Can Large Language Models Translate Unseen Languages in Underrepresented Scripts?

Dianqing Lin[1] **Aruukhan**[1] **Hongxu Hou**[1]*
Shuo Sun[2] **Wei Chen**[1] **Yichen Yang**[1] **Guodong Shi**[1]
[1]College of Computer Science, Inner Mongolia University, China
[2]College of Intelligent Science and Technology,
Inner Mongolia University of Technology, China
lindian7ing@163.com, csaruukhan@mail.imu.edu.cn, cshhx@imu.edu.cn

Abstract

Large language models (LLMs) have demonstrated impressive performance in machine translation, but still struggle with unseen low-resource languages, especially those written in underrepresented scripts. To investigate whether LLMs can translate such languages with the help of linguistic resources, we introduce Lotus, a benchmark designed to evaluate translation for Mongolian (in traditional script) and Yi. Our study shows that while linguistic resources can improve translation quality as measured by automatic metrics, LLMs remain limited in their ability to handle these languages effectively. We hope our work provides insights for the low-resource NLP community and fosters further progress in machine translation for underrepresented script low-resource languages. Our code and data are available[1].

1 Introduction

Large language models (LLMs) have demonstrated remarkable performance in NLP tasks, particularly in machine translation (Brown et al., 2020; Raunak et al., 2023; Vilar et al., 2023; Zhang et al., 2023; Zhu et al., 2024a; Wang et al., 2024). Although these models were not specifically designed for machine translation, they are capable of performing effective translations based on natural language instructions and a few prompt examples, due to the large scale multilingual corpora, predominantly in English, used during their training (Zhu et al., 2024b). However, among the more than 7,000 known languages globally (Joshi et al., 2020; van Esch et al., 2022; Zhang et al., 2024b), the majority of low-resource languages are not represented in the training data of these LLMs. As a result, LLMs often exhibit limited performance in machine translation tasks for low-resource languages, especially those that are unseen.

Figure 1: LLMs struggle to translate unseen, underrepresented languages. Can external linguistic knowledge help?

To further investigate the reasoning abilities of LLMs in the context of machine translation for unseen languages, recent research (Zhang et al., 2024b; Tanzer et al., 2024; Zhang et al., 2024a; Pei et al., 2025) has proposed incorporating external linguistic resources such as bilingual dictionaries, grammar descriptions, and parallel corpora into prompts that emulate human second language learning processes. Although previous studies have made impressive progress, they have primarily focused on unseen languages with Latin scripts.

We argue that machine translation for unseen languages written in underrepresented scripts presents greater challenges for LLMs than for unseen languages in Latin scripts. This is mainly due to three reasons. Firstly, using bilingual dictionaries as prompts to enhance machine translation for unseen languages with underrepresented scripts is not as straightforward as for languages written in Latin

* Corresponding author.
[1]https://github.com/csdq777/Lotus

Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, pages 23149–23162
November 4-9, 2025

scripts with natural word boundaries. Many of these scripts lack explicit word boundary markers and do not have readily available word-level processing tools. Second, from the perspective of resource acquisition, Latin script languages often benefit from mature OCR tools that facilitate the extraction of data from digitized linguistic materials. Although recent work suggests that grammar books alone may have limited impact on improving translation for unseen languages (Aycock et al., 2025), the availability of OCR tools still allows Latin script languages to transform such resources into usable parallel corpora. In contrast, languages with underrepresented scripts often lack OCR support and digitized resources (Ignat et al., 2022), and manual annotation may require learning input methods for those scripts (Ding et al., 2019). Third, most available parallel corpora are concentrated in Latin script languages, leaving those written in underrepresented scripts largely absent from the pretraining of LLMs, which in turn results in degraded generation performance (Bang et al., 2023; Ahuja et al., 2023). Therefore, we pose the following interesting question: Can LLMs translate unseen languages in underrepresented scripts? As illustrated in Figure 1.

In this paper, we introduce Lotus (Large **L**anguage Models **O**n **T**ranslation of Unseen and **U**nderrepresented **S**cripts) , a benchmark designed to explore whether LLMs can translate languages written in underrepresented scripts that they have not seen before. Specifically, Lotus focuses on two challenging cases: Mongolian in traditional script and Yi, both of which lack representation in mainstream NLP resources. Our experiments demonstrate that, similar to unseen Latin script languages, these underrepresented script languages can also benefit from external linguistic resources, leading to improved translation performance. However, the overall translation ability of LLMs remains limited. In particular, the model shows only basic translation capability for Mongolian, while effective translation for Yi is not yet achievable.

Our contributions are as follows:(1) We introduce Lotus, a benchmark for exploring whether LLMs can translate unseen languages written in underrepresented scripts. (2) We conduct evaluations across multiple LLMs and provide detailed analysis of the results, offering insights for the low-resource NLP community, especially for languages with underrepresented scripts.

2 Linguistic Resources and Word-Level Processing Tools

2.1 Bilingual Dictionaries

The Mongolian-Chinese dictionary was sourced from Hitoshi Kuribayashi's website[2], which provides a comprehensive collection of bilingual entries. To obtain the reverse direction, we inverted the dictionary pairs to construct a Chinese-Mongolian dictionary. After processing and cleaning, we obtained 15,014 entries in the Mongolian-Chinese direction and 30,069 entries in the Chinese-Mongolian direction.

The Yi-Chinese dictionary was collected from the Glosbe website[3], a collaborative dictionary platform with contributions from native speakers. Glosbe supports bidirectional lookup between language pairs. However, due to inconsistencies in the bidirectional data, certain entries appear in only one direction. To address this asymmetry, we collected both the Yi-Chinese and Chinese-Yi dictionaries separately and then converted each into its reverse direction. After merging and deduplication, we obtained 14,854 entries in the Yi-Chinese direction and 28,626 in the Chinese-Yi direction.

2.2 Parallel Corpora

Given the scarcity of low-resource languages and the difficulty of collecting their data, we constructed parallel sentence pairs from multiple sources, including news websites, WeChat public accounts, bilingual dictionary example sentences, language learning books, and a small amount of manually annotated data. In total, we obtained 5,461 Mongolian-Chinese sentence pairs and 5,421 Yi-Chinese sentence pairs. More detailed statistics are provided in the appendix B.

2.3 Word-Level Processing Tools

To incorporate bilingual dictionary entries into LLM prompts, we first perform word-level processing on the source sentence. Unlike Latin script languages, which mark word boundaries with whitespace, the two languages in our study lack explicit delimiters, requiring additional preprocessing for segmentation.

Mongolian Tokenizer Mongolian is an agglutinative language, with words typically composed of a root followed by one or more suffixes that

[2] http://hkuri.cneas.tohoku.ac.jp/
[3] https://glosbe.com/

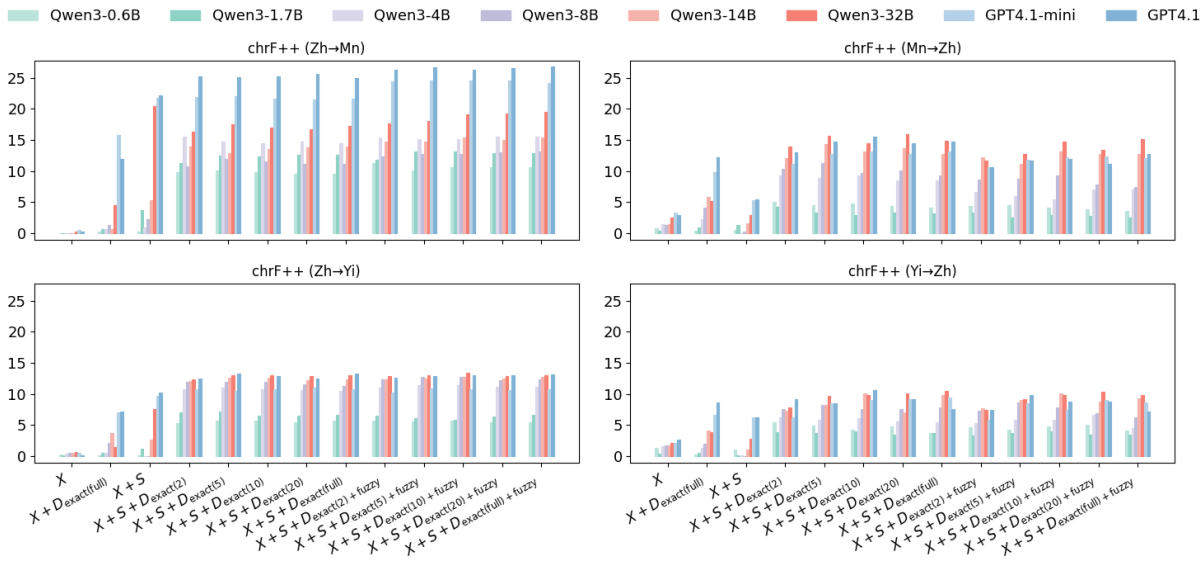

Figure 2: This table compares different methods across a range of models on the Lotus benchmark test set, with chrF++ as the evaluation metric.We denote X as the source sentence, S as the top 3 retrieved parallel examples, $D_{\text{exact}(n)}$ as exact matches with n entries per word, and $D_{\text{fuzzy}(k,n)}$ as fuzzy matches with n entries for the top k unmatched words.

indicate grammatical functions. In traditional Mongolian script, a specific type of suffix known as "Mongolian word additional components" appears visually separated by whitespace-like characters, although they differ in Unicode encoding. If not properly removed, these components can interfere with dictionary-based lookup. To address this, we developed a custom tokenizer, implemented by one of the authors who is a native speaker of Mongolian. This tokenizer segments words, removes additional components, and preserves stems, improving the overall accuracy of tokenization.

Yi Word Segmenter Yi is an isolating language with limited morphological variation. However, like other isolating languages such as Chinese and Thai, its script does not explicitly mark word boundaries, which makes word-level segmentation challenging. To address this, we followed the approach of Muxia (2018) by first extracting terms from their Yi segmentation thesis, and then expanded the dictionary with bilingual word pairs from our data. Using this dictionary as a lexicon, we implemented a basic Yi word segmenter designed to support our experiments.

3 Prompt Formalization

We formalize our linguistic resources as follows:

Source sentence X The input consists solely of a source language sentence.

Bilingual Dictionary D The bilingual dictionary D consists of bilingual words, each with multiple entries (translations or meanings). We adopt the following two retrieval strategies for the entries in the dictionary:

- **Exact matching** ($D_{exact(n)}$): For each word in the source sentence that appears in the dictionary, we retrieve and expose the top n entries. For instance, $D_{exact(2)}$ means showing two dictionary entries per matched word.
- **Fuzzy matching** ($D_{\text{fuzzy}(k,n)}$): Given that both languages under study are low-resource, and that Mongolian exhibits rich morphological complexity as an agglutinative language, exact dictionary coverage is often insufficient. To address this, we following DIPMT++ (Zhang et al., 2024a) employ a two-stage retrieval strategy: exact matching followed by fuzzy matching via BM25 (Robertson and Zaragoza, 2009). We treat each unmatched word as a query and each dictionary word as a document in the retrieval process. The top k most similar entries are retrieved and n dictionary translations per match are exposed in the prompt. To control the noise introduced by fuzzy retrieval, we restrict it to $k = 2$ and $n = 2$.

Parallel sentence S We retrieve a small number of parallel sentence examples from a parallel cor-

pus using BM25 (Robertson and Zaragoza, 2009). We adopt a 3-shot setup, where the top 3 most relevant sentence pairs are selected and prepended to the prompt as translation demonstrations.

We formalize our prompt using a structured template, and adopt the prompt template designed by (Zhang et al., 2024a,b), where the input consists of linguistic resources and the output Y is generated by a LLM. Specifically, the input includes:

(1) X, where only the source sentence is provided. (2) $X + D_{exact(full)}$, where the source sentenc is provided along with each word in the sentence that can be exactly retrieved, along with all its entries. (3) $X + S$, where 3-shot most relevant sentence pairs are retrieved from the parallel corpus and then the source sentence is presented for translation. (4) $X + S + D_{exact(n)}$, The 3-shot most relevant sentence pairs are retrieved from the parallel corpus, and for each source sentence, the top n dictionary entries for words that can be exactly matched are retrieved. (5) $X + S + D_{exact(n)+fuzzy}$, which extends (4) by applying fuzzy retrieval: if a word has no exact match in the dictionary, the top k = 2 most similar words are retrieved, with n = 2 entries provided for each. See the prompt template in Appendix F.

4 Experiments

4.1 Experimental Setup

Models We use nine models in our experiments: GPT-4.1, GPT-4.1-mini (Achiam et al., 2023), and the full Qwen3-Dense series (Yang et al., 2025) with models of 0.6B, 1.7B, 4B, 8B, 14B, and 32B parameters.

Metrics Our main evaluation metric is chrF++ (Popović, 2017), which is more appropriate for low-resource languages (Aycock et al., 2025). We also report SacreBLEU (Post, 2018) scores for completeness. Further details can be found in the appendix E.

Test Set We construct our test set by randomly sampling from our collected corpus, following the design of the ZhuangBench (Zhang et al., 2024a). Sentences are categorized into three levels low, medium, and high based on their length, with 200 sentences sampled for each level.

4.2 Results and Analyses

We present our experimental results in Figure 2. The experimental results confirm that incorporating external linguistic resources improves the translation performance of both Mongolian and Yi, two languages with underrepresented scripts. Using chrF++ as the primary evaluation metric, we observe that for Mongolian to Chinese translation, chrF++ increases from 2.5 with the X to 16 with the $X + S + D_{exact(20)}$ on the Qwen3-32B model. For Chinese to Mongolian, GPT-4.1 shows an increase from 0.3 with the X to 26.8 with the $X + S + D_{exact(full)+fuzzy}$. For Yi to Chinese, GPT-4.1 improves from 2.3 with the X to 10.6 with the $X + S + D_{exact(10)}$, while for Chinese to Yi, Qwen3-32B sees an increase from 0.7 with the X to 13.4 with the $X + S + D_{exact(10)+fuzzy}$. However, upon analyzing specific cases, we find that successful translations for Yi often involve simple sentences or those closely related to in-context examples. While linguistic resources clearly provide improvements for Yi, they do not yet support reliable and general-purpose translation. In contrast, Mongolian can be considered to have achieved basic machine translation capabilities. BLEU scores are also reported in the Appendix E.

Dictionary Retrieval Strategies In Mongolian-Chinese translation with basic translation capabilities, for variants of the $X + S + D$, we observe that for both target languages (Chinese and Mongolian), the better performance of each variant is often influenced by a particular retrieval strategy. Chinese is an isolating language, and when used as the target language, exact matching generally outperforms the combination of exact and fuzzy matching. In contrast, Mongolian is an agglutinative language with complex morphological variations. In this case, exact matching alone often fails to cover all word forms, and incorporating fuzzy matching leads to better translation performance. These results suggest that the choice of dictionary retrieval strategy should be informed by the morphological characteristics of the target language.

Dictionary Entry Coverage In Mongolian-Chinese translation with basic translation capabilities, for variants of the $X + S + D$, we observe that when the model has a large number of parameters and the target language is Chinese, increasing the number of exact match entries leads to more significant improvements. In contrast, for smaller models with lower baseline chrF++ scores, even slight improvements in automatic metrics do not result in meaningful improvements in translation quality. We attribute this to the severe lack of lin-

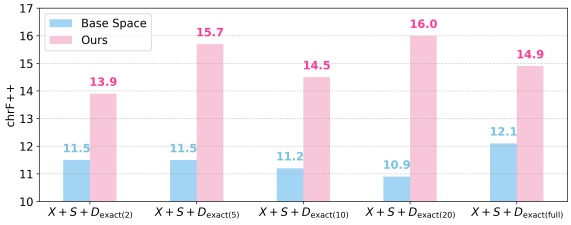

Figure 3: Comparison of different Mongolian word segmentation strategies.

guistic resources for Mongolian, specifically the Mongolian written in the traditional script used in our study, which makes it difficult for LLMs to acquire basic knowledge of Mongolian and to accurately generate or select appropriate target words.

Word-Level Processing For Mongolian tokenization, we conducted comparative experiments between two approaches: the standard space-based tokenization method and our proposed method, which performs tokenization by removing Mongolian word additional components. In the translation direction from Mongolian to Chinese, where the source language is a morphologically rich agglutinative language, effective tokenization is particularly important. Based on our previous experimental results and analysis, we find that in this setting, where the target language is high-resource Chinese, dictionary strategies based on exact matching yield the most substantial improvements. Therefore, we conducted experiments using the Qwen3-32B model under the $X + S + D_{exact(n)}$ setting, which combines tokenization and multi-entry dictionary prompts. As shown in the Figure 3, our method achieves better performance across multiple evaluation metrics, demonstrating the effectiveness of affix removal followed by tokenization for Mongolian.

However, for Yi language segmentation, even though Yi has not achieved basic machine translation capabilities in our experiments, we find that by comparing the $X + S$ and $X + S + D$, when segmentation is not performed, the $X+S$ primarily relies on pattern matching and repeats the most recent examples as output. In contrast, the $X+S+D$ method is able to achieve correct translation at least when handling simple test sentences, by combining the words in D that are correctly segmented and exactly missing, and the examples in S that are highly relevant to the simple test sentence. Given the additional complexity and challenges faced by Yi segmentation, where the Yi script lacks explicit word boundary markers, even a person who has never studied Yi would get worse results when translating by looking up a dictionary, compared to languages with natural word boundaries, As stated in the human evaluation section of Appendix D. We believe this situation further highlights the severe limitations faced by languages like Yi, which are both extremely low-resource and lack natural word boundary cues. These challenges underscore the urgent need for future research to focus on such languages and develop dedicated segmentation tools and preprocessing techniques tailored to their unique characteristics.

5 Conclusion

In this paper, we investigate whether LLMs can translate unseen and underrepresented script languages. Through experiments with state-of-the-art LLMs and advanced techniques, we find that, similar to previous studies, these models can improve automatic translation metrics by incorporating external linguistic resources. However, unlike prior work, we observe that machine translation for these unrepresentative scripts still faces significant limitations: for Mongolian, only basic translation capabilities are achieved, while for Yi, effective machine translation remains unattainable. We hope our work offers a new perspective for future research and contributes to advancing NLP support for the low-resource language community.

Limitations

Corpus Scope Given the scarcity of available resources, our study focuses on two underrepresented and unseen languages in LLMs, rather than including other underrepresented and unseen script languages. We encourage future work to further expand the inclusion of such underrepresented languages in the NLP community..

Yi Word Segmenter Although we adopted a method proposed by a Yi scholar (Muxia, 2018), we acknowledge that there may be issues in the Yi language segmentation process, as there were no Yi speakers involved in the experiment. At the same time, we call on the community to pay more attention to low-resource languages like Yi, which lack natural word boundaries, in order to promote linguistic diversity in NLP.

Acknowledgments

The fund of Supporting the Reform and Development of Local Universities (Disciplinary Construction) and the special research project of First-class Discipline of Inner Mongolia A. R. of China under Grant YLXKZX-ND-036.

We would like to sincerely thank Pusheng An, Asur, Haifeng Bai, Zhiyu Dou, Dengyun Liu, Tian Lan, Xue Wang, and Xiangyue Zhang for their valuable contributions. The names are listed in alphabetical order (based on the conventions in Chinese), without implying any ranking of contributions.

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, and 1 others. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Kabir Ahuja, Harshita Diddee, Rishav Hada, Millicent Ochieng, Krithika Ramesh, Prachi Jain, Akshay Nambi, Tanuja Ganu, Sameer Segal, Mohamed Ahmed, Kalika Bali, and Sunayana Sitaram. 2023. MEGA: Multilingual evaluation of generative AI. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 4232–4267, Singapore. Association for Computational Linguistics.

Seth Aycock, David Stap, Di Wu, Christof Monz, and Khalil Sima'an. 2025. Can LLMs really learn to translate a low-resource language from one grammar book? In *The Thirteenth International Conference on Learning Representations*.

Yejin Bang, Samuel Cahyawijaya, Nayeon Lee, Wenliang Dai, Dan Su, Bryan Wilie, Holy Lovenia, Ziwei Ji, Tiezheng Yu, Willy Chung, Quyet V. Do, Yan Xu, and Pascale Fung. 2023. A multitask, multilingual, multimodal evaluation of ChatGPT on reasoning, hallucination, and interactivity. In *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 675–718, Nusa Dua, Bali. Association for Computational Linguistics.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, and 12 others. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

Chenchen Ding, Masao Utiyama, and Eiichiro Sumita. 2019. MY-AKKHARA: A Romanization-based Burmese (Myanmar) input method. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations*, pages 157–162, Hong Kong, China. Association for Computational Linguistics.

Oana Ignat, Jean Maillard, Vishrav Chaudhary, and Francisco Guzmán. 2022. OCR improves machine translation for low-resource languages. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 1164–1174, Dublin, Ireland. Association for Computational Linguistics.

Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. 2020. The state and fate of linguistic diversity and inclusion in the NLP world. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6282–6293, Online. Association for Computational Linguistics.

Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph Gonzalez, Hao Zhang, and Ion Stoica. 2023. Efficient memory management for large language model serving with pagedattention. In *Proceedings of the 29th symposium on operating systems principles*, pages 611–626.

Abie Muxia. 2018. Research on yi script dictionary-based word segmentation technology using python. Master's thesis, Southwest Minzu University, Chengdu, China. In Chinese.

Renhao Pei, Yihong Liu, Peiqin Lin, François Yvon, and Hinrich Schuetze. 2025. Understanding in-context machine translation for low-resource languages: A case study on Manchu. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8767–8788, Vienna, Austria. Association for Computational Linguistics.

Maja Popović. 2017. chrF++: words helping character n-grams. In *Proceedings of the Second Conference on Machine Translation*, pages 612–618, Copenhagen, Denmark. Association for Computational Linguistics.

Matt Post. 2018. A call for clarity in reporting BLEU scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Brussels, Belgium. Association for Computational Linguistics.

Vikas Raunak, Arul Menezes, and Hany Awadalla. 2023. Dissecting in-context learning of translations in GPT-3. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 866–872, Singapore. Association for Computational Linguistics.

Stephen Robertson and Hugo Zaragoza. 2009. The probabilistic relevance framework: Bm25 and beyond. *Found. Trends Inf. Retr.*, 3(4):333–389.

Garrett Tanzer, Mirac Suzgun, Eline Visser, Dan Jurafsky, and Luke Melas-Kyriazi. 2024. A benchmark for learning to translate a new language from one grammar book. In *The Twelfth International Conference on Learning Representations*.

Daan van Esch, Tamar Lucassen, Sebastian Ruder, Isaac Caswell, and Clara Rivera. 2022. Writing system and speaker metadata for 2,800+ language varieties. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 5035–5046, Marseille, France. European Language Resources Association.

David Vilar, Markus Freitag, Colin Cherry, Jiaming Luo, Viresh Ratnakar, and George Foster. 2023. Prompting PaLM for translation: Assessing strategies and performance. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15406–15427.

Shanshan Wang, Derek Wong, Jingming Yao, and Lidia Chao. 2024. What is the best way for ChatGPT to translate poetry? In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 14025–14043, Bangkok, Thailand. Association for Computational Linguistics.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, and 41 others. 2025. Qwen3 technical report. *Preprint*, arXiv:2505.09388.

Biao Zhang, Barry Haddow, and Alexandra Birch. 2023. Prompting large language model for machine translation: A case study. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 41092–41110. PMLR.

Chen Zhang, Xiao Liu, Jiuheng Lin, and Yansong Feng. 2024a. Teaching large language models an unseen language on the fly. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 8783–8800, Bangkok, Thailand. Association for Computational Linguistics.

Kexun Zhang, Yee Choi, Zhenqiao Song, Taiqi He, William Yang Wang, and Lei Li. 2024b. Hire a linguist!: Learning endangered languages in LLMs with in-context linguistic descriptions. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 15654–15669, Bangkok, Thailand. Association for Computational Linguistics.

Shaolin Zhu, Menglong Cui, and Deyi Xiong. 2024a. Towards robust in-context learning for machine translation with large language models. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 16619–16629, Torino, Italia. ELRA and ICCL.

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

Figure 4: Different scripts in different languages

A Background

To explore issues related to script diversity, we focus on two underrepresented script languages: Mongolian and Yi. Both are spoken by ethnic minorities in China and are largely unseen by current LLMs. As a result, they are significantly underrepresented in NLP research. Their unique linguistic and script characteristics make them valuable for studying challenges in multilingual language processing. Figure 4 provides a comparative view of different scripts used in these languages.

The Mongols An ethnic minority in northern China and the predominant ethnic group in Mongolia, the Mongols primarily speak Mongolian (ISO 639-1: mn; ISO 639-3: mon). This language, belonging to the Mongolic language family, is agglutinative and features two major writing systems: the Cyrillic Mongolian script, widely used in Mongolia, and the traditional Mongolian script, primarily employed in Inner Mongolia, China. This study focuses on the traditional Mongolian script. Mongolian has clear word segmentation boundaries, making it structurally distinct from Yi. Figure 1 shows the visual representation of the Mongolian script on computers, while Figure 4 illustrates its actual vertical writing form.

The Yi People An ethnic minority in southwestern China, the Yi people are primarily concentrated in Sichuan, Yunnan, and Guizhou, with a population of 9.8 million. The Yi language (ISO 639-1: ii; ISO 639-3: iii), part of the Sino-Tibetan language family, is an analytic language that uses the Standard Yi script as its official writing system. Yi encompasses multiple dialects, and this study focuses on the Northern dialect, particularly the Liangshan Standard Yi. The Yi language is spoken by a minority ethnic group in China, and it lacks explicit word segmentation boundaries, which poses challenges for NLP tasks.

Target	Dict	Rev. Dict	Train	Test
Mongolian	30,069	15,014	5,261	200
Yi	28,626	14,854	5,221	200

Table 1: Statistics of the Lotus benchmark. **Dict** refers to the size of the bilingual dictionary. **Rev. Dict** refers to the reverse dictionary. **Train** and **Test** indicate the number of sentence pairs in the parallel data.

B Data statics

In this section, we provide detailed statistics on the data we collected.

Statistic of the Lotus Data In Table 1, we report the data statistics of the Lotus Data Components.

Statistic of the Train Set In Table 2, we report the data statistics of the Train set. In Table 4, we report the sources of the parallel corpus data.

Statistic of the Test Set Table 3 presents the statistics of the Lotus test set.

C Implementation Details

For the open-source models, including Qwen3-32B, Qwen3-14B, Qwen3-8B, Qwen3-4B, Qwen3-1.7B, and Qwen3-0.6B, we deployed them on 2 NVIDIA H20 GPUs with 96 GB of memory, using vLLM (Kwon et al., 2023) for efficient inference. Decoding was performed using a temperature of 0. For the GPT series models, we accessed them via API and used the default settings.

D Human Evaluation

We conducted human evaluation, and the results are presented in Table 9. Four participants who had never studied Mongolian or Yi were recruited as non-native evaluators, and two native Mongolian speakers participated as native evaluators. For GPT-4.1, we adopt $X + S + D$, where Chinese and Yi are target languages with $X + S + D_{exact(full)}$, while Mongolian is $X + S + D_{exact(n)} + fuzzy$.

For the non-native evaluators, we provided all linguistic resources, except for the test set, along with a brief grammar guide of no more than 500 words to facilitate understanding of the target languages. During the evaluation, participants were allowed to use search engines to look up information but were not permitted to use any translation software or LLMs. After the evaluation, the first author conducted short interviews with the participants.

Statistic	Mongolian-Chinese	Yi-Chinese
Training instances	5,261	5,221
Avg. Chinese length (char)	35.3	30.9
Avg. Chinese length (word)	19.2	17.7
Avg. Target length (word)	32.4 (Mongolian)	24.8 (Yi)
Avg. Target length (char)	–	37.3

Table 2: Statistics of the Mongolian-Chinese and Yi-Chinese parallel training sets. Character- and word-level statistics are computed separately for source (Chinese) and target languages.

Category	Easy	Medium	Hard
Mongolian-Chinese			
Test Set Instances	75	60	65
Average Chinese Length (char)	10.7	24.6	42.0
Average Chinese Length (word)	6.8	14.5	25.0
Average Mongolian Length (word)	7.9	21.0	30.5
Yi-Chinese			
Test Set Instances	75	60	65
Average Chinese Length (char)	11.0	24.3	41.3
Average Chinese Length (word)	7.2	15.2	25.8
Average Yi Length (char)	12.3	28.6	49.1
Average Yi Length (word)	8.9	19.1	31.0

Table 3: Test set statistics in the Lotus benchmark. **Easy**, **Medium**, and **Hard** denote difficulty levels. Lengths are measured by character and word for both source (Chinese) and target (Mongolian or Yi) sentences.

We found that the scripts of underrepresented languages differ significantly from Latin-based scripts. For most *ACL readers and authors, English is at least readable, and its 26-letter alphabet is relatively easy to recognize. In contrast, some participants were entirely unfamiliar with the Mongolian script. The Yi script posed even greater challenges, as it lacks explicit word boundaries, making segmentation and comprehension more difficult. These challenges slowed down the evaluation process. The first author recorded the time each annotator spent on the task. Even for only 30 sentences, the fastest participant required more than 90 minutes, excluding preparation time. By contrast, GPT-4.1 completed bidirectional translation for the same set in under five minutes.

For the native evaluators, results showed that native speakers substantially outperformed GPT-4.1. In terms of automatic metrics, chrF++ appeared more suitable for Chinese–Mongolian translation, while BLEU provided a more reasonable assessment for Mongolian–Chinese. The first author, a native speaker of Chinese, carefully compared the Mongolian–Chinese translations produced by native speakers with the reference translations. Although the wording and word order often differed, the meanings were consistently preserved.

E Evaluation Metrics

We additionally report SacreBLEU scores for all four translation directions. The results are shown in Tables 5,6,7,8.

Target	Type	Source	Description	Count
Mongolian	News	Gov. reports	Chinese government work reports	3402
Mongolian	Dictionary	Lexicon	Mongolian-Chinese dictionary[4]	1652
Mongolian	Book	Textbook	*A Practical Handbook for Learning Mongolian*	207
Mongolian	Manual	Annotation	200 sentences generated by GPT-4.0, manually annotated	200
Yi language	News	Gov. reports	Chinese government work reports	2712
Yi language	Social	WeChat	WeChat public platforms[5]	1491
Yi language	Dictionary	Lexicon	Glosbe website	239
Yi language	Book	Textbook	*600 Sentences in Liangshan Yi Conversation*[6]	167

Table 4: Sources and statistics of the Mongolian-Chinese and Yi-Chinese parallel corpora in the Lotus benchmark.

Method	Qwen3-32B	Qwen3-14B	Qwen3-8B	Qwen3-4B	Qwen3-1.7B	Qwen3-0.6B	GPT-4.1	GPT-4.1 mini
X	1.2 / 2.5	0.3 / 1.5	0.1 / 1.3	0.3 / 1.5	0.0 / 0.4	0.1 / 0.8	1.0 / 3.0	1.7 / 3.3
$X + D_{\text{exact(full)}}$	4.4 / 5.2	3.4 / 5.9	3.3 / 4.1	1.1 / 2.3	0.7 / 1.0	0.0 / 0.4	11.8 / 12.2	9.0 / 9.9
$X + S$	1.7 / 2.9	0.7 / 1.6	0.0 / 0.3	0.0 / 0.0	0.0 / 1.3	0.0 / 0.6	5.3 / 5.5	5.4 / 5.3
$X + S + D_{\text{exact(2)}}$	14.7 / 13.9	12.5 / 12.1	11.5 / 10.4	**7.8 / 9.3**	**1.4 / 4.3**	**2.1 / 5.0**	15.5 / 13.0	12.2 / 11.1
$X + S + D_{\text{exact(5)}}$	20.2 / 15.7	**17.9 / 14.3**	**13.2 / 11.3**	5.9 / 8.9	0.9 / 3.3	1.7 / 4.5	17.7 / **14.7**	14.9 / 12.8
$X + S + D_{\text{exact(10)}}$	18.3 / 14.5	14.1 / 13.2	7.0 / 9.7	6.5 / **9.3**	0.8 / 3.0	1.9 / 4.8	**19.5** / 15.5	**15.6** / 13.1
$X + S + D_{\text{exact(20)}}$	**20.7 / 16.0**	17.1 / 13.7	9.5 / 10.1	5.2 / 8.5	0.9 / 3.3	1.5 / 4.4	17.8 / 14.5	**15.6** / 12.8
$X + S + D_{\text{exact(full)}}$	15.6 / 14.9	13.3 / 12.8	6.5 / 9.3	5.0 / 8.5	0.8 / 3.2	1.4 / 4.1	18.6 /**14.7**	15.3 / **13.2**
$X + S + D_{\text{exact(2)+fuzzy}}$	11.6 / 11.7	14.7 / 12.2	5.8 / 8.6	3.2 / 6.7	1.0 / 3.4	1.8 / 4.4	12.0 / 10.6	10.0 / 10.6
$X + S + D_{\text{exact(5)+fuzzy}}$	12.7 / 12.8	8.9 / 11.2	5.7 / 8.8	2.3 / 6.0	0.7 / 2.6	**2.1** / 4.6	13.5 / 11.7	11.6 / 11.8
$X + S + D_{\text{exact(10)+fuzzy}}$	15.5 / 14.7	15.7 / 13.1	6.5 / 9.3	1.9 / 5.4	0.8 / 2.9	1.6 / 4.2	13.6 / 11.9	12.5 / 12.2
$X + S + D_{\text{exact(20)+fuzzy}}$	11.7 / 13.4	15.8 / 12.8	4.2 / 7.9	3.3 / 7.0	0.8 / 2.8	1.5 / 3.9	13.8 / 11.1	12.7 / 12.4
$X + S + D_{\text{exact(full)+fuzzy}}$	18.3 / 15.1	15.5 / 12.7	3.9 / 7.5	3.4 / 7.1	0.6 / 2.5	1.3 / 3.6	15.4 / 12.8	12.3 / 12.1

Table 5: BLEU / chrF++ scores for Mongolian to Chinese (mn→zh) translation across different prompting strategies and model sizes.

Method	Qwen3-32B	Qwen3-14B	Qwen3-8B	Qwen3-4B	Qwen3-1.7B	Qwen3-0.6B	GPT-4.1	GPT-4.1 mini
X	0.0 / 0.3	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.3	0.0 / 0.6
$X + D_{\text{exact(full)}}$	0.0 / 4.5	0.0 / 0.7	0.0 / 1.3	0.0 / 0.7	0.0 / 0.7	0.0 / 0.3	0.0 / 11.9	0.1 / 15.8
$X + S$	**3.1 / 20.4**	0.0 / 5.3	0.0 / 2.3	0.0 / 1.0	0.0 / 3.7	0.0 / 0.3	**2.9** / 22.2	**2.9** / 21.8
$X + S + D_{\text{exact(2)}}$	1.3 / 16.3	0.7 / 13.9	0.3 / 10.8	**0.6** / 15.5	0.4 / 11.3	0.3 / 9.9	2.5 / 25.2	0.6 / 21.9
$X + S + D_{\text{exact(5)}}$	1.9 / 17.5	0.7 / 12.9	0.3 / 12.0	0.4 / 14.7	0.5 / 12.5	0.4 / 10.1	2.4 / 25.1	0.9 / 22.1
$X + S + D_{\text{exact(10)}}$	2.0 / 17.0	0.8 / 13.5	0.3 / 11.5	0.5 / 14.5	0.4 / 12.4	0.3 / 9.9	2.4 / 25.2	0.9 / 21.6
$X + S + D_{\text{exact(20)}}$	1.8 / 16.8	0.8 / 13.8	0.3 / 11.1	0.5 / 14.8	0.4 / 12.6	0.3 / 9.6	**2.9** / 25.6	0.7 / 21.5
$X + S + D_{\text{exact(full)}}$	1.9 / 17.2	0.8 / 13.9	0.3 / 11.1	0.5 / 14.5	0.4 / 12.6	0.3 / 9.6	2.5 / 24.9	0.7 / 21.7
$X + S + D_{\text{exact(2)+fuzzy}}$	1.9 / 17.7	0.6 / 14.7	0.4 / 12.3	0.5 / 15.4	0.4 / 11.8	0.3 / **11.3**	2.2 / 26.3	0.7 / 24.4
$X + S + D_{\text{exact(5)+fuzzy}}$	2.1 / 18.1	0.8 / 14.7	0.4 / 12.7	0.4 / 15.2	0.4 / **13.2**	0.3 / 10.1	2.5 / 26.7	0.8 / 24.5
$X + S + D_{\text{exact(10)+fuzzy}}$	2.4 / 19.1	**0.9 / 15.4**	0.4 / 12.7	**0.6** / 15.2	0.5 / **13.2**	0.4 / 10.6	2.4 / 26.3	1.0 / 24.5
$X + S + D_{\text{exact(20)+fuzzy}}$	2.6 / 19.2	0.8 / 15.0	0.4 / 13.0	0.5 / 15.5	0.4 / 12.9	0.4 / 10.7	2.1 / 26.5	0.6 / **24.6**
$X + S + D_{\text{exact(full)+fuzzy}}$	2.7 / 19.5	0.8 / **15.4**	0.4 / **13.2**	0.5 / 15.5	0.4 / 12.9	0.3 / 10.6	2.8 / **26.8**	1.0 / 24.1

Table 6: BLEU / chrF++ scores for Chinese to Mongolian (zh→mn) translation across different prompting strategies and model sizes.

Method	Qwen3-32B	Qwen3-14B	Qwen3-8B	Qwen3-4B	Qwen3-1.7B	Qwen3-0.6B	GPT-4.1	GPT-4.1 mini
X	0.5 / 2.2	0.3 / 1.7	0.2 / 1.8	0.3 / 1.6	0.0 / 0.4	0.1 / 1.3	0.7 / 2.7	0.5 / 2.2
$X + D_{\text{exact(full)}}$	2.8 / 3.9	1.8 / 4.1	1.3 / 2.0	0.5 / 1.3	0.0 / 0.5	0.1 / 0.3	7.2 / 8.6	4.8 / 6.6
$X + S$	2.5 / 2.8	0.0 / 1.1	0.0 / 0.0	0.0 / 0.2	0.0 / 0.2	0.1 / 1.1	7.8 / 6.2	8.0 / 6.2
$X + S + D_{\text{exact(2)}}$	8.0 / 7.8	7.1 / 7.3	**8.8** / 7.6	**3.6** / 6.3	1.3 / 3.9	**3.5 /5.5**	10.0 / 9.2	6.6 / 6.2
$X + S + D_{\text{exact(5)}}$	13.5 / 9.7	8.4 / 8.2	6.1 / 8.2	2.5 / 5.9	1.4 / 3.7	2.3 / 4.9	10.1 / 8.5	8.6 / 8.5
$X + S + D_{\text{exact(10)}}$	13.8 / 9.9	10.7 / **10.1**	4.6 / 7.6	2.7 / 6.1	**1.6 / 4.0**	2.3 / 4.3	**11.4 / 10.6**	8.4 / 9.0
$X + S + D_{\text{exact(20)}}$	11.3 / 10.1	4.5 / 7.0	4.8 / 7.6	2.2 / 5.6	1.2 / 3.5	2.2 / 4.8	10.6 / 9.2	8.8 / 9.2
$X + S + D_{\text{exact(full)}}$	13.7 / **10.5**	8.6 / 9.8	5.1 / 7.8	2.1 / 5.4	1.5 / 3.8	1.5 / 3.8	9.8 / 7.6	**9.3 / 9.5**
$X + S + D_{\text{exact(2)+fuzzy}}$	8.5 / 7.4	9.8 / 7.7	4.5 / 7.3	2.9 / 5.3	1.0 / 3.4	2.3 / 4.6	8.9 / 7.5	5.9 / 5.9
$X + S + D_{\text{exact(5)+fuzzy}}$	10.6 / 9.2	**11.7** / 9.1	6.6 / **8.6**	2.5 / 5.8	1.2 / 3.8	2.3 / 4.3	10.2 / 9.8	7.6 / 8.5
$X + S + D_{\text{exact(10)+fuzzy}}$	13.0 / 9.8	11.0 / **10.1**	5.2 / 7.8	3.4 / 5.9	1.2 / 4.0	2.4 / 4.8	10.4 / 8.8	7.5 / 7.4
$X + S + D_{\text{exact(20)+fuzzy}}$	**14.2** / 10.4	9.3 / 8.8	3.6 / 6.9	3.1 / **6.6**	1.1 / 3.5	2.6 / 5.0	10.3 / 8.8	8.1 / 9.0
$X + S + D_{\text{exact(full)+fuzzy}}$	12.2 / 9.9	6.9 / 9.3	3.1 / 6.3	1.9 / 4.5	1.2 / 3.5	2.0 / 4.2	9.4 / 7.2	7.7 / 8.7

Table 7: BLEU / chrF++ scores for Yi to Chinese (yi→zh) translation across different prompting strategies and model sizes.

Method	Qwen3-32B	Qwen3-14B	Qwen3-8B	Qwen3-4B	Qwen3-1.7B	Qwen3-0.6B	GPT-4.1	GPT-4.1 mini
X	0.0 / 0.7	0.0 / 0.5	0.0 / 0.6	0.0 / 0.4	0.0 / 0.2	0.0 / 0.3	0.0 / 0.2	0.0 / 0.5
$X + D_{\text{exact(full)}}$	0.0 / 1.5	0.1 / 3.8	0.1 / 2.1	0.0 / 0.6	0.0 / 0.6	0.0 / 0.1	0.3 / 7.2	0.3 / 7.1
$X + S$	2.4 / 7.6	0.6 / 2.7	0.0 / 0.0	0.0 / 0.0	0.3 / 1.2	0.0 / 0.2	3.6 / 10.2	**3.5** / 9.7
$X + S + D_{\text{exact(2)}}$	3.9 / 12.4	3.7 / 12.1	4.0 / 11.9	3.3 / 10.7	1.9 / 7.1	1.5 / 5.3	4.0 / 12.5	2.7 / 10.7
$X + S + D_{\text{exact(5)}}$	**4.7** / 13.0	3.8 / 12.6	3.9 / 12.0	3.1 / 11.0	**2.5** / **7.2**	1.7 / 5.7	4.2 / **13.3**	2.6 / 10.5
$X + S + D_{\text{exact(10)}}$	4.3 / 13.0	**3.9** / 12.6	4.0 / 12.0	3.3 / 10.8	2.3 / 6.5	1.8 / 5.7	4.0 / 12.9	2.6 / 10.8
$X + S + D_{\text{exact(20)}}$	4.2 / 12.9	3.6 / 12.2	3.9 / 11.6	3.0 / 10.6	2.2 / 6.5	1.8 / 5.5	4.0 / 12.5	2.8 / **11.0**
$X + S + D_{\text{exact(full)}}$	4.3 / 13.0	3.6 / 12.3	3.7 / 11.3	3.0 / 10.5	2.3 / 6.7	**1.9** / 5.7	**4.5** / **13.3**	2.7 / 10.8
$X + S + D_{\text{exact(2)+fuzzy}}$	4.4 / 12.9	3.4 / 12.3	4.0 / 12.4	3.0 / 11.0	2.0 / 6.5	**1.9** / 5.7	4.2 / 12.6	1.8 / 10.2
$X + S + D_{\text{exact(5)+fuzzy}}$	4.3 / 13.0	3.7 / 12.5	3.9 / **12.8**	3.5 / **11.4**	2.3 / 6.1	1.4 / 5.6	3.9 / 12.9	2.6 / 10.9
$X + S + D_{\text{exact(10)+fuzzy}}$	**4.7** / **13.4**	3.7 / **12.8**	**4.2** / **12.8**	**3.7** / **11.4**	2.1 / 5.9	1.6 / 5.7	4.4 / 13.0	2.4 / 10.8
$X + S + D_{\text{exact(20)+fuzzy}}$	4.3 / 12.9	3.8 / 12.5	3.9 / 12.2	3.3 / 11.1	2.3 / 6.4	1.7 / 5.4	4.4 / 13.0	2.5 / 10.6
$X + S + D_{\text{exact(full)+fuzzy}}$	4.5 / 13.0	**3.9** / 12.7	4.0 / 12.3	3.3 / 11.2	2.3 / 6.7	1.6 / 5.4	4.4 / 13.1	2.6 / 10.7

Table 8: BLEU / chrF++ scores for Chinese to Yi (zh→yi) translation across different prompting strategies and model sizes.

Evaluation Type	zh→mn		mn→zh		zh→yi		yi→zh	
	BLEU	chrF++	BLEU	chrF++	BLEU	chrF++	BLEU	chrF++
GPT-4.1	1.4	27.5	18.4	15.6	7.4	7.6	4.8	13.2
Non-native annotators	4.8	28.5	17.8	14.1	1.9	11.9	4.6	6.8
Native annotators	37.8	60.8	49.0	36.2	–	–	–	–

Table 9: Automatic and human evaluations of GPT-4.1 on Chinese↔Mongolian and Chinese↔Yi translation.

F Prompt Construction

X Prompt:

请帮我把下面的句子从{source_language}翻译成{target_language}：{src_sentence}
请尽你所能进行翻译，即使翻译不好也没关系，不要拒绝尝试，我不会责怪你的。
请将你的翻译用###括起来。比如，如果你的翻译是“你好，世界”，那么你输出的最后部分应该是###你好，世界###

Please translate the following sentence from {source_language} into {target_language}: {src_sentence}.
Do your best to provide a translation. It is acceptable if the translation is not perfect; please do not refuse to try, and I will not blame you.
Enclose your translation with ###. For example, if your translation is "Hello, world", then your final output should be ###Hello, world###.

X+S Prompt:

请仿照样例，将{source_language}句子翻译成{target_language}句子。请尽你所能进行翻译，即使翻译不好也没关系，不要拒绝尝试，我不会责怪你的。请将你的翻译用###括起来。比如，如果你的翻译是“你好，世界”，那么你输出的最后部分应该是###你好，世界###
{source_language}: {source_example_sentence1}
{target_language}: {target_example_sentence1}
{source_language}: {source_example_sentence2}
{target_language}: {target_example_sentence2}
{source_language}: {source_example_sentence3}
{target_language}: {target_example_sentence3}
{source_language}: {source_sentence}
{target_language}:

Please follow the example above and translate the sentence in {source_language} into the sentence in {target_language}. Try your best to translate, even if the translation is not perfect, it's okay. Don't refuse to try, I won't blame you. Please enclose your translation with ###. For example, if your translation is "Hello, World," the last part of your output should be ###Hello, World###.
{source_language}: {source_example_sentence1}
{target_language}: {target_example_sentence1}
{source_language}: {source_example_sentence2}
{target_language}: {target_example_sentence2}
{source_language}: {source_example_sentence3}
{target_language}: {target_example_sentence3}
{source_language}: {source_sentence}
{target_language}:

X+*D* **Prompt:**

请仿照样例，将{source_language}句子翻译成{target_language}句子。请尽你所能进行翻译，即使翻译不好也没关系，不要拒绝尝试，我不会责怪你的。请将你的翻译用###括起来。比如，如果你的翻译是“你好，世界”，那么你输出的最后部分应该是###你好，世界###
{source_language}: {source_sentence}
在上面的句子中，{source_language} 词语 “{word}” 在 {target_language} 中可能的翻译是 {word_meanings}；...
{target_language}:

Please follow the example above and translate the sentence in {source_language} into the sentence in {target_language}. Try your best to translate, even if the translation is not perfect, it's okay. Don't refuse to try, I won't blame you. Please enclose your translation with ###. For example, if your translation is "Hello, World," the last part of your output should be ###Hello, World###.
{source_language}: {source_sentence}
In the sentence above, the word "{word}" in {source_language} may have the following translations in {target_language} is {word_meanings};...
{target_language}:

X+*S*+*D* **Prompt:**

请仿照样例，参考给出的词汇和语法，将 {source_language} 句子翻译成 {target_language}：

请将 {source_language} 句子翻译成 {target_language}：{source_example_sentence1}。
在上面的句子中，{source_language} 词语“{word}”在 {target_language} 中可能的翻译是 {word_meanings}；...
所以，该 {source_language} 句子完整的 {target_language} 翻译是：{target_example_sentence1}

请将 {source_language} 句子翻译成 {target_language}：{source_example_sentence2}。
在上面的句子中，...
所以，该 {source_language} 句子完整的 {target_language} 翻译是：{target_example_sentence2}

请将 {source_language} 句子翻译成 {target_language}：{source_example_sentence3}。
在上面的句子中，...
所以，该 {source_language} 句子完整的 {target_language} 翻译是：{target_example_sentence3}

请将 {source_language} 句子翻译成 {target_language}：{src_sentence}。
在上面的句子中，...（词语“{word}”的可能翻译为 {word_meanings}）
所以，该 {source_language} 句子完整的 {target_language} 翻译是：{target_sentence}

Please follow the examples and refer to the dictionary hints to translate the following {source_language} sentence into {target_language}.

Translate {source_language} sentence: {source_example_sentence1}.
In this sentence, the word {word} may be translated as {word_meanings} in {target_language}; ...
So the full translation in {target_language} is: {target_example_sentence1}

Translate {source_language} sentence: {source_example_sentence2}.
...
So the full translation in {target_language} is: {target_example_sentence2}

Translate {source_language} sentence: {source_example_sentence3}.
...
So the full translation in {target_language} is: {target_example_sentence3}

Translate {source_language} sentence: {src_sentence}.
In this sentence, the word {word} may be translated as {word_meanings}.
So the full translation in {target_language} is: {target_sentence}