VerbaNexAI at SemEval-2025 Task 2: Enhancing Entity-Aware Translation with Wikidata-Enriched MarianMT

Daniel Peña Gnecco
Juan Carlos Martinez-Santos
Edwin Puertas
Universidad Tecnológica de Bolívar, Cartagena, Colombia
dgnecco@utb.edu.co jcmartinezs@utb.edu.co epuerta@utb.edu.co

Abstract

This paper presents the VerbaNexAi Lab system for SemEval-2025 Task 2: Entity-Aware Machine Translation (EA-MT), focusing on translating named entities from English to Spanish across categories such as musical works, foods, and landmarks. Our approach integrates detailed data preprocessing, enrichment with 240,432 Wikidata entity pairs, and fine-tuning of the MarianMT model to enhance entity translation accuracy. Official results reveal a COMET score of 87.09, indicating high fluency, an M-ETA score of 24.62, highlighting challenges in entity precision, and an Overall Score of 38.38, ranking last among 34 systems. While Wikidata improved translations for familiar entities like "Águila de San Juan," our static methodology underperformed compared to dynamic LLM-based approaches (Yuksel et al., 2025).

1 Introduction

Translating named entities such as proper nouns, geographic locations, and culturally significant references across languages remains a persistent challenge in natural language processing (NLP). This difficulty is particularly evident in the English-to-Spanish language pair, where lexical and cultural disparities often hinder accurate translation (Conia et al., 2025). For instance, a literal translation of "The Shawshank Redemption" as "La redención de Shawshank" fails to convey its identity as a well-known film, potentially confusing Spanish-speaking audiences. Similarly, "Eagle of St. John" requires translation to "Águila de San Juan" to retain its cultural and religious significance rather than an erroneous "Águila de Jhon." These examples underscore the importance of entity-aware machine translation (EA-MT), the focus of SemEval-2025 Task 2, which seeks to enhance precision in translating such entities for applications, including content localization, cross-cultural communication, and user-facing services.

Human translators excel at navigating these nuances by leveraging cultural knowledge and external resources, such as glossaries, to adapt entities appropriately (Vishwakarma, 2023). For example, rendering "Thanksgiving" as "Día de Acción de Gracias" demands understanding its North American cultural context. This task challenges automation without advanced systems. Neural machine translation (NMT) has markedly improved fluency in general translation tasks. Yet, it often struggles with named entities due to insufficient training data for rare or culturally specific terms and the absence of real-time contextual adaptation (Zeng et al., 2023). These shortcomings motivated our participation in SemEval-2025 Task 2, aiming to improve entity translation accuracy.

We propose an EA-MT system that combines MarianMT, an efficient model for English-to-Spanish translation, with enrichment of 240,432 bilingual entity pairs from Wikidata. This approach balances computational scalability, suitable for resource-constrained environments, with precision for culturally significant entities such as movie titles, foods, and landmarks. MarianMT's lightweight architecture facilitates fine-tuning with limited resources, while Wikidata's structured data addresses data scarcity challenges (Hu et al., 2022). We aim to narrow the divide between human-like cultural adaptation and automated scalability, addressing a pressing need in contemporary NLP.

Our contributions are threefold: (1) a scalable pipeline for data integration using Wikidata, adaptable to various languages and domains; (2) an empirical analysis comparing our static approach to dynamic LLM-based systems (Yuksel et al., 2025); and (3) insights into the limitations of static EA-MT and the demand for real-time, culturally sensitive adaptation. We intend to release our code and enriched dataset to support further research. This paper is structured as follows: Section 2 reviews related work, Section 3 details our methodology,

Proceedings of the The 19th International Workshop on Semantic Evaluation (SemEval-2025), pages 1255–1262
July 31 - August 1, 2025

Section 4 evaluates performance, and Section 5 summarizes findings and outlines future directions.

2 Related Work

Entity-Aware Machine Translation (EA-MT) has emerged as a critical subfield in NLP, addressing the limitations of traditional Neural Machine Translation (NMT) in handling named entities that require cultural and contextual precision (Conia et al., 2025). Recent advancements, such as SemEval-2025 Task 2, introduce specialized metrics like M-ETA for entity-specific accuracy and COMET for overall translation quality. These developments build upon approaches like Yuksel et al.'s dynamic LLM-based methods (Yuksel et al., 2025). Unlike these dynamic strategies, our system leverages Wikidata as a static knowledge base for entity enrichment.

Machine translation has evolved significantly over the years. Early rule-based systems relied on manually crafted linguistic rules, offering limited scalability and adaptability. Statistical machine translation (SMT), exemplified by tools like Moses, improved upon this by leveraging parallel corpora. However, it struggled with rare entities and context-dependent translations due to their reliance on statistical alignments rather than semantic understanding. The advent of transformer-based NMT models (Yang et al., 2020) marked a significant leap in translation quality. Yet, limitations persist in entity translation, as traditional approaches often lack mechanisms for cultural adaptation and real-time knowledge integration. Modern neural approaches like CroCoAlign (Molfese et al., 2024) refine sentence alignment, optimizing training data for NMT systems.

Introducing models like BERT (Devlin et al., 2018) brought pre-trained language representations, further enhancing NMT capabilities. However, challenges persist, particularly in translating rare named entities (Saadany et al., 2024). Efforts such as MOSAICo (Conia et al., 2024) address data scarcity by providing large-scale, multilingual, semantically annotated corpora. Other techniques have contributed to improved entity translation, including entity projection via MT (Jain et al., 2019) and denoising pre-training with monolingual data (Hu et al., 2022). Our system builds on these advancements by integrating entity enrichment through Wikidata.

Handling named entities remains a significant challenge in NMT. Zeng et al.'s Extract-and-Attend method dynamically extracts entity candidates, reducing errors by up to 35% (Zeng et al., 2023). Similarly, Lee et al. (Lee et al., 2021) employ NER post-processing to refine translation outputs, an approach we adapt statically via fine-tuning. Our system enhances entity translation by leveraging Wikidata to ensure contextual accuracy across languages.

Cultural adaptation is a crucial aspect of EA-MT. Challenges such as preserving culturally significant titles (e.g., Breaking Bad) align with our focus on entities like Águila de San Juan (Vishwakarma, 2023). Wang et al. (Wang et al., 2024) highlight the issue of cultural dominance in LLMs, which we mitigate through the integration of multilingual data in Wikidata. Named Entity Recognition (NER) plays a foundational role in this effort, with surveys like Li et al. (Li et al., 2022) guiding our enrichment strategy.

Evaluation remains a significant challenge in EA-MT. Traditional metrics like BLEU fail to capture entity accuracy, leading to the adoption of M-ETA, which reflects the limitations of our static approach. Jung et al. (Jung et al., 2023) propose fine-grained error analysis for deeper quality assessment. This approach could further refine our evaluation.

Our work aims to address existing gaps in EA-MT by incorporating structured knowledge bases and static entity enrichment, enhancing translation accuracy and cultural relevance.

3 System Description

Our EA-MT system enhances English-to-Spanish entity translation through a three-stage process tailored for scalability and precision in SemEval-2025 Task 2. As detailed below, our methodology adapts a static approach using MarianMT and Wikidata, explicitly discarding dynamic alternatives due to resource constraints.

3.1 Data Preprocessing

We preprocessed the EA-MT dataset to remove noise and standardize text, ensuring semantic focus. It involved (1) removing emojis, URLs, mentions (e.g., @username), and hashtags (e.g., #tag); (2) eliminating non-standard special characters (retaining ., !, ?); and (3) replacing accented characters in general text (e.g., "Á" to "A") while preserving entity names with accents to maintain integrity (e.g., "Águila" unchanged). We preserved the original case to avoid obscuring entity boundaries (Jurafsky and Martin, 2025). Accent replacement in non-entity text risked degrading contextual translation (Naveen and Trojovský, 2024), but entity preservation ensured outputs like "Águila de San Juan" remained accurate.

3.2 Wikidata Enrichment

To address entity data scarcity, we enriched training with 240,432 Wikidata pairs across categories like musical works (Q2188189, 3766 labels; Q105543609, 70176 labels), foods (Q2095, 2575 labels; Q25403900, 445 labels), plants (Q756, 15037 labels), books (Q571, 2396 labels), book series (Q1667921, 750 labels), fictional entities (Q14897293, 17865 labels), landmarks (Q570116, 6157 labels; Q2319498, 620 labels), movies (Q11424, 67293 labels), places of worship (Q24398318, 12275 labels), natural places (Q1286517, 18387 labels), and TV series (Q5398426, 16282 labels). These pairs, extracted via Wikidata API queries, enhanced coverage for familiar entities like "Águila de San Juan," though rare entity representation remained limited.

While our enrichment improved precision for frequent entities, the static nature of this approach limits its effectiveness for rare or emerging entities. A more curated version of Wikidata, focusing on task-specific entities, could further enhance M-ETA scores, though the inherent limitation of static systems' incapability to adapt to new or context-specific entities absent from pre-enriched data would persist, underscoring the need for dynamic retrieval methods.

3.3 MarianMT Fine-Tuning

We selected MarianMT (`Helsinki-NLP/opus-mt-en-es`) for its efficiency and suitability for English-to-Spanish translation. Unlike larger models like NLLB-200 and M2M-100, which are designed for broad multilingual coverage and require significantly more computational resources, MarianMT offers a balanced trade-off between performance and resource efficiency. Given our hardware constraints (NVIDIA RTX 3050 GPU, 4GB), fine-tuning a larger model would have been impractical.

While NLLB-200 and M2M-100 may outperform in general multilingual translation, their advantage in entity-specific tasks remains uncertain, particularly in combination with our entity enrichment strategy. We fine-tuned MarianMT on our dataset using a learning rate of 3×10^{-5}, a batch size of 4 with gradient accumulation, and four epochs, optimizing with AdamW ($\beta_1 = 0.9$, $\beta_2 = 0.999$) (Yang et al., 2020). With more significant computational resources, increasing the number of epochs could further improve entity translation accuracy. Hardware limitations dictated this static approach, as dynamic LLM-based methods, such as recovery-augmented generation (RAG), required more VRAM, impractical for our setup. See Appendix 6 (Table 4) for a comparison of the EA-MT approaches considered.

Additionally, we explicitly discarded dynamic LLM-based approaches, such as retrieval-augmented generation (RAG), due to their high computational demands. Instead, we prioritized a static, resource-efficient solution better suited to our constraints. See Appendix 6 (Table 4) for a comparison of EA-MT approaches considered.

3.4 Prediction Generation

Predictions used the fine-tuned MarianMT statically, applying identical preprocessing steps. We formatted outputs as JSONL per SemEval requirements. Unlike dynamic LLM-based systems (Yuksel et al., 2025), our approach prioritized efficiency over adaptability.

4 Results and Analysis

We assess performance using SemEval-2025 Task 2 metrics: COMET (general quality), M-ETA (entity accuracy), and Overall Score:

$$\text{Overall Score} = 2 \times \frac{\text{COMET} \times \text{M-ETA}}{\text{COMET} + \text{M-ETA}}$$

All scores range from 0 to 100.

4.1 Performance Metrics

Our system achieved a COMET of 87.09, M-ETA of 24.62, and an Overall Score of 38.38, ranking 34th out of 34 systems. The high COMET reflects fluency, but the low M-ETA indicates struggles with rare entities (Naveen and Trojovský, 2024).

Metric	Validation	Test
COMET	87.24	87.09
M-ETA	27.74	24.62
Overall Score	45.12	38.38

Table 1: Validation vs. test metrics.

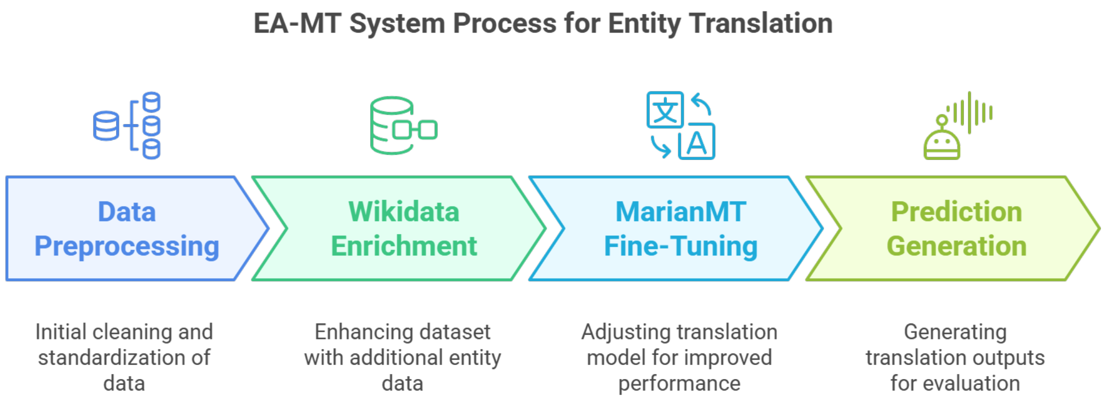

Figure 1: Pipeline diagram of the EA-MT system, illustrating data preprocessing, Wikidata enrichment, MarianMT fine-tuning, and prediction generation.

4.2 Comparison with Top Systems

We now compare our system's performance against top-performing models in the SemEval-2025 Task 2 leaderboard across Overall, M-ETA, and COMET metrics. Leading systems, such as Qwen2.5-Max (Overall: 92.63), Llama-3.3-70B-Instruct + DeepSeek-R1 (M-ETA: 90.50), and GPT-4o (COMET: 95.31), leverage large-scale LLMs, significantly outperforming our static MarianMT-based approach. Table 2 summarizes these results.

System	Overall	M-ETA	COMET
GPT-4o	92.42	89.88	95.31
Qwen2.5-Max	92.63	90.26	95.09
Qwen2.5-72B	92.54	90.13	95.09
Phi-4	92.50	90.09	95.04
Llama-3.3-70B-Instruct	91.72	88.42	95.28
Qwen2.5-32B	91.72	88.42	92.77
Llama-3.3-70B-Instruct + DeepSeek-R1	92.17	90.50	93.91
Ours	38.38	24.62	87.09

Table 2: Comparison of our system with top-performing systems in SemEval-2025 Task 2 across Overall, M-ETA, and COMET metrics

The top Overall scores, led by Qwen2.5-Max at 92.63, reflect a balanced performance in fluency and entity accuracy, far surpassing our 38.38. In M-ETA, systems like Llama-3.3-70B-Instruct + DeepSeek-R1 (90.50) and Qwen2.5-Max (90.26) demonstrate exceptional entity precision, while our 24.62 highlights a significant gap in handling named entities. For COMET, GPT-4o (95.31) and Llama-3.3-70B-Instruct (95.28) set the benchmark for fluency. Yet, our 87.09 remains competitive, indicating that our system's strength lies in general translation quality rather than entity-specific accuracy. These leading systems utilize large language models (LLMs) with retrieval-augmented generation (RAG) techniques, enabling them to access and incorporate external knowledge during translation dynamically. This dynamic approach allows models to handle a range of entities, including rare or domain-specific ones, by retrieving relevant information in real time. In contrast, our reliance on static Wikidata enrichment, while effective for familiar entities, fails to adapt to new or less frequent entities, explaining our low M-ETA score. It underscores the advantage of dynamic methods, as discussed in recent work on RAG in machine translation, such as Yuksel et al. (2025) (Yuksel et al., 2025).

4.3 Qualitative Analysis

The M-ETA metric, an exact-match evaluation for named entity translation accuracy, considers a translation correct only if it precisely matches the reference, offering no partial credit for approximate matches. This strict standard penalizes any deviation, be it a mistranslation, typographical error, or cultural misinterpretation. Our system's low M-ETA score of 24.62 indicates that many named entities were not translated accurately under this metric, reflecting challenges with rare entities and context-specific adaptations.

Successes included "Eagle of St. John" as "Águila de San Juan," showcasing Wikidata's strength with well-documented entities. However, frequent failures reveal cultural and contextual deficits (Saadany et al., 2024). Examples include "Breaking Bad," which was mistranslated as "Rompiendo Malo" instead of retaining its title, "Star Wars" as "Guerras Estelares" rather than "La Guerra de las Galaxias," and "Empire State Building" as "Edificio del Estado del Imperio" instead of preserving its name. These errors highlight limitations in handling culturally significant titles and landmarks, penalized heavily by M-ETA's exact match requirement. See Appendix 6 (Table 5) for a detailed list of translation examples with error types.

5 Conclusion

Despite Neural Machine Translation (NMT) advancements, our results highlight fundamental shortcomings in entity-aware translation when relying solely on static knowledge sources. While our system achieved a COMET score of 87.09, demonstrating strong fluency, its M-ETA score of 24.62 exposed severe deficiencies in entity precision, ultimately leading to an Overall Score of 38.38, the lowest among competing systems. These results confirm that a static enrichment approach, even when incorporating a large-scale structured knowledge base like Wikidata, is insufficient for handling the complexity of named entity translation. Static methods offer scalability and efficiency in low-resource settings (e.g., 4GB GPU) compared to RAG's demands, but the COMET-M-ETA gap shows fluency prioritization over precision, misaligned with EA-MT goals.

One of the main issues observed was the rigid dependency on Wikidata, which, while useful for well-documented entities, failed to capture emerging terms, domain-specific references, and subtle cultural nuances. The absence of real-time retrieval mechanisms also resulted in translation errors for ambiguous or context-sensitive entities. Compared to retrieval-augmented systems (Yuksel et al., 2025), our approach could not dynamically adjust translations, leading to cases where named entities were either mistranslated or omitted entirely. Exploring more curated or updated versions of structured knowledge bases like Wikidata could enhance entity translation accuracy. However, the fundamental limitation of static approaches' inability to adapt to new or context-specific entities would remain, reinforcing the need for dynamic retrieval methods.

Another critical limitation was our preprocessing pipeline, which, although effective in text normalization, introduced unintended side effects. For example, the replacement of accented characters (e.g., "Águila" to "Aguila") compromised entity integrity, further reducing translation accuracy (Naveen and Trojovský, 2024). Moreover, our constrained hardware (NVIDIA RTX 3050, 4GB) restricted fine-tuning to only four epochs, potentially limiting the model's ability to leverage the enriched dataset effectively (Yang et al., 2020). Traditional approaches, such as rule-based systems and statistical machine translation (e.g., Moses), suffer from similar limitations, poor scalability, lack of semantic understanding, and inadequate entity handling, rendering them obsolete for modern EA-MT tasks.

Our metrics provide clear evidence for the need for alternative solutions. The stark contrast between our M-ETA score 24.62 and the top systems' scores (e.g., 90.50 for Llama-3.3-70B-Instruct + DeepSeek-R1) indicates a significant gap in entity translation accuracy. In contrast, our competitive COMET score (87.09 vs. 95.31 for GPT-4o) suggests fluency is less of a bottleneck. This disparity underscores the inadequacy of static methods for entity-specific tasks. It justifies the adoption of dynamic, retrieval-augmented approaches capable of addressing rare and context-dependent entities. A hybrid approach, like lightweight RAG caching frequent entities, could balance efficiency and adaptability.

5.1 Limitations

Our approach revealed several constraints affecting performance:

1. **Over-reliance on Wikidata:** While structured knowledge bases offer valuable entity translations, their static nature prevents adaptation to emerging or domain-specific terms, reducing overall system robustness (Li et al., 2022).

2. **Lack of Contextual Adaptation:** Unlike retrieval-augmented LLM-based approaches, our system could not adjust entity translations dynamically, leading to rigid and often incorrect outputs (Yuksel et al., 2025).

3. **Preprocessing-induced Errors:** The aggressive normalization of text removed diacritics, impacting the accuracy of culturally significant named entities and altering their intended meaning (Naveen and Trojovský, 2024).

4. **Computational Constraints:** Limited hardware resources severely restricted the fine-tuning depth, potentially capping the model's ability to leverage the enriched dataset (Yang et al., 2020) fully.

5.2 Future Work

Our results strongly indicate that static knowledge bases alone are insufficient for robust entity translation. Future work must focus on integrating retrieval-augmented generation (RAG) (Yuksel et al., 2025) and adaptive entity-linking techniques to incorporate contextual information dynamically. Additionally, improving preprocessing strategies to preserve linguistic integrity (Jurafsky and Martin, 2025) and increasing computational resources to enable deeper fine-tuning (Yang et al., 2020) will be critical in overcoming current limitations. Finally, exploring meta-learning (Deb et al., 2022) and heuristic reasoning (Aoki et al., 2024) could enhance adaptability, reducing errors in domain-specific and low-resource entity translations.

6 Acknowledgments

We dedicate this work to the master's degree scholarship program in Engineering at the Universidad Tecnológica de Bolívar (UTB) in Cartagena, Colombia. We extend our deepest gratitude to the VerbaNex AI Lab team for their dedication, collaboration, and continuous support of our research endeavors.

References

Yoichi Aoki, Keito Kudo, Tatsuki Kuribayashi, Shusaku Sone, Masaya Taniguchi, Keisuke Sakaguchi, and Kentaro Inui. 2024. First heuristic then rational: Dynamic use of heuristics in language model reasoning. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 14255–14271.

Simone Conia, Edoardo Barba, Abelardo Carlos Martinez Lorenzo, Pere-Lluís Huguet Cabot, Riccardo Orlando, Luigi Procopio, and Roberto Navigli. 2024. Mosaico: a multilingual open-text semantically annotated interlinked corpus. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 7990–8004. Association for Computational Linguistics.

Simone Conia, Min Li, Roberto Navigli, and Saloni Potdar. 2025. SemEval-2025 task 2: Entity-aware machine translation. In *Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025)*. Association for Computational Linguistics.

Budhaditya Deb, Guoqing Zheng, and Ahmed Hassan Awadallah. 2022. Boosting natural language generation from instructions with meta-learning. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022*, pages 6792–6808.

Jacob Devlin, Ming Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 1:4171–4186.

Junjie Hu, Hiroaki Hayashi, Kyunghyun Cho, and Graham Neubig. 2022. Deep: Denoising entity pre-training for neural machine translation. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 1:1753–1766.

Alankar Jain, Bhargavi Paranjape, and Zachary C. Lipton. 2019. Entity projection via machine translation for cross-lingual ner. *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, pages 1083–1092.

Dahyun Jung, Chanjun Park, Sugyeong Eo, and Heuiseok Lim. 2023. Enhancing machine translation quality estimation via fine-grained error analysis and large language model. *Mathematics 2023, Vol. 11, Page 4169*, 11:4169.

Daniel Jurafsky and James H. Martin. 2025. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*, 3rd edition edition.

Jangwon Lee, Jungi Lee, Minho Lee, and Gil Jin Jang. 2021. Named entity correction in neural machine translation using the attention alignment map. *Applied Sciences 2021, Vol. 11, Page 7026*, 11:7026.

Jing Li, Aixin Sun, Jianglei Han, and Chenliang Li. 2022. A survey on deep learning for named entity recognition. *IEEE Transactions on Knowledge and Data Engineering*, 34:50–70.

Francesco Molfese, Andrei Bejgu, Simone Tedeschi, Simone Conia, and Roberto Navigli. 2024. Crocoalign: A cross-lingual, context-aware and fully-neural sentence alignment system for long texts. In *Proceedings of the 18th Conference of the European Chapter* *of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2209–2220. Association for Computational Linguistics.

Palanichamy Naveen and Pavel Trojovský. 2024. Overview and challenges of machine translation for contextually appropriate translations. *iScience*, 27.

Hadeel Saadany, Ashraf Tantawy, and Constantin Orasan. 2024. Cyber risks of machine translation critical errors : Arabic mental health tweets as a case study. *arXiv preprint arXiv:2405.11668*.

Dr. Vimal Kumar Vishwakarma. 2023. Translating Cultural Nuances: Challenges and Strategies.

Wenxuan Wang, Wenxiang Jiao, Jingyuan Huang, Ruyi Dai, Jen-Tse Huang, Zhaopeng Tu, Michael R Lyu, and Tencent Ai Lab. 2024. Not all countries celebrate thanksgiving: On the cultural dominance in large language models. *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1:6349–6384.

Shuoheng Yang, Yuxin Wang, and Xiaowen Chu. 2020. A survey of deep learning techniques for neural machine translation. *arXiv preprint arXiv:2002.07526*.

Kamer Ali Yuksel, Ahmet Gunduz, Abdul Baseet Anees, and Hassan Sawaf. 2025. Efficient machine translation corpus generation: Integrating human-in-the-loop post-editing with large language models. *arXiv preprint arXiv:2502.12755*.

Zixin Zeng, Rui Wang, Yichong Leng, Junliang Guo, Xu Tan, Tao Qin, and Tie Yan Liu. 2023. Extract and attend: Improving entity translation in neural machine translation. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 1697–1710.

Appendix

This appendix provides supplementary tables and details about the VerbaNexAI system for SemEval-2025 Task 2, which we removed from the main paper to comply with the 5-page limit.

A. Comparison of EA-MT Approaches

The following table, originally in Section 3, summarizes the primary EA-MT approaches considered in our methodology:

Our static approach leverages MarianMT and Wikidata for scalability under resource constraints. At the same time, we discarded dynamic RAG systems due to hardware limitations. Traditional SMT methods, like Moses, were not considered due to their obsolescence and poor performance on entity translation tasks.

B. Translation Examples with Error Types

The following table, originally in Section 4, provides examples of entity translations with identified error types:

These examples illustrate both successes (e.g., "Eagle of St. John") and frequent failures (e.g., "Breaking Bad"), highlighting limitations in cultural adaptation and entity precision.

C. Hardware Constraints

Our experiments were conducted on an NVIDIA RTX 3050 GPU with 4GB VRAM, which limited batch sizes and fine-tuned epochs. This constraint likely impacted our ability to fully leverage the enriched dataset, suggesting that future work with higher-capacity hardware could yield improved results.

Approach	Advantages	Limitations	Adaptation/Discard
Static (Ours)	Efficient, scalable	Low M-ETA, no adaptability	Adapted with MarianMT/Wikidata
Dynamic (RAG)	High M-ETA, adaptable	Resource-intensive	Discarded due to hardware
Traditional (SMT)	Simple alignment	Poor entity accuracy	Discarded, outdated

Table 3: Comparison of relevant EA-MT approaches, highlighting adaptation or discard decisions in our system.

Approach	Advantages	Limitations	Adaptation/Discard
Static (Ours)	Efficient, scalable	Low M-ETA, no adaptability	Adapted with MarianMT/Wikidata
Dynamic (RAG)	High M-ETA, adaptable	Resource-intensive	Discarded due to hardware
Traditional (SMT)	Simple alignment	Poor entity accuracy	Discarded, outdated

Table 4: Comparison of relevant EA-MT approaches, highlighting adaptation or discard decisions in our system.

Entity	Correct Translation	Our Output	Error Type
Eagle of St. John	Águila de San Juan	Águila de San Juan	Correct
Breaking Bad	Breaking Bad	Rompiendo Malo	Mistranslation
The Room	La Habitación	La Sala	Cultural Error
Darth Vader	Darth Vader	Dar Vader	Typographical
Star Wars	La guerra de las galaxias	Guerras Estelares	Mistranslation
Sushi	sushi	Sushí	Typographical
Empire State Building	Empire State Building	Edificio del Estado del Imperio	Mistranslation

Table 5: Translation examples with error types.