句读前置与多智能体协同的古籍文白翻译方法

张蕊[1]，邵明凡[2]，何彦青[1*]
(1.中国科学技术信息研究所 北京 100038; 2.山东齐鲁书社出版有限公司 山东 济南 250003)

摘要 [目的]针对无标点古籍文本中句读边界模糊性、古今语言的语义鸿沟、平行语料噪声分布不均以及单一阶段生成模型缺乏自校验机制等挑战，本研究旨在构建一种高保真翻译质量的古文—白话文翻译框架。**[方法]**本研究提出了结合句读预处理与多智能体协同机制的 ACMT 模型。利用独立的句读专项模型将无标点古文转换为结构化文本，以此降低后续任务的长距离依赖复杂度,通过由“初译”、“深度思考”与“参考反思”三个智能体构成的协作闭环，执行“翻译—推敲—校验”的迭代优化过程。**[结果]**在人工构建的高质量文白翻译评测集上，ACMT 模型生成的译文在自动评价指标（ROUGE、BLEU）和人工评价指标（忠实度、可读性及知识正确性维度）中均优于基准对比模型，消融实验证实了各模块的有效性。**[结论]**句读前置策略提升了译文与参考文本在词项、短语及序列结构层面的一致性。多智能体协同的迭代译校机制有效地保证了各项自动化指标的稳定提升。

关键词：古籍文白翻译；自动句读；大语言模型；多智能体

中图分类号：TP391　　**文献标识码**：A

Ancient Chinese-to-Modern Chinese Translation with Punctuation Restoration and Multi-Agent Collaboration

ZHANG Rui[1], SHAO Mingfan[2], HE Yanqing[1*]
(1. Institute of Scientific and Technical Information of China, Beijing, China 100038; 2. Shandong Qilu Press Company Limited, Shandong Jinan 250003)

Abstract [Objective] Addressing challenges such as the ambiguity of sentence boundaries in unpunctuated ancient texts, the semantic gap between ancient and modern languages, the uneven distribution of noise in parallel corpora, and the lack of self-verification mechanisms in single-stage generative models, this study aims to construct a high-fidelity translation framework for Classical-to-Modern Chinese. **[Methods]** This study proposes the ACMT model, which integrates punctuation preprocessing with a multi-agent collaborative mechanism. An independent, specialized sentence segmentation model is utilized to convert unpunctuated Classical Chinese into structured text, thereby mitigating the complexity of long-distance dependencies in downstream tasks. Furthermore, a collaborative loop comprising three agents “Initial Translator,” “Deep Thinker,” and “Reference Reflector” is established to execute an iterative optimization process of “translation, refinement, verification.” **[Results]** On a manually curated, high-quality evaluation dataset for Classical-to-Modern Chinese translation, the ACMT model significantly outperforms baseline models across both automatic metrics (ROUGE, BLEU) and human evaluation metrics (faithfulness, readability, and knowledge correctness). Ablation studies further confirm the effectiveness of each proposed module. **[Conclusion]** The pre-punctuation strategy significantly enhances consistency between the generated translations and reference texts at the lexical, phrasal, and sequential structural levels. Additionally, the multi-agent collaborative iterative translation and verification mechanism effectively ensures the stable improvement of various automated evaluation metrics.

Keywords: Ancient Chinese-to-Modern Chinese translation; punctuation restoration; large language model; multi-agent collaboration

收稿日期：2026-09-05　　录用日期：2026-10-19
基金项目：山东出版传媒股份有限公司融合创新发展项目中华武术典籍知识服务平台建设(一期);山东省数字融合出版创新创业共同体科技计划项目中华武术典籍知识服务关键技术研究及平台建设（一期）;中国科学技术信息研究所重点工作 (项目编号：ZD2026-09, ZD2026-10)
*通信作者：heyq@istic.ac.cn

1 引言

古籍文白翻译旨在把无标点或少标点的古文转化为带现代标点的原文及流畅、准确的现代白话译文，是古籍数字化、文言文教学和数字人文研究中的基础任务。与通用机器翻译相比，该任务不仅需要完成古今语言转换，还需要识别句读边界、解析文言句法、辨析古今词义，并结合历史文化语境处理专名和典故。因此，构建兼具语言理解、知识储备和稳定生成能力的文白翻译模型，对于降低古籍整理成本、扩大传统文化传播范围具有重要意义。

现有研究采用注释信息知识增强的方式[1]、重叠滑动窗口机制[2]、多维审校[3]等方法增强文白翻译技术的性能。然而，仍存在许多问题：（1）公开古文—白话平行语料分散、重复且质量参差，限制了领域知识学习和指令对齐；（2）古文无标点、文言语法深奥且高度依赖语境，端到端直译会使句读错误级联为误译，且错误来源难以定位；（3）单一模型一次生成缺少针对专名、典故、虚词和特殊句式的复核机制。其直译往往停留在字词层面的直接映射，未能解析典故、隐喻等深层文化概念；过度显化则因阐释冗余而损害译文的简洁性与可读性。如何在形式对应与概念明晰之间求得平衡，构成该任务的核心挑战。

为了解决上述问题，本文提出句读前置与多智能体协同的古籍文白翻译模型 ACMT，从领域知识显式利用、文化解释控制和多维质量评价的角度优化方法。首先，以独立句读模块显式生成可检查的带标点文本，再完成古今语义转换，并通过初翻、深度思考和参考反思三个智能体形成“翻译—推敲—校验”闭环。本文的主要贡献包括：①构建分级古文—白话平行语料，采用继续预训练与 LoRA 指令微调相结合的训练策略；②构建“先句读、后翻译”的两级解耦框架，使句读错误可定位、可独立修正，并设计多智能体迭代译校机制，提高译文准确性与可读性；③经过实验验证，本文模型的性能优于基线模型。

2 国内外相关研究

2.1 古籍智能处理与文白翻译研究

国内古籍智能处理研究经历了由规则和统计方法向神经网络、预训练语言模型、领域大模型演进的过程。早期方法依赖词典、句法规则和人工特征，在特定语料上具有一定可解释性，但面对词类活用、古今词义迁移和跨时代表达时泛化能力有限。领域大模型较通用基座模型具有更强的古文处理能力[4]。Xunzi 系列模型进一步通过古今混合语料继续预训练和指令微调开展古籍翻译研究。何佳怡等研究人员[1]以注释信息知识增强的方式，采用 LoRA 参数高效微调策略提出一种中医古籍今译模型（TCM Classic-Translator）。郑小英等研究人员[2]采用重叠滑动窗口机制，使用句子位置感知编码技术捕获篇章结构特征，从而提出一种针对东巴古籍的基于上下文感知的翻译框架。李家欢等研究人员[5]提出一种一个词典释义选择-融合的二阶段机器翻译框架，其采用词典注释资源来增强译文质量。宋熹玥等研究人员[6]提出一种基于注意力机制的 Seq2Seq 模型的中医古籍智能机器翻译模型。吴梦成等研究人员[7]提出一种基于神经机器翻译 OpenNMT 模型，将词性知识、词边界知识、实体知识和依存句法知识融入翻译模型。然而，单一模型方法采用的直译往往仅保留了表面形式，而忽略了潜在的概念解析，而过度明确则损害了译文的简洁性和易读性，因此，如何在这两者之间实现有效权衡，构成了该任务的核心挑战。

2.2 机器翻译与多智能体译校研究

多智能体协同为复杂的翻译任务提供了新的思路。TransAgents[8]模拟翻译机构中的编辑、译者、本地化和校对角色，将长篇文学翻译分解为准备、翻译、润色和质量检查等环节。在迭代推理方面，Wang 等研究人员[9]提出一种 DRT 模型，利用译者、顾问和评价者三个智能体生成长思维链翻译数据，使模型学习文学翻译中的多轮推敲过程。Di 等研究人员[10]提出面向术语约束翻译的 AIDAterm，以分析、翻译、后编辑和审校四类智能体组成顺序流水线，并通过消融实验验证术语预

分析、词表注入和推理审校的作用。因此，角色分工、批评反馈和多轮修订能够弥补单一模型一次生成缺少复核的不足，尤其适合文化负载词、上下文一致性和文体表达要求较高的文本。为此，多智能体应运而生。

近年来，面向古籍翻译领域，基于多智能体的相关工作被提出。Lv 等研究人员[3]提出面向文言翻译的多智能体框架，将词级释义、段落生成和多维审校结合，并引入关键词释义库、检索增强生成与迭代反馈。Chen 等研究人员[11]构建古典诗歌英译基准 PoetMT，从准确、流畅和雅致三个维度评价大语言模型，并提出融合诗歌知识的检索增强翻译方法 RAT。Jiao 等研究人员[12]构建《诗经》中—英—德逐行平行语料及古语词汇知识库，并提出结合知识规则与大模型裁判的混合评价框架。He 等研究人员[13]提出多智能体文化感知翻译框架 MACAT，通过文化负载词检测、知识注入、候选重排和多轮评价处理中医典籍与《论语》中的隐含文化知识。然而，对于原始无标点古文，若缺少可检查的句读中间结果，模型仍难以区分边界识别错误与语义转换错误，也不利于针对不同朝代、文体和专业领域进行模块化迭代，为文白翻译技术带来了挑战。

2.3 古籍翻译数据集

无论是基于 Transformer 架构的专用翻译模型，还是基于大语言模型（LLM）的方法，现有文白翻译技术在训练阶段均深度依赖于规模充足、对齐精准的平行语料。Yang 等研究人员[14]构建了一个古今汉语句级平行语料库，并提出一种 Guwen-BERT 的翻译方法。宋熹玥等研究人员[15]经过筛选代表性古籍、数字化处理 、古籍原文与译文对齐等环节，构建了一个大规模中医古籍古今翻译平行语料库。然而，现有研究表明[16]，现有 LLM 由于其文言文语料覆盖率不足，导致其语义解析和语境推理性能不稳定，且现有古籍领域专用模型仍然存在较大波动性。

为缓解古文标注数据不足的问题，国内研究开始利用大规模古文单语语料进行领域预训练。An-chiBERT[17]在古文理解、古今翻译及古诗生成等任务中表现良好。SikuBERT 和 SikuRoBERTa[18]利用《四库全书》语料开展预训练，并应用于分词、断句标点、词性标注和命名实体识别等任务。然而，这类模型多为仅编码器（encoder-only）架构，其预训练目标本质上服务于古文的理解类判别任务，与文白翻译所需的条件生成能力之间存在目标错位；同时，古文单语预训练仅在源语言内部强化表示，并未显式建模古今语言之间的语义对应，因此，翻译性能所能获得的增益有限。

3 方法

3.1 任务定义

古籍文白翻译以无标点的原始古文为输入，对文本中的句读边界、句法结构、语义逻辑和上下文关联进行解析，自动添加现代标点符号，并在此基础上将文言语句转换为流畅、准确的现代白话文，最终形成"带标点古文＋现代白话译文"对照、可直接服务于阅读、教学与研究的结构化结果。

3.2 总体框架

ACMT 采用“先句读、后翻译”的两阶段，将文白翻译任务解耦为标点恢复与译文生成两个子任务，分别由两个独立的大语言模型承担。在推理流程上，古文句读模型首先对原始无标点文本执行标点恢复，生成句法边界明确、结构层次清晰的带标点文本，作为供下游阶段表示的中间表征；多智能体协同文白翻译模型随后以该中间表征为条件，经多角色分工协作生成忠实、流畅的白话译文。为此，句读前置为翻译阶段提供显式的句子边界与结构先验，使模型能够更准确地建模句内成分关系与句间逻辑衔接，而且将两类异质任务解耦可规避联合建模中的目标冲突，并有效抑制断句歧义经级联路径向下游译文的传导与放大。ACMT 的整体框架如图 1 所示。

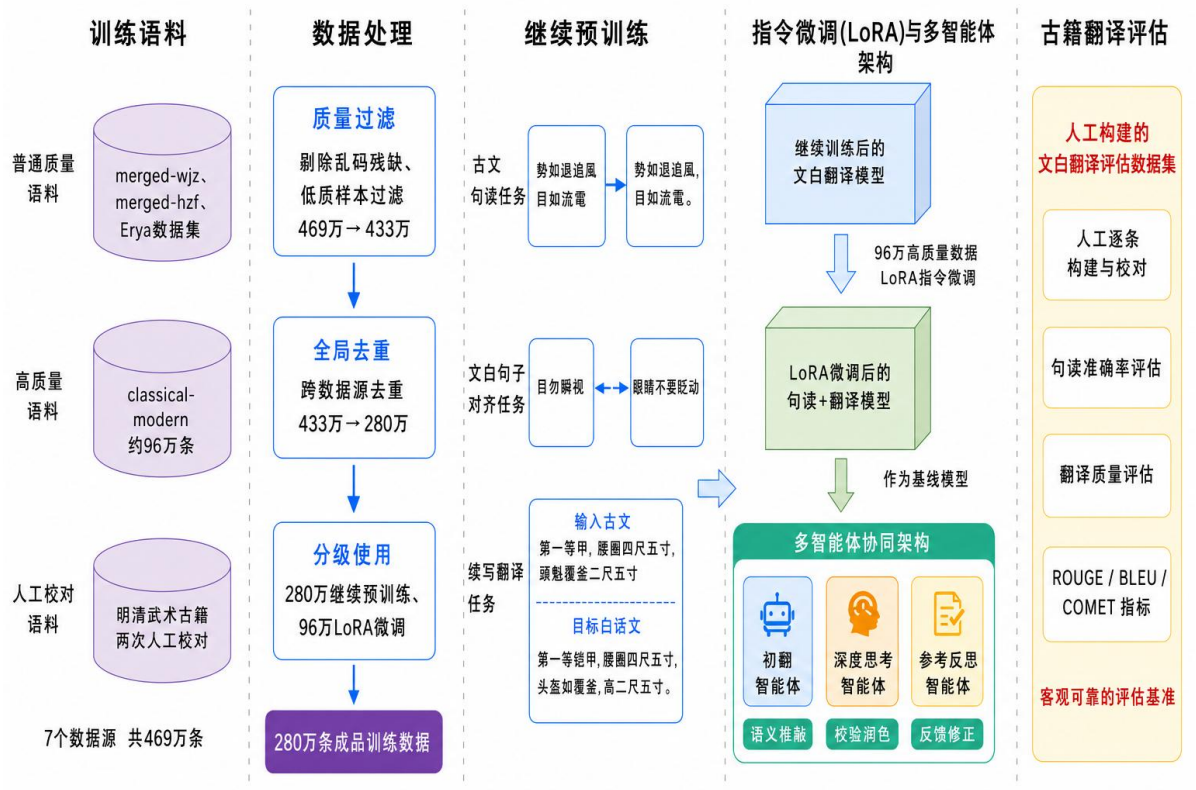


图 1　ACMT 模型总体架构

Fig.1 Overall framework of ACMT model

3.3 自动句读模型

古文自动句读模块通过在句读标注语料上的监督微调，使模型学习古文的句法结构、语义逻辑与篇章级上下文依赖，进而以自回归生成的方式，为输入的原始无标点古文预测并插入逗号、句号、问号等现代标点符号。原始无标点古文为 x，句读模型 $\mathcal{M}_{\mathrm{j}}$ 生成带标点文本 $p = \mathcal{M}_{\mathrm{j}}(x)$，该结果作为中间表示传递至下游翻译阶段。其显式输出中间结果 p，使研究者能够在翻译前检查句读边界，并在发现错误时仅重训或替换句读模型，避免端到端架构中错误来源难以辨识的问题。

3.4 多智能体协同的文白翻译模型

在微调后的模型之上设计三个专业化智能体，各司其职、紧密协作，构成"翻译—推敲—校验"的闭环：

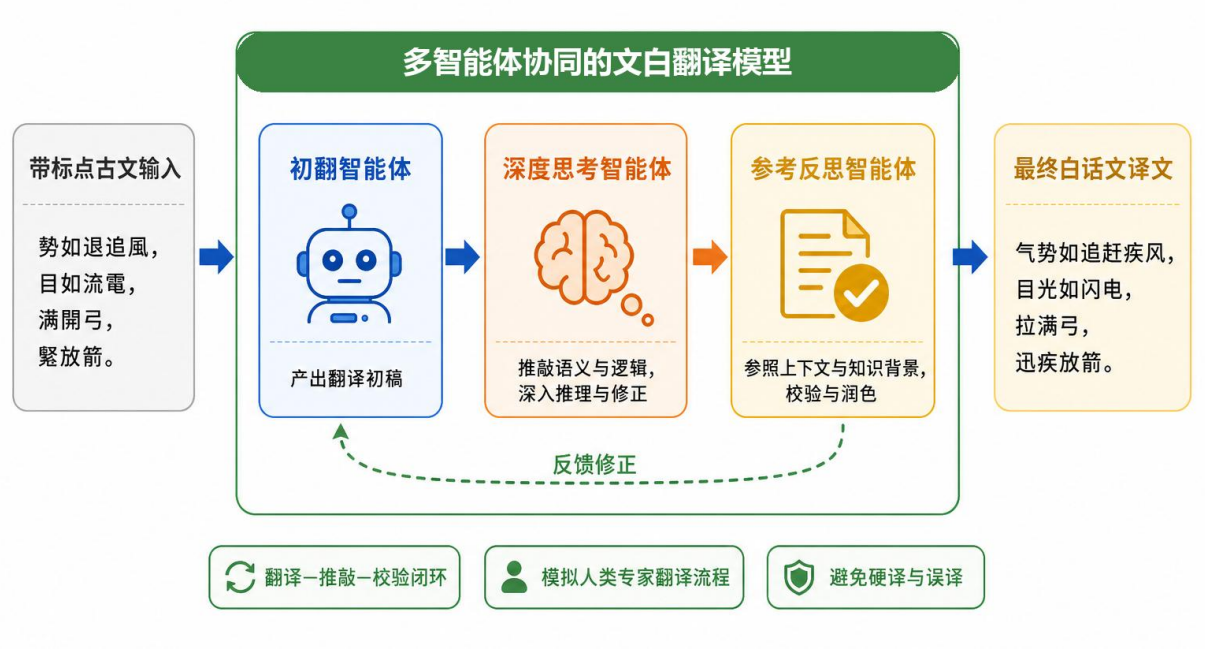


图 2　多智能体协同的文白翻译模型

Fig.2 A Multi-Agent Cooperative Model for Ancient Chinese-to-Modern Chinese Translation

表 1　多智能体角色及其职责

Tab.1 Multi-agent roles and responsibilities

智能体	职责
初翻智能体	完成古文到白话文的初步翻译
深度思考智能体	对初稿进行语义逻辑层面的深入推理与修正
参考反思智能体	参照上下文与知识背景进行反思、校验与润色，并将问题反馈前端修正

具体而言，初翻智能体生成候选译文；深度思考智能体检查句法关系、词义和逻辑连贯性；参考反思智能体结合上下文与知识背景复核专名、典故及表达风格，并将问题反馈给前序环节。最终结果由多轮修订后的候选译文确定。值得一提的是，多智能体协作采用固定上限与质量阈值相结合的策略来终止迭代：最大迭代轮数设为 N（如 N=3），即"翻译—推敲—校验"最多执行 N 轮；若某一轮深度思考未发现任何问题，则提前终止。此外，每轮迭代产生一个候选译文，系统保留全部候选。最终，优先选取深度思考未发现问题的候选；若不存在，则选取问题计数最少的候选；若问题计数相同，则选取最后一轮的候选，以确保信息的充分性。

因此，得益于前置句读处理，翻译模型能更清晰地理解句子边界和逻辑关系，提升准确性与可读性，避免因断句错误导致的"硬译"或"误译"。

3.5 两阶段训练策略

本文采用“领域自适应继续预训练 → 参数高效指令微调”的两阶段训练范式：第一阶段通过多任务继续预训练向基础模型注入古汉语领域的语言知识，第二阶段以低秩适配方式习得指令遵循能力。

阶段一：多任务领域自适应继续预训练

在经严格清洗与去重后的 280 万条全量古文语料上，构造以下生成式任务的混合训练，以最小分布偏移的方式将古文领域知识注入模型参数：

(1) 古文句读任务。以无标点文言文x为输入、恢复句读标注的文本y为目标，建模条件分布$p_\theta(y \mid x)$。该任务要求模型在缺乏显式分隔信号的条件下习得文言句法边界与语义停顿的内在规律，本质上等价于面向古汉语的序列切分与标点重建，同时为下游翻译提供高质量的句级切分基础。

(2) 文白句子对齐任务。在其中 180 万条子集上构建“文言文—白话”平行对，并施加对齐目标，显式建模两种语体在词汇、句法与语义层面的对应关系，弥合跨语体转换的表示鸿沟。

(3) 续写翻译任务。以文言片段及其上文c为条件，令模型自回归生成目标白话译文，即最大化$p_\theta(y \mid x, c)$，将翻译建模为上下文条件下的受限文本生成，使模型学习篇章级文白转换能力。

阶段二：基于 LoRA 的参数高效指令微调

继续预训练后的模型虽具备领域知识，但对指令化任务范式的遵循能力不足。为此，从全量语料中筛选质量最高的 classical–modern 平行子集，辅以人工校验的高质量样本，共约 96 万条，统一改写为“指令—响应”格式的监督数据，覆盖句读、翻译等核心任务。

鉴于全参数微调的高显存开销与灾难性遗忘风险，本文采用低秩适配方法 LoRA 进行微调。仅需存储与更新极低秩矩阵，降低梯度与优化器状态的显存占用；推理时增量矩阵可与原权重合并，不引入额外延迟，以习得指令遵循能力的同时保留模型的通用语言能力。

4 实验与结果分析

4.1 数据集

模型训练依托由 7 个数据源汇集而成的古文—白话平行语料：原始数据共 469 万条，经逐数据集质量过滤（剔除乱码、残缺段落与低质句对）后保留 433 万条，再经跨源全局去重最终沉淀 280 万条成品训练数据，避免训练数据污染。

表 2　所用数据集概况

Tab.2 Overview of Datasets

数据集	过滤后数量	领域	质量
merged-wjz	347,262	古文、诗歌	中
zhushi-wjz	4,063	注释	良
merged-hzf	1,235,284	古文	良
erya-dataset1	169,242	古籍	优
erya-dataset2	1,623,427	古籍	优
classical-modern	939,257	古籍	优
明清古籍（武术类）	14,363	武术古籍	优（经两次人工校对）
合计（过滤后）	4,332,898	—	含跨源重复样本
合计（去重后）	2,803,098	—	成品训练数据

数据处理完成后，各阶段训练数据如下：

表 3　数据集详情

Tab.3 Details of the Statistics of Datasets

		训练集	验证集	测试集	共计
文白翻译	普通质量	1,677,457	186,384	408	1,864,249
	高质量	865,332	93,925		939,665
	武术类	11,850	1,000		13,258
句读生成		2,760,757	21,182	21,159	2,803,098

语料治理遵循“先源内清洗、后跨源去重、再按质量分级使用”的流程。普通质量数据主要扩展领域覆盖，高质量数据用于强化指令对齐，武术类人工校对数据用于补充专业领域表达。

4.2 实验设置与评价指标

本研究选取了涵盖不同参数量级与架构系列的 8 个基线模型进行对比。其中，古籍翻译基线（Xunzi-Qwen2-7B[19]、TongGu-7B-Instruct[20]、PoetMT[11]）；监督微调基线则囊括了主流开源大语言模型，不仅包含 Llama 系列（chinese-llama-2-7b-16k[21]）、Qwen 系列[22]（Qwen1.5-14B-Chat、Qwen1.5-MoE-A2.7B-Chat、QwQ-32B）、DeepSeek 系列[23]（DeepSeek-R1-Distill-Qwen-32B），以及 Transformer 模型[24]。评价指标采用 ROUGE-1、ROUGE-2、ROUGE-L[25]、BLEU-4[26]与 COMET[27]，分别衡量词项重合、短语匹配、最长公共子序列、n 元组精确度及基于预训练模型的语义质量。

此外，ACMT 模型基于 Qwen3-32B 基座模型进行实验。自动句读模型和多智能体协同的文白翻译模型各自以张量并行方式分配至 2 张 GPU，共需 4 张 80GB 级 GPU（如 NVIDIA A100 80GB 或 H100 80GB），单卡供电能力不低于 400W，并配备强力风冷或水冷散热；两个模型部署于相互隔离的 GPU 集群，确保高负载下流水线互不抢占、稳定顺畅运行。

4.3 实验结果

自动评价结果见表 4。根据实验结果表明，ACMT 模型在五项指标上均取得最高值，以验证模型的有效性。

表 4　ACMT 模型的性能

Tab.4 The results of the ACMT model

类型	基座模型	方法	ROUGE-1	ROUGE-2	ROUGE-L	BLEU-4	COMET
通用模型	Transformer	SFT	/	/	/	20.94	48.8066
	Qwen1.5-14B-Chat		39.6577	19.0041	38.3500	21.7294	47.2539
	Qwen1.5-MoE-A2.7B-Chat		40.7722	19.3264	39.5908	21.4303	46.7740
	chinese-llama-2-7b-16k		16.7981	5.3437	7.6016	3.1064	37.5533
	DeepSeek-R1-Distill-Qwen-32B		28.5668	13.9292	25.8866	14.5547	45.9530
	QwQ-32B		9.9124	2.6048	4.2259	1.3948	46.4827
专用模型	Qwen2-7B	荀子大模型（Xunzi-Qwen2-7B ）	28.6902	11.5236	18.5014	9.1391	46.3199
	Baichuan2-7B-Base	TongGu-7B-Instruct	27.0139	9.8068	25.744	10.4616	39.2697
	Qwen3-32B	PoetMT	42.6488	21.0292	41.6897	22.1770	48.2565
	Qwen3-32B	**ACMT**	**72.0286**	**59.1951**	**71.8685**	**60.7218**	**49.0941**

4.4 人工评价

为弥补自动评价指标难以充分判断专名、典故、史实和译文风格的局限，本文设置专家盲评。按照古籍年代、文体和篇章长度从测试集中分层抽取样本共 300 条，隐去模型名称并随机排列不同系统的输出，由五位古籍整理和古汉语专家独立评分，取平均分作为最终分数。

评价维度包括：①忠实度，即译文是否完整、准确地传达原文语义；②可读性，即译文是否符合现代汉语表达习惯且行文连贯；③知识正确性，即专名、典故和历史事实是否解释准确；各维度分值范围为 1（最差）到 5（最优），对分歧较大的样本进行复核。实验结果如表 5 所示。

表 5 人工评价结果

Tab.5 Human evaluation results

专用模型	忠实度	可读性	知识正确性
荀子大模型（Xunzi-Qwen2-7B）	3.3893	3.2973	3.386
TongGu-7B-Instruct	3.314	3.2313	3.356
PoetMT	3.3766	3.2166	3.3966
ACMT	**3.4953**	**3.492**	**3.498**

根据实验结果显示，ACMT 模型在忠实度、可读性和知识正确性三个维度上的分数均高于现有古籍翻译模型的分数。其中，可读性维度的增幅最大，说明句读前置与多智能体推敲、复核有助于改善译文的连贯性和现代汉语表达；忠实度与知识正确性得分的同步提高，则表明 ACMT 模型在保持原文语义和处理专名、典故及历史文化知识方面具有较好的描述性优势。

4.5 消融实验

为进一步验证 ACMT 各模块的贡献程度，本文进行了一系列消融实验，通过移除句读前置模型（w/o judou）和多智能体结构（w/o agent）来量化其对整体性能的影响。其中，去掉句读前置模块（w/o judou）旨在验证句读前置模块对于辅助生成译文的作用，去掉多智能体结构（w/o agent）旨在验证多智能体结构的重要性。实验结果如表 6 所示。

表 6 消融实验结果

Tab.6 The results of the ablation study

专用模型	ROUGE-1	ROUGE-2	ROUGE-L	BLEU-4	COMET
Xunzi-Qwen2-7B	28.6902	11.5236	18.5014	9.1391	46.3199
w/o judou	49.7767	26.2569	49.4733	29.25626	47.3014
w/o agent	55.7032	34.0303	55.0785	38.624	48.9457
ACMT	**72.0286**	**59.1951**	**71.8685**	**60.7218**	**49.0941**

消融结果表明，句读前置与多智能体译校均能改善 ACMT 性能。与 w/o judou 相比，ACMT 模型的 ROUGE-1、ROUGE-2、ROUGE-L 和 BLEU-4 分别提高 22.2519、32.9382、22.3952 和 31.4655 个点，COMET 由 47.3014 提高至 49.0941，增幅为 1.7927 个点。五项指标均有提升，说明显式句读不仅增强了译文与参考文本在词项、短语和序列结构上的一致性，也对语义评价指标产生了正向作用。与 w/o agent 相比，ACMT 模型的 ROUGE-1、ROUGE-2、ROUGE-L、BLEU-4 和 COMET 分别提高 16.3254、25.1648、16.7900、22.0978 和 0.1484 个点，说明多智能体推敲与复核对译文质量具有稳定的正向作用，其中对词面匹配和结构完整性的改善尤为明显。

5 结论

面对海量未经整理的无标点古籍文献，需要高效可靠的数字化整理手段，将其准确地翻译为白话文。然而，现有方法在断句与翻译的联合处理上仍存在精度瓶颈与领域适配困难。为了解决上述问题，本文提出了句读前置与多智能体协同的古籍文白翻译模型 ACMT。该模型将无标点古文的句读分析与现代白话翻译解耦，利用多源分级语料开展继续预训练和 LoRA 指令微调，并以三个专业

化智能体完成初译、推敲与复核。经过实验验证，ACMT 在统一评估集上的五项自动指标和三项人工评价指标均优于现有模型。消融实验证实了句读前置对词面、结构一致性及语义评价指标均产生了正向作用，多智能体迭代译校具有稳定效果。因此，本文模型能够将古籍文本翻译为高保真的白话文。

参考文献：

[1]何佳怡,谢佳东,李芹,等. 融合监督微调和知识增强的中医古籍今译方法研究[J]. 世界科学技术-中医药现代化,2026,28(6):2043-2051.

[2]郑小英,刘文涛,周爱民,等. 基于上下文位置感知的东巴古籍文本翻译[J]. 中文信息学报,2026,40(2):71-81.

[3] Lv W, Cao Q, Liu X. A Multi Agent Classical Chinese Translation Method Based on Large Language Models[J]. Scientific Reports, 2025, 15: 40160.

[4] Zhao Z, Sun G, Liu C, et al. Research on Machine Translation of Ancient Books in the Era of Large Language Model[J]. npj Heritage Science, 2025, 13: 122.

[5]李家欢,吴若纯,黄书剑,等. 词典释义增强的古文机器翻译[J]. 中文信息学报,2025,39(4):85-95.

[6]宋熹玥,周净,刘伟. 中医古籍智能机器翻译模型构建研究[J]. 中国中医药图书情报杂志,2024,48(6):130-135.

[7]吴梦成,林立涛,吴娜,等. 融合不同语义知识的中国古代典籍机器翻译研究[J]. 情报资料工作,2024,45(2):97-104.

[8] Wu M, Xu J, Yuan Y, et al. (Perhaps) Beyond Human Translation: Harnessing Multi-Agent Collaboration for Translating Ultra-Long Literary Texts[J]. Transactions of the Association for Computational Linguistics, 2025, 13: 901-922.

[9] Wang J, Meng F, Liang Y, et al. DRT: Deep Reasoning Translation via Long Chain-of-Thought[C]//Findings of the Association for Computational Linguistics: ACL 2025. 2025: 6770-6782.

[10] Di Rosa E. Multi-Agent Orchestration for Terminology-Constrained Machine Translation in Industrial Localization[C]//Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics: Industry Track. 2026: 917-926.

[11] Chen A, Lou L, Chen K, et al. Benchmarking LLMs for Translating Classical Chinese Poetry: Evaluating Adequacy, Fluency, and Elegance[C]//Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing. 2025.

[12] Jiao Y, Sun M. From Shijing to English and German: Resources and Evaluation for LLM Translation of Early Chinese Poetry[C]//Findings of the Association for Computational Linguistics: ACL 2026. 2026.

[13] He X, Lan K, You M, et al. Worlds Within Words: Translating Culture in Ancient Chinese Texts with Multi-Agent Coordination[EB/OL]. arXiv:2606.01276, 2026.

[14]Yang, Z., Chen, Kj., Chen, J. (2021). Guwen-UNILM: Machine Translation Between Ancient and Modern Chinese Based on Pre-Trained Models. In: Wang, L., Feng, Y., Hong, Y., He, R. (eds) Natural Language Processing and Chinese Computing. NLPCC 2021. Lecture Notes in Computer Science(), vol 13028. Springer, Cham.

[15]宋熹玥,邹荣辉,刘伟. 大规模中医古籍古今翻译平行语料库的构建和应用[J]. 中国数字医学,2025,20(5):98-103.

[16]文玉锋,林伟杰,夏翠娟,等. 面向古籍文献智能处理的大语言模型效能测评[J]. 图书馆论坛,2025,45(8):52-60.

[17] Tian H, Yang K, Liu D, et al. AnchiBERT: A Pre-Trained Model for Ancient Chinese Language Understanding and Generation[C]//2021 International Joint Conference on Neural Networks (IJCNN). IEEE, 2021: 1-8.

[18]王东波,刘畅,朱子赫,等.SikuBERT 与 SikuRoBERTa：面向数字人文的《四库全书》预训练模型构建及应用研究[J].图书馆论坛,2022,42(6):31-43.

[19]荀子大语言模型团队. Xunzi-Qwen2-7B[EB/OL]. [2026-08-20] https://modelscope.cn/models/Xunzillm4cc/Xunzi-Qwen2-7B.

[20] Cao, J., Peng, D., Zhang, P., et al (2024). TongGu: Mastering Classical Chinese Understanding with Knowledge-Grounded Large Language Models. EMNLP 2024.

[21] Cui, Y., Yang, Z., & Yao, X. (2023). Efficient and Effective Text Encoding for Chinese LLaMA and Alpaca. arXiv preprint arXiv:2304.08177.

[22] Qwen Team. (2023). Qwen Technical Report.

[23] Guo, D., Yang, D., Zhang, H. et al. DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning. Nature 645, 633–638 (2025).

[24]Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need[C]//Advances in Neural Information Processing Systems. 2017, 30: 5998-6008.

[25] Lin C Y. ROUGE: A Package for Automatic Evaluation of Summaries[C]//Text Summarization Branches Out. 2004: 74-81.

[26] Papineni K, Roukos S, Ward T, et al. BLEU: A Method for Automatic Evaluation of Machine Translation[C]//Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. 2002: 311-318.

[27] Rei R, Stewart C, Farinha A C, et al. COMET: A Neural Framework for MT Evaluation[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. 2020: 2685-2702.