

藏汉篇章机器翻译研究及语料库构建

田佳乐¹, 江静¹, 李亚超^{1,2,*}

¹西北民族大学 语言与文化计算教育部重点实验室, 730030

²西北民族大学 青藏高原人文环境数据智能实验室, 730030

y242430342@stu.xbmu.edu.cn, jiangj@xbmu.edu.cn, liyc@xbmu.edu.cn

摘要

篇章机器翻译旨在使用计算机将源语言篇章自动翻译为具有相同语义的目标语言篇章, 是机器翻译的前沿研究热点。相对于传统的句子级翻译, 以篇章作为翻译单位, 模型能够更有效地利用上下文信息, 提升翻译的一致性与连贯性, 具有广阔的应用前景和研究价值。与资源丰富语言(如汉语、英语、法语等)机器翻译研究相比, 藏语机器翻译资源稀缺, 公开可用的数据集数量有限, 在篇章级机器翻译方面的探索尚无公开论文发表。鉴于此, 本文首先构建一个藏汉翻译数据集, 标注了句子级、段落级和篇章级的边界, 为藏汉篇章翻译任务提供高质量的多粒度标注数据集。然后, 本文基于该数据集研究了藏汉篇章机器翻译, 并对比机器翻译在句子层面、段落层面和篇章层面翻译效果的差异。本文对所构建的藏汉篇章翻译语料库予以开源, 希望能推动相关研究的发展。链接: <https://github.com/liyc7711/tb-zh-mt>。

关键词: 藏汉篇章机器翻译; 语料库构建; 大语言模型

A Study on Tibetan-Chinese Document-Level Machine Translation and Corpus Construction

Jiale Tian¹, Jing Jiang¹, Yachao Li^{1,2,*}

¹The Key Laboratory of Linguistic and Cultural Computing of Ministry of Education, Northwest Minzu University, 730030

²Data Intelligence Laboratory of Tibetan Plateau Humanistic Environment, Northwest Minzu University, 730030

y242430342@stu.xbmu.edu.cn, jiangj@xbmu.edu.cn, liyc@xbmu.edu.cn

Abstract

Document-level machine translation (MT) aims to automatically translate a source-language document into a target-language document with equivalent semantics using computational methods. As a cutting-edge topic in MT research, document-level translation enables models to utilize broader contextual information compared to traditional sentence-level approaches, thereby improving translation consistency and coherence. This makes it particularly promising in terms of practical applications and research value. In contrast to MT studies involving resource-rich languages such as Chinese, English, and French, Tibetan MT faces significant resource scarcity, with very limited publicly available datasets. To date, no published work has systematically explored Tibetan-Chinese document-level MT. To address this gap, this study

* 为通讯作者

本文得到国家自然科学基金(62266038)、中央高校基本科研业务费专项资金(31920230122)、西北民族大学兰山人才工程(xbmulsr202415)和引进人才科研项目(xbmuyjrc202424)等资助。

©2025 中国计算语言学大会

根据《Creative Commons Attribution 4.0 International License》许可出版

constructs a Tibetan–Chinese translation dataset with annotated boundaries at the sentence, paragraph, and document levels, providing high-quality, multi-granularity aligned data for document-level translation research. Based on this dataset, we conduct experiments on Tibetan–Chinese document-level MT and compare the translation quality at the sentence, paragraph, and document levels. The constructed dataset has been made publicly available to support and advance further research in this area. The link is: <https://github.com/liyc7711/tb-zh-mt>.

Keywords: Tibetan-Chinese Document-level Translation , Corpus Construction , Large Language Models

1 引言

机器翻译旨在利用计算机实现不同语言之间的自动转换，是促进不同语言间信息交流的重要技术。近年来，随着神经机器翻译技术(Sutskever et al., 2014)的突破，汉语、英语、法语等资源丰富语言之间的翻译质量取得了显著提升。然而，主流研究多集中于句子级翻译，忽略了篇章内上下文信息的建模，导致在指代消解、语义连贯、篇章结构保持等方面仍存在明显不足(吕星林 et al., 2024)。篇章机器翻译通过引入跨句上下文信息，不仅能够提升句子的翻译质量，更能在整篇译文中实现语义的一致性和结构的完整性(亢晓勉 and 宗成庆, 2022)。从引入额外上下文信息实现篇章翻译(Wang et al., 2017; Tu et al., 2018)，到篇章到篇章的直接翻译(Bao et al., 2021; Li et al., 2023)，篇章机器翻译研究得到了极大的发展。与此同时，以ChatGPT为代表的大语言模型在机器翻译任务中表现出强大的跨语言迁移能力与上下文理解能力(Wang et al., 2023)，相比于传统的端到端神经机器翻译系统，大语言模型突破了一般篇章的输入长度限制，拥有更强的长文本建模能力，在篇章机器翻译任务上表现尤为显著。

藏汉机器翻译相关研究目前主要集中在句子级翻译(李亚超 et al., 2017; 慈祯嘉措 et al., 2019; 沙九 et al., 2020; 慈祯嘉措 et al., 2020; 沙九 et al., 2021; 桑杰端珠 and 才让加, 2023)。藏语作为一种资源稀缺语言，在机器翻译领域的研究起步较晚，缺乏高质量、细粒度的篇章级数据资源。这种局限性使得藏语机器翻译的研究难以向篇章级翻译扩展。

本文针对上述问题系统开展了藏汉篇章机器翻译的研究工作。首先从网络文本中收集藏汉双语语料，并对原始语料进行清洗、筛选与结构化处理，构建了一个多粒度的藏汉篇章翻译语料库，其中训练集来自新闻领域，标注了段落级与篇章级边界；测试集来自新闻、畜牧养殖、农业种植、医疗健康和科普教育五个不同领域，标注了句子级、段落级和篇章级边界。此外，本文基于该语料库在不同大语言模型上实现藏汉篇章翻译，采用了两种不同的微调策略，并设计了段落级和篇章级微调指令。实验结果表明，通过该语料库微调的大模型，在特定领域的翻译质量得到了显著的提升。

本文的主要贡献为：

(1) 本文构建了一个包含6769个篇章、48910个段落、约11万个句子的高质量新闻领域藏汉翻译训练集，标注段落级与篇章级的边界，为低资源的藏汉篇章机器翻译研究提供了重要的数据基础。根据已公开发表的工作，该数据集是第一个规模较大的藏汉篇章翻译数据集。

(2) 本文构建了一个大规模的覆盖新闻、畜牧养殖、农业种植、医疗健康和科普教育五个不同领域的多领域测试集，共包含19570个句子，并对该测试集进行了句子对齐、段落对齐和篇章对齐，为藏汉机器翻译系统的综合翻译能力评估提供了高质量、细粒度的数据支持。

(3) 本文基于构建的藏汉语料库，采用两种不同微调策略，进行了基于大语言模型微调实验，实现了高质量藏汉篇章机器翻译。实验中采用了段落级和篇章级指令微调，并系统对比分析了不同大语言模型的翻译性能，并在多领域语料上进行了广泛测试。实验结果显示，篇章翻译能够有效提升翻译质量。

2 相关工作

本文从三个方面讨论相关工作：篇章机器翻译、基于大语言模型的机器翻译和藏汉机器翻译。

2.1 篇章机器翻译

传统的篇章级神经机器翻译研究主要集中在源侧局部上下文扩展与整体文档建模两大思路之上。早期工作通过将当前句子与其前后若干句拼接输入，使模型能够利用更多源端信息来改善翻译质量。例如，Tiedemann and Scherrer(2017)提出简单的多句拼接策略，Bawden et al.(2018)、Zhang et al.(2018)与Voita et al.(2018)在Transformer中分别引入多头上下文注意力以赋予不同句子可学习的权重，Maruf and Haffari(2018)、Maruf et al.(2019)则通过记忆网络缓存历史句子特征来辅助生成，之后Tan et al.(2021)、Lyu et al.(2021)及Xu et al.(2021)又进一步探索了跨句联合训练与动态上下文窗口策略，尽管这些方法在源侧上下文建模上取得了显著进展，但对目标侧译文历史的利用仍显不足(Li et al., 2023)。与此同时，另一类研究直接将整篇文档作为输入单元进行端到端翻译。Tiedemann 和Scherrer(2017)、Agrawal et al.(2018)、Ma et al.(2020)以及Zhang et al.(2020)率先通过将多句简单串联来实现篇章到篇章翻译，Junczys-Dowmunt(2019)、Lupo et al.(2022)与Sun et al.(2022)则进一步利用大规模增强数据集对原始Transformer进行篇章级训练，显著提升长文本的整体连贯性；Liu et al.(2020)则借助预训练语言模型（如BERT、mBART）在文档级语料上微调，以实现译文一致性与流畅度的双重优化。为更好地捕捉篇章级的长距离依赖，最新研究在自注意力机制中引入了结构感知与位置感知策略：Bao et al.(2021)通过层次化注意力融合句内与句间信息，Li et al.(2023)则提出位置感知注意力，通过动态选择上下文信息，从而进一步提升译文在一致性和连贯性方面的表现。根据已公开的工作，本文首次系统开展了藏汉篇章机器翻译研究。

2.2 基于大语言模型的机器翻译

基于大语言模型的机器翻译研究主要可分为两大方向：一是“提示学习”(prompt-based learning)，二是“模型微调”(fine-tuning)。在提示学习范式下，系统无需更新模型参数，只需借助设计精巧的自然语言提示或示例，便能让大语言模型产生符合要求的译文。Hendy(2023)等人系统评估了ChatGPT、GPT3.5、GPT4和text-davinci002在18个翻译方向上的性能，并分析了提示策略对翻译质量的影响，证明了提示设计对翻译质量的关键作用；Lyu et al.(2024)提出了一种解码增强的多阶段提示微调方法(DeMPT)，有效增强了大语言模型对于句子信息的理解，提高了大语言模型的翻译质量。另一种思路则是对大语言模型进行有针对性的微调，以便更好地适应特定领域或语言对。Schioppa et al.(2023)通过持续预训练提升了大语言模型的零样本翻译能力；Li et al.(2024a)在翻译指令层面开展微调，同样获得了显著的性能改善。微调可采用“全量更新”策略，也可以利用LoRA等参数高效技术，仅调整少量低秩矩阵以保留预训练知识并降低资源消耗(Hu et al., 2022)。Reinauer et al.(2023)的对比实验表明，相较于提示学习，微调方法在领域适配上更为稳定，且几乎不依赖于上下文示例或后处理手段。不过，微调往往需要大量的计算资源投入，且在特定任务上对模型进行过度微调可能削弱模型在其他领域的泛化能力。

2.3 藏汉机器翻译

藏汉机器翻译作为一种典型的低资源翻译任务，到目前为止，针对藏语到汉语的机器翻译工作相对较少。随着神经机器翻译技术的不断发展，李亚超等人(2017)开展了基于注意力机制的神经网络机器翻译实验，针对藏汉平行语料匮乏的问题，引入迁移学习方法以提升模型性能。针对藏汉平行语料稀缺的问题，慈祯嘉措等人(2019)则提出将藏语单语语言模型融合至神经网络翻译系统的方法，显著提升了翻译质量。沙九等人(2020)从藏语翻译中的粒度问题出发，提出融合音节、词语及其组合的多粒度训练方法，并在解码器中引入自注意力机制，显著提升了神经机器翻译效果。慈祯嘉措等人(2020)通过迭代式回译策略，为低资源条件下提升机器翻译质量提供了有效路径。桑杰端珠和才让加(2023)提出了一种基于词典注入的预训练方法，通过利用双语词典和混合语言现象，提高了藏汉机器翻译模型的性能。这些研究为提升藏汉机器翻译的质量和性能提供了宝贵的经验和理论支持。与这些工作不同，本文首次开展藏汉篇章机器翻译相关研究，促进藏汉翻译进一步迈向实用。

3 藏汉篇章机器翻译研究

3.1 语料库构建

本数据集的原始数据来源于国内权威的藏汉双语网站，该网站包含了涵盖多个领域的中文

与藏文平行语料。为了构建高质量的藏汉篇章语料库，本文通过抓取该藏汉网站上的藏汉对应语料，采用基于时间轴与网页结构匹配的篇章对齐、基于BLEU得分的段落规则匹配与人工校正、以及借助大语言模型进行句子切分与核对等对齐策略，最终实现了语料库训练集的句子对齐和段落对齐，测试集的句子对齐、段落对齐和篇章对齐。

原始语料清洗与过滤 原始语料主要来源于权威的藏汉双语网站，这些语料虽具备较高的语言质量，但是由于网站的排版问题导致的藏汉语料之间出现的部分不对应问题，本文进行了清洗与过滤，剔除了网页格式混乱、内容缺失或对齐错误的样本。

训练集构建 对于训练集，本文主要进行段落级和篇章级对齐。对于篇章对齐，本文首先通过抓取语料发布时间和网站链接进行排序匹配，通过确保藏语篇章与汉语篇章在时间轴和网页结构上的一一对应，实现了藏汉篇章的初步对齐。随后，采用了基于机器翻译的自动评估指标—BLEU值，来进行翻译文本与原始文本的匹配，并通过设定阈值筛选具有更高对齐质量的藏汉篇章数据，最后经过人工校对从而完成了篇章级的对齐；对于段落对齐，本文同样采用了基于BLEU值的规则匹配方法，通过计算汉语和藏语文本在翻译内容上的相似度，初步实现了段落对齐。随后，对初步对齐的结果进行人工校正，确保了每一段的内容在两种语言间的精确对应，从而达到高质量的段落对齐效果。训练集在对齐后按一定比例分为了训练集和验证集，训练数据的详细统计信息如表1所示。

	大小	Token数	段落数	篇章数
训练集	81.07M	3497.8万	46818	6487
验证集	3.5M	150.8万	2092	282

表 1. 训练数据的详细统计信息

测试集构建 对于测试集，本文基于切分好的段落进行更细化的对齐，实现了句子级、段落级与篇章级的多粒度对齐，并覆盖了新闻、畜牧养殖、农业种植、医疗健康、科普教育五个领域，构建了一个大规模的多粒度多领域测试集，其中新闻领域测试集与构建的训练集领域一致。对于句子对齐，本文采用了DeepSeek大语言模型进行段落的句子切分。通过设计特定的句子切分命令，DeepSeek模型能够较为准确地识别出每个段落中的句子边界，并对其进行切分标注。切分后的结果经过人工校对，确保句子在中文与藏文之间的对应关系准确无误。最终，经过多轮人工修正与优化，达到了准确的句子对齐。段落级与篇章级测试集的对齐方式分别延用了训练集中对应粒度的对齐策略。各领域测试集的详细统计信息如表2所示。

领域	大小	Token数	句子数	段落数	篇章数
新闻	4.30M	187.1万	4218	2135	295
农业种植	2.50M	110.3万	5027	2823	292
畜牧养殖	2.41M	106.1万	4072	2170	174
医疗健康	1.84M	78.8万	3390	2357	222
科普教育	1.83M	79.5万	2863	1629	167
总计	12.88M	561.8万	19570	11114	1150

表 2. 各领域测试集的详细统计信息

3.2 指令构建

为了增强大语言模型理解机器翻译任务的能力，本文设计了两种微调指令。本文的指令微调格式遵循Jiao et al.(2023)的训练数据格式，参照Li et al.(2024b)的构造策略。

间的一一对应关系恢复与评估。

对于使用这种相同结构来构造段落级微调指令和篇章级微调指令的方法，可以在保持实验一致性的情况下，支持模型的多粒度训练。

3.3 模型微调方法

LoRA微调 LoRA微调是一种参数高效微调方法，旨在在不修改原始预训练模型大部分参数的前提下，实现对大语言模型的高效适配。LoRA的核心思想是在模型中的部分权重矩阵 W 引入两个低秩可训练矩阵，通过低秩重构的方式学习微调所需的增量更新。具体而言，LoRA将原始全连接层的权重更新形式从直接训练 $W \in \mathbb{R}^{d \times k}$ ，替代为：

$$W' = W + \Delta W$$

其中 $\Delta W = A \cdot B$ ， $A \in \mathbb{R}^{r \times k}$ ， $B \in \mathbb{R}^{d \times r}$ ， $r \ll \min(d, k)$ 表示低秩的秩参数， A 和 B 是唯一可训练的矩阵，而原始模型参数 W 保持冻结状态。这种方式大幅减少了训练参数数量，同时保持了模型的表达能力和迁移性能。

本文在藏汉翻译场景中采用LoRA微调方法，不仅显著降低了训练资源需求，还通过LoRA微调的参数控制能力，实现对模型行为的精确调节，从而在保证效率的同时，具有一定的翻译质量。虽然在本文中LoRA微调方法在一定程度上提升了篇章翻译质量，但相较于下文所提出的部分参数微调方法，LoRA微调在藏汉篇章翻译任务中表现较弱。

部分参数微调 在本文前期实验中发现基于LoRA微调的模型翻译质量较弱，实验的详细结果见表3，借鉴Li et al.(2024b)的微调方法，本文提出了一种部分参数微调方法。部分参数微调是一种通过仅更新预训练模型中选定参数子集的方式进行任务适配的方法。

在此策略中，设模型包含 L 个层，参数集合为：

$$\Theta = \{\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(L)}\}$$

选定一组索引 $I \subset \{1, 2, \dots, L\}$ ，仅对其中的层 $\theta^{(l)}$ ($l \in I$) 进行参数更新，其余层保持不变。按层选择的微调策略通过仅训练预训练模型中少数关键层，显著降低参数开销，同时保留原模型语义能力，提升适配效率与稳定性。本文的具体的选层策略是在最底层、中间层和最高层中的各选取一层进行微调，其余层参数保持冻结。

部分参数微调在藏汉篇章翻译任务中兼顾了参数效率、模型性能与适配能力，是探索基于大语言模型的低资源藏汉篇章翻译中表现的一种有效方法。

4 主要实验结果

4.1 实验设置

实验数据 在本文实验中，实验数据采用本文构建的藏汉翻译数据集。选用其中的训练数据作为训练集，并使用相同新闻领域的测试集作为主要实验测试集。

为了系统评估藏汉篇章机器翻译在跨领域条件下的泛化能力，实验采用本文构建的一个大规模的藏汉机器翻译测试集。具体使用畜牧养殖、农业种植、医疗健康和科普教育四个不同领域对模型的泛化能力进行评测。

模型设置 本文基于Transformers框架与DeepSpeed分布式训练技术进行模型训练，以提升显存利用效率与训练稳定性。实验选取Llama-3.1-8B-Instruct与Qwen-2.5-7B-Instruct作为基线模型，并分别采用LoRA微调和部分参数微调两种策略进行实验对比。在LoRA微调中，将LoRA的秩参数 r 设置为8，缩放因子 α 设置为16，并将LoRA作用于注意力模块的Query与Value子模块，以实现高效的低秩参数适应。在部分参数微调中，固定除部分Transformer层外的所有参数，仅更新部分层参数以实现轻量化的训练方案：对Llama-3.1-8B-Instruct，选取第0层、第15层与第31层进行微调；对Qwen-2.5-7B-Instruct，选取第0层、第13层与第27层进行微调，其余层保持冻结状态。

翻译质量评估 在模型性能评估中，本文采用BLEU(Liu et al., 2020)和COMET(Rei et al., 2020)指标来衡量模型在藏语到汉语翻译任务上的翻译表现，其中COMET分数的计算使用了wmt22-comet-da模型。在汉语到藏语的翻译任务上只采用BLEU指标来衡量模型翻译表现。对于中文的分词选用了Jieba工具包进行分词，对于藏文的分词则是使用了西北民族大学所开源的藏文分词工具TIP-LAS(李亚超et al., 2015)。

4.2 实验结果与分析

本文实验结果部分展示了基于所构建的藏汉语料库，采用不同大语言模型进行藏汉翻译任务的翻译性能表现。实验分别从句子级、段落级和篇章级三个层次进行系统对比分析，以全面评估各模型在藏汉篇章翻译任务中的实际表现。为了进一步验证模型的翻译能力，实验引入了当前主流大语言模型Qwen与Llama的两个蒸馏版本—DeepSeek-R1-Distill-Qwen-14B与DeepSeek-R1-Distill-Llama-8B，以评估其在藏汉篇章翻译任务中的性能。所有实验结果详见表3，表中三种BLEU指标（s-BLEU，p-BLEU，d-BLEU）对应的COMET指标分别反映句子级、段落级和篇章级的翻译质量，以揭示其在实际藏汉翻译任务中的表现差异。

模型	s-BLEU	s-COMET	p-BLEU	p-COMET	d-BLEU	d-COMET
Deepseek-R1-Distill-Llama-8B	0.21	0.2198	0.23	0.2281	0.31	0.2686
Deepseek-R1-Distill-Qwen-14B	0.20	0.2143	0.28	0.2566	0.36	0.2775
Qwen-2.5-7B-Instruct-LoRA	8.68	0.6708	8.81	0.7012	9.13	0.7107
Llama-3.1-8B-Instruct-LoRA	34.55	0.8245	34.18	0.8386	33.34	0.8343
Qwen-2.5-7B-Instruct	51.60	0.8572	53.98	0.8675	59.41	0.8881
Llama-3.1-8B-Instruct	60.63	0.8760	61.71	0.8824	61.53	0.8794

表 3. 各大语言模型实验结果表

从表3的实验结果看出，采用两种微调策略后，Llama-3.1-8B-Instruct与Qwen-2.5-7B-Instruct模型的翻译性能均明显优于蒸馏模型DeepSeek-R1-Distill-Qwen-14B与DeepSeek-R1-Distill-Llama-8B，后者在BLEU和COMET等评估指标上得分偏低。经过对译文的分析发现其存在指令理解偏差、语言输出混乱等问题，尚难以满足藏汉翻译任务的实际需求。

在采用LoRA微调策略后，Llama-3.1-8B-Instruct-LoRA与Qwen-2.5-7B-Instruct-LoRA模型的翻译性能明显优于蒸馏模型，但远远落后于采用部分参数微调策略后的Llama-3.1-8B-Instruct与Qwen-2.5-7B-Instruct模型。其中，Qwen-2.5-7B-Instruct-LoRA模型提升较小，在所有BLEU指标上均未超过10，但是篇章级的BLEU与COMET分数仍优于段落级与句子级，表明LoRA微调策略虽未能显著提升其在藏汉翻译任务中的表现，但篇章翻译模型仍能从篇章级信息中受益。相较之下，Llama-3.1-8B-Instruct-LoRA模型的翻译性能提升明显，BLEU维持在约33的水平，具有一定的翻译能力，但与部分参数微调后的Llama-3.1-8B-Instruct仍有较大差距，说明LoRA微调策略能有效提升该模型在藏汉篇章翻译中的表现，但部分参数微调策略更具优势。

在采用部分参数微调策略后，Llama-3.1-8B-Instruct与Qwen-2.5-7B-Instruct模型均展现出良好的适应能力，尤其是Llama-3.1-8B-Instruct，其三项BLEU指标均超过60分，且d-COMET达到0.8794，显示出强大的藏汉篇章翻译能力。同时，篇章级和段落级的BLEU和COMET分数优于句子级，尽管篇章级与段落级在指标上的差异不大，但从译文分析来看，篇章翻译在上下文衔接、术语一致性等方面表现更优，进一步验证了篇章信息在提升藏汉翻译质量方面的重要作用。Qwen-2.5-7B-Instruct同样表现良好，尤其在d-COMET上达到0.8881，略高于Llama，表明其在语义一致性方面具有一定优势，并且篇章级的BLEU和COMET在分数上优于段落级与句子级，证明了篇章信息的有效性。实验结果不仅从侧面验证了所构建语料库的价值，也说明部分参数微调在提升大模型翻译性能方面具有良好效果。

领域	s-BLEU	s-COMET	p-BLEU	p-COMET	d-BLEU	d-COMET
农业种植	10.85	0.7042	10.41	0.7300	10.35	0.7342
畜牧养殖	8.51	0.6604	8.38	0.6789	8.21	0.6546
医疗健康	9.88	0.6910	9.77	0.7030	9.42	0.7255
科普教育	22.87	0.7740	23.31	0.7920	23.41	0.7989

表 4. 多领域机器翻译实验结果

本文还进一步评估了部分参数微调后的Llama-3.1-8B-Instruct模型在跨领域任务中的泛化

- Yachao Li, Junhui Li, Jing Jiang, Shimin Tao, Hao Yang, and Min Zhang. 2023. P-transformer: Towards better document-to-document neural machine translation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 31:3859–3870.
- Jiahuan Li, Hao Zhou, Shujian Huang, Shanbo Cheng, and Jiajun Chen. 2024a. Eliciting the translation ability of large language models via multilingual finetuning with translation instructions. *Transactions of the Association for Computational Linguistics*, 12:576–592.
- Yachao Li, Junhui Li, Jing Jiang, and Min Zhang. 2024b. Enhancing document-level translation of large language model via translation mixed-instructions. *arXiv preprint arXiv:2401.08088*.
- Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. 2020. Multilingual denoising pre-training for neural machine translation. *Transactions of the Association for Computational Linguistics*, 8:726–742.
- Lorenzo Lupo, Marco Dinarelli, and Laurent Besacier. 2022. Focused concatenation for context-aware neural machine translation. In *Proceedings of the Seventh Conference on Machine Translation (WMT)*, pages 830–842.
- Xinglin Lyu, Junhui Li, Zhengxian Gong, and Min Zhang. 2021. Encouraging lexical translation consistency for document-level neural machine translation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 3265–3277.
- Xinglin Lyu, Junhui Li, Yanqing Zhao, Min Zhang, Daimeng Wei, Shimin Tao, and Hao Yang. 2024. Dempt: Decoding-enhanced multi-phase prompt tuning for making llms be better context-aware translators. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 20280–20295.
- Shuming Ma, Dongdong Zhang, and Ming Zhou. 2020. A simple and effective unified encoder for document-level machine translation. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 3505–3511.
- Sameen Maruf and Gholamreza Haffari. 2018. Document context neural machine translation with memory networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1275–1284.
- Sameen Maruf, André FT Martins, and Gholamreza Haffari. 2019. Selective attention for context-aware neural machine translation. In *North American Association for Computational Linguistics 2019*, pages 3092–3102. Association for Computational Linguistics (ACL).
- Ricardo Rei, Craig Stewart, Ana C Farinha, and Alon Lavie. 2020. Comet: A neural framework for mt evaluation. *arXiv preprint arXiv:2009.09025*.
- Raphael Reinauer, Patrick Simianer, Kaden Uhlig, Johannes EM Mosig, and Joern Wuebker. 2023. Neural machine translation models can learn to be few-shot learners. *arXiv preprint arXiv:2309.08590*.
- Andrea Schioppa, Xavier Garcia, and Orhan Firat. 2023. Cross-lingual supervision improves large language models pre-training. *arXiv preprint arXiv:2305.11778*.
- Zewei Sun, Mingxuan Wang, Hao Zhou, Chengqi Zhao, Shujian Huang, Jiajun Chen, and Lei Li. 2022. Rethinking document-level neural machine translation. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 3537–3548.
- Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, 27.
- Xin Tan, Longyin Zhang, and Guodong Zhou. 2021. Coupling context modeling with zero pronoun recovering for document-level natural language generation. In *Proceedings of the 2021 conference on empirical methods in natural language processing*, pages 2530–2540.
- Jörg Tiedemann and Yves Scherrer. 2017. Neural machine translation with extended context. In *Proceedings of the Third Workshop on Discourse in Machine Translation*, pages 82–92.
- Zhaopeng Tu, Yang Liu, Shuming Shi, and Tong Zhang. 2018. Learning to remember translation history with a continuous cache. *Transactions of the Association for Computational Linguistics*, 6:407–420.

- Elena Voita, Pavel Serdyukov, Rico Sennrich, and Ivan Titov. 2018. Context-aware neural machine translation learns anaphora resolution. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1264–1274.
- Longyue Wang, Zhaopeng Tu, Andy Way, and Qun Liu. 2017. Exploiting cross-sentence context for neural machine translation. *arXiv preprint arXiv:1704.04347*.
- Longyue Wang, Chenyang Lyu, Tianbo Ji, Zhirui Zhang, Dian Yu, Shuming Shi, and Zhaopeng Tu. 2023. Document-level machine translation with large language models. *arXiv preprint arXiv:2304.02210*.
- Mingzhou Xu, Liangyou Li, Derek F Wong, Qun Liu, and Lidia S Chao. 2021. Document graph for neural machine translation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 8435–8448.
- Jiacheng Zhang, Huanbo Luan, Maosong Sun, Feifei Zhai, Jingfang Xu, Min Zhang, and Yang Liu. 2018. Improving the transformer translation model with document-level context. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 533–542.
- Pei Zhang, Boxing Chen, Niyu Ge, and Kai Fan. 2020. Long-short term masking transformer: A simple but effective baseline for document-level neural machine translation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1081–1087.
- 亢晓勉 and 宗成庆. 2022. 基于篇章结构多任务学习的神经机器翻译. *软件学报*, 33(10):3806–3818.
- 吕星林, 李军辉, 陶仕敏, 杨浩, and 张民. 2024. 文档级神经机器翻译综述. *软件学报*, pages 1–33.
- 慈禛嘉措, 桑杰端珠, 孙茂松, 色差甲, and 周毛先. 2019. 融合单语语言模型的藏汉机器翻译方法研究. *中文信息学报*, 33(12):61–66.
- 慈禛嘉措, 桑杰端珠, 孙茂松, 周毛先, and 色差甲. 2020. 基于迭代式回译策略的藏汉机器翻译方法研究. *中文信息学报*, 34(11):67–73+83.
- 李亚超, 江静, 加羊吉, and 于洪志. 2015. Tip-las: 一个开源的藏文分词词性标注系统. *中文信息学报*, 29(6):203–207.
- 李亚超, 熊德意, 张民, 江静, 马宁, and 殷建民. 2017. 藏汉神经网络机器翻译研究. *中文信息学报*, 31(6):103–109.
- 桑杰端珠 and 才让加. 2023. 基于词典注入的藏汉机器翻译模型预训练方法. *中文信息学报*, 37(08):43–51.
- 沙九, 冯冲, 张天夫, 郭宇航, and 刘芳. 2020. 多策略切分粒度的藏汉双向神经机器翻译研究. *厦门大学学报: 自然科学版*, 59(2):213–219.
- 沙九, 冯冲, 周鹭琴, 李洪政, 张天夫, and 慧慧. 2021. 面向司法领域的高质量开源藏汉平行语料库构建. *中文信息学报*, 35(11):51–59.