Authors' Response to Reviews of

Neurocomputing

Dinh-Truong Do

***Manuscript Title**: Enhancing Zero-Shot Multilingual Semantic Parsing: A Framework Leveraging Large Language Models for Data Augmentation and Advanced Prompting Techniques*

***Authors**: Dinh-Truong Do, Minh-Phuong Nguyen, and Le-Minh Nguyen,*

RC: *Reviewers' Comment*, AR: Authors' Response

Sincerest thanks for your comments on our manuscript. We appreciated your constructive criticisms. We have addressed each of your concerns as outlined below.

1. Reviewer #1

(The text modification color is blue)

RC: ***Q1. The paper's content requires further clarification. For instance, 1) it claims to focus on augmenting multilingual datasets of target languages, yet it discusses using an English-augmented dataset instead; 2) It describes leveraging the cross-lingual abilities of LLMs to expand data from existing English datasets into new languages. However, the actual method involves translating English utterances into target languages using standard translation tools, leading to a methodological inconsistency.***

AR: 1) Following your suggestion, we have clarified the scope of our work in the paper to better reflect our approach. Specifically, our work focuses on zero-shot multilingual semantic parsing, aiming to build multilingual semantic parsers without having annotated data in the target languages, using only annotated data in English. In this context, "zero-shot" means we do not require any annotated data in the target languages, which aligns with the definition used in previous research [1, 2].

- *Abstract Section*: "In recent years, significant progress ... zero-shot multilingual semantic parsing)".
- *Introduction Section (page 2)*: "Existing semantic parsing datasets ... NLP capabilities."

2) Building semantic parsing data is challenging, and making it available in a wide range of languages is even more difficult. In this paper, we focus on expanding data from English to other languages. While we use off-the-shelf translation tools to convert English utterances into target languages, this is consistent with previous works [1, 2, 4]. However, aligning the logical tokens (intentions, slots) with the translated utterance remains a significant challenge due to structural differences between languages. In this paper, we propose Chain-of-Thought prompting methods to leverage the power of LLMs in aligning logical tokens with span text in the translated utterances. Annotated alignment data is labor-intensive, and using only machine translation techniques cannot accomplish this task for the target languages. This alignment is a key issue we explore in this paper.

To further analyze the potential of using LLMs for translation, we conducted an intuitive experiment by translating English sentences into target languages using Llama-2-70b-chat. The results indicated that this

LLM struggled with low-resource languages like Thai, failing to produce accurate translations, as shown in Figure 1. This finding supports our decision to use specialized translation tools for this task, aligning with the findings from previous work [4]. Our focus remains on effectively leveraging these tools to facilitate the translation process while addressing the inherent challenges of cross-linguistic alignment.

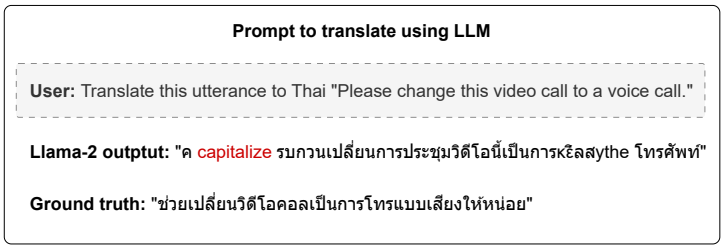

Figure 1: An example of Llama-2-70b-chat fails to return the correct translation.

References

[1] Nicosia, M., Qu, Z., Altun, Y. (2021). Translate fill: Improving zero-shot multilingual semantic parsing with synthetic data. In Findings of the Association for Computational Linguistics: EMNLP 2021 (pp. 3272–3284). Punta Cana, Dominican Republic: Association for Computational Linguistics.

[2] Sherborne, T., Lapata, M. (2022). Zero-shot cross-lingual semantic parsing. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 4134–4153). Dublin, Ireland: Association for Computational Linguistics.

[3] Moradshahi, M., Tsai, V., Campagna, G., Lam, M. (2023). Contextual semantic parsing for multilingual task-oriented dialogues. In Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics (pp. 902–915). Dubrovnik, Croatia: Association for Computational Linguistics.

[4] Wang, Z., Cuenca, G., Zhou, S., Xu, F. F., Neubig, G. (2023). MCoNaLa: A benchmark for code generation from multiple natural languages. In Findings of the Association for Computational Linguistics: EACL 2023 (pp. 265–273). Dubrovnik, Croatia: Association for Computational Linguistics.

[5] Son, J., Kim, B. (2023). Translation performance from the user's perspective of large language models and neural machine translation systems. Information, 14(10), 574.

RC: ***Q2. There is a significant discrepancy between the methods proposed in the paper and its title. The paper introduces a zero-shot approach for multilingual semantic parsing, but the authors have already constructed a multilingual dataset for training, which contradicts the zero-shot setting.***

AR: Thank you for your insightful comment. We understand the confusion and would like to clarify that our work does not imply zero-shot learning in the sense of having no training process at all. Instead, our focus is on a zero-shot setting where we aim to build multilingual semantic parsers without having annotated data in the target languages. In this context, "zero-shot" means we do not require any annotated data in the target languages, which aligns with the definition used in previous research [1, 2].

To ensure clarity, we have revised the abstract to include a clear definition of "zero-shot multilingual semantic parsing." The revised abstract now states: "One promising approach to bridge this gap is ... (known as zero-shot multilingual semantic parsing)". This addition aims to prevent any further confusion regarding our methodology.

References

[1] Nicosia, M., Qu, Z., Altun, Y. (2021). Translate fill: Improving zero-shot multilingual semantic parsing with synthetic data. In Findings of the Association for Computational Linguistics: EMNLP 2021 (pp. 3272–3284). Punta Cana, Dominican Republic: Association for Computational Linguistics.

[2] Sherborne, T., Lapata, M. (2022). Zero-shot cross-lingual semantic parsing. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 4134–4153). Dublin, Ireland: Association for Computational Linguistics.

RC: ***Q3. The rationale behind proposing three novel multilingual semantic parsing CoT prompting techniques is unclear. Why not focus on the most effective technique instead of introducing multiple methods akin to the three semantic filtering strategies?***

AR: The introduction of three novel CoT prompting techniques is not about redundancy but about exploring diverse approaches to tackle the complexities of multilingual semantic parsing. Each technique offers unique advantages that address different aspects of the task, allowing us to determine the most effective strategy through comparative analysis and empirical evaluation.

For instance, the "Single-turn CoT" technique is efficient as it requires only a single inference call to the LLM, making it resource-friendly. The "Multi-turn CoT" technique helps mitigate the issue of incorrect logical tokens by iteratively refining the output. Lastly, the "Multi-turn CoT Symmetry" technique enhances the alignment between the source and target languages, improving translation accuracy and consistency.

Before conducting experiments, it is challenging to predict which method will perform best or be most suitable for a given dataset. Therefore, evaluating multiple techniques allows us to identify the optimal approach for different multilingual contexts. We have added a discussion of this rationale to *Section 3.1.1 LLM-based Augmentation* (page 11), to provide further clarity on our approach.

RC: ***Q4. The necessity of using a more complex approach involving dataset augmentation followed by parser training is questionable. Why not directly apply LLMs for multilingual semantic parsing?***

AR: Our framework aims to improve the performance of zero-shot multilingual semantic parsing by leveraging the power of LLMs in data augmentation. While our approach includes multiple steps, each has demonstrated its effectiveness through ablation studies (Table 4), and our framework achieves state-of-the-art performance compared to other zero-shot multilingual semantic parsers. Additionally, our method is extendable to any target language, including low-resource languages.

One reason we don't directly apply LLMs for zero-shot semantic parsing is their performance is not very high in this context. To illustrate, we added a new section analyzing the use of direct LLMs on the MTOP dataset (Section 5.6, page 27). In this analysis, given an utterance in the target language, we find the most similar English annotated utterance, prompt the LLM with the English utterance, its logical form, and the target language utterance, and have the LLM output the logical form in the target language. The results, shown in Table 9 (page 28), indicate that direct LLM performance is lower compared to our proposed framework.

Another reason for training a smaller semantic parsing model using data augmented with LLMs is resource efficiency. Deploying an LLM requires substantial computational resources, making it impractical for resource-constrained environments such as IoT devices. In contrast, a smaller, fine-tuned semantic parsing model can operate efficiently on these limited-resource devices, ensuring broader applicability and practicality in real-world scenarios. We have added these key points to *Section 2.2 Advancing Semantic Parsing with LLMs* (page 7) to further clarify the motivation behind our approach.

References

[1] Roy, S., Thomson, S., Chen, T., Shin, R., Pauls, A., Eisner, J., Van Durme, B. (2024). BenchCLAMP: A Benchmark for Evaluating Language Models on Syntactic and Semantic Parsing. Advances in Neural Information Processing Systems, 36.

[2] Mekala, D., Wolfe, J., Roy, S. (2023). ZEROTOP: Zero-shot task-oriented semantic parsing using large language models. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (pp. 5792–5799). Singapore: Association for Computational Linguistics.

RC: ***Q5. Section 3.2 lacks smooth integration with the rest of the text, creating confusion about how it relates to other sections.***

AR: Thank you for your suggestion. To address this issue, we have clarified how Section 3.2 integrates with the rest of the manuscript. We have revised *the Abstract*, *Section 1 Introduction (page 4)*, and *Section 3 Methodology (pages 8, 16)* to ensure better coherence and integration.

Specifically, we now explain that after applying our framework to the full English annotated data, we analyzed its effectiveness under more stringent constraints. These constraints involve not only the absence of data in target languages but also the availability of only a limited annotated dataset in English. To simulate a low-resource scenario, we created subsets from the full annotated English dataset using three selective compression methods: random selection, label-coverage-based sampling, and a hybrid of random and label-coverage-based sampling (detailed in Section 3.2). By evaluating the performance of our framework on these subsets, we demonstrate its robustness under limited data conditions. Furthermore, analyzing the results of each selective method provides valuable insights for annotators on constructing a small yet effective annotated dataset, ensuring an efficient system.

RC: ***Q6. The paper's writing needs to improve clarity and logical flow.***

AR: We have thoroughly revised the manuscript to enhance its clarity and logical flow. This revision includes restructuring the content of various sections, refining the language, and ensuring that each part of the manuscript transitions smoothly to the next. Changes made for clarity and improved flow are highlighted in blue throughout the text.

2. Reviewer #2

(The text modification color is magenta)

RC: ***Q1. The LLM is so powerful that it can be directly applied into semantic parsing. So what's the point of training a small semantic parsing model using data augmented with LLM?***

AR: Our framework aims to improve the performance of zero-shot multilingual semantic parsing by leveraging the power of LLMs in data augmentation. While our approach includes multiple steps, each has demonstrated its effectiveness through ablation studies (Table 4), and our framework achieves state-of-the-art performance compared to other zero-shot multilingual semantic parsers. Additionally, our method is extendable to any target language, including low-resource languages.

One reason we don't directly apply LLMs for zero-shot semantic parsing is their performance is not very high in this context. To illustrate, we added a new section analyzing the use of direct LLMs on the MTOP dataset (Section 5.6, page 27). In this analysis, given an utterance in the target language, we find the most similar

English annotated utterance, prompt the LLM with the English utterance, its logical form, and the target language utterance, and have the LLM output the logical form in the target language. The results, shown in Table 9 (page 28), indicate that direct LLM performance is lower compared to our proposed framework.

Another reason for training a smaller semantic parsing model using data augmented with LLMs is resource efficiency. Deploying an LLM requires substantial computational resources, making it impractical for resource-constrained environments such as IoT devices. In contrast, a smaller, fine-tuned semantic parsing model can operate efficiently on these limited-resource devices, ensuring broader applicability and practicality in real-world scenarios. We have added these key points to *Section 2.2 Advancing Semantic Parsing with LLMs* (page 7) to further clarify the motivation behind our approach.

References

[1] Roy, S., Thomson, S., Chen, T., Shin, R., Pauls, A., Eisner, J., Van Durme, B. (2024). BenchCLAMP: A Benchmark for Evaluating Language Models on Syntactic and Semantic Parsing. Advances in Neural Information Processing Systems, 36.

[2] Mekala, D., Wolfe, J., Roy, S. (2023). ZEROTOP: Zero-shot task-oriented semantic parsing using large language models. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (pp. 5792–5799). Singapore: Association for Computational Linguistics.