



Translating Literature with LLMs

Maintaining Consistency on Named Entities and Integrating External Knowledge

Jiarui Liu

University of Massachusetts Amherst

Contents

1. Introduction
2. Methodology
3. Evaluation

Committee

- Committee Chair: Mohit Nagaraja Iyyer
- Committee Members: James Allan, Benjamin Marlin
- Project Advisor: Marzena Karpinska

Introduction

Machine Translation Using LLMs

- Machine translation offers a cost-effective, rapid, and multilingual pathway to high-quality translation
- LLMs perform very well on translation tasks without any specific training or fine-tuning.

Translate the sentence "Lolita, light of my life, fire of my loins. My sin, my soul." from English to Chinese.



The translation of the sentence "Lolita, light of my life, fire of my loins. My sin, my soul." into Chinese is:

洛丽塔, 我生命之光, 我欲火之焰。我的罪孽, 我的灵魂。



LLMs vs Traditional MT Models

- Different from traditional MT models that translate at the **sentence level**, LLMs can utilize their long input/output text window to translate all input text together at **document-level**.
- LLMs are able to **utilize the context** to make more comprehensive translations.
- Many of us already use LLM translators to help us read foreign papers and articles.

LLMs vs Traditional MT Models

- Different from traditional MT models that translate at the **sentence level**, LLMs can utilize their long input/output text window to translate all input text together at **document-level**.
- LLMs are able to **utilize the context** to make more comprehensive translations.
- Many of us already use LLM translators to help us read foreign papers and articles.

Question

Can they translate literature novels effectively?

Challenges in LLM Translation

There are still issues with LLM translation; one is that LLMs cannot handle named entity translation themselves.

- Translating named entities often necessitates different approaches and methods.
- LLMs don't translate named entities consistently, which can be very confusing, especially in literature.

Complexity of Translating Named Entities

Translating named entities often necessitates different approaches and methods:

- Some require retrieving official translations from a knowledge base.
- **Easy:**
 - Donald Trump → 唐纳德·特朗普 (phonological translation)
 - 安倍晋三 (Shinzo Abe) → 安倍晋三 (transliteration)
 - WYSIWYG → 所见即所得 (abbreviation)

Complexity of Translating Named Entities

Translating named entities often necessitates different approaches and methods:

- Some require retrieving official translations from a knowledge base.
- **Easy:**
 - Donald Trump → 唐纳德·特朗普 (phonological translation)
 - 安倍晋三 (Shinzo Abe) → 安倍晋三 (transliteration)
 - WYSIWYG → 所见即所得 (abbreviation)
- **Medium:**
 - John King Fairbank → 费正清 (Chinese name created separately)

Complexity of Translating Named Entities

Translating named entities often necessitates different approaches and methods:

- Some require retrieving official translations from a knowledge base.
- **Easy:**
 - Donald Trump → 唐纳德·特朗普 (phonological translation)
 - 安倍晋三 (Shinzo Abe) → 安倍晋三 (transliteration)
 - WYSIWYG → 所见即所得 (abbreviation)
- **Medium:**
 - John King Fairbank → 费正清 (Chinese name created separately)
- **Hard:**
 - Hippolytus Bosuiou → 苏念澄 (less famous than Fairbank)
 - David Moser → 莫大伟 (official name, can't translate to 莫大卫)

Complexity of Translating Named Entities

- Some require Additional information and knowledge.

need to know which Character to use:

- Wei Zhang → 张伟, 张维, 张威, 张薇, 张玮...
- みゆき (Miyuki) → 美雪, 美由紀, 美幸...

Complexity of Translating Named Entities

- Some require Additional information and knowledge.

need to know which Character to use:

- Wei Zhang → 张伟, 张维, 张威, 张薇, 张玮...
- みゆき (Miyuki) → 美雪, 美由紀, 美幸...

Need to follow conventions

- FBI → FBI/ FBI (联邦调查局)
- Row, Column (Math):
 - In Mainland China: 行 (row), 列 (column)
 - In Taiwan: 列 (row), 行 (column)

Complexity of Translating Named Entities

- Some require Additional information and knowledge.

need to know which Character to use:

- Wei Zhang → 张伟, 张维, 张威, 张薇, 张玮...
- みゆき (Miyuki) → 美雪, 美由紀, 美幸...

Need to follow conventions

- FBI → FBI/ FBI (联邦调查局)
- Row, Column (Math):
 - In Mainland China: 行 (row), 列 (column)
 - In Taiwan: 列 (row), 行 (column)

Text may presuppose subtle intentional meanings:e.g., poetry

Consistency is Critical in Long-text Translation

Inconsistency Example, in a book:

- **Chapter 1:**

- EN: *Miyuki* touched her chin while seriously analyzing the situation. ...
- ZH: 美雪摸着下巴，一本正经地分析道。 ...

- **Chapter 3:**

- EN: But what was *Miyuki* doing right now? ...
- ZH: 但是美由紀现在在做什么呢? ...

Very confusing! Are 美雪 and 美由紀 same person?

Input/Output length Limitations

Due to computational resource and financial limitations, **we can't send the entire book to LLMs and translate in a single run**—too long!

- Book *In Search of Lost Time*: 4,211 pages, 9,609,000 characters!



Then we need to cut the book into smaller chunk, and translate by chunk.

Issue:

- Given different contexts in each chunk, LLMs may map the **same named entity** to **different reasonable translations**.
- Especially problematic for fiction and web novels having many fictional concepts and names.

This can make translated text barely readable, not due to translation quality but because inconsistent named entities continuously confuse the reader. **Consistency is essential!**

Example: During book translation, *Qwen-Plus* mapped the name ***Lin, Jiage*** to 4 different reasonable translations

Translations of Lin, Jiage	Frequency
林嘉格	113
林嘉阁	111
林嘉歌	37
林稼歌	23

Methodology

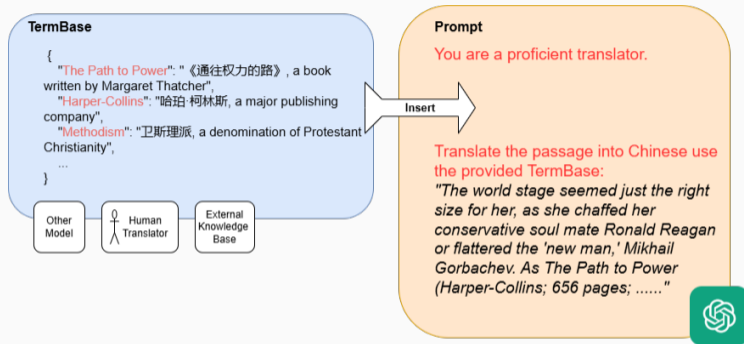
TermBase-Augmented Method

We proposed a simple yet effective TermBase-augmented in-context learning method to address the issue.

Approach:

As LLMs can't handle named entities effectively on their own, we **separate named entity translation from main text translation.**

TermBase-Augmented Method

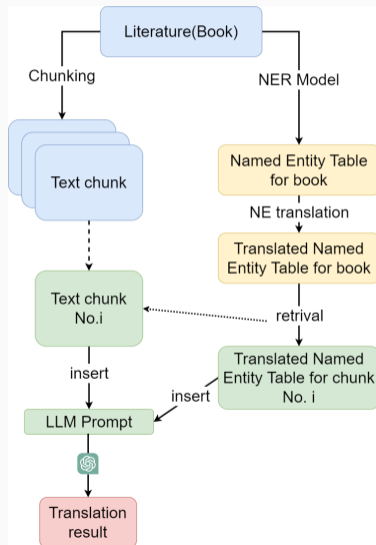


- First, get named entities translated and generate **TermBase**.
- Then, insert **TermBase** in the prompt and translated with LLM.

Translation Pipeline

Steps:

1. Named Entity Recognition (NER)
2. Translation of Named Entities (and brief explanations)
3. Merge & Store Named Entity Translations and Human Proofreading
4. TermBase Generation and Final Translation



Named Entity Recognition and Translation

The first step involves extracting relevant named entities from the entire source text and translating them.

- Utilize models like Stanza and LLMs for NER tasks.

We focus on specific types of named entities: (for real and fictional)

- **Person Name:** individuals, for both Real and fictional.
- **Location:** Geographical locations, landmarks
- **Organization:** Companies, institutions, groups, etc.
- **Creative Work:** Books, films, songs, etc.
- **Events:** Historical, cultural or fictional events

Excluded Named Entities

We are NOT interested in named entities of type:

- LANGUAGE, DATE, TIME, PERCENT, MONEY, QUANTITY, ORDINAL, CARDINAL.

Reason: They are

- either well-known nouns that can be handled by LLMs themselves
- or do not influence translation quality.

Translating Named Entities

Methods:

- Use translation models like LLMs.
- Or **translate them with human translators** for higher quality.
 - **Benefits:** A high-quality TermBase could significantly increase translation quality, terminology consistency, and readability.

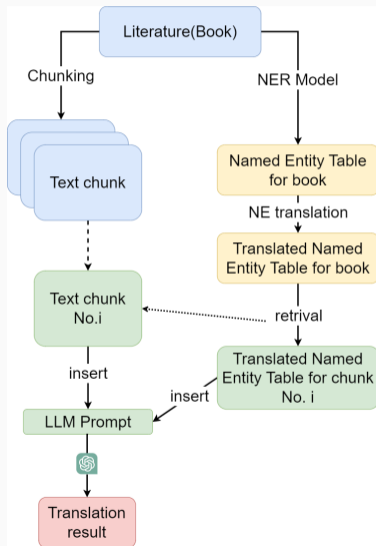
Optional Step:

- Provide brief **descriptions** of each named entity to help the model understand and translate them in the main text.

Main Text Translation

Process:

- The entire book is **too long**—we cut text into chunks and translate by chunk.
- **Filter named entities appearing in each chunk** and **retrieve their translations from the database** to create the TermBase.
- **Insert the TermBase** into the prompt and translate using the LLM.



Evaluation

Evaluation Dataset

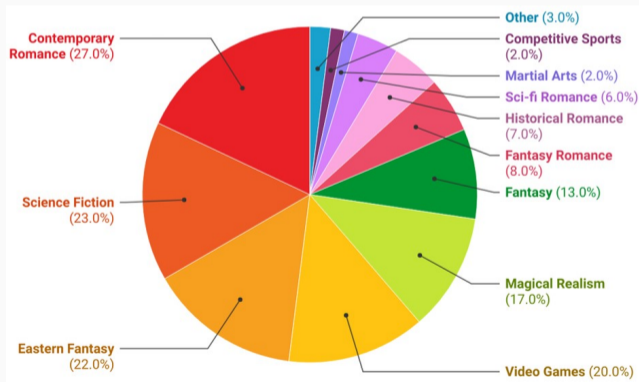
Dataset: WMT23 Shared Task - GuoFeng Webnovel.

- A publicly copyrighted, high-quality, discourse-level, and multilingual corpus of web fiction.



Features of GuoFeng Webnovel

- Rich Linguistic and Cultural Phenomena
- Long-Range Context
- Rich Genres and Named Entities



Initial Experiments 1

Objective: Test GPT-4o's translation capability using 3 prompt:

- **No TermBase:** Baseline prompt like *Translate below text:...*
- **Low-quality TermBase:** Terms directly translated by Google MT
- **High-quality TermBase:** Use Term translation of the reference text

Measurement

- **Measure Translation Quality:** *BLEU*, *COMET* score
- **Measure Terminology Consistency Metric:** *cTT* Score (Wang et al., 2023; Lyu et al., 2021)

$$cTT = \frac{\sum_{w \in TB} \frac{\sum_{i=1}^n \sum_{j=i+1}^n 1(t_{wi}=t_{wj})}{C_n^2}}{|TB|} \quad (1)$$

Results on Test Set 1

Experiment	BLEU ↑	COMET ↑	cTT ↑
Baseline	26.892	0.756	0.811
With TermBase (Low Quality)	26.070	0.751	-
With TermBase (High Quality)	29.328	0.785	0.921

表 1: Evaluation results on WMT23 test set 1

Findings:

- Low-quality TermBase decreased quality scores very little.
- High-quality TermBase increased all 3 scores.

Results on Test Set 1

Experiment	BLEU ↑	COMET ↑	cTT ↑
Baseline	17.696	0.8460	0.7179
With TermBase (High Quality)	18.539	0.8577	0.9078

表 2: Results of first 10 chapters of *Virtual World: Close Combat Mage*

Findings:

- Low-quality TermBase decreased quality scores very little.
- High-quality TermBase increased all 3 scores.

Further Evaluation

Evaluated using 27 books from the dataset. Evaluate on Models:

- **GPT-4o**: One of the most advanced LLM.
- **GPT-4o-Mini**: Smaller, cost-efficient GPT4o
- **Qwen-Plus**: With strong training on Chinese data.
- **Qwen-Turbo**: Smaller, cost-efficient Qwen

Use 1-shot paragraph-level demonstration prompt.

Baseline Prompt Templates

You will be given a chunk from a book in English. Your task is to translate the passage into fluent and natural Chinese.

INSTRUCTIONS:

- Fluency and Naturalness: The overall translation should be fluent, natural, and should read as if it were originally written in Chinese.
- Contextual Integrity: Ensure that the translation maintains the context and meaning of the original passage.
- Formatting: Preserve the original formatting, ensure that each line in the translation corresponds to the same line in the original text.

EXAMPLE:

CHUNK:

It has now been five years since Margaret Thatcher resigned as Britain's Prime Minister. In her heyday she strode the international headlines with such bravura that she seemed inevitable, a natural force.

The world stage seemed just the right size for her, as she chaffed her conservative soul mate Ronald Reagan or flattered the 'new man,' Mikhail Gorbachev. As *The Path to Power* (Harper-Collins; 656 pages; \$30), the second volume of her autobiography, makes clear, Thatcher was probably too simple and direct for the Tories, with their heavy baggage of class and compromise. She traveled light, proud of her roots as a grocer's daughter from the small town of Grantham but never tethered by working-class resentments or delusions of inferiority. Her parents taught her the verities they believed in: Methodism, hard work, thrift and the importance of the individual. She has never wavered from them, and they run through the book.

TRANSLATION OUTPUT, PRESERVING NEWLINE CHARACTER:

玛格丽特·撒切尔从英国首相位置上退下来已经五年了。

在她政治生涯的鼎盛时期，她以光彩照人的风格而成为国际上的新闻人物，她好像必然如此，她是一股自然的力量。在她跟她的保守的精神伙伴罗纳德·里根打趣时，或是在奉承新人米哈伊尔·戈尔巴乔夫时，这个世界看来恰好是适合她驰骋的舞台。正如她的第二本自传《通往权力的路》（哈珀·柯林斯出版社；656页；30美元）清楚指明的那样，对于那些阶级意识很强、善于折中的英国保守党党员，也许撒切尔过于简单，过于直来直去。她没有负担。她以自己是格兰瑟姆小镇一个杂货商的女儿而感到自豪，但她没有被劳动阶级因为地位低下而产生的怨恨或迷惑所束缚。她的父母教她懂得了他们所信仰的真理：卫斯理派。刻苦，节俭，以及个人的重要性。对于这些信仰，她从未动摇过。这些内容贯穿着全书。

YOUR TASK:

TermBase Augment Prompt

You will be given a dictionary of named entities and a chunk from a book in English. Your task is to translate the passage into fluent and natural Chinese, ensuring that each named entity in the text is accurately translated according to the meanings provided in the dictionary.

INSTRUCTIONS:

- Named Entities: The named entities in the passage must be translated exactly as specified in the provided dictionary. The translation should respect the meanings and nuances given for each entity.
- Fluency and Naturalness: The overall translation should be fluent, natural, and should read as if it were originally written in Chinese.
- Contextual Integrity: Ensure that the translation maintains the context and meaning of the original passage while accurately integrating the named entities.
- Formatting: Preserve the original formatting, ensure that each line in the translation corresponds to the same line in the original text.

EXAMPLE:

NAMED ENTITIES TABLE:

```
{
  \"Margaret Thatcher\": \"玛格丽特·撒切尔, former Prime Minister of the United Kingdom\",
  \"Ronald Reagan\": \"罗纳德·里根, former President of the United States\",
  \"Mikhail Gorbachev\": \"米哈伊尔·戈尔巴乔夫, former leader of the Soviet Union\",
  \"The Path to Power\": \"《通往权力的路》, a book written by Margaret Thatcher\",
  \"Harper-Collins\": \"哈珀·柯林斯, a major publishing company\",
  \"Grantham\": \"格兰瑟姆, a town in England\",
  \"Methodism\": \"卫斯理派, a denomination of Protestant Christianity\"
}
```

CHUNK:

It has now been five years since Margaret Thatcher resigned as Britain's Prime Minister. In her heyday she strode the international headlines with such bravura that she seemed inevitable, a natural force. The world stage seemed just the right size for her, as she chaffed her conservative soul mate Ronald Reagan or flattered the 'new man,' Mikhail Gorbachev. As *The Path to Power* (Harper-Collins; 656 pages; \$30), the second volume of her autobiography, makes clear, Thatcher was probably too simple and direct for the Tories, with their heavy baggage of class and compromise. She traveled light, proud of her roots as a grocer's daughter from the small town of Grantham but never tethered by working-class resentments or delusions of inferiority. Her parents taught her the verities they believed in: Methodism, hard work, thrift and the importance of the individual. She has never wavered from them, and they run through the book.

TRANSLATION OUTPUT, PRESERVING NEWLINE CHARACTER:

玛格丽特·撒切尔从英国首相位置上退下来已经五年了。
在她政治生涯的鼎盛时期，她以光彩照人的风格而成为国际上的新闻人物，她好像必然如此，她是一股自然的力量。在她跟她的保守的精神伙伴罗纳德·里根打趣时，或是在奉承新人米哈伊尔·戈尔巴乔夫时，这个世界看来恰好是适合她驰骋的舞台。正如她的第二本自传《通往权力的路》（哈珀·柯林斯出版社：656页；30美元）清楚指明的那样，对于那些阶级意识很强、善于折中的英国保守党党员，也许撒切尔过于简单，过于直来直去。她没有负担，她以自己是格兰瑟姆小镇一个杂货商的女儿而感到自豪，但她没有被劳动阶级因为地位低下而产生的怨恨或迷惑所束缚。她的父母教她懂得了他们所信仰的真理：卫斯理派，刻苦，节俭，以及个人的重要性。对于这些信仰，她从未动摇过。这些内容贯穿着全书。

YOUR TASK:

Comparison of Models

Model	BLEU ↑	COMET ↑	cTT ↑
GPT-4o Mini (Baseline)	20.90	0.8089	0.6127
GPT-4o (Baseline)	23.24	0.8236	0.6453
GPT-4o Mini (TermBase)	23.97	0.8384	0.8660
GPT-4o (TermBase)	25.78	0.8450	0.8750
Qwen-Turbo (Baseline)	21.87	0.8126	0.5762
Qwen-Plus (Baseline)	23.92	0.8236	0.5842
Qwen-Turbo (TermBase)	24.20	0.8360	0.8246
Qwen-Plus (TermBase)	27.18	0.8499	0.8685

Human Analysis

Findings:

- Majority ($\geq 95\%$) of named entities were consistently translated when use TermBase.
- Baseline models had multiple inconsistent translations.

Translations of Lin, Jiage	Frequency
林嘉格	113
林嘉阁	111
林嘉歌	37
林稼歌	23

Human Analysis

Findings:

- Majority ($\geq 95\%$) of named entities were consistently translated when use TermBase.
- Baseline models had multiple inconsistent translations.

Translations of Lin, Jiage	Frequency
林嘉格	0
林嘉阁	0
林嘉歌	301
林稼歌	0

Consistency in Specialized Terms

TermBase maintained consistency in specialized fiction terms like Weapon names, skill names, fictional concepts.

- Very useful for fiction and web-novel translation.

Named Entity	Golden Standard Translation	Translations of Termbase Augment	Translations of Baseline
Frenzied Devil Blade Technique	疯魔刀法	疯魔刀法	疯魔刀法, 狂魔刀法
Psystrike	精神冲击	精神冲击	精神冲击, 精神打击, 模糊的伤害, 心灵打击
the Seven Luminaries Mage Association	七曜法师协会	七曜法师协会	七光辉法师协会, 七位光辉法师协会, 七辉法师协会, 七辉魔法协会

Failure Cases

Some context-specific common nouns were inconsistently translated.

Example:

- *Elementalist* (a specific class)
- *Card* (a category of unreal weapons)

Named Entity	Golden Standard Translation	Translations of Termbase Augment
Elementalist	元素操控师	元素操控师, 元素师, 法术师

Phrases like mantras or catchphrases are less likely to be inconsistently translated. Example

- I Don't Know Everything, I Just Know What I Know. (*Monogatari Series*)
- It's Time To Duel! (*Yu-Gi-Oh!*)

Conclusion

Key Points:

- TermBase augmentation method reduce most ($\geq 95\%$) Terminology inconsistency
- High quality TermBase augmentation improves translation quality, consistency, and readability.
- Extremely useful for fiction and web-novel translation.

A good practice:

- Human translation of named entities followed by LLM translation of the full text.

Sharing of our work

We Shared TermBase-augment method with the SakuraLLM team.

- SakuraLLM is a LLM specify in Japanese to Chinese Translation in Light Novel & Galgame Domain
 - SakuraLLM TermBase-augment support in their v1.0 models.
- 支持术语表(GPT字典), 以保持专有名词和人称的一致性。

Note: Pending comprehensive evaluations

Continue:

- Evaluate more models (e.g. on smaller 7B models).
- Explore other translation directions.
- Thoroughly test the automatic TermBase-augmented pipeline.

Explore:

- Explore other human translation practices (e.g., Translation Memory, Search Engine) for better translation quality and style consistency of literature translation.

Acknowledgments

Thank You:

- Committee Chair: *Mohit Nagaraja Iyyer*
- Committee Members: *James Allan, Benjamin Marlin*
- Project Advisor: *Marzena Karpinska*
- *Sakayanagi* in SakuraLLM Team
- Presentation theme: `github.com/matze/mtheme`

Thank you