多领域翻译中语义消歧的话题方向盘方法

满志博[1,2]，**张玉洁**[1,2†]，**陈圆梦**[1,2]，**陈钰枫**[1,2]，**徐金安**[1,2]
[1]北京交通大学，交通大数据与人工智能教育部重点实验室，北京，100044
[2]北京交通大学，计算机科学与技术学院，北京，100044
{zhiboman,yjzhang,yuanmengchen,jaxu,chenyf}@bjtu.edu.cn

摘要

近年，大语言模型（Large Language Models, LLMs）在通用文本翻译任务上的翻译质量取得大幅的提升，但是在面对多领域文本时，翻译质量呈现明显下降。如何利用有限的领域双语平行语料增强领域翻译知识成为主要的研究目标，已有方法大多使用人为设置的领域标签学习语义表示，导致其在消歧知识的获取上受到限制，如何构建有效的消歧知识成为一种挑战。为此，本文提出一种多领域翻译中语义消歧的话题方向盘方法，旨在增强大语言模型在多领域上的语义消歧能力，具体包括：**（1）基于话题模型的语义表示获取机制**：我们首先利用ETM自动聚类算法获取细小颗粒度的话题语义表示用于之后构建消歧知识，这种话题的表示更贴近语义，也更适合作为语义单元来构建语义表示。然后，我们设计TopicModel函数将大模型的表示转换成话题的语义表示。**（2）基于话题方向盘的领域消歧知识获取机制**：我们设计可学习的变换矩阵，通过建模不同领域下话题分布的投影方向获取多领域上的语义消歧知识。话题的语义表示经过领域方向投影的再次变换后，有效的语义消歧特征得到强化，从而提升大语言模型在不同领域下的语义消歧能力。我们选取Qwen-2.5-1.5B作为基础模型，在英语-汉语以及德语-汉语两个多领域翻译任务上进行实验验证。实验结果表明，该方法在平均BLEU值和COMET均超出基线模型，进一步我们对于翻译质量的提升与消歧效果之间的关系进行了分析，并通过翻译实例给出详细说明。

关键词： 大语言模型；多领域翻译；语义消歧；话题方向盘

Word Semantic Disambiguation via Topic Steering in Multi-Domain Translation

Zhibo Man[1,2], **Yujie Zhang**[1,2], **Yuanmeng Chen**[1,2], **Yufeng Chen**[1,2], **Jinan Xu**[1,2],
[1]Key Laboratory of Big Data & Artificial Intelligence in Transportation, Ministry of Education, Beijing, 100044
[2]School of Computer Science and Technology, Beijing Jiaotong University, Beijing, 100044
{zhiboman,yjzhang,yuanmengchen,jaxu,chenyf}@bjtu.edu.cn

Abstract

In recent years, Large Language Models (LLMs) have achieved remarkable progress in general translation tasks, yet their performance significantly drops in multi-domain scenarios. Enhancing domain-specific translation using limited parallel data remains a key challenge, especially as existing methods rely on manually assigned domain labels, which limit their ability to capture disambiguation knowledge. To address this, we propose a Topic Steering Wheel approach to improve semantic disambiguation in

† 通讯作者

第二十四届中国计算语言学大会论文集，第709页-第721页，济南，中国，2025年8月11日至14日。

(c) 2025 中国中文信息学会计算语言学专业委员会

multi-domain translation. It includes: (1) a topic-based semantic representation mechanism, where fine-grained topic embeddings are obtained using the ETM model and mapped from LLM outputs via a TopicModel function; and (2) a domain-aware disambiguation mechanism that learns projection matrices to model topic shifts across domains, enhancing domain-specific semantic cues. We evaluate our method on English–Chinese and German–Chinese translation tasks using Qwen-2.5-1.5B. Results show consistent improvements in BLEU and COMET over strong baselines, with further analysis demonstrating the link between disambiguation and translation quality.

Keywords: Large language models , Multi-domain translation , Word semantic disambiguation , Topic steer

1 引言

最近，大语言模型（Large Language Models, LLMs）凭借其强大的文本生成能力，在通用领域的机器翻译（Machine Translation, MT）任务中取得了显著成果(Feng et al., 2024; Feng et al., 2025; Zhang et al., 2025; Tan and Monz, 2025)。然而，当面对多领域翻译（Multi-domain Translation, MDT）任务时，其翻译性能仍不尽如人意。如图1所示，我们在英语-汉语翻译方向上，初步评估了Qwen-3.2-3B在六个领域[1]（教育、法律、新闻、科学、口语以及通用领域）上的表现。结果显示，相较于通用领域，该模型在多个特定领域上的BLEU与COMET指标明显下降，表明其对领域语义的建模能力仍显不足。已有研究表明，这种性能差异主要源于跨领域之间存在单词歧义(Man et al., 2025)。为进一步验证这一问题，我们对多领域数据集中“一词多义”现象进行了统计分析。结果显示（如表7所示），在多领域文本上有大量的多义词存在，这进一步表明在多领域文本上，大语言模型对于词汇歧义的处理能力尚且不足。由此可见，提升大语言模型在多领域文本上的语义消歧能力，已成为提升其翻译质量的关键所在。

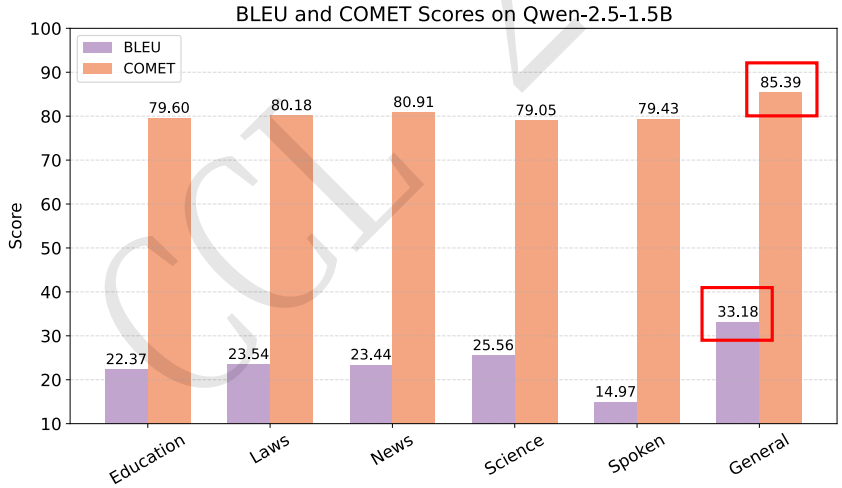


Figure 1: 大模型Qwen-2.5-1.5B模型在多领域文本翻译上的表现

长期以来，如何利用领域信息对单词进行语义消歧，一直是多领域翻译任务中的核心问题。早期在基于统计的机器翻译系统中，研究人员通常借助外部领域词典来限制单词在不同领域中的译语(Hiroya and Honda, 2013)。随着神经机器翻译（NMT）框架的发展，研究人员提出了“领域比例”的概念，将单词在各个领域中的语义视为一种分布比例，从而刻画其在不同领域中的语义偏向(Jiang et al., 2020; Man et al., 2023)。然而，我们观察到，数据不仅在各个领域内部具有分布独立性，在不同领域之间也存在语义交叉，仅依赖领域比例的方法难以全面捕

©2025 中国计算语言学大会
根据《Creative Commons Attribution 4.0 International License》许可出版

[1]多领域测试数据来自*UM-Corpus*，通用领域测试集使用WMT22的general数据集。

单词	领域		单词	领域	
	教育领域	教育领域		教育领域	法律领域
power	能量	力量	manageable	易处理	易处理

Table 1: 单词的语义并非总是完全由领域决定：power在同一个领域上译语不同；manageable在不同领域上译语相同

捉词语的语义变化。如表1所示，同一个单词在同一领域中可能对应多种译语，而在不同领域中则可能共享相同译语，这表明单词的语义并非总是完全的由领域所决定，仅利用领域标签学习语义表示无法精确划定词义的语义边界。

为了解决上述领域占比方法在建模细粒语义分布时的不足，我们提出了一种“话题方向盘”框架，用于多领域翻译中的语义消歧。与传统的领域划分不同，话题模型不再将数据硬性归类到预设的领域标签下，而是通过对文本的潜在语义结构建模，将每个文档表示为多个话题的混合分布，每个话题则由一组语义相关的词汇组成(Dieng et al., 2020; Wang et al., 2021; Man et al., 2023; Xezonaki et al., 2023)。因此，话题模型能够从训练数据中自动发现更加细腻的语义共现规律，挖掘出跨领域但语义一致的话题，或是在同一领域内部进一步划分出具有语义差异的子话题。具体地，在训练阶段，我们将话题模型生成的单词语义表示作为指导信号，对大语言模型的词嵌入层进行线性变换，以提升其在多领域环境下对细粒度语义差异的辨识与建模能力。在推理阶段，我们基于经过线性变换的单词语义表示权重，引导模型感知输入句子的潜在领域属性，从而实现对领域信息的有效控制，提升其在不同领域下的语义消歧能力。

综上，本文的贡献主要有以下三方面：（1）我们设计一种基于话题模型的语义表示获取机制，能够从多领域数据中自动挖掘潜在的语义结构，提升模型对词义在不同语境中变化的感知与建模能力。（2）我们提出一种基于话题方向盘的领域消歧知识获取机制，通过建模不同领域下话题分布的投影方向获取多领域上的语义消歧知识，提升大语言模型在不同领域下的语义消歧能力。（3）基于开源大语言模型，我们在英-中与德-英两个语言对上构建了多领域翻译实验。结果表明，所提方法在无需全量微调的情况下，显著提升了译文质量。示例分析进一步验证了话题方向盘在缓解语义粒度过粗问题方面的有效性。

2 相关工作

2.1 多领域翻译

现有多领域翻译中围绕词语消歧展开的研究工作，按照领域信息的建模粒度，主要可分为两大类：（1）基于句子级别的多领域翻译方法：该类方法通常将整个句子视为领域建模的基本单元，常见策略包括在模型输入中引入显式的领域标签(Kobus et al., 2017)，为不同领域构建专属参数(Britz et al., 2017; Tars and Fishel, 2018; Bapna and Firat, 2019; Man et al., 2024b)。然而，由于这些方法仅关注整体句子层面的领域信息，难以精确区分句中出现的多义词。（2）基于单词级别的多领域翻译方法：相较之下，单词级方法更关注词汇层面的领域敏感建模(Zeng et al., 2018)。Jiang 等人(2020) 提出了“领域比例”方法，将单词在各个领域中的译法建模为一种分布，捕捉其领域偏向性。Man 等人(2023) 则通过上下文动态感知的词向量表示增强消歧能力。此外，还有研究探索使用注意力机制(Zhang et al., 2021) 或上下文感知编码器对词义进行细粒度建模。尽管这类方法在词语消歧任务中具有更强表现，但在上下文较弱或语境不明确时，往往存在建模不稳定的问题。

2.2 基于大语言模型的机器翻译

随着大语言模型在自然语言处理中的广泛应用，研究者开始将其引入机器翻译任务，以借助其强大的语言建模能力和知识迁移能力，提升译文在多样语境下的鲁棒性与一致性。当前，基于LLM 的多领域翻译方法主要可分为两类：（1）基于微调的机器翻译方法：该类方法通过对大语言模型进行全参数或参数高效的微调，例如LoRA (Hu et al., 2021)，以实现模型在多领域翻译场景中的适配与优化(Zhu et al., 2024; Qian et al., 2024; Hu et al., 2024a)。此类方法在中高资源条件下能够显著提升领域内部的翻译质量，尤其对术语一致性和上下文建模表现出较强能力。（2）基于提示学习的机器翻译领域翻译方法：为降低计算成本并提升迁移灵活

(c) 2025 中国中文信息学会计算语言学专业委员会

性，另一类方法通过自然语言提示（prompt）引导大语言模型按照目标领域风格生成译文。例如，(Vilar et al., 2022; Jiao et al., 2023; Zhang et al., 2023; Moslem et al., 2023; He et al., 2024; Man et al., 2024a) 通过设计领域标签或任务描述提示，引导模型在低资源领域中生成更符合语境的译文。该类方法具备较强的可扩展性，适用于少样本或零样本的领域迁移场景。

综上所述，相较于现有研究，本文方法具有如下显著区别与优势：（1）**更细粒度的语义建模能力**：相比传统多领域翻译方法，我们引入话题模型，从潜在语义结构中挖掘细粒度的领域信息，有效解决了同一领域内语义歧义及跨领域语义重合的问题。（2）**无需全量微调，控制机制更灵活**：相较于基于大语言模型的翻译方法，我们在不进行全参数微调的前提下，通过话题方向盘机制引导语义表示学习，无需人工设计复杂提示模板。

3 语义消歧的话题方向盘方法

为了增强大语言模型在多领域翻译任务中的语义消歧能力，尤其在处理高歧义性专业术语时的上下文适应性，受到Han(2024)工作的启发，如图2所示，我们提出一种用于多领域翻译消歧的话题方向盘（Topic Steer）的翻译方法。该方法不对语言模型主体参数进行修改，仅在输出词嵌入空间施加一个可控的线性变换，结合语境中的话题分布信息，引导语言模型的词义建模方向。整体结构如图所示，分为两个核心模块：**（1）基于话题模型的语义表示获取机制** 和 **（2）基于话题方向盘的领域消歧知识获取机制**，算法的总体流程如算法表1所示。

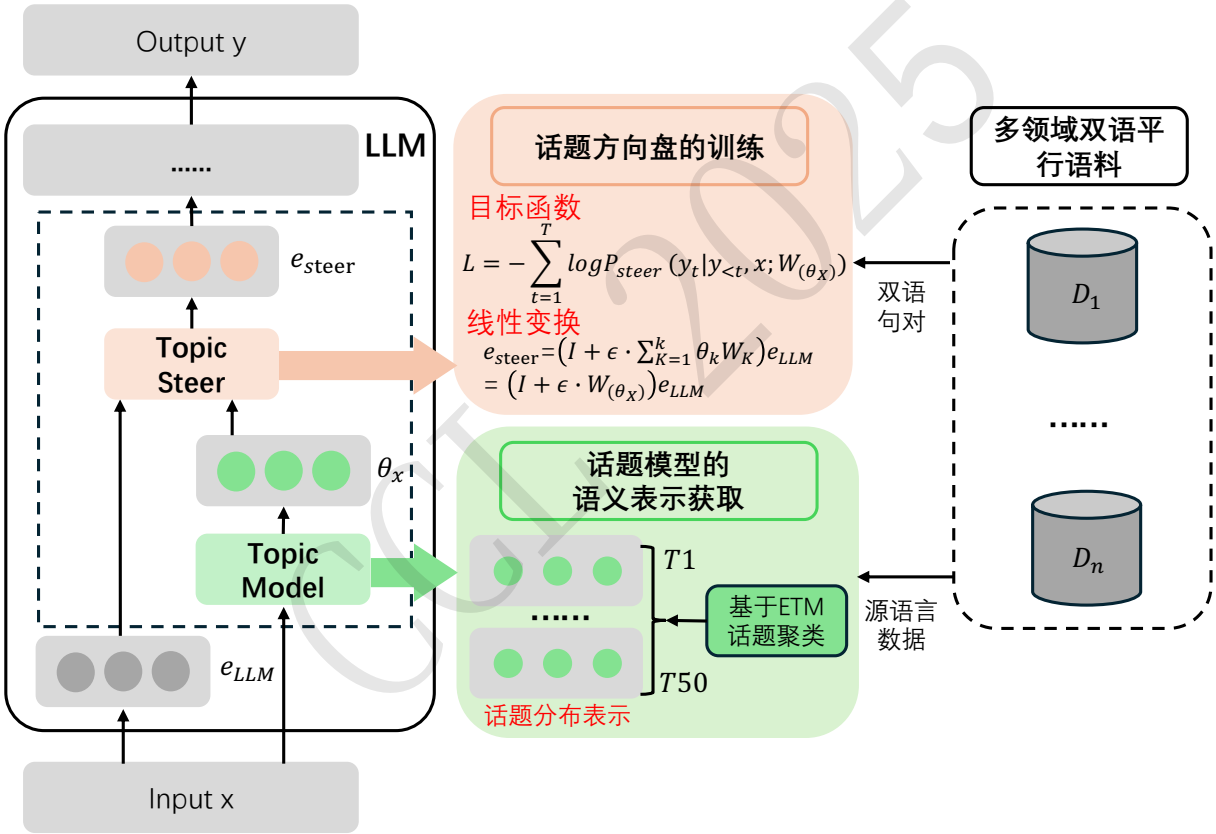


Figure 2: 多领域翻译中语义消歧的话题方向盘方法框架

3.1 基于话题模型的语义表示获取机制

为了获取有效的术语消歧翻译知识，我们引入话题模型ETM(Dieng et al., 2020)学习细粒度的语义表示，为获取消歧知识做准备，具体地，我们利用多个领域的双语平行语料中的源语言文本，基于话题模型ETM设置50个话题数量进行聚类，从而获取多领域数据上的话题语义表示。如此，相比较于已有工作(Jiang et al., 2020)，我们的方法具备以下优势：（1）该方法能够获取更大维度的语义表示，进而可以表示颗粒度更加细小的语义，为后续获取丰富的语义消歧知识奠定基础。已有方法则使用领域数目的维度（5或者7个维度）作为领域的语义表示，过

(c) 2025 中国中文信息学会计算语言学专业委员会

粗的语义颗粒度难以表达消歧知识的细微变化。（2）该方法通过自动聚类获取语义颗粒度的表示，能够规避利用领域标签获取语义表示时所带来的问题。具体地，已有方法借助领域标签作为学习目标，一方面会导致某个单词在同一领域中的不同语义的表示学习趋于接近，而另一方面会导致某个单词在不同领域上的相同语义的表示学习趋于不同。而本文的方法可以规避这些问题。（3）该方法获得的话题更贴近语义，因此更适合作为最小语义单元构建语义表示。已有方法则使用领域类别表示语义，对于“Economic”、“IT”这样的领域类别可以在某种程度上表达语义，但是语义颗粒度过粗；而对于“Spoken”这样的领域类别表达的是一种语体，并不是严格意义上的语义；再有对于“News”这样的领域类别，实际上包含了“Economic”、“IT”，这种领域的交叉性会导致语义表示学习上的混淆。

对于如图2 所示，我们将每个源句$\mathbf{x}$ 映射为一个K 维的主题分布向量：

$$\boldsymbol{\theta}_{\mathbf{x}} = \text{TopicModel}(\mathbf{x}) \in \mathbb{R}^{K}, \tag{1}$$

其中，$\sum_{k=1}^{K} \theta_k = 1$，$\theta_k$ 表示句子在第k 个话题上的归属概率。我们采用ETM (Dieng et al., 2020) 来建模源句的语义分布，其通过结合预训练词嵌入和主题分布，实现对句子细粒度语义的建模。每个话题以一个嵌入向量表示，句子则通过变分推断获得其在话题空间中的分布$\boldsymbol{\theta}_{\mathbf{x}}$。

3.2 基于话题方向盘的领域消歧知识获取机制

我们设计话题方向盘机制将大语言模型在3.1节中获得的话题语义表示转换成领域分布下的话题语义表示，用于大模型在生成多领域译文上的语义消歧。具体地，对于源语言句子的$\theta_{\mathbf{x}}$，我们设计一组可学习的变换矩阵$W_k{}_{k=1}^{K}$通过建模不同领域下话题分布的投影方向，来学习多领域上的语义消歧知识。其中$W_k \in \mathbb{R}^{d\times d}$，$d$ 为词嵌入维度，基于话题方向盘的语义表示变换计算公式如下所示：

$$\mathbf{e}'_{steer} = \left(I + \epsilon \cdot \sum_{k=1}^{K} \theta_k W_k \right) \mathbf{e}_{LLM} = \left(I + \epsilon \cdot W(\boldsymbol{\theta}_{\mathbf{x}}) \right) \mathbf{e}_{LLM}, \tag{2}$$

其中，$\mathbf{e}_{LLM}$是语言模型生成的初始词向量表示，$\mathbf{e}'_{steer}$是经过话题方向盘变化后的词向量，I 是单位矩阵，ϵ 为标量超参数，用于控制话题方向盘强度。W为领域下的话题投影方向。经过上述变换，话题语义表示在获得消歧知识后，在语义消歧的有效话题特征上得到强化，进而会影响大语言模型对译语候选词的偏好选择，从而实现不同领域下单词的语义消歧。

3.3 话题方向盘的训练方法

为了训练话题方向盘矩阵$\{W_k\}_{k=1}^{K}$，我们冻结大语言模型的参数部分，仅学习每个话题对应的方向盘矩阵$\{W_k\}_{k=1}^{K}$。训练过程采用标准翻译任务中的目标函数优化话题方向盘机制，使其在语义空间中学习到在领域下话题分布的变化梯度。

给定训练数据对$(\mathbf{x}, \mathbf{y})$，源句$\mathbf{x}$ 的话题分布表示为$\boldsymbol{\theta}_{\mathbf{x}}$，我们定义训练目标为最小化经话题方向盘变换后大语言模型的输出分布与参考译文之间的交叉熵损失：

$$\mathcal{L} = -\sum_{t=1}^{T} \log P_{\text{LM-steer}} \left(y_t \mid y_{<t}, \mathbf{x};\ W(\boldsymbol{\theta}_{\mathbf{x}}) \right). \tag{3}$$

其中，输出词分布$P_{\text{LM-steer}}$ 是加入话题方向盘变换后的语义表示$\mathbf{e}_{steer}$最终所得到的softmax输出。

4 评测实验与分析

在实验中，我们主要探究以下两方面的问题：（1）相比已有的多领域翻译消歧工作，我们提出方法的翻译效果如何（§4.4）？（2）相比已有的多领域翻译消歧工作，我们提出方法的消歧准确率如何（§5.1）？

(c) 2025 中国中文信息学会计算语言学专业委员会

Algorithm 1: 多领域翻译中语义消歧话题方向盘的训练

Input: 多领域平行语料对$(\mathbf{x}, \mathbf{y})$，话题数K，嵌入维度d，大语言模型LLM，学习率η，Steer 强度ϵ

Output: 训练好的话题Steer 矩阵$\{W_k\}_{k=1}^{K}$

初始化 $\{W_k\}_{k=1}^{K} \in \mathbb{R}^{d\times d}$，全为零矩阵;

冻结语言模型LM 参数;

foreach 训练轮次 **do**

  foreach 训练样本$(\mathbf{x}, \mathbf{y})$ **do**

    使用ETM获取源句$\mathbf{x}$ 的话题分布$\boldsymbol{\theta}_x \in \mathbb{R}^K$;

    构造加权Steer 矩阵: $W(\boldsymbol{\theta}_x) \leftarrow \sum_{k=1}^{K} \theta_{x,k} W_k$;

    foreach 大模型单词嵌入$\mathbf{e}_{LLM}$ **do**

      计算扰动后嵌入: $\mathbf{e}'_{steer} \leftarrow (I + \epsilon W(\boldsymbol{\theta}_x))\mathbf{e}_{LLM}$;

    使用$\mathbf{e}'_{steer}$ 计算翻译损失: $\mathcal{L} \leftarrow \text{CrossEntropy}(P_{\text{LM}}(\mathbf{y} \mid \mathbf{x}))$;

    更新$\{W_k\}$: $W_k \leftarrow W_k - \eta \cdot \frac{\partial \mathcal{L}}{\partial W_k}$;

推理阶段:

Input: 测试句子$\mathbf{x}_{\text{test}}$

Output: 译文$\mathbf{y}_{\text{gen}}$

使用ETM获得的预训练好的话题模型语义$\boldsymbol{\theta}_{\text{test}}$;

构造$W(\boldsymbol{\theta}_{\text{test}}) \leftarrow \sum_k \theta_{\text{test},k} W_k$;

foreach 大模型单词嵌入$\mathbf{e}_{LLM}$ **do**

  应用steer: $\mathbf{e}_{steer} \leftarrow (I + \epsilon W(\boldsymbol{\theta}_{\text{test}}))\mathbf{e}_{LLM}$;

使用调整后的嵌入进行译文生成;

return $\mathbf{y}_{gen}$

ccl 2025

4.1 数据集

我们分别在英语-汉语和英语-德语两个翻译任务上进行实验：（1）英语-汉语：针对英-汉翻译任务，我们使用多领域数据集UM-Corpus上的五个领域(Tian et al., 2014) [2] ：新闻、口语、科学、教育、法律（News, Spoken, Science, Education, Laws）。（2）英语-德语：针对英-德翻译任务，我们使用多领域数据集OPUS[3]上的五个领域，分别是：法律、互联网、古兰经、医学、字幕（Laws, It, Koran, Medical, Subtitles）。如表2和3所示，我们列出了在英语-汉语和德语-英语翻译任务上训练集、验证集和测试集的句对统计信息。

领域	训练集	验证集	测试集
Education	445K	2K	462
Laws	208K	2K	456
News	444K	2K	1500
Science	263K	2K	503
Spoken	216K	2K	455

Table 2: 英语-汉语数据集统计信息

领域	训练集	验证集	测试集
Laws	467K	2K	2K
It	233K	2K	2K
Koran	17K	2K	2K
Medical	248K	2K	2K
Subtitles	1445K	2K	2K

Table 3: 德语-英语数据集统计信息

4.2 实验设置

在实验中，我们选择Qwen-2.5-1.5B作为基础模型进行验证，关于话题数量的选择我们选择{10-100}进行验证。针对实验数据的预处理，首先，本文使用Moses脚本(Koehn et al., 2007)对英语和德语句子进行分词等处理。关于实验评测指标，我们使用BLEU (Papineni et al., 2002)和COMET[4](Guerreiro et al., 2024)值来评估模型性。在优化器选择上，我们使用Adam 优

[2]http://nlp2ct.cis.umac.mo/um-corpus/

[3]https://opus.nlpl.eu/

[4]https://github.com/Unbabel/COMET

(c) 2025 中国中文信息学会计算语言学专业委员会

英语-汉语多领域翻译任务							
方法	Education	Laws	News	Science	Spoken	AVG	Para
PT	22.37/79.60	23.54/80.18	23.44/80.91	25.56/79.05	14.97/79.43	21.98/79.83	-
PTDT	23.00/79.93	24.27/81.49	23.60/80.36	26.28/80.96	15.12/80.21	22.45/80.59	-
DP	23.69/80.10	24.77/81.25	23.81/80.56	26.57/81.22	15.32/80.40	22.83/80.17	231M
WDSRL	23.82/80.42	24.85/81.30	23.91/80.60	26.85/81.08	15.44/80.53	22.97/80.79	482M
Ours	24.01/80.67	25.20/81.65	24.10/80.98	27.12/81.43	15.72/80.83	**23.23**/**81.11**	325M
德语-英语多领域翻译任务							
方法	IT	Koran	Laws	Medical	Subtitles	AVG	
PT	25.57/72.93	10.81/66.92	21.04/76.96	29.05/77.22	19.98/72.73	21.29/73.35	-
PTDT	26.56/73.99	11.05/67.31	22.70/77.09	29.68/78.10	20.66/72.95	22.13/73.89	-
DP	26.68/73.87	11.51/67.54	22.81/77.56	29.74/78.64	20.94/73.01	22.34/74.12	148M
WDSRL	26.69/73.44	11.54/67.64	22.90/77.89	29.71/78.65	21.18/73.26	22.40/74.18	355M
Ours	27.03/74.10	11.88/68.13	23.40/77.92	30.11/79.25	22.07/73.69	**22.90**/**74.62**	268M

Table 4: 英语-汉语和德语-英语多领域翻译任务上的实验结果。"/"前后分别表示BLEU和COMET分数。

化器(Kingma and Ba, 2014)，学习率设为1×10^{-2}，训练1000步。引导矩阵W 的初始值服从均值为0、方差为1×10^{-3} 的高斯分布。在所有实验中，我们使用随机种子0、1、2 进行三次训练。我们使用的实验设备是一块具有16GB CUDA显存的Tesla V100 GPU。

4.3 对比方法

PT（Prompt Translation）(Jiao et al., 2023)：直接对大语言模型输入一个通用提示，形式为："请将以下句子从源语言翻译为目标语言：[源语言文本]。"

PTDT（Prompt Translation with Domain Tag）(Hu et al., 2024b)：在PT 的基础上加入领域标签信息，形式为："请将以下句子从源语言翻译为目标语言，并符合[领域标签]领域的专业风格：[源语言文本]。"

DP（Domain Proportion）：为了对比已有的多领域建模方法，我们基于Jiang 等人 (2020) 提出的领域比例策略，将其拓展到大语言模型的词嵌入控制上。该方法假设每个句子可能同时涉及多个领域，并通过预训练的领域分类器预测其领域分布，即一个概率向量$\boldsymbol{\delta}_{\mathbf{x}} \in \mathbb{R}^D$，其中$D$ 为领域数，δ_d 表示输入句子$\mathbf{x}$ 属于第d 个领域的概率，满足$\sum_{d=1}^{D} \delta_d = 1$。具体地，我们为每个领域$d$ 设置一个独立的可学习线性变换矩阵$W_d \in \mathbb{R}^{d \times d}$，对语言模型输出的词嵌入$\mathbf{e}_v$ 进行如下加权变换：

$$\mathbf{e}'_v = \left(I + \epsilon \cdot \sum_{d=1}^{D} \delta_d W_d \right) \mathbf{e}_v, \tag{4}$$

其中I 是单位矩阵，ϵ 是超参数，用于调节嵌入调整的强度。

WDSRL(Man et al., 2023)：基于表示学习的多领域翻译方法，基于Qwen-2.5-1.5B模型进行微调，学习不同领域的表示，从而得到不同领域的翻译结果。

Ours（Topic Steer）：不同于显式使用领域标签的方法，我们提出的方法利用源语言句子的潜在话题分布来自适应调节输出嵌入，从而实现更细粒度的语义消歧，具体实现公式如第3节。

4.4 主要实验结果

如表 4所示，我们给出英语-汉语以及德语-英语多领域翻译任务上的实验结果，我们的方法在两个翻译任务上的BLEU与COMET均超过了基线，达到了最高性能。具体地，我们对比所有的方法，得出如下分析：在英语-汉语任务中，相比直接对大语言模型利用提示学习生成译文的PT方法，引入领域标签的PTDT方法在所有领域均带来了提升，说明添加领域信息的提示有助于大模型进行多领域翻译。然而，PTDT的提升幅度相对有限，表明其在建模语义歧义时仍存在不足。进一步引入领域比例控制的DP方法，使得模型更加关注领域的

(c) 2025 中国中文信息学会计算语言学专业委员会

英语-汉语多领域翻译任务						
方法	Education	Laws	News	Science	Spoken	AVG
PT	33.14/88.10	50.82/88.94	30.04/84.51	28.76/84.82	19.20/77.00	32.39/84.67
PTDT	33.46/88.21	51.39/89.20	30.36/84.92	28.78/86.13	20.89/77.46	32.98/85.18
DP	33.88/88.41	52.14/89.67	31.68/85.66	29.08/86.52	21.39/77.88	33.63/85.63
Ours	34.15/88.97	53.02/90.68	32.55/86.47	30.60/87.95	22.80/78.12	34.62/86.44

Table 5: 英语-汉语在Qwen-2.5-7B上的实验结果。“/”前后分别表示BLEU和COMET分数。

分布概率，相比PT和PTDT方法，DP在五个领域上的BLEU与COMET指标均有提升，尤其在Science和News领域上效果更为显著。在Science领域，DP相比BLEU和COMET分数PT分别提升了1.01和2.17，相比PTDT也有0.29和0.26的提升，表明该方法对于语义歧义强、术语密集的文本具有更强的建模能力。在News领域，DP相较PTBLEU分数提升了0.37，尽管COMET分数下降0.35，但相较PTDT，DP在BLEU和COMET上仍分别提升了0.21和0.20，体现出基于领域分布的消歧策略对新闻类文本的翻译仍具有一定效果。因此，DP方法能够根据单词在不同领域分布建模多领域语义特征，提升翻译质量。

在此基础上，我们提出的Topic Steer方法在所有领域上均取得最高性能，平均BLEU和COMET分数分别达到23.23 和81.11。相比DP，Topic Steer在五个领域中均有进一步提升，平均BLEU和COMET分数分别提升了0.40和0.94。其中，在专业性较强的Science和Laws领域，Topic Steer的提升尤为显著：在Science领域，BLEU和COMET分别提升了0.55和0.21；在Laws领域，分别提升了0.43和0.40。我们发现，Topic Steer能更有效地控制模型输出的语义方向，从而显著提升专业领域文本中的翻译质量。说明相比直接利用人工标注的标签，话题模型能够学习到更细粒度的语义信息，得到了更准确的表示。

在德语-英语翻译任务中，整体结果与英语-汉语任务呈现出一致的趋势。Topic Steer在多个专业性强的领域（如IT、Medical 和Laws）均实现显著性能提升，即使在资源极为稀缺的Koran领域，也较基线方法取得了提升。与英语-汉语结果相似，Topic Steer在BLEU和COMET两个指标上达到最高性能，进一步验证了方法在多种语言上的适应性。

如表5所示，不同方法在Qwen-2.5-7B模型下的表现。相比于基础方法PT和加入领域提示的PTDT，我们的方法在所有五个领域上均取得了最好的BLEU和COMET分数，平均提升明显，尤其在Science和News领域表现突出，说明所提方法在提升多领域翻译质量方面具有显著优势。综上所述，Topic Steer通过引入话题先验引导与动态调控机制，显著增强了大模型对语义歧义的建模能力，相比直接利用人工标注的标签，话题模型能够学习到更细粒度的语义信息。

5 分析与讨论

在这一部分，我们对英语-汉语翻译任务进一步探究消歧准确率（§5.1）、话题模型个数对结果的影响（§5.2）以及示例分析（§5.3），通过这些分析说明提出的方法能够有效的捕捉细粒度的语义信息，将领域的单位进一步精细化，使得译语得到正确的翻译，从而提升翻译质量。

5.1 消歧准确率对比

我们提出方法的主要目标是对多领域翻译中的单词消歧，进而提升翻译质量。因此，如何量化模型语义歧义的能力成为了关键问题。为了获得歧义单词的数据，在我们的工作中，我们对多领域翻译数据进行词对齐，利用词对齐工具*Awesome-Align*[5] (Dou and Neubig, 2021)在多个领域的训练语料上获得单词对之间的关系，以源语言为核心得到一对多的训练集上的多义词词表，进一步根据训练集上的词表对应出测试集的词表，具体的规模如下表9所示。此外，我们进一步统计了在这些多义词中对应歧义个数的数据信息，包括一个源语言单词对应两个、三个、四个、五个以及五个以上译语的情况，如表7所示。这些数据统计信息都进一步说明了在多领域翻译上歧义现象的复杂程度，进一步反应了我们研究的意义和目的。

[5]https://github.com/neulab/awesome-align

(c) 2025 中国中文信息学会计算语言学专业委员会

	Education	Laws	News	Science	Spoken
训练集	30K	54K	33K	50K	21K
测试集	492	686	720	471	422

Table 6: 英语-汉语多领域翻译任务的训练集和测试集上的歧义词表

两个译语	三个译语	四个译语	五个译语	五个以上译语
1120	1060	724	23	4

Table 7: 英语-汉语多领域翻译任务的训练集和测试集上歧义单词的统计结果

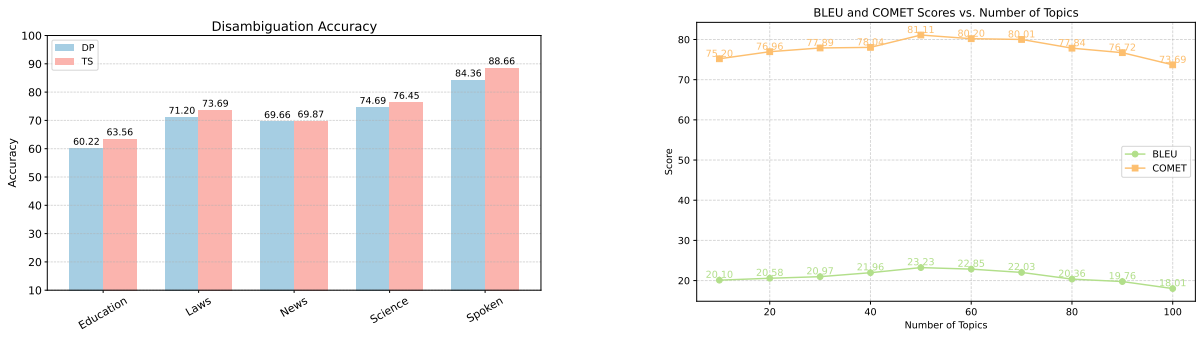
如3a所示，我们根据歧义词表计算了DP方法和我们提出方法的消歧准确率的对比，我们发现在五个领域上我们方法消歧的准确率都高于DP，这说明我们提出的话题方向盘得到了最小语义的表示，不依赖于人工设计标签对于领域知识的学习，从而解决单词歧义的问题。

5.2 话题模型数量对实验结果的影响

为了进一步探究在主题模型生成的主题个数对翻译性能的影响，我们在实验中设计不同的主题数量训练主题模型，如图3b所示，我们发现当主题模型数量设置为50时，BLEU和COMET的分数均达到了最高的性能。此外，我们观察到当主题模型数量设置在10-40之间时，在编码过程中生成的单词主题向量的潜在语义信息不足以表征其领域信息导致翻译性能较低；当主题模型数量设置在60-100之间时，这说明过多的主题模型会导致领域信息过于细化，从而导致模型性能翻译下降。

5.3 示例分析

我们列出两个示例说明提出方法的有效性，如表 8所示，单词“power”在同一教育领域中具有语义歧义，其可被译为“能量”或“能力”，分别对应物理活动与能力赋予等不同语义。DP无法准确区分这一细微差别，在不同上下文中给出统一或不够贴切的翻译，将“power”一律译为力量，造成语义模糊或不自然。而我们提出的话题方向盘（TS）方法通过引入更细粒度的话题信息，在语义建模过程中对上下文话题进行引导，从而更有效地区分同一领域内部的细粒度语义差异。如表 9所示，单词“*manageable*”在教育与法律两个不同领域中尽管语义差异显著，但其语义表达高度一致，均可译为“易处理”或“可处理”。DP根据领域标签学习到属于每个领域单独的知识，无法建立领域间的语义表示。相较而言，我们提出的话题放向盘通过话题引导增强模型对细粒度语义的表达。



(a) (b)

Figure 3: (a)英语-汉语翻译任务上消歧准确率的比较结果；(b)话题模型中话题数量对翻译质量的影响

(c) 2025 中国中文信息学会计算语言学专业委员会

领域	教育领域
源语言 目标语言	Human generates power through biochemical processes to perform physical. 人体通过生化过程产生能量，以执行体力活动。
DP TS	人体通过生化过程产生力量来完成身体任务。 人体通过生化过程产生能量从而执行体力活动。
领域	教育领域
源语言 目标语言	Education gives people the power to change their lives and achieve dreams. 教育赋予人们能力去改变人生并实现梦想。
DP TS	教育给予人们改变生活、实现梦想的力量。 教育赋予人们能力去改变人生以及实现梦想。

Table 8: 翻译示例：单词在同一领域上译语不同的情况

领域	教育领域
源语言 目标语言	Narrow the question down to a coherent and manageable set of issues. 把问题具体到一系列相关易处理的主题。
DP TS	将问题缩小为一系列连贯且易管理的问题。 问题具体为一系列有关且易处理的问题
领域	法律领域
源语言 目标语言	Ensure that all terms are manageable within the legal framework. 确保所有条款在法律范围内均可处理
DP TS	确保所有条款在法律框架内均可管理。 确保所有条款在法律范围内均可处理。

Table 9: 翻译示例：单词在不同领域上译语相同的情况

6 结论与未来研究工作

随着大语言模型在自然语言处理领域的广泛应用，其在通用场景下的翻译能力已达到较高水平。然而，在面对多领域文本时，现有模型在语义消歧方面仍存在明显短板。针对这一问题，本文提出了一种基于话题方向盘的多领域翻译方法，设计基于话题模型的语义表示机制和基于话题方向盘的领域消歧知识获取机制来学习细粒度的语义表示，增强大语言模型的在多领域翻译文本上的消歧能力。实验结果验证了所提方法在多个标准多领域翻译基准上的优越性能，证明了话题方向盘在提升语义消歧能力与整体翻译质量方面的有效性。在未来的工作中，我们计划在规模更大的语言模型以及更多的语言对上进行实验验证提出方法的有效性。

致谢

本研究受国家自然科学基金面上项目（No.62376019，62476023，61976015，61976016，61876198，61370130）以及国家重点研发计划（2020AAA0108001）的资助。作者们还对匿名评审专家给予的宝贵建议表示衷心的感谢。

参考文献

Ankur Bapna and Orhan Firat. 2019. Simple, scalable adaptation for neural machine translation. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1538–1548, Hong Kong, China, November. Association for Computational Linguistics.

(c) 2025 中国中文信息学会计算语言学专业委员会

Denny Britz, Quoc Le, and Reid Pryzant. 2017. Effective domain mixing for neural machine translation. In Ondřej Bojar, Christian Buck, Rajen Chatterjee, Christian Federmann, Yvette Graham, Barry Haddow, Matthias Huck, Antonio Jimeno Yepes, Philipp Koehn, and Julia Kreutzer, editors, *Proceedings of the Second Conference on Machine Translation*, pages 118–126, Copenhagen, Denmark, September. Association for Computational Linguistics.

Adji B Dieng, Francisco JR Ruiz, and David M Blei. 2020. Topic modeling in embedding spaces. *Transactions of the Association for Computational Linguistics*, 8:439–453.

Zi-Yi Dou and Graham Neubig. 2021. Word alignment by fine-tuning embeddings on parallel corpora. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2112–2128.

Zhaopeng Feng, Ruizhe Chen, Yan Zhang, Zijie Meng, and Zuozhu Liu. 2024. Ladder: A model-agnostic framework boosting llm-based machine translation to the next level. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 15377–15393.

Zhaopeng Feng, Shaosheng Cao, Jiahan Ren, Jiayuan Su, Ruizhe Chen, Yan Zhang, Zhe Xu, Yao Hu, Jian Wu, and Zuozhu Liu. 2025. Mt-r1-zero: Advancing llm-based machine translation via r1-zero-like reinforcement learning. *arXiv preprint arXiv:2504.10160*.

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

Chi Han, Jialiang Xu, Manling Li, Yi Fung, Chenkai Sun, Nan Jiang, Tarek Abdelzaher, and Heng Ji. 2024. Word embeddings are steers for language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 16410–16430.

Zhiwei He, Tian Liang, Wenxiang Jiao, Zhuosheng Zhang, Yujiu Yang, Rui Wang, Zhaopeng Tu, Shuming Shi, and Xing Wang. 2024. Exploring human-like translation strategy with large language models. *Transactions of the Association for Computational Linguistics*, 12:229–246.

S. Hiroya and M. Honda. 2013. A multi-domain translation model framework for statistical machine translation. *European Journal of Operational Research*, 215(2):832–840.

J. Hu, Y. Zhang, Z. Liu, Y. Lin, T. K. Chia, and S. L. Lee. 2021. Lora: Low-rank adaptation of large language models. In *Proceedings of the 38th International Conference on Machine Learning (ICML 2021)*. PMLR.

Tianxiang Hu, Pei Zhang, Baosong Yang, Jun Xie, Derek Wong, and Rui Wang. 2024a. Large language model for multi-domain translation: Benchmarking and domain cot fine-tuning. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 5726–5746.

Tianxiang Hu, Pei Zhang, Baosong Yang, Jun Xie, Derek F. Wong, and Rui Wang. 2024b. Large language model for multi-domain translation: Benchmarking and domain CoT fine-tuning. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 5726–5746, Miami, Florida, USA, November. Association for Computational Linguistics.

Haoming Jiang, Chen Liang, Chong Wang, and Tuo Zhao. 2020. Multi-domain neural machine translation with word-level adaptive layer-wise domain mixing. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1823–1834, Online, July. Association for Computational Linguistics.

Wenxiang Jiao, Wenxuan Wang, Jen-tse Huang, Xing Wang, and Zhaopeng Tu. 2023. Is chatgpt a good translator? a preliminary study. *arXiv preprint arXiv:2301.08745*.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Catherine Kobus, Josep Crego, and Jean Senellart. 2017. Domain control for neural machine translation. In Ruslan Mitkov and Galia Angelova, editors, *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017*, pages 372–378, Varna, Bulgaria, September. INCOMA Ltd.

CCL 2025

(c) 2025 中国中文信息学会计算语言学专业委员会

Philipp Koehn, Marcello Federico, Wade Shen, Nicola Bertoldi, Ondrej Bojar, Chris Callison-Burch, Brooke Cowan, Chris Dyer, Hieu Hoang, Richard Zens, et al. 2007. Open source toolkit for statistical machine translation: Factored translation models and confusion network decoding. In *CLSP Summer Workshop Final Report WS-2006, Johns Hopkins University*.

Zhibo Man, Zengcheng Huang, Yujie Zhang, Yu Li, Yuanmeng Chen, Yufeng Chen, and Jinan Xu. 2023. Wdsrl: Multi-domain neural machine translation with word-level domain-sensitive representation learning. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 32:577–590.

Zhibo Man, Kaiyu Huang, Yujie Zhang, Yuanmeng Chen, Yufeng Chen, and Jinan Xu. 2024a. ICL: Iterative continual learning for multi-domain neural machine translation. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 7732–7743, Miami, Florida, USA, November. Association for Computational Linguistics.

Zhibo Man, Yujie Zhang, Yu Li, Yuanmeng Chen, Yufeng Chen, and Jinan Xu. 2024b. An ensemble strategy with gradient conflict for multi-domain neural machine translation. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 23(2):1–22.

Zhibo Man, Yuanmeng Chen, Yujie Zhang, Yufeng Chen, and Jinan Xu. 2025. Dmdteval: An evaluation and analysis of llms on disambiguation in multi-domain translation. *arXiv preprint arXiv:2504.20371*.

Yasmin Moslem, Rejwanul Haque, John Kelleher, and Andy Way. 2023. Adaptive machine translation with large language models. In *Proceedings of the 24th Annual Conference of the European Association for Machine Translation*, pages 227–237.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Shenbin Qian, Archchana Sindhujan, Minnie Kabra, Diptesh Kanojia, Constantin Orasan, Tharindu Ranasinghe, and Fred Blain. 2024. What do large language models need for machine translation evaluation? In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 3660–3674, Miami, Florida, USA, November. Association for Computational Linguistics.

Shaomu Tan and Christof Monz. 2025. Remedy: Learning machine translation evaluation from human preferences with reward modeling. *arXiv preprint arXiv:2504.13630*.

Sander Tars and Mark Fishel. 2018. Multi-domain neural machine translation. In Juan Antonio Pérez-Ortiz, Felipe Sánchez-Martínez, Miquel Esplà-Gomis, Maja Popović, Celia Rico, André Martins, Joachim Van den Bogaert, and Mikel L. Forcada, editors, *Proceedings of the 21st Annual Conference of the European Association for Machine Translation*, pages 279–288, Alicante, Spain, May.

Liang Tian, Derek F. Wong, Lidia S. Chao, Paulo Quaresma, Francisco Oliveira, Yi Lu, Shuo Li, Yiming Wang, and Longyue Wang. 2014. UM-corpus: A large English-Chinese parallel corpus for statistical machine translation. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 1837–1842, Reykjavik, Iceland, May. European Language Resources Association (ELRA).

David Vilar, Markus Freitag, Colin Cherry, Jiaming Luo, Viresh Ratnakar, and George Foster. 2022. Prompting palm for translation: Assessing strategies and performance. *arXiv preprint arXiv:2211.09102*.

Weixuan Wang, Wei Peng, Meng Zhang, and Qun Liu. 2021. Neural machine translation with heterogeneous topic knowledge embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 3197–3202.

Danai Xezonaki, Talaat Khalil, David Stap, and Brandon Denis. 2023. Improving domain robustness in neural machine translation with fused topic knowledge embeddings. In Masao Utiyama and Rui Wang, editors, *Proceedings of Machine Translation Summit XIX, Vol. 1: Research Track*, pages 209–221, Macau SAR, China, September. Asia-Pacific Association for Machine Translation.

CCL 2025

(c) 2025 中国中文信息学会计算语言学专业委员会

Jiali Zeng, Jinsong Su, Huating Wen, Yang Liu, Jun Xie, Yongjing Yin, and Jianqiang Zhao. 2018. Multi-domain neural machine translation with word-level domain context discrimination. In Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii, editors, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 447–457, Brussels, Belgium, October-November. Association for Computational Linguistics.

Shiqi Zhang, Yan Liu, Deyi Xiong, Pei Zhang, and Boxing Chen. 2021. Domain-aware self-attention for multi-domain neural machine translation. *Proc. Interspeech 2021*, pages 2047–2051.

Biao Zhang, Barry Haddow, and Alexandra Birch. 2023. Prompting large language model for machine translation: A case study. In *International Conference on Machine Learning*, pages 41092–41110. PMLR.

Hongbin Zhang, Kehai Chen, Xuefeng Bai, Xiucheng Li, Yang Xiang, and Min Zhang. 2025. Exploring translation mechanism of large language models. *arXiv preprint arXiv:2502.11806*.

Wenhao Zhu, Hongyi Liu, Qingxiu Dong, Jingjing Xu, Shujian Huang, Lingpeng Kong, Jiajun Chen, and Lei Li. 2024. Multilingual machine translation with large language models: Empirical results and analysis. In Kevin Duh, Helena Gomez, and Steven Bethard, editors, *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 2765–2781, Mexico City, Mexico, June. Association for Computational Linguistics.

CCL 2025

(c) 2025 中国中文信息学会计算语言学专业委员会