所能济，故以四逆回厥，茵陈宣湿退黄也。	如果伴有面部发黄、四肢怕冷，不是上方所能治疗，必须使用四逆汤温经散寒，缓解四肢怕冷症状，配合茵陈宣化湿邪消除黄疸。	如果再出现面黄肢厥，就不单是前文的茵陈蒿汤主治，所以用四逆汤回阳恢复厥冷，茵陈蒿汤清湿热退黄。	如果再出现面色发黄、四肢厥冷，就不是前面的方剂所能治疗的了，所以用四逆汤回阳救逆，茵陈蒿汤宣化湿邪、退黄。	Better recovery of historical sense for “宣”; SFT over-normalises to “清湿热”
金匮要略	咳而上气，此为肺胀，其人喘，目如脱状，脉浮大者，越婢加半夏汤主之。	咳嗽而气逆，这是肺胀，患者气喘，两目外突像要掉出来一样，脉象浮大的，用越婢加半夏汤主治。	咳而喘息不能平卧，此为肺胀，病人还有气喘，目睛下垂，脉浮大的，用越婢加半夏汤主治。	咳嗽而呼吸急促，这是肺胀的表现，病人气喘，眼睛好像要脱出的样子，脉象浮大，用越婢加半夏汤主治。	“脱” is historically “protruding,” not “drooping”
金匮要略	师曰：尺脉浮，目睛晕黄，衄未止，晕黄去，目睛慧了，知衄今止。	老师说：尺脉呈浮脉，眼睛昏晕呈黄色，说明仍有衄血，如果眼睛不黄而视物清爽，就可知衄血已经停止了。	老师说:尺脉浮，眼睛发黄，说明衄血未止，如果眼睛的黄色消失，瞳孔明亮，就知道衄血已经停止了。	老师说:尺脉浮，眼睛发黄，表明衄血未止，如果黄色消失，眼睛清明，就知道衄血已经止住了。	Both models partially compress diachronic semantics for “晕黄/慧了”

Case analysis indicates that SFT+GRPO handled semantic-shift phenomena better, especially for polysemy and historical sense disambiguation. However, difficult diachronic terms remain partially unresolved in both models, indicating an ongoing preservation risk in high-ambiguity contexts.

Overall, this dimension supports a heritage-oriented conclusion: gains are not merely lexical, but linked to improved preservation of historically grounded meaning. Future work may further strengthen this line through

three directions. The first direction is diachronic semantic distance modelling [30, 31, 32]. The second direction is retrieval- enhanced historical term disambiguation [33, 34, 35]. The third direction is explicit generation of contextual historical definitions [36].

5.4.2. Cultural-epistemic context preservation

Translation of classical TCM texts involves not only linguistic transfer but also preservation of culturally embedded epistemic structures. This dimension evaluates whether model outputs maintain doctrinally grounded meaning across core TCM concepts, including yin-yang relations, zang-fu theory, meridian theory, and pathogenesis logic, rather than merely producing fluent modern-language paraphrases.

Figure 8 reports model performance on this subset.

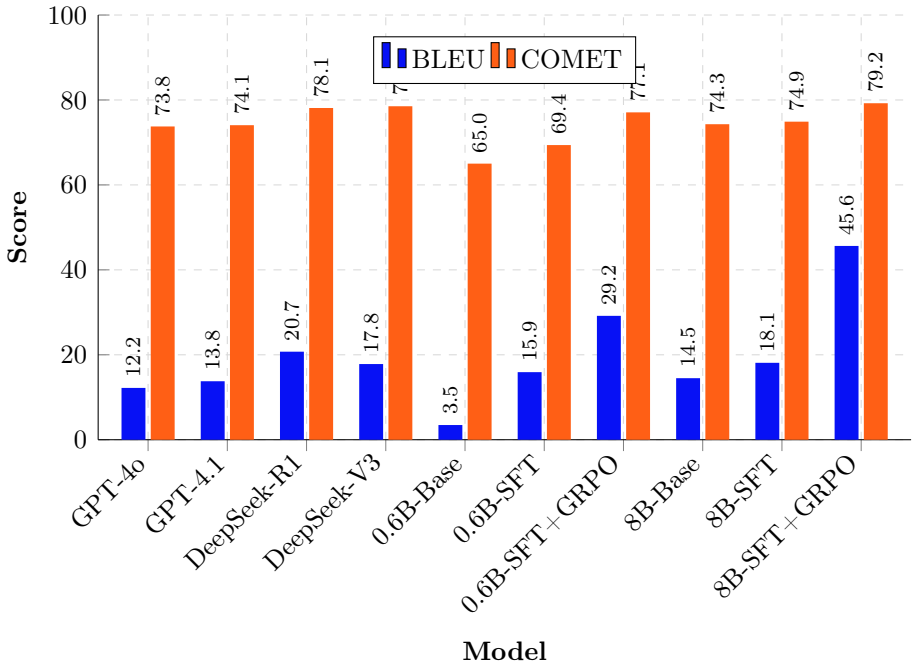


Figure 8. Results on the cultural-epistemic context subset, with BLEU and COMET used as proxy metrics.

Key findings are as follows:

- **The SFT+GRPO 8B model performs strongest overall** on this subset, with BLEU at 45.6 and COMET at 79.2.

- **GRPO delivered substantial gains** over SFT in both 0.6B and 8B settings, indicating improved preservation of culturally embedded epistemic relations.
- **Frontier general models remain competitive on COMET**, but case-level analysis reveals persistent doctrinal misreadings in highly compressed classical expressions.
- **Model capacity contributes to robustness**: 8B variants generally show stronger stability than 0.6B variants on cultural-epistemic items.

Table 9 presents case-level comparisons between SFT and SFT+GRPO.

Table 9. Case study: cultural-epistemic context translation. Red marks key doctrinal concepts.

Category	Original	Reference	SFT Output	SFT+GRPO Output	Interpretive Note
阴阳学说	岐伯对曰：阳气衰于下，则为寒厥；阴气衰于下，则为热厥。	岐伯答道：阳气衰竭于下，发为寒厥；阴气衰竭于下，发为热厥。	岐伯回答说：阳气衰于下焦，不能温养四肢，所以发生寒逆厥冷；阴气衰于下焦，不能控制阳气，发生热逆厥冷。	岐伯回答说：阳气衰弱于下，则发生寒厥；阴气衰弱于下，则发生热厥。	SFT over-expands beyond source scope; SFT+GRPO better preserves yin-yang correspondence
藏象学说	肝为刚脏，内寄相火，非纯刚所能折。	肝为刚脏，内寄相火，并不是单纯使用刚药就能奏效。	肝为刚脏，寄相火于其中，若仅仅只是以纯刚之法加以剥削，则反而不利于肝的养生。	肝为刚脏，内寄相火，并不是单纯的刚劲之力所能制服的。	Better semantic control on "折" under SFT+GRPO
经络学说	平旦寅初，肺气流布，起于寸口，运行十二经中，周而复始。	在凌晨3时到5时，肺气在寸口部位开始，推动经脉中的气血运行，循环不休。	早晨，也就是一日的凌晨时分，这个时辰肺经主气，这时进行针刺取穴寸口，可以更好地使气血运行于经脉之中。	清晨寅时初刻，肺气开始流行布散，从寸口开始，运行于十二经脉 之中，循环往复。	SFT misinterprets as acupuncture operation; SFT+GRPO preserves physiological description
五运六气	见少阳，左阳明，右太阴。	少阳司天，左间是阳明，右间是太阴。	在左手脉象应是少阳，右脉宜为阳明或太阴。	可见少阳脉，左手见阳明脉，右手见太阴脉。	Both models struggle; both drift toward pulse-diagnosis interpretation
藏象比喻	脾为之使，胃为之市。	脾主运化，水谷精微赖以转输，胃主受纳，饮食水谷汇聚于此。	脾为胃的使官，胃为脾的市场。	脾是它们的使者，胃是它们的市场。	Both remain overly literal; doctrinal functional meaning is under-preserved

Case-level evidence suggests that SFT+GRPO more consistently preserves core doctrinal relations and reduces over-interpretive expansion beyond source scope. However, metaphorical doctrinal language and highly specialised conceptual systems, such as Five Movements and Six Qi, remain difficult for both SFT and SFT+GRPO.

Overall, this dimension indicates that cultural-epistemic context is a major preservation bottleneck in classical medical translation. The observed gains are best interpreted as improved interpretive stability under heritage-relevant constraints, rather than as complete resolution of doctrinal understanding. Future work may further strengthen this direction through three directions. The first direction is knowledge-graph-assisted cultural grounding [37, 38]. The second direction is culture-aware data augmentation [39, 40]. The third direction is culture-sensitive evaluation and preference modelling [41, 42].

5.4.3. Zero-anaphora recovery and referential coherence

Zero anaphora, which frequently involves omission of subjects and objects, is pervasive in classical Chinese and poses a direct risk to epistemic fidelity, because omitted referents often encode causal agents in diagnostic or therapeutic propositions. This dimension therefore evaluates **referential coherence preservation** rather than surface fluency alone.

Figure 9 reports model performance on the zero-anaphora subset.

Key observations are as follows:

- **The SFT+GRPO 8B model performs strongest overall**, with BLEU at 37.3 and COMET at 78.0.
- **GRPO improved over SFT**, especially in explicit recovery of omitted conditional markers.
- **Frontier general models remain competitive on COMET**, but often show weaker lexical alignment in omission-heavy examples.

Table 10 presents representative cases, including omitted conjunctions and omitted antecedents.

Case-level evidence indicates that SFT and SFT+GRPO remain close on hard antecedent-recovery cases, while GRPO shows clearer gains in conjunction recovery and logical explicitness. This suggests that zero-anaphora remains a major unresolved interpretive risk, especially for omitted referential agents that require broader discourse grounding.

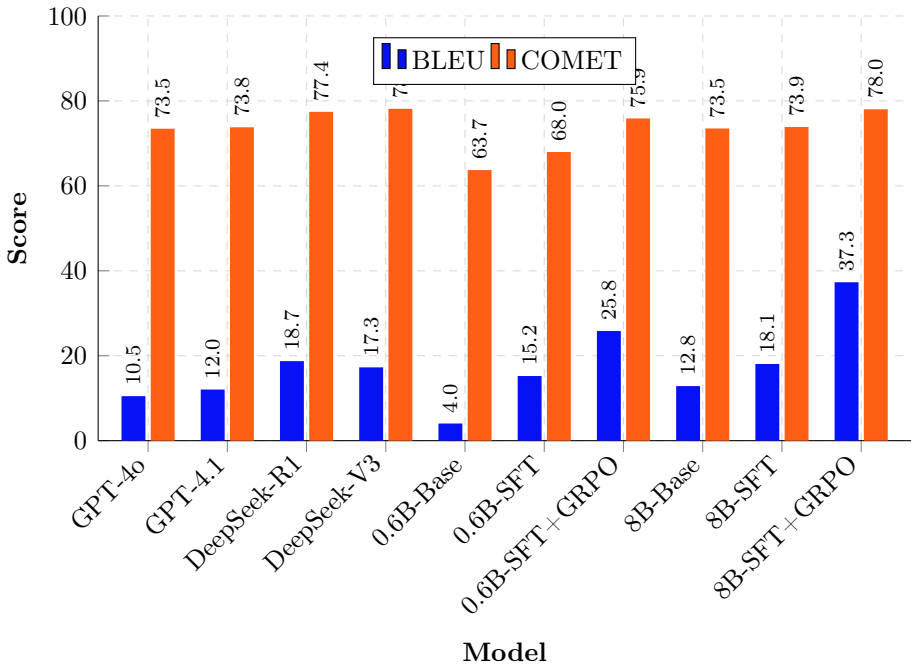


Figure 9. Results on the zero-anaphora subset, with BLEU and COMET used as proxy metrics.

Table 10. Case study: zero-anaphora translation. Red marks omitted elements; Ø marks omission positions.

Omission Type	Original	Reference	SFT Output	SFT+GRPO Output	Interpretive Note
连词+主语省略	此春气之应，养生之道也。Ø逆之则伤肝，夏为寒变。	这是春气的感应，是养生的道理。人若违背了春气，就会损伤肝气，夏天就会发生寒变。	这是春气的感应，养生的道理。违背它就会损伤肝气，夏天变成寒变，奉养长气不足。	这是春气的感应，养生的道理。如果违背它，就会损伤肝气，夏天就会发生寒变。	GRPO better recovers omitted conditional marker, but subject recovery remains incomplete
主语承前省	所以Ø肺热叶焦，皮毛虚弱急薄，Ø著则生痿也。	所以热邪使肺热叶焦，皮毛虚弱急薄，热邪不去，则变生痿。	所以肺热叶焦，皮毛虚弱急薄，停留就生痿。	所以肺热叶焦，皮毛虚弱急薄，停留就生痿。	Both models fail to restore omitted agent “热邪”
条件连词省略	此必因虚邪之风与其身形，两虚相得，乃客其形。Ø两实相逢，众人肉坚。	这必定是因为虚邪之风与虚弱的身体相遇，两虚相合，邪气才能侵入人体。若两实相逢，人的肌肉就会坚实。	这必定是因为虚邪之风与虚弱的身体相遇，两虚相合，邪气才能侵入人体。两实相逢，人的肌肉就会坚实。	这必定是因为虚邪之风与虚弱的身体相遇，两虚相合，邪气才能侵入人体。如果两实相逢，人的肌肉就会坚实。	GRPO more reliably explicitates omitted conditional relation

Overall, this dimension supports a conservative conclusion: GRPO improved referential explicitness in some omission types, but robust restoration of omitted antecedents is still limited. Future work may improve this through three directions. The first direction is mention-aware semantic modelling [43, 44]. The second direction is joint modelling of translation and zero-anaphora recovery [45]. The third direction is contrastive evaluation protocols that are more sensitive to referential errors [46, 47].

5.4.4. Paratactic logic preservation and explicit relation recovery

Classical TCM prose is strongly paratactic: logical relations are often encoded implicitly through sequence and semantic juxtaposition rather than explicit connectors. In heritage translation, this creates a major epistemic risk, because missing relation recovery can alter diagnostic causality and therapeutic logic.

Figure 10 reports model performance on the parataxis subset.

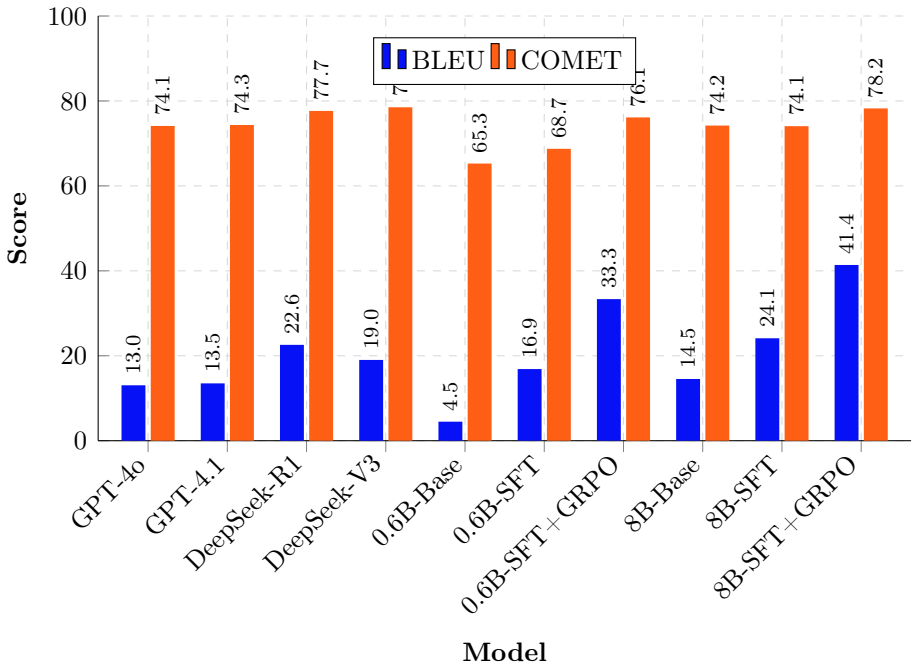


Figure 10. Results on the parataxis subset, with BLEU and COMET used as proxy metrics.

Key findings are as follows:

- **The SFT+GRPO 8B model performs strongest overall**, with

BLEU at 41.4 and COMET at 78.2.

- **GRPO produced clear gains** over SFT in both 0.6B and 8B settings, especially on instances requiring explicit recovery of conditional and causal links.
- **Frontier general models remain competitive on COMET**, but show lower BLEU in connector-sensitive examples.
- **Capacity effect remains visible**: 8B models generally outperform 0.6B models on implicit-logic recovery.

Table 11 provides representative case comparisons.

Table 11. Case study: parataxis and implicit-logic translation. Red marks logic-relevant expressions.

Relation Type	Original	Reference	SFT Output	SFT+GRPO Output	Interpretive Note
因果关系	因于露风，乃生寒热。	如果触冒风邪，就会发生寒热。	感受了寒邪，就会产生寒热。	如果感受了风邪，就会发生寒热病。	SFT confuses "露风"; SFT+GRPO better recovers causal explicitness
转化关系	故重阴必阳，重阳必阴。	阴气过盛会转化为阳病，阳气过盛也会转变为阴病。	所以阳极则阴，阴极则阳。	所以重阴必转为阳，重阳必转为阴。	SFT+GRPO preserves transformation pattern better, though pathology nuance remains compressed
时间条件	所谓痹者，各以其时重感于风寒湿之气也。	总之，这些痹证是各脏在所主季节里重复感受了风、寒、湿气所造成的。	所谓痹病，是指邪气在病人身体的相应部位反复侵犯所引起的病变。	所谓痹病，都是在不同的时间里，再次感受风、寒、湿邪所致的。	SFT misreads temporal scope; SFT+GRPO recovers time-conditioned interpretation more reliably

Case-level evidence indicates that SFT+GRPO improves explicit recovery of implicit relations, particularly conditional and causal relations, and reduces unsupported explanatory expansion. However, both models remain vulnerable on doctrinally compressed expressions, where logic reconstruction and medical-theory interpretation are tightly coupled.

Overall, this dimension supports a heritage-oriented conclusion: GRPO showed stronger paratactic logic preservation, but high-fidelity reconstruction of deeply compressed doctrinal semantics remains an open challenge. Future work may benefit from three directions. The first direction is explicit discourse-structure modelling [48, 49, 50]. The second direction is joint implicit-relation classification and connector generation [51, 52]. The third direction is discourse-sensitive evaluation protocols tailored to classical medical texts [53, 54].

5.4.5. Composite-risk cases across multiple phenomena

Many classical TCM sentences contain overlapping interpretive risks, including semantic shift, cultural-epistemic context, and zero-anaphora, where errors in one layer can propagate to others. We therefore analyse **composite-risk cases** to assess whether gains observed in single dimensions transfer to structurally entangled instances.

Table 12 presents representative examples.

Table 12. Case study: composite-risk translation. Red marks lexemes linked to multiple interpretive phenomena.

Composite Phenomena	Original	Reference	SFT Output	SFT+GRPO Output	Interpretive Note
古今异义+文化语境	病生胁，气归于左，善太息，甚则心痛否满，腹胀而泄。	人们的疾病生于胁，燥气偏于左侧，时时叹息，甚则心痛痞满，腹胀而泄泻。	胆病发于胳胁部，气归于左右两侧，因之常多长吁短叹，病情严重时，痛连心腹而胀满。	病生于胁，邪气归于左边，常常叹息，严重时会出现心痛、痞满、腹部胀满而腹泻。	SFT loses historical interchangeability of “否/痞”; SFT+GRPO improves part of lexical recovery but still compresses doctrinal context, especially “气归于左”
古今异义+零形回指+文化语境	太阳有余，病骨痹身重；Ø不足，病肾痹；滑则病肾风疝；涩则病积，善时巅疾。	太阳之气有余，可以发生骨痹、身体沉重；Ø不足则发生肾痹；气血过于滑利则患肾风疝；气血运行涩滞则病积聚，且不时发生巅顶部疾病。	太阳有余则病肿痹，身体重着，肾不足则病肾痹，脉滑则病肾风疝，脉涩则病成积，晨起头巅疾病。	太阳邪有余，会发生骨痹病，身体沉重；邪气不足，则患肾痹病；脉见滑象，则患肾风疝病；脉见涩象，则患积病，时常发作头痛。	SFT misreads “善时” as “晨起”; SFT+GRPO recovers frequency meaning but introduces referential drift from “太阳之气” to “邪气” and mechanism drift by treating “滑/涩” as pulse signs

These cases indicate that composite phenomena remain the most difficult setting for heritage-text translation. Improvements on single dimensions do not fully transfer to multi-phenomenon contexts, where terminology, reference, and doctrinal interpretation interact tightly.

From a heritage-science perspective, this finding is important: preservation risk is often cumulative rather than isolated. Therefore, future optimisation should target integrated epistemic consistency across coupled phenomena, not only dimension-wise improvements.

6. Discussion

This study examines R1-TCM-Translator as a heritage-oriented computational framework for classical medical text translation, with emphasis on epistemic fidelity rather than metric optimisation alone. The empirical results show consistent improvements across multiple evaluation settings, while also revealing clear boundary conditions for preservation-oriented deployment.

6.1. Complementary roles of SFT and GRPO

The results suggest a complementary mechanism between SFT and GRPO. SFT primarily contributes domain grounding through terminology exposure and basic concept alignment, whereas GRPO yields stronger structure-aware translation behaviour under implicit logic and diachronic semantics. In practical terms, SFT functions as lexical and conceptual acquisition, whereas GRPO promotes reason-then-translate behaviour under reward-guided optimisation.

Case analyses indicate that this combination is particularly useful in three aspects. First, it reduces modern-language over-interpretation in semantic-shift cases. Second, it preserves doctrinal correspondences in cultural-epistemic contexts. Third, it recovers implicit conditional and causal relations in paratactic structures. We therefore interpret the observed gains as consistent with improved interpretive stability, rather than claiming a universal or causal proof of "deep reasoning activation."

From a heritage perspective, this mechanism matters because it improves the semantic faithfulness of digital renderings used for downstream archival access, retrieval, and scholarly interpretation [6, 20, 21].

6.2. Terminology sparsity: challenge and opportunity

Low-frequency terminology remains a central bottleneck. Although GRPO delivered substantial gains in sparse-term settings, very rare items still exhibit unstable translation behaviour. This is expected in heritage corpora where key concepts are historically specific, distributionally sparse, and context-dependent.

A practical implication is that preservation quality depends on both model behaviour and corpus quality control. Future improvements are likely to come from tighter integration of structured lexical resources, retrieval-augmented grounding, and sparsity-aware optimisation strategies, together with stronger document-quality governance in heritage corpora [55, 20].

From a heritage-science perspective, improved rare-term handling directly affects whether digitised archives retain medically meaningful vocabulary rather than only surface- level paraphrases.

6.3. Differentiated difficulty across four interpretive risks

The four challenge categories show differentiated difficulty profiles. Semantic-shift and parataxis cases benefit most from GRPO, suggesting that reward-guided reasoning is effective for recovering implicit logic and diachronic meanings. Cultural-epistemic cases improve but remain sensitive to doctrine-specific metaphors and highly specialised conceptual mappings. Zero-anaphora remains the most persistent bottleneck, especially for subject recovery under long-distance context dependencies. These patterns indicate that future gains will require tighter integration of discourse-level modelling, terminology resources, and challenge-specific supervision.

This differentiated pattern is important for heritage practice: not all interpretive risks are equally tractable, and preservation pipelines should prioritise high-risk categories such as zero-anaphora and doctrine-compressed passages for targeted quality control [20].

6.4. Limitations and boundary conditions

This study has several limitations. First, the corpus covers eight classics and may not fully represent the full diversity of TCM textual traditions. Second, evaluation is primarily based on automatic metrics, namely BLEU and COMET, which are proxies and do not directly quantify epistemological integrity. Third, complex doctrine-dense passages, such as Five Movements and Six Qi, remain difficult and can still produce interpretive drift.

Fourth, despite three-run decoding averaging, very long and highly compressed sentences remain challenging, indicating unresolved limits in long-range discourse-level preservation. Fifth, we have not yet conducted blinded expert evaluation with inter-rater agreement reporting; future work will include at least two independent TCM experts, model-identity masking, and agreement statistics such as Cohen's κ or ICC for doctrinal fidelity assessment.

These limitations define clear boundary conditions: the current framework improves consistency under many heritage-relevant scenarios, but it does not yet guarantee full doctrinal fidelity in all high-complexity contexts.

6.5. Transferability and constraints

The method has potential transferability to other classical medical traditions, including Ayurveda, Greco-Roman medicine, and Arabic medical texts, but this should be treated as a constrained hypothesis rather than a demonstrated result. What is likely transferable is the methodological skeleton, namely structured reasoning scaffolds and reward decomposition. What is not directly transferable are domain lexicons, expert conventions, and corpus-specific curation rules, which must be rebuilt and validated on target traditions.

Accordingly, future cross-tradition studies should treat transfer as an empirical question and re-establish annotation protocols and preservation criteria in each new heritage corpus.

7. Conclusion

This study introduces R1-TCM-Translator, a heritage-oriented framework that combines structured reasoning traces with GRPO for classical TCM translation. Rather than claiming a complete solution, our results indicate consistent improvements in handling major interpretive risks in heritage texts, including semantic shift, cultural-epistemic context, parataxis, and zero-anaphora.

Empirically, R1-TCM-Translator-8B achieves strong performance on both BLEU as a lexical proxy and COMET as a semantic-fidelity proxy. On BLEU, it substantially outperforms strong general-purpose baselines, for example GPT-4o and DeepSeek-R1. On COMET, it remains competitive with frontier models while maintaining better cross-text interpretive consistency in multiple challenge-specific analyses.

Case-level findings are consistent with the following mechanism: SFT primarily improves domain grounding, while GRPO supports better recovery of implicit logic and historically situated meanings under structured reward alignment. At the same time, difficult doctrine-dense passages and composite multi-risk cases remain challenging, especially when omitted referents and highly specialised conceptual mappings co-occur.

From a heritage-science perspective, the framework provides a practical computational pathway for improving semantic fidelity in digitised medical heritage archives, with potential downstream value for knowledge organisation and scholarly access. However, current results do not guarantee doctrinally complete fidelity in all high-complexity passages.

Future work will focus on three directions: first, stronger discourse-level and referential modelling for zero-anaphora and long-range causal chains; second, tighter integration of knowledge-grounded disambiguation mechanisms; and third, broader validation under cross-tradition settings. The framework may provide a testable methodological hypothesis for other classical medical traditions, but transfer requires rebuilding annotation standards on each target corpus to ensure internationally comparable heritage-science evidence.

Acknowledgments

The authors thank the domain experts and annotators who contributed to corpus curation and quality control.

Author Contributions

S.X. designed the study, conducted all experiments, and performed the main experimental analyses. X.S. and W.S. aligned and curated the dataset and carried out the case-study analyses. J.L. provided computational resources and guidance for part of the experiments. M.O. and S.X. drafted the manuscript and revised the text. W.H. provided technical support and assisted with manuscript polishing. W.L. coordinated and supervised the project as principal advisor. All authors reviewed the manuscript, approved the final version, and agreed to be accountable for the work.

Funding

This research was supported by the Natural Science Foundation of Hunan Province (Grant Number 2022JJ30438), the Natural Science Foundation of

Changsha (Grant Number kq2202260), and the Research Project of Hunan Provincial Administration of Traditional Chinese Medicine (Grant Number B2023039).

Data Availability

The processed datasets supporting this study (including aligned samples, split metadata, and evaluation references) are publicly available at https://huggingface.co/datasets/xs12345/R1-TCM-Data. Owing to copyright restrictions on source editions, raw source books are not redistributed.

Code Availability

Code for data processing, training, and evaluation is available at https://github.com/xiesi1234/R1-TCM-Translator. Model checkpoints are available at https://huggingface.co/xs12345/R1-TCM-Translator-0.6B and https://huggingface.co/xs12345/R1-TCM-Translator-8B.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] X. Nie, M. Ma, J. Ji, L. Zheng, The spatial distribution of traditional intangible cultural heritage medicine of China and its influencing factors, Heritage Science 11 (2023) 90. doi:10.1186/s40494-023-00929-0.

[2] R. Nichols, Understanding East Asian holistic cognitive style and its cultural evolution: a multi-disciplinary case study of Traditional Chinese Medicine, Journal of Cultural Cognitive Science 5 (1) (2021) 17–36. doi:10.1007/s41809-021-00075-8.

[3] J. Curran, The Yellow Emperor's Classic of Internal Medicine, BMJ 336 (7647) (2008) 777. doi:10.1136/bmj.39527.472303.4E.

[4] Y. Tu, The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine, Nature Medicine 17 (10) (2011) 1217–1220. doi:10.1038/nm.2471.

[5] X.-z. Su, L. H. Miller, The discovery of artemisinin and the Nobel Prize in Physiology or Medicine, Science China Life Sciences 58 (11) (2015) 1175–1179. doi:10.1007/s11427-015-4948-7.

[6] UNESCO, Recommendation concerning the preservation of, and access to, documentary heritage including in digital form, UNESCO, Paris, recommendation text, adopted in 2015 (2015).

[7] Y. Y. Huang, S. Yu, J. Chu, H. Fan, B. Du, Using knowledge graphs and deep learning algorithms to enhance digital cultural heritage management, Heritage Science 11 (2023) 204. doi:10.1186/s40494-023-01042-y.

[8] T. Fan, H. Wang, T. Hodel, CICHMKG: a large-scale and comprehensive chinese intangible cultural heritage multimodal knowledge graph, Heritage Science 11 (2023) 115. doi:10.1186/s40494-023-00927-2.

[9] B. Bai, W. Hou, The application of knowledge graphs in the chinese cultural field: the ancient capital culture of beijing, Heritage Science 11 (2023) 77. doi:10.1186/s40494-023-00922-7.

[10] M. L. Nappi, M. Buono, C. Chivaran, R. M. Giusto, Models and tools for the digital organisation of knowledge: accessible and adaptive narratives for cultural heritage, Heritage Science 12 (2024) 112. doi:10.1186/s40494-024-01219-z.

[11] M. Skublewska-Paszkowska, M. Milosz, P. Powroznik, E. Lukasik, 3D technologies for intangible cultural heritage preservation—literature review for selected databases, Heritage Science 10 (2022) 3. doi:10.1186/s40494-021-00633-x.

[12] B. Liu, C. Wu, W. Xu, Y. Shen, F. Tang, Emerging trends in GIS application on cultural heritage conservation: a review, Heritage Science 12 (2024) 139. doi:10.1186/s40494-024-01265-7.

[13] Q. Xie, M. Cui, Z. Wu, H. Zhao, Traditional chinese medicine information digitalization discussion, Journal of Alternative and Complementary Medicine 16 (11) (2010) 1207–1209. doi:10.1089/acm.2009.0700.

[14] X. Zhou, S. Chen, B. Liu, R. Zhang, Y. Wang, P. Li, Y. Guo, H. Zhang, Z. Gao, X. Yan, Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support, Artificial Intelligence in Medicine 48 (2-3) (2010) 139–152. doi:10.1016/j.artmed.2009.07.012.

[15] B. Zhang, R. Li, K. Yin, S. Hua, S. Li, M. Jiang, H. An, P. Li, Construction and application of traditional chinese medicine knowledge graph based on large language model, Interdisciplinary Sciences: Computational Life Sciences (online first), published online 2025-07-02 (2025). doi:10.1007/s12539-025-00735-1.

[16] X. Qu, A review of knowledge graph in traditional chinese medicine: Analysis, construction, application and prospects, Computers, Materials and Continua 81 (2024) 3583–3616.

[17] J. Ødemark, E. Engebretsen, Challenging medical knowledge translation: convergence and divergence of translation across epistemic and cultural boundaries, Humanities and Social Sciences Communications 9 (2022) 71. doi:10.1057/s41599-022-01088-6.

[18] Y. Chen, S. Li, Y. Li, M. Atari, Surveying the dead minds: Historical-psychological text analysis with contextualized construct representation (CCR) for classical Chinese, in: Y. Al-Onaizan, M. Bansal, Y.-N. Chen (Eds.), Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Miami, Florida, USA, 2024, pp. 2597–2615. doi:10.18653/v1/2024.emnlp-main.151.
URL https://aclanthology.org/2024.emnlp-main.151/

[19] N. Borenstein, P. Rust, D. Elliott, I. Augenstein, PHD: Pixel-based language modeling of historical documents, in: H. Bouamor, J. Pino, K. Bali (Eds.), Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Singapore, 2023, pp. 87–107. doi:10.18653/v1/2023.emnlp

-main.7.
URL https://aclanthology.org/2023.emnlp-main.7/

[20] Y. Chen, L. Zhang, Q. Dong, Using natural language processing to evaluate local conservation text: A study of 624 documents from 303 sites of the world heritage cities programme, Journal of Cultural Heritage 70 (2024) 259–270. doi:10.1016/j.culher.2024.09.011.

[21] M. J. Hill, S. Hengchen, Quantifying the impact of dirty OCR on historical text analysis: Eighteenth century collections online as a case study, Digital Scholarship in the Humanities 34 (4) (2019) 825–843. doi:10.1093/llc/fqz024.

[22] C. Booth, R. Shoemaker, R. Gaizauskas, A language modelling approach to quality assessment of OCR'ed historical text, in: N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, J. Odijk, S. Piperidis (Eds.), Proceedings of the Thirteenth Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2022, pp. 5859–5864.
URL https://aclanthology.org/2022.lrec-1.630/

[23] K. Bennett, Epistemic translation: towards an ecology of knowledges, Perspectives 34 (1) (2026) 16–31. doi:10.1080/0907676X.2023.2294123.
URL https://doi.org/10.1080/0907676X.2023.2294123

[24] Y. Assael, T. Sommerschield, B. Shillingford, M. Bordbar, J. Pavlopoulos, M. Chatzipanagiotou, I. Androutsopoulos, J. Prag, N. de Freitas, Restoring and attributing ancient texts using deep neural networks, Nature 603 (2022) 280–283.

[25] K. Papineni, S. Roukos, T. Ward, W.-J. Zhu, BLEU: a method for automatic evaluation of machine translation, in: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, 2002, pp. 311–318. doi:10.3115/1073083.1073135.
URL https://aclanthology.org/P02-1040/

[26] R. Rei, C. Stewart, A. C. Farinha, A. Lavie, COMET: A neural framework for MT evaluation, in: B. Webber, T. Cohn, Y. He, Y. Liu (Eds.), Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Online, 2020, pp. 2685–2702. doi:10.18653/v1/2020.emnlp-main.213.
URL https://aclanthology.org/2020.emnlp-main.213/

[27] T. Luong, I. Sutskever, Q. Le, O. Vinyals, W. Zaremba, Addressing the rare word problem in neural machine translation, in: C. Zong, M. Strube (Eds.), Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Association for Computational Linguistics, Beijing, China, 2015, pp. 11–19. doi:10.3115/v1/P15-1002.
URL https://aclanthology.org/P15-1002/

[28] R. Sennrich, B. Haddow, A. Birch, Neural machine translation of rare words with subword units, in: K. Erk, N. A. Smith (Eds.), Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Berlin, Germany, 2016, pp. 1715–1725. doi:10.18653/v1/P16-1162.
URL https://aclanthology.org/P16-1162/

[29] P. Koehn, R. Knowles, Six challenges for neural machine translation, in: Proceedings of the First Workshop on Neural Machine Translation, Association for Computational Linguistics, Vancouver, 2017, pp. 28–39. doi:10.18653/v1/W17-3204.

[30] F. Periti, S. Montanelli, Lexical semantic change through large language models: a survey, ACM Computing Surveys 56 (11) (2024) 1–38. doi:10.1145/3672393.
URL https://doi.org/10.1145/3672393

[31] A. Yüksel, B. Uğurlu, A. Koç, Semantic change detection with gaussian word embeddings, IEEE/ACM Transactions on Audio, Speech, and Language Processing 29 (2021) 3349–3361. doi:10.1109/TASLP.2021.3120645.

[32] X. Tang, Y. Zhou, T. Aida, P. Sen, D. Bollegala, Can word sense distribution detect semantic changes of words?, in: H. Bouamor, J. Pino, K. Bali (Eds.), Findings of the Association for Computational Linguistics: EMNLP 2023, Association for Computational Linguistics, Singapore, 2023, pp. 3575–3590. doi:10.18653/v1/2023.findings-emnlp.231.
URL https://aclanthology.org/2023.findings-emnlp.231/

[33] F. Periti, S. Picascia, S. Montanelli, A. Ferrara, N. Tahmasebi, Studying word meaning evolution through incremental semantic shift detection, Language Resources and Evaluation 59 (2025) 1363–1399. doi:10.1007/s10579-024-09769-1.

[34] P. Mandikal, Ancient wisdom, modern tools: Exploring retrieval-augmented LLMs for ancient indian philosophy, in: J. Pavlopoulos, T. Sommerschield, Y. Assael, S. Gordin, K. Cho, M. Passarotti, R. Sprugnoli, Y. Liu, B. Li, A. Anderson (Eds.), Proceedings of the 1st Workshop on Machine Learning for Ancient Languages (ML4AL 2024), Association for Computational Linguistics, Hybrid in Bangkok, Thailand and online, 2024, pp. 224–250. doi:10.18653/v1/2024.ml4al-1.23.
URL https://aclanthology.org/2024.ml4al-1.23/

[35] P. Wang, Z. Ren, The uncertainty-based retrieval framework for Ancient Chinese CWS and POS, in: R. Sprugnoli, M. Passarotti (Eds.), Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages, European Language Resources Association, Marseille, France, 2022, pp. 164–168.
URL https://aclanthology.org/2022.lt4hala-1.25/

[36] M. Fedorova, A. Kutuzov, Y. Scherrer, Definition generation for lexical semantic change detection, in: L.-W. Ku, A. Martins, V. Srikumar (Eds.), Findings of the Association for Computational Linguistics: ACL 2024, Association for Computational Linguistics, Bangkok, Thailand, 2024, pp. 5712–5724. doi:10.18653/v1/2024.findings-acl.339.
URL https://aclanthology.org/2024.findings-acl.339/

[37] S. Conia, D. Lee, M. Li, U. F. Minhas, S. Potdar, Y. Li, Towards cross-cultural machine translation with retrieval-augmented generation from

multilingual knowledge graphs, in: Y. Al-Onaizan, M. Bansal, Y.-N. Chen (Eds.), Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Miami, Florida, USA, 2024, pp. 16343–16360, outstanding Paper Award. doi:10.18653/v1/2024.emnlp-main.914.
URL https://aclanthology.org/2024.emnlp-main.914/

[38] C. C. Liu, I. Gurevych, A. Korhonen, Culturally aware and adapted NLP: A taxonomy and a survey of the state of the art, Transactions of the Association for Computational Linguistics 13 (2025) 652–689. doi:10.1162/tacl_a_00760.
URL https://aclanthology.org/2025.tacl-1.31/

[39] C. Li, M. Chen, J. Wang, S. Sitaram, X. Xie, CultureLLM: Incorporating cultural differences into large language models, in: A. Globerson, et al. (Eds.), Advances in Neural Information Processing Systems, Vol. 37, Curran Associates, Inc., 2024, pp. 84799–84838. doi:10.52202/079017-2693.

[40] C. Li, et al., CulturePark: Boosting cross-cultural understanding in large language models, in: A. Globerson, et al. (Eds.), Advances in Neural Information Processing Systems, Vol. 37, Curran Associates, Inc., 2024, pp. 65183–65216. doi:10.52202/079017-2082.

[41] B. Yao, M. Jiang, T. Bobinac, D. Yang, J. Hu, Benchmarking machine translation with cultural awareness, in: Y. Al-Onaizan, M. Bansal, Y.-N. Chen (Eds.), Findings of the Association for Computational Linguistics: EMNLP 2024, Association for Computational Linguistics, Miami, Florida, USA, 2024, pp. 13078–13096. doi:10.18653/v1/2024.findings-emnlp.765.
URL https://aclanthology.org/2024.findings-emnlp.765/

[42] Y. Nakatani, T. Kajiwara, T. Ninomiya, Comparing BERT-based reward functions for deep reinforcement learning in machine translation, in: Proceedings of the 9th Workshop on Asian Translation, International Conference on Computational Linguistics, Gyeongju, Republic of Korea, 2022, pp. 37–43.
URL https://aclanthology.org/2022.wat-1.2/

[43] K. Wang, X. Zhao, Y. Li, W. Peng, PROSE: A pronoun omission solution for Chinese-English spoken language translation, in: H. Bouamor, J. Pino, K. Bali (Eds.), Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Singapore, 2023, pp. 2297–2311. doi:10.18653/v1/2023.emnlp-main.141.
URL https://aclanthology.org/2023.emnlp-main.141/

[44] Q. Yin, Y. Zhang, W. Zhang, T. Liu, W. Y. Wang, Zero pronoun resolution with attention-based neural network, in: E. M. Bender, L. Derczynski, P. Isabelle (Eds.), Proceedings of the 27th International Conference on Computational Linguistics, Association for Computational Linguistics, Santa Fe, New Mexico, USA, 2018, pp. 13–23.
URL https://aclanthology.org/C18-1002/

[45] L. Wang, S. Liu, M. Xu, Y. Feng, Y. Zhang, A survey on zero pronoun translation, in: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, 2023, pp. 3321–3338. doi:10.18653/v1/2023.acl-long.187.
URL https://aclanthology.org/2023.acl-long.187/

[46] N. Skachkova, Multilingual coreference resolution as text generation, in: M. Ogrodniczuk, A. Nedoluzhko, M. Poesio, S. Pradhan, V. Ng (Eds.), Proceedings of the Seventh Workshop on Computational Models of Reference, Anaphora and Coreference, Association for Computational Linguistics, Miami, 2024, pp. 114–122. doi:10.18653/v1/2024.crac-1.11.
URL https://aclanthology.org/2024.crac-1.11/

[47] M. Müller, A. Rios, E. Voita, R. Sennrich, A large-scale test set for the evaluation of context-aware pronoun translation in neural machine translation, in: O. Bojar, R. Chatterjee, C. Federmann, M. Fishel, Y. Graham, B. Haddow, M. Huck, A. J. Yepes, P. Koehn, C. Monz, M. Negri, A. Névéol, M. Neves, M. Post, L. Specia, M. Turchi, K. Verspoor (Eds.), Proceedings of the Third Conference on Machine Translation: Research Papers, Association for Computational Linguistics, Brussels, Belgium, 2018, pp. 61–72. doi:10.18653/v1/W18-6307.
URL https://aclanthology.org/W18-6307/

[48] A. Mabona, L. Rimell, S. Clark, A. Vlachos, Neural generative rhetorical structure parsing, in: K. Inui, J. Jiang, V. Ng, X. Wan (Eds.), Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 2284–2295. doi:10.18653/v1/D19-1233.
URL https://aclanthology.org/D19-1233/

[49] X. Tan, L. Zhang, F. Kong, G. Zhou, Towards discourse-aware document-level neural machine translation, in: L. De Raedt (Ed.), Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22, International Joint Conferences on Artificial Intelligence Organization, 2022, pp. 4383–4389. doi:10.24963/ijcai.2022/608.

[50] X. Tan, L. Zhang, D. Xiong, G. Zhou, Hierarchical modeling of global context for document-level neural machine translation, in: K. Inui, J. Jiang, V. Ng, X. Wan (Eds.), Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 1576–1585. doi:10.18653/v1/D19-1168.
URL https://aclanthology.org/D19-1168/

[51] W. Liu, M. Strube, Annotation-inspired implicit discourse relation classification with auxiliary discourse connective generation, in: A. Rogers, J. Boyd-Graber, N. Okazaki (Eds.), Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Toronto, Canada, 2023, pp. 15696–15712. doi:10.18653/v1/2023.acl-long.874.
URL https://aclanthology.org/2023.acl-long.874/

[52] W. Xiang, B. Wang, A survey of implicit discourse relation recognition (2022). arXiv:2203.02982, doi:10.48550/arXiv.2203.02982.
URL https://arxiv.org/abs/2203.02982

[53] Y. E. Jiang, T. Liu, S. Ma, D. Zhang, M. Sachan, R. Cotterell, Discourse-centric evaluation of document-level machine translation with a new

densely annotated parallel corpus of novels, in: A. Rogers, J. Boyd-Graber, N. Okazaki (Eds.), Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Toronto, Canada, 2023, pp. 7853–7872. doi:10.18653/v1/2023.acl-long.435.
URL https://aclanthology.org/2023.acl-long.435/

[54] L. Wang, S. Liu, C. Lyu, W. Jiao, X. Wang, J. Xu, Z. Tu, Y. Gu, W. Chen, M. Wu, L. Zhou, P. Koehn, A. Way, Y. Yuan, Findings of the WMT 2024 shared task on discourse-level literary translation, in: B. Haddow, T. Kocmi, P. Koehn, C. Monz (Eds.), Proceedings of the Ninth Conference on Machine Translation, Association for Computational Linguistics, Miami, Florida, USA, 2024, pp. 699–700. doi:10.18653/v1/2024.wmt-1.58.
URL https://aclanthology.org/2024.wmt-1.58/

[55] Y. Li, J. Liang, X. Huang, Ancient chinese lexicon construction based on unsupervised algorithm of minimum entropy and CBDB optimization, in: Q. Zu, Y. Tang, V. Mladenović (Eds.), Human Centered Computing, Vol. 12634 of LNCS, Springer International Publishing, Cham, 2021, pp. 143–149. doi:10.1007/978-3-030-70626-5_15.

Appendix A. Authoritative Edition Sources

This appendix lists the authoritative source editions used for corpus construction.

Table A.13. Authoritative editions used for corpus construction

Work	Author	Publisher and edition	ISBN
Annotated Huangdi Neijing Translation	Guo Aichun	Guizhou Education Press, 1st ed., Aug 2010	978-7-5456-0125-1/R·25
Maijing Translation	Shen Yannan	People's Medical Publishing House, 1st ed., Mar 1993	7-117-01779-1/R·1780
Practical Jingui Yaolue Explanation	Liu Guantao, Liu Yisong, Shi Xiangqian	Military Medical Science Press, 1st ed., Aug 2005	978-7-8012-7134-0
Plain-Language Nanjing Interpretation	Wang Zhuxing	Tianjin Science and Technology Press, 1st ed., May 2010	978-7-5308-5361-0
Collated Plain-Language Shanghan Lun	Guo Aichun	China Press of Traditional Chinese Medicine, 1st ed., Nov 2012	978-7-5132-1153-6
Illustrated Plain-Language Bianque Xinshu	Zhang Jiawei	People's Medical Publishing House, 1st ed., May 2020	978-7-117-29946-6
Plain-Language Wenbing Tiaobian	Wu Yufeng, Kou Xinyun	China Medical Science and Technology Press, 1st ed., Aug 2016	978-7-5067-7869-5
Plain-Language Sisheng Xinyuan	Cheng Chuanhao	Henan Science and Technology Press, 1st ed., Jul 2020	978-7-5725-0187-6

Appendix B. Six-Step Reasoning Chain Example

This appendix provides a worked six-step reasoning-chain example for the pattern of channel-qi counterflow in the Foot Taiyang Bladder Channel.

Table B.14. Six-step reasoning chain example for "Foot Taiyang Bladder Channel qi-counterflow syndrome"

推理步骤	内容
1. 原文	厥，挟脊而痛者，至顶，头沉沉然，目荒荒然，腰脊强，取足太阳中血络。
2. 术语解析	厥：此处指经气厥逆、气机逆乱的病理状态。挟脊：沿着脊柱两侧，与足太阳膀胱经循行路线相关。沉沉然：形容头部沉重、昏蒙的感觉。目荒荒然：荒（*huāng*），视物不清（《玉篇·目部》：荒，目不明）。中：即委中穴，足太阳膀胱经合穴。
3. 医理推演	足太阳膀胱经循行于背脊、头顶及腰部。经气厥逆，气血壅滞于经脉，导致沿脊柱疼痛上达头顶；清阳不升，故头重昏沉；目赖膀胱经阳气上注，经气逆乱则目失所养，故视物不清；经气不利，筋脉拘急，故腰脊僵硬强直。治疗取本经合穴委中刺血，意在疏通经气、泻除瘀滞，使气血调和。
4. 句法结构	先陈述病证"厥"（病机），引出"挟脊而痛者，至顶，头沉沉然，目荒荒然，腰脊强"（症状序列），后陈述治法"取足太阳中血络"。整体结构为"病机+ 症状，后接治法"，逻辑关系清晰。
5. 文化注释	体现中医经络辨证与"上病下取"的治疗思想。症状全部出现在足太阳膀胱经循行部位，治疗选取该经远端合穴委中刺血，符合"经脉所过，主治所及"的原则。刺血疗法反映"菀陈则除之"的治则（《黄帝内经·针解》）。
6. 现代译文	经气厥逆，出现沿着脊柱两侧疼痛并上达头顶，头部沉重昏蒙，眼睛视物不清，腰部和脊柱僵硬。

Appendix C. Prompt Template for Six-Step Reasoning Translation

The following prompt template was used to elicit structured reasoning traces and final modern-Chinese translation output.

【角色与背景定位】
你是一位精通中医古籍与古汉语的专家，具备训诂学、临床中医知识与古今术语转换能力。

【总体任务要求】
当用户提供中医古籍原文时，先按六步在<think> 中完成分析，再在<answer> 中输出现代汉语译文。译文需准确、通顺，并保留医理与文化内涵。

【详细分析步骤及每步目的】

1. **[原文]**：列出用户提供的古籍原文，确立分析对象。
2. **[术语解析]**：解析核心术语的本义、语境义、古今差异与易误读点，确保用词准确。
3. **[医理推演]**：基于中医理论（阴阳五行、脏腑经络、气血津液、病因病机）推演原文医理；可参考现代医学，但以中医为主。
4. **[句法解构]**：处理倒装、互文、省略等结构，补足省略成分，明确句意关系。
5. **[文化注释]**：说明涉及的中医哲学与文化语境（如天人相应、整体观）及关键比喻含义。
6. **[现代译文]**：给出你生成的现代汉语译文。

【输入】 用户将提供{古文原文}。

【输出格式模板】
严格按以下格式输出，不要增删标签:

<think>
[原文]
{用户提供的古文原文}
[术语解析]
{核心术语的训诂、语境义、古今差异与易混淆点}
[医理推演]
{基于中医理论对原文生理、病理机制的解释}
[句法解构]
{古文句法结构与省略补全，明确句意}
[文化注释]
{原文的中医哲学、文化背景与象征意义}
[现代译文]
{你生成的现代汉语译文}
</think>
<answer>
你生成的现代汉语译文
</answer>

Appendix D. SFT Hyperparameters

Table D.15. SFT hyperparameters retained from training logs

Parameter	**Value**
Base model checkpoint	`Qwen/Qwen3-8B`
Model class	`Qwen3ForCausalLM`
Numeric precision	`torch.bfloat16`
Fine-tuning method	LoRA
LoRA rank r	32
LoRA alpha	64 ($= 2 \times r$)
Learning-rate scheduler	cosine
Num epochs	2
Batch size	8
Gradient accumulation steps	2
Initial learning rate	1.0×10^{-4}

Appendix E. GRPO Hyperparameters

Table E.16. GRPO training hyperparameters (appendix version)

Parameter	**Value**
Fine-tuning method	LoRA
LoRA rank r	32
LoRA alpha	64 ($= 2 \times r$)
Learning rate	5×10^{-5}
Weight decay	0.1
Warmup ratio	0.01
Scheduler	cosine
Optimizer	AdamW 8-bit
Logging steps	1
Temperature	0.95
`num_generations`	8
Prompts per device	4
Per-device train batch size	$32 = 8 \times 4$
Gradient accumulation steps	1
Max prompt length	1500
Max completion length	2596
Num train epochs	2
Save steps	200
Max grad norm	0.2

Appendix F. Detailed Word-Alignment Accuracy Data

This appendix reports detailed word-alignment translation accuracy across word-frequency bins and alignment categories for different models.

Table F.17. Word-alignment accuracy by word-frequency bin for reproducible open-source model settings. Frequency bins are estimated from the GRPO training corpus.

Frequency Bin	Count	0.6B-SFT	0.6B-SFT+GRPO	8B-SFT	8B-SFT+GRPO
1–2	2957	17.10	25.72	22.80	23.37
2–4	1972	18.00	26.99	24.71	25.32
4–8	2609	17.80	26.69	25.44	26.82
8–16	3281	18.84	26.80	27.95	29.33
16–32	4090	19.45	25.02	29.01	29.93
32–64	4655	24.19	32.75	35.06	35.59
64–128	4786	25.07	32.78	37.42	38.60
128–256	4134	29.44	37.95	41.72	42.61
256–512	2992	34.48	49.00	46.93	47.49
Low-freq avg (<16)	–	17.93	26.55	25.23	26.21
Mid-freq avg (16–512)	–	26.53	35.50	38.03	38.84
Overall (rare-word eval)	10335	19.78	28.09	25.98	27.50

Appendix G. Long-Sentence Translation Examples

To further illustrate the challenges of long-sentence translation, Table G.18 presents representative examples in the 40–55 character range. These sentences contain complex medical reasoning, multiple symptom descriptions, and conditional relations, requiring strong interpretive capacity from the model.

Table G.18. Long-sentence translation examples from 40 to 55 characters. Comparison of original text, reference translation, and R1-TCM-Translator-8B output.

Length	Original, Classical Chinese	Reference, Modern Chinese	Model Output
55 chars	太阳中者，发热恶寒，身重而疼痛，其脉弦细芤迟，小便已，洒洒然毛耸，手足逆冷，小有劳，身即热，口开，前板齿燥。	太阳中暑证，出现发热，怕冷，身体沉重疼痛，脉象弦细芤迟，解了小便后，就毛骨悚然、怕冷更甚，手足冰凉，稍微劳动，身体就发热，口就张开呼吸，门齿干燥。	太阳中暑的病人，发热恶寒，身体沉重而疼痛，其脉弦细芤迟，小便之后，浑身发冷，毛竖耸立，手足逆冷，稍微活动一下，身体就发热，张口呼吸，前牙干燥。
43 chars	伤寒脉浮，医以火迫劫之，亡阳，必惊狂，起卧不安者，桂枝去芍药加蜀漆牡蛎龙骨救逆汤主之。	太阳伤寒证，脉象浮，本应当发汗解表，医生却用火治法强迫发汗，导致心阳外亡、神气浮越，出现惊恐狂乱、坐卧不安的，用桂枝去芍药加蜀漆牡蛎龙骨救逆汤主治。	患伤寒病，脉象浮，医生用火法强行发汗，以致亡失阳气，病人必然出现惊狂，坐卧不安等症状，用桂枝去芍药加蜀漆牡蛎龙骨救逆汤主治。
44 chars	伤寒八九日，下之，胸满烦惊，小便不利，谵语，一身尽重，不可转侧者，柴胡加龙骨牡蛎汤主之。	外感病八九天，误用攻下，出现胸部满闷、烦躁惊惕不安、小便不通畅，谵语、全身沉重、不能转侧的，用柴胡加龙骨牡蛎汤主治。	患伤寒病八九天，误用攻下后，出现胸胁胀满，心烦惊悸，小便不通利，胡言乱语，身体沉重，不能转侧的，用柴胡加龙骨牡蛎汤主治。
40 chars	伤寒大下后，复发汗，心下痞，恶寒者，表未解也，不可攻痞，当先解表，表解乃可攻痞。	伤寒表证，用峻泻药攻下后，再发其汗，导致心下痞塞，如果有发热畏寒等见证的，是表证还未解除，不能先泄热消痞，而应先解表，表证解除以后才能泄热消痞。	患伤寒病，曾用峻猛泻法治疗，又再发汗，出现心下痞塞，怕冷的，这是表邪未解，不可攻心下痞，应当先解表，表邪解除后才可以攻心下痞。

From Table G.18, we observe the following patterns:

- **Accurate capture of multiple symptoms**: for the 55-character sentence, the model correctly renders several coordinated symptoms and preserves source completeness.
- **Understanding of therapeutic logic**: in sentences such as "表未解也，不可攻痞，当先解表", the model correctly handles procedural order and causality.
- **Preservation of formula names**: the model retains complex formula names with high accuracy.
- **Remaining challenges**: some fine-grained expressions still differ from the reference; for example, "洒洒然毛耸" is rendered with slight semantic variation, reflecting persistent ambiguity in classical-text interpretation.

These examples indicate that, although R1-TCM-Translator has made clear progress on long-sentence translation, very long sentences with complex therapeutic logic and multi-clause structure still require further improvement in deep reasoning.