Improving Low-Resource Translation with Dictionary-Guided Fine-Tuning and RL: A Spanish-to-Wayuunaiki Study

Manuel Mosquera[1], **Melissa Verónica Robles**[1], **Johan R. Portela**[1], **Rubén Manrique**[1]
[1]Universidad de Los Andes, Bogotá, Colombia
{ma.mosquerao, mv.robles, jd.rodriguez1234, rf.manrique}@uniandes.edu.co

Abstract

Low-resource machine translation remains a significant challenge for large language models (LLMs), which often lack exposure to these languages during pretraining and have limited parallel data for fine-tuning. We propose a novel approach that enhances translation for low-resource languages by integrating an external dictionary tool and training models end-to-end using reinforcement learning, in addition to supervised fine-tuning. Focusing on the Spanish–Wayuunaiki language pair, we frame translation as a tool-augmented decision-making problem in which the model can selectively consult a bilingual dictionary during generation. Our method combines supervised instruction tuning with Group Relative Policy Optimization (GRPO), enabling the model to learn both when and how to use the tool effectively. BLEU similarity scores are used as rewards to guide this learning process. Preliminary results show that our tool-augmented models achieve up to +3.37 BLEU improvement over previous work and an 18% relative gain compared to a supervised baseline without dictionary access, on the Spanish–Wayuunaiki test set from the AmericasNLP 2025 Shared Task. We also conduct ablation studies to assess the effects of model architecture and training strategy, comparing Qwen2.5-0.5B-Instruct with other models such as LLaMA and a prior NLLB-based system. These findings highlight the promise of combining LLMs with external tools and the role of reinforcement learning in improving translation quality in low-resource language settings.

Code — https://github.com/Manuel-2011/RLTranslator/

Datasets — https://github.com/Manuel-2011/RLTranslator/tree/main/datasets

Introduction

Natural language processing (NLP) has witnessed remarkable progress in recent years, yet such advances have largely bypassed low-resource languages, especially Indigenous languages, due to the scarcity of high-quality parallel corpora and the predominance of oral over written traditions (Ignat et al. 2024; Robles et al. 2024). As a result, even state-of-the-art generative AI systems struggle to produce reliable output: a BIDLab study found that AI responses in Indigenous languages are correct only 54% of the time, with answers on average four times shorter and noticeably degraded in fluency and adequacy (Lucas et al. 2025).

Against this backdrop, community-driven and academic initiatives have begun to address the gap. Notably, the AmericasNLP Shared Task (2025) introduced translation benchmarks covering 14 Indigenous languages from North, Central, and South America, catalyzing new efforts in corpus compilation, data curation, and evaluation protocols tailored for severely data-scarce contexts (De Gibert et al. 2025). These efforts not only facilitate digital access for largely marginalized language communities but also reinforce ongoing programs in language revitalization, educational outreach, and cultural heritage preservation.

Methodologically, most prior work on Indigenous language translation employs supervised fine-tuning of large language models (LLMs) on small, carefully curated parallel datasets (De Gibert et al. 2025; Hus, Anastasopoulos, and Krasner 2025). While such approaches have yielded promising gains in some low-resource scenarios, they remain fundamentally constrained by the availability of annotated data and tend to generalize poorly to out-of-distribution inputs (Ignat et al. 2024; Hettiarachchi et al. 2025; Khade et al. 2025; De Gibert et al. 2025). Consequently, purely supervised paradigms struggle to capture the linguistic richness and variability inherent to Indigenous languages, which often exhibit complex morphology, dialectal variation, and limited orthographic standardization.

Recently, reinforcement learning (RL) has emerged as a promising post-training strategy, requiring far fewer annotated examples and capable of both complementing and supplanting traditional supervised techniques. These RL methods such as Proximal Policy Optimization (PPO) (Schulman et al. 2017) and the more recent Group Relative Policy Optimization (GRPO) (Shao et al. 2024; Liu et al. 2025) have gained popularity in LLM training. These techniques have proven effective in aligning model outputs with human preferences, as demonstrated in Reinforcement Learning from Human Feedback (RLHF) (Ouyang et al. 2022), and have subsequently been employed to enhance the reasoning abilities of LLMs (DeepSeek-AI et al. 2025). In contrast to supervised fine-tuning, RL enables models to learn policies over sequences of actions, facilitating dynamic interaction with an environment and enabling better adaptation to sparse or delayed feedback. However, RL methods have yet to be ex-

plored in the context of machine translation, particularly in low-resource settings.

Furthermore, RL has been used to extend model capabilities through the integration of external tools that help the model with different tasks like executing code, performing math calculations, or searching the web (Jin et al. 2025; Feng et al. 2025; Goldie et al. 2025). These agent-like abilities enhance model performance in domains where specialized tools can provide meaningful support. A key advantage of RL in tool usage is that it enables models to learn autonomously how to use tools effectively to improve task performance. Despite its effectiveness, little work has focused on developing or leveraging such tools specifically for machine translation (Briva-Iglesias 2025), especially in the low-resource context.

In this paper, we propose an alternative to traditional fine-tuning strategies for improving machine translation performance in Wayuunaiki, the most widely spoken Indigenous language in Colombia. Our approach builds on the instruction-tuned model Qwen2.5-0.5B-Instruct (Qwen et al. 2025), which we further train using reinforcement learning. Unlike standard methods, we frame the model as an agent capable of interacting with an external Wayuunaiki–Spanish dictionary. To support this interaction, we adopted the GRPO framework introduced by DeepSeek (DeepSeek-AI et al. 2025), enabling the model to learn when and how to call the dictionary. This agent-based formulation facilitates tool-augmented translation and reduces reliance on large annotated corpora. To the best of our knowledge, this is the first work to incorporate a dictionary as an interactive tool in low-resource machine translation, and the first to apply RL to adapt LLMs in the translation context. By framing the model as an agent, our methodology opens new avenues for research into tool-augmented translation strategies for underrepresented languages.

Paper organization

This paper is divided into four main sections. The Related Work section reviews existing approaches to machine translation for low-resource and Indigenous languages, emphasizing the challenges of data scarcity and highlighting recent efforts to incorporate reinforcement learning into translation. The Methods section presents our framework for tool-augmented translation, describing both the supervised fine-tuning pipeline and the reinforcement learning setup, including the GRPO algorithm, the construction of our parallel corpus, model selection, and training protocols. In the Results section, we present our experimental findings, followed by the Discussion section, which reflects on the implications of tool-augmented machine translation in low-resource settings, addresses limitations, and outlines directions for future research.

Related Work

Wayuunaiki is an Arawakan language primarily used within the Wayuu indigenous community and is spoken by approximately 420,000 people across northern Colombia and Venezuela. Additionally, in contrast to English, it features a predominant subject–object–verb (SOV) word order and exhibits agglutinative morphology, in which words are formed by combining morphemes, each contributing distinct semantic or grammatical information. However, despite its relatively large number of speakers compared to other indigenous languages in the region, Wayuunaiki remains underrepresented in the NLP field, with few applications and datasets available.

Most efforts to date have focused on developing linguistic resources—such as aligned sentence-pair corpora and descriptive analyses—and on building Wayuunaiki–Spanish translation systems. Notable examples include Rafael José Negrette Amaya's bilingual Wayuunaiki–Spanish dictionary, which contains over 74,000 entries (Amaya 2021), and the aligned translations of religious and institutional texts, ranging from the Bible and the Colombian Constitution to various educational materials and linguistic studies of Wayuunaiki (Prieto et al. 2024). In terms of translation systems, key developments include the first Wayuunaiki–Spanish neural machine translation system built in 2023 (Graichen, Van Genabith, and España-bonet 2023); the fine-tuning of large Finnish-language pretrained models selected for their structural parallels to Wayuunaiki; and adaptations of multilingual frameworks such as Meta's No Language Left Behind (NLLB) model, which supports numerous low-resource languages (Robles et al. 2024; Prieto et al. 2024; Hus, Anastasopoulos, and Krasner 2025; NLLBTeam 2022).

While these efforts demonstrate that contemporary architectures can be adapted to Wayuunaiki–Spanish translation, published evaluations report modest performance, primarily due to the scarcity of parallel data and the narrow topical coverage of existing corpora (Graichen, Van Genabith, and España-bonet 2023; Hus, Anastasopoulos, and Krasner 2025). Moreover, training data frequently fail to reflect the language as it is actively spoken: in the AmericasNLP Shared Task, BLEU scores on up-to-date, carefully curated test sets differ markedly from those on standard validation sets, highlighting the need for novel, data-efficient modeling techniques and for resources that better capture real-world linguistic variation (De Gibert et al. 2025).

Recently, researchers have found that adopting RL techniques as an additional training stage for LLMs can significantly improve their performance, while requiring substantially less data than in the pre-training phase. Specifically, these advancements have been driven by two RL algorithms, PPO (Schulman et al. 2017), which was used in the popular RLHF method (Ouyang et al. 2022) to better align the output of models with user preferences; and GRPO (Liu et al. 2025; DeepSeek-AI et al. 2025; Shao et al. 2024), introduced by DeepSeek to further enhance memory efficiency during RL-based training and to allow models to improve their coding, math, and reasoning capabilities.

In 2024, Zhan et. al (Zhang et al. 2024) introduced a reinforcement learning domain adaptation approach for neural machine translation, utilizing in-domain monolingual data to mitigate overfitting and reinforce domain-specific knowledge acquisition. Their method involves training a ranking-based model with a small-scale in-domain parallel corpus,

which serves as a reward model to select higher-quality generated translations during fine-tuning.

Apart from the promise of RL techniques, agent-based frameworks have also been proposed to address the complexities of translation tasks. For instance, inspired by traditional human translation workflows, Briva-Iglesias (2025) presented a multi-agent system for translating ultra-long literary texts, where specialized agents collaborate to handle different aspects of the translation process—such as adequacy review and fluency enhancement—resulting in translations that better maintain contextual fidelity and cultural nuances. While not specifically designed for translation tasks, other agent-based solutions have shown great potential by integrating external tools into LLMs, thus extending their abilities to perform more complex tasks. Recent approaches such as Search-R1 (Jin et al. 2025), ReTool (Feng et al. 2025), and SWiRL (Goldie et al. 2025) even employ reinforcement learning to teach models when and how to use these external tools, which include code interpreters, calculators, or web search. In the translation domain, some works incorporate external resources such as dictionaries or translation examples (Shu et al. 2024; Merx et al. 2024; Court and Elsner 2024). However, in these approaches, the model does not learn when or how to utilize the external information. Tool usage is either predetermined or statically appended to the input. In contrast, our approach leverages RL to train the model to adaptively decide when to query the dictionary tool, enabling it to use external resources only when beneficial to translation quality.

Methods

Figure 1 summarizes our methodology. To develop our translation system, we start with an already pretrained large language model capable of following user instructions. After selecting this base model, we perform supervised fine-tuning using an artificially augmented dataset that consist of Wayuunaiki-Spanish translation pairs and automatically generated examples of dictionary lookups. Finally, we use RL to boost the translation performance of our system.

Supervised fine-tuning phase

The supervised fine-tuning stage serves two key purposes: (1) to train the model to produce outputs in a structured format using predefined tags, and (2) to enable the model to learn how to properly invoke the dictionary tool. In this stage, we train the model on Spanish–Wayuunaiki translation examples using a prompt template that instructs the model how to invoke the dictionary tool and how to format its final translation.

As is common practice, the Spanish text and its corresponding Wayuunaiki translation are concatenated to the previous prompt to illustrate the translation task. To teach the model how to use the external dictionary tool, we insert artificial examples of dictionary calls immediately before the Wayuunaiki translation. To generate these examples, between zero and four words are randomly selected from the Spanish side to be queried using the dictionary tool. Then, for each lookup, the output of the dictionary—which consists of the first five matches from the dictionary entries—is also appended to the prompt.

Although these examples are randomly generated and are probably useless to achieve the correct translation, recent findings on the cognitive behaviors underlying self-improving reasoning in language models (Gandhi et al. 2025) suggest that acquiring structured habits, such as proper tool usage, can further enhance the performance achieved in the reinforcement learning stage. This benefit arises because the reinforcement learning phase can focus only on refining its tool usage rather than having to learn it entirely from scratch.

Reinforcement learning phase

Once the model has been fine-tuned to follow the structured prompt format and correctly use the dictionary tool, we proceed to the reinforcement learning stage. We adopt the GRPO framework (DeepSeek-AI et al. 2025), which is designed to align LLM behavior with complex tasks. In this setup, the language model itself acts as the policy. At each training step, we sample a Spanish–Wayuunaiki sentence pair and generate multiple candidate translations. Specifically, we generate 8 different translations for the same input prompt as defined during fine-tuning, which potentially include different combinations of dictionary tool invocations.

For each prediction, only the text enclosed within the `<answer>` tags is extracted and used for evaluation. Each generated output is then evaluated against a reference translation using BLEU (Papineni et al. 2002), which serves as the reward signal for GRPO to update the policy based on translation quality. Additionally, tool outputs are masked to ensure they do not contribute to the policy loss (Jin et al. 2025). This process enables the model to iteratively refine its translation strategy, improving overall performance while learning when and how to use the dictionary tool more effectively. To monitor progress during training, we evaluate the model every 50 steps on a fixed set of 640 sentence pairs sampled from the training dataset.

Since our task involves translating into Wayuunaiki, a language that differs significantly from the original training distribution of the model, we adopt the approach used in DAPO (Yu et al. 2025) and Dr.GRPO (Liu et al. 2025), which relax the traditional GRPO constraint based on KL-divergence penalties. This adjustment is essential because the model must undergo substantial behavioral changes to produce coherent Wayuunaiki translations. Standard regularization methods that constrain the model to remain close to its initial policy would limit its ability to adapt effectively.

Datasets and models

For training, we use the Spanish–Wayuunaiki parallel corpus introduced by Prieto et. al (Prieto et al. 2024), which was included in the AmericasNLP 2025 Shared Task (De Gibert et al. 2025). This dataset was chosen because it provides a more natural and modern context for evaluation, rather than relying on translations of formal documents such as the Bible.

To support tool-augmented translation, we incorporate a bilingual dictionary compiled by Rafael Jose Negrette

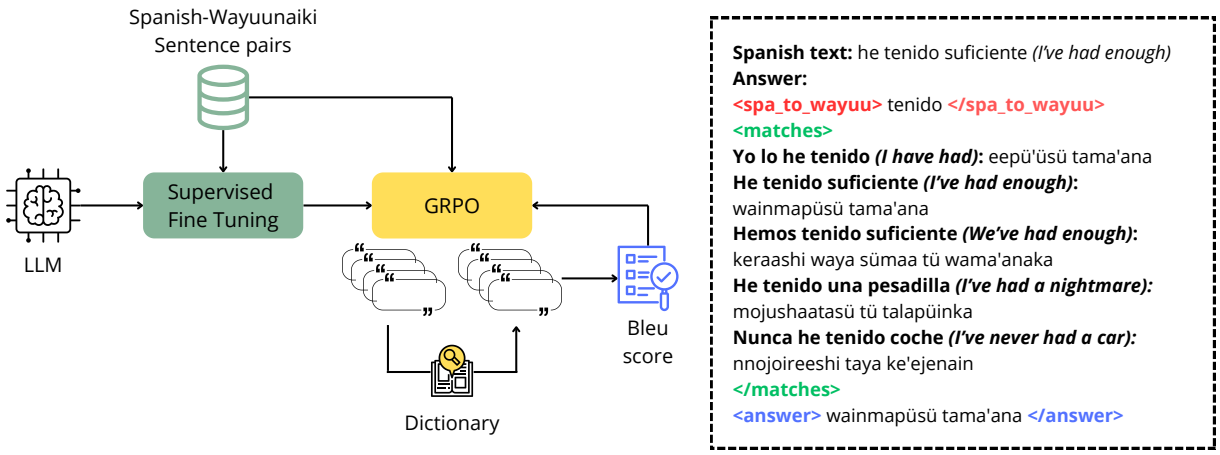

Figure 1: Overview of the training pipeline. A large language model is first finetuned using supervised learning on Spanish–Wayuunaiki sentence pairs. The finetuned model is then further optimized using GRPO, where the reward is based on BLEU scores computed against reference translations. During this phase, the model can optionally use a dictionary tool to assist translation. The right-hand side illustrates an example of how the model interacts with the dictionary during the generation process. English translations were added to the image to help the reader follow along more easily.

Amaya (Amaya 2021), which originally contains approximately 74,000 Spanish–Wayuunaiki word and phrase pairs. To ensure tool responses remain concise and manageable, we filter this dictionary to retain only entries with five words or fewer on the Spanish side, resulting in a final dictionary of approximately 29,000 entries.

For testing, we employ a curated translation dataset consisting of the opening pages of Jules Verne's *Journey to the Center of the Earth* (Verne 1864), translated into Wayuunaiki by the company Wayuunaiki Translation Services and funded by the Universidad de Los Andes. This dataset was used as the official test set in the AmericasNLP 2025 Shared Task, underscoring the importance of employing up-to-date, native-speaker translations, since training corpora (e.g., the Bible, the Colombian Constitution) often differ substantially from contemporary spoken usage.

As a base instruction model, we use Qwen2.5-0.5B-Instruct (Qwen et al. 2025), which offers multilingual support across more than 20 languages and is specifically optimized for cross-lingual tasks. One of the key design choices behind this model is its ability to generalize across languages through a cross-lingual transfer mechanism. This is achieved by translating instructions from high-resource languages into low-resource ones and generating corresponding response candidates. This training strategy makes Qwen2.5-0.5B-Instruct particularly well-suited for tasks involving low-resource languages such as Wayuunaiki, where robust generalization and instruction-following are essential.

Training

To evaluate model performance during training, we use the BLEU score (Papineni et al. 2002), which measures translation quality by comparing overlapping n-grams between the generated output and a reference. For parameter-efficient adaptation, we apply LoRA (Low-Rank Adaptation) (Hu et al. 2021) in both supervised fine-tuning and reinforcement learning. In the RL phase, we further optimize for efficiency and stability by (1) leveraging vLLM (Kwon et al. 2023) for faster inference and trajectory sampling, (2) accumulating gradients over eight steps to balance memory footprint and effective batch size, (3) integrating DeepSpeed (Rasley et al. 2020) to reduce memory usage and boost throughput, and (4) omitting clipping in the policy loss, which allows us to keep only a single model instance in memory throughout training. All models are optimized with AdamW at a fixed learning rate of 5×10^{-6}.

Experimental setup

Our experiments systematically evaluate three key factors: training approach (zero-shot, supervised fine-tuning, reinforcement learning), dictionary access (available vs. unavailable), and model architecture (instruction-tuned vs. translation-specific models).

We begin by establishing baselines using the instruction-tuned model Qwen2.5-0.5B-Instruct in zero-shot settings. To test whether tool awareness alone is beneficial, we also include a variant where the model is informed that a dictionary is available but receives no examples of how to use it.

We then explore supervised fine-tuning to assess whether explicit demonstrations improve performance. One set of experiments uses standard parallel sentence pairs without tool interaction, serving to isolate the benefits of exposure to target-domain data. A second set extends this by introducing synthetic demonstrations that show the model how to use the dictionary tool. These examples are automatically constructed and illustrate when and how to query the tool during translation, allowing us to test whether models can learn tool-augmented behaviors from examples alone. For both settings, models were fine-tuned for one epoch on 59,715 paired sentences, using a learning rate of 1×10^{-4}, the AdamW optimizer, and prompt masking to ensure training focused only on the target completions.

We then evaluate a combined approach where SFT is followed by RL, in order to assess whether reinforcement learning can further refine tool usage and translation quality after initial supervised adaptation. These experiments are run both with and without tool access, allowing us to isolate the impact of the dictionary in the context of policy optimization. Notably, RL training for the tool-enabled model is performed on an SFT-trained version that incorporates tool usage, whereas for the tool-free model, RL is applied to an SFT-trained version that was not exposed to the tool.

Within the RL framework, we explore two reward strategies: sentence-level BLEU scores (Papineni et al. 2002) and character-level edit-based rewards (Morris, Maier, and Green 2004). Additionally, we examine the effect of RL training duration by directly comparing the performance of models trained for 400 steps versus those trained for 1400 steps.

Finally, to assess the generality of our approach, we replicate key experiments across different model architectures. We apply our full methodology—involving SFT and RL with dictionary access—to Llama-3.2-1B-Instruct, enabling a comparison over different pretraining bases. We also test a larger model, Qwen2.5-7B-Instruct, to explore whether scale offers measurable gains in low-resource translation. In parallel, we test our RL framework on a translation-specific model, NLLB (NLLBTeam 2022), which is not instruction-tuned and cannot utilize the tool. For this setup, we use the Wayuunaiki-specific checkpoint from (Prieto et al. 2024) and apply GRPO without tool access or prompting, thereby isolating the effects of reinforcement learning on a model with strong translation priors.

To evaluate all our models, we use the average BLEU score computed between sentences on the 503 samples from the test set. Additionally, we measure different metrics to analyze tool usage. To ensure cost efficiency, we cap the number of allowed dictionary calls at a maximum of four.

Results

This section presents the experimental results evaluating the performance of different models and training approaches for Spanish-to-Wayuunaiki translation, primarily using the BLEU score as the evaluation metric.

Figure 2 presents the main results for the Qwen model under three configurations: without any fine-tuning (Base), with supervised fine-tuning (SFT), and with an additional reinforcement learning (RL) stage comprising 1,400 steps, using BLEU as the reward signal. The base Qwen-0.5B model achieved very low BLEU scores, underscoring the need for training on Wayuunaiki data. Performance improved consistently at each stage of training, with SFT contributing the largest gain, and RL delivering an additional 11% improvement, both with and without dictionary access. Additionally, the external dictionary tool provided a relative performance boost of approximately 6% in both the SFT and SFT+RL stages. While prior work reported an average BLEU score of 10.54 on a test set similar to their training set (Robles et al. 2024), their model achieved only 0.93 BLEU on the curated test set used in our evaluation (De Gibert et al. 2025). These results demonstrate the effectiveness of our combined SFT and RL training pipeline, particularly when enhanced by access to an external dictionary tool.

Table 1 offers a detailed breakdown of performance and tool usage across our training pipeline with the dictionary enabled. Notably, **the best-performing model (Qwen-0.5B+SFT+RL) makes the most extensive use of the dictionary**, employing it in every case and averaging 3.94 calls per sample, close to the allowed maximum of 4. The SFT stage plays a key role in enhancing performance by providing examples that teach the model both accurate translation pairs and effective tool usage. This is reflected in a success rate of almost 90% when querying the dictionary, i.e., receiving valid matches for the queried word. These capabilities were further reinforced during the RL stage, which enabled the model to fully exploit the external tool, achieving a 95% success rate.

Model	Avg. BLEU	Answers w/ Tools	Avg. Tool Calls	Succ. Tool Calls
Base	0.06	45.72%	1.00	0.02%
Base+SFT	3.08	99.00%	2.13	89.76%
Base+SFT+RL	**3.42**	**100.00%**	**3.94**	**95.23%**

Table 1: Tool usage and BLEU scores for different variants of the Qwen-0.5B model. The results indicate that better-performing models make more extensive use of the dictionary tool. Notably, the Qwen-0.5B+SFT+RL model invokes the tool in every response and approaches the maximum allowed number of calls per translation, averaging 3.94 out of 4.

Moreover, in Table 2, we evaluate our proposed method using different model architectures: Qwen2.5, LLaMA3.2, and NLLB. We also assess its effectiveness across different sizes of the Qwen model (0.5B and 7B parameters). For NLLB, which is not instruction-tuned, the dictionary tool is disabled. Additionally, the base NLLB model cannot be tested, as it does not natively support Wayuunaiki.

The results indicate that instruction-tuned models (Qwen and LLaMA) benefit significantly from both the SFT and SFT+RL stages when tool access is enabled. All instruction-tuned models achieve their best performance when trained using the complete pipeline. In contrast, the RL stage does not appear to enhance the performance of the NLLB model, which remains below that of the other tested models. Notably, with the exception of NLLB, larger models tend to achieve better results. Qwen2.5-7B reaches the highest average BLEU score of 4.45, outperforming all other models.

Tool usage also becomes more frequent and sophisticated across training stages, as models learn to more effectively leverage the dictionary. Since base larger models like Qwen2.5-7B are already capable of using the tool properly, tool usage does not necessarily increase in volume but becomes more refined, contributing to improved performance. A more detailed analysis of tool usage is provided in the following subsection.

In Table 3, we analyze the impact of different reward signals (BLEU versus CharacTer Error Rate) and the number

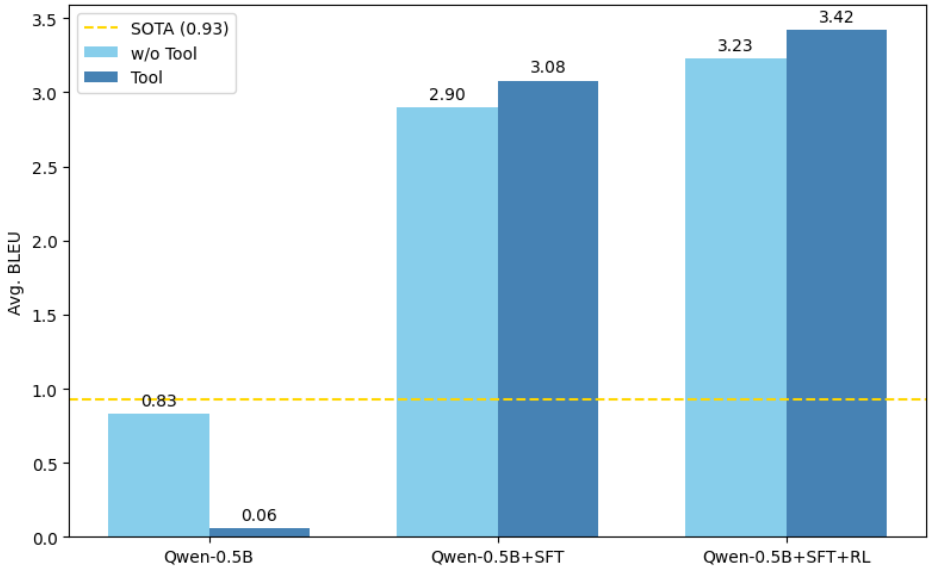

Figure 2: Average BLEU scores for different Qwen model variants, with and without tool usage. The results show that SFT effectively imparts basic translation capabilities, while RL yields a modest improvement on top of it. Enabling the dictionary tool provides an estimated 6% relative gain.

Model	Avg. BLEU	Answers w/ Tools	Avg. Tool Calls
Base Models			
Qwen-0.5B	0.06	45.72%	1.00
Llama3.2-1B	0.11	59.05%	2.31
Qwen-7B	2.10	94.04%	**4.22**
NLLB-3B	–	–	–
+ SFT			
Qwen-0.5B	3.08	99%	2.13
Llama3.2-1B	3.15	99%	2.98
Qwen-7B	4.33	97.81%	2.97
NLLB-3B	0.93	–	–
+ RL			
Qwen-0.5B	3.16	**100%**	2.97
Llama3.2-1B	3.48	**100%**	3.88
Qwen-7B	**4.45**	98.01%	2.78
NLLB-3B	0.93	–	–

Table 2: Performance comparison across base models, SFT, and RL stages. Instruction-tuned models show significant improvements through both SFT and RL, partly due to their increasing use of external tools, as analyzed in the subsequent results subsection. Larger instruction-tuned models tend to perform better, with Qwen7B+SFT+RL achieving the highest score (4.45 Avg. BLEU), effectively doubling its base performance.

of RL steps (400 vs. 1400) during the final RL training stage of the Qwen2.5-0.5B model. The results indicate that the BLEU metric is the only effective signal for improving the translation performanceof the model, yielding a 2.6% improvement after 400 steps and achieving an 11% relative gain with 1400 steps. In contrast, using the CharacTer metric leads to a 10.4% performance degradation. Although there is some improvement after the initial 400 steps, the performance does not recover after 1400 steps of training.

Despite the divergence in translation quality, both reward signals lead to increased tool usage over the course of RL training. The average number of tool calls per use rises from 2.13 to 3.94, and tool usage frequency increases from 99% to 100% after 1400 steps with both metrics.

Reward Signal	Avg. BLEU	Answers w/ Tools	Avg. Tool Calls
Qwen-0.5+SFT	3.08	99%	2.13
400 RL Steps			
BLEU	3.16	**100%**	2.97
CharacTer	2.59	**100%**	3.02
1400 RL Steps			
BLEU	**3.42**	**100%**	**3.94**
CharacTer	2.76	**100%**	**3.94**

Table 3: Effect of reward signal type and RL training duration on BLEU scores and tool usage. The results show that BLEU scores outperform CharacTer scores as the reward signal. Increasing the number of RL training steps significantly improves performance and encourages more intensive tool usage.

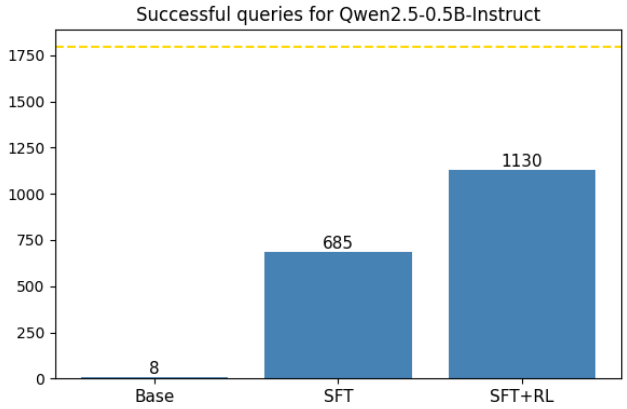

Figure 3: Number of dictionary lookups that returned results, referred to as successful queries. The results indicate that successful queries increase across training stages. The yellow horizontal line marks the theoretical upper bound of successful queries in our setup, which limited each sample to a maximum of 4 dictionary calls. Considering this constraint and filtering only for words present in the dictionary, the maximum achievable number of successful queries is 1798.

Dictionary Usage Analysis

To evaluate how effectively the models leverage the dictionary tool, we measured the number of successful dictionary lookups for three versions: the base model, the model fine-tuned with SFT, and the model trained with both SFT and RL. As an upper bound, we defined a successful query as one where the Spanish word appears in the dictionary. Since the model is limited to querying a maximum of four words per sample, the theoretical maximum number of successful queries is 1,798.

As shown in Figure 3, the number of successful lookups increases substantially at each training stage. The model trained with both SFT and RL achieved 1,130 successful lookups, a 65% improvement over the model trained with SFT alone. These results highlight the effectiveness of each training phase in teaching the model to better utilize the dictionary tool to enhance translation performance. The fully trained model reaches 63% of the theoretical maximum. However, it is important to note that, due to knowledge already acquired during the SFT phase, querying every word may not be necessary, as some words may already be known by the model.

Furthermore, we assessed the impact of dictionary integration on translation quality by comparing, for each lookup, the maximum BLEU score attainable using only the dictionary's best suggestion against the BLEU score of the final output of the model. For the SFT model, the mean "dictionary-only" BLEU is 0.109, whereas the mean BLEU of the model reaches 3.07; a paired two-sided t-test yields a p-value of $p = 6 \times 10^{-17}$, and 64% of examples have a better BLEU score for the model output than for the best dictionary result. Similarly, the SFT+RL model attains a mean "dictionary-only" BLEU of 0.21 and a mean BLEU of 3.42 for the model's output ($p = 4.6 \times 10^{-15}$), with improvements in 63.2% of cases. These results demonstrate that the trained models, when using the dictionary, produce translations that are statistically significantly better than simply selecting the best dictionary result. This suggests that the models do not merely copy from the dictionary but effectively refine and enhance suggestions using their learned language knowledge.

Nevertheless, we identified important limitations in the dictionary itself. Only 10.4% of the unique Spanish words in the test set appear as entries in the dictionary, and of these, just 16.3% provide a Wayuunaiki translation that matches the reference.

Discussion and Future Work

Our findings provide strong evidence that LLMs trained using SFT and RL to leverage external lexical resources, such as dictionaries, significantly improve translation performance in low-resource settings. These results were consistent across different model architectures, including LLaMA and Qwen, and across various model sizes.

Although our experiments focused exclusively on the Wayuunaiki language, the methodology is broadly applicable, as it does not rely on any language-specific techniques. As long as a dictionary is available, our approach can be readily extended to other languages. In fact, for non-agglutinative languages, the benefits could be even greater, since words in such languages are typically easier to translate independently. This contrasts with agglutinative languages like Wayuunaiki, where words are often formed by chaining multiple subwords, complicating the translation process.

Importantly, the improvements from our method are complementary to those achieved through traditional SFT on parallel corpora. This suggests a promising research direction for enhancing translation performance beyond the limitations imposed by the scarcity of parallel data.

Despite our success, we observed that the effectiveness of the dictionary tool was significantly constrained by both its limited coverage of the Wayuunaiki language and its overall quality. In many cases, the suggestions of the tool did not align with our reference translations. This underscores the critical need to develop high-quality, reliable external resources that can support language models in future work.

Our experiments also revealed that the effectiveness of the RL stage is highly dependent on the type of reward signal employed. This raises important questions about why the CharacTer reward signal (which focuses on character-level matches rather than word-level matches, like BLEU) was insufficient to drive improvements and, in some cases, even led to performance regressions. Future research could investigate the properties that make a reward function effective in the context of machine translation.

Another crucial consideration is the use of evaluation datasets with multiple reference translations. Such datasets can account for the various valid ways to express the same content, thereby enabling the design of more robust and representative reward signals.

References

Amaya, R. J. N. 2021. OSF spanish-wayuunaki.

Briva-Iglesias, V. 2025. Are AI agents the new machine translation frontier? Challenges and opportunities of single- and multi-agent systems for multilingual digital communication. arXiv:2504.12891.

Court, S.; and Elsner, M. 2024. Shortcomings of LLMs for Low-Resource Translation: Retrieval and Understanding are Both the Problem. arXiv:2406.15625.

De Gibert, O.; Pugh, R.; Marashian, A.; Vazquez, R.; Ebrahimi, A.; Denisov, P.; Rice, E.; Gow-Smith, E.; Prieto, J.; Robles, M.; Manrique, R.; Moreno, O.; Lino, A.; Coto-Solano, R.; Alvarez, A.; Agüero-Torales, M.; Ortega, J. E.; Chiruzzo, L.; Oncevay, A.; Rijhwani, S.; Von Der Wense, K.; and Mager, M. 2025. Findings of the AmericasNLP 2025 Shared Tasks on Machine Translation, Creation of Educational Material, and Translation Metrics for Indigenous Languages of the Americas. In Mager, M.; Ebrahimi, A.; Pugh, R.; Rijhwani, S.; Von Der Wense, K.; Chiruzzo, L.; Coto-Solano, R.; and Oncevay, A., eds., *Proceedings of the Fifth Workshop on NLP for Indigenous Languages of the Americas (AmericasNLP)*, 134–152. Albuquerque, New Mexico: Association for Computational Linguistics. ISBN 979-8-89176-236-7.

DeepSeek-AI; Guo, D.; Yang, D.; Zhang, H.; Song, J.; Zhang, R.; Xu, R.; Zhu, Q.; Ma, S.; Wang, P.; Bi, X.; Zhang, X.; Yu, X.; Wu, Y.; Wu, Z. F.; Gou, Z.; Shao, Z.; Li, Z.; Gao, Z.; Liu, A.; Xue, B.; Wang, B.; Wu, B.; Feng, B.; Lu, C.; Zhao, C.; Deng, C.; Zhang, C.; Ruan, C.; Dai, D.; Chen, D.; Ji, D.; Li, E.; Lin, F.; Dai, F.; Luo, F.; Hao, G.; Chen, G.; Li, G.; Zhang, H.; Bao, H.; Xu, H.; Wang, H.; Ding, H.; Xin, H.; Gao, H.; Qu, H.; Li, H.; Guo, J.; Li, J.; Wang, J.; Chen, J.; Yuan, J.; Qiu, J.; Li, J.; Cai, J. L.; Ni, J.; Liang, J.; Chen, J.; Dong, K.; Hu, K.; Gao, K.; Guan, K.; Huang, K.; Yu, K.; Wang, L.; Zhang, L.; Zhao, L.; Wang, L.; Zhang, L.; Xu, L.; Xia, L.; Zhang, M.; Zhang, M.; Tang, M.; Li, M.; Wang, M.; Li, M.; Tian, N.; Huang, P.; Zhang, P.; Wang, Q.; Chen, Q.; Du, Q.; Ge, R.; Zhang, R.; Pan, R.; Wang, R.; Chen, R. J.; Jin, R. L.; Chen, R.; Lu, S.; Zhou, S.; Chen, S.; Ye, S.; Wang, S.; Yu, S.; Zhou, S.; Pan, S.; Li, S. S.; Zhou, S.; Wu, S.; Ye, S.; Yun, T.; Pei, T.; Sun, T.; Wang, T.; Zeng, W.; Zhao, W.; Liu, W.; Liang, W.; Gao, W.; Yu, W.; Zhang, W.; Xiao, W. L.; An, W.; Liu, X.; Wang, X.; Chen, X.; Nie, X.; Cheng, X.; Liu, X.; Xie, X.; Liu, X.; Yang, X.; Li, X.; Su, X.; Lin, X.; Li, X. Q.; Jin, X.; Shen, X.; Chen, X.; Sun, X.; Wang, X.; Song, X.; Zhou, X.; Wang, X.; Shan, X.; Li, Y. K.; Wang, Y. Q.; Wei, Y. X.; Zhang, Y.; Xu, Y.; Li, Y.; Zhao, Y.; Sun, Y.; Wang, Y.; Yu, Y.; Zhang, Y.; Shi, Y.; Xiong, Y.; He, Y.; Piao, Y.; Wang, Y.; Tan, Y.; Ma, Y.; Liu, Y.; Guo, Y.; Ou, Y.; Wang, Y.; Gong, Y.; Zou, Y.; He, Y.; Xiong, Y.; Luo, Y.; You, Y.; Liu, Y.; Zhou, Y.; Zhu, Y. X.; Xu, Y.; Huang, Y.; Li, Y.; Zheng, Y.; Zhu, Y.; Ma, Y.; Tang, Y.; Zha, Y.; Yan, Y.; Ren, Z. Z.; Ren, Z.; Sha, Z.; Fu, Z.; Xu, Z.; Xie, Z.; Zhang, Z.; Hao, Z.; Ma, Z.; Yan, Z.; Wu, Z.; Gu, Z.; Zhu, Z.; Liu, Z.; Li, Z.; Xie, Z.; Song, Z.; Pan, Z.; Huang, Z.; Xu, Z.; Zhang, Z.; and Zhang, Z. 2025. DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning. arXiv:2501.12948.

Feng, J.; Huang, S.; Qu, X.; Zhang, G.; Qin, Y.; Zhong, B.; Jiang, C.; Chi, J.; and Zhong, W. 2025. ReTool: Reinforcement Learning for Strategic Tool Use in LLMs. arXiv:2504.11536.

Gandhi, K.; Chakravarthy, A.; Singh, A.; Lile, N.; and Goodman, N. D. 2025. Cognitive Behaviors that Enable Self-Improving Reasoners, or, Four Habits of Highly Effective STaRs. arXiv:2503.01307.

Goldie, A.; Mirhoseini, A.; Zhou, H.; Cai, I.; and Manning, C. D. 2025. Synthetic Data Generation and Multi-Step RL for Reasoning and Tool Use. arXiv:2504.04736.

Graichen, N.; Van Genabith, J.; and España-bonet, C. 2023. Enriching Wayúunaiki-Spanish Neural Machine Translation with Linguistic Information. In Mager, M.; Ebrahimi, A.; Oncevay, A.; Rice, E.; Rijhwani, S.; Palmer, A.; and Kann, K., eds., *Proceedings of the Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP)*, 67–83. Toronto, Canada: Association for Computational Linguistics.

Hettiarachchi, H.; Ranasinghe, T.; Rayson, P.; Mitkov, R.; Gaber, M.; Premasiri, D.; Tan, F. A.; and Uyangodage, L., eds. 2025. *Proceedings of the First Workshop on Language Models for Low-Resource Languages*. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics.

Hu, E. J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; and Chen, W. 2021. LoRA: Low-Rank Adaptation of Large Language Models. arXiv:2106.09685.

Hus, J.; Anastasopoulos, A.; and Krasner, N. 2025. Machine Translation Using Grammar Materials for LLM Post-Correction. In Mager, M.; Ebrahimi, A.; Pugh, R.; Rijhwani, S.; Von Der Wense, K.; Chiruzzo, L.; Coto-Solano, R.; and Oncevay, A., eds., *Proceedings of the Fifth Workshop on NLP for Indigenous Languages of the Americas (AmericasNLP)*, 92–99. Albuquerque, New Mexico: Association for Computational Linguistics. ISBN 979-8-89176-236-7.

Ignat, O.; Jin, Z.; Abzaliev, A.; Biester, L.; Castro, S.; Deng, N.; Gao, X.; Gunal, A. E.; He, J.; Kazemi, A.; Khalifa, M.; Koh, N.; Lee, A.; Liu, S.; Min, D. J.; Mori, S.; Nwatu, J. C.; Perez-Rosas, V.; Shen, S.; Wang, Z.; Wu, W.; and Mihalcea, R. 2024. Has It All Been Solved? Open NLP Research Questions Not Solved by Large Language Models. In Calzolari, N.; Kan, M.-Y.; Hoste, V.; Lenci, A.; Sakti, S.; and Xue, N., eds., *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, 8050–8094. Torino, Italia: ELRA and ICCL.

Jin, B.; Zeng, H.; Yue, Z.; Wang, D.; Zamani, H.; and Han, J. 2025. Search-R1: Training LLMs to Reason and Leverage Search Engines with Reinforcement Learning. ArXiv:2503.09516 [cs].

Khade, O.; Jagdale, S.; Phaltankar, A.; Takalikar, G.; and Joshi, R. 2025. Challenges in Adapting Multilingual LLMs to Low-Resource Languages using LoRA PEFT Tuning. In Sarveswaran, K.; Vaidya, A.; Krishna Bal, B.; Shams, S.; and Thapa, S., eds., *Proceedings of the First Workshop on Challenges in Processing South Asian Languages (CHiP-*

SAL 2025), 217–222. Abu Dhabi, UAE: International Committee on Computational Linguistics.

Kwon, W.; Li, Z.; Zhuang, S.; Sheng, Y.; Zheng, L.; Yu, C. H.; Gonzalez, J. E.; Zhang, H.; and Stoica, I. 2023. Efficient Memory Management for Large Language Model Serving with PagedAttention. In *Proceedings of the ACM SIGOPS 29th Symposium on Operating Systems Principles*.

Liu, Z.; Chen, C.; Li, W.; Qi, P.; Pang, T.; Du, C.; Lee, W. S.; and Lin, M. 2025. Understanding R1-Zero-Like Training: A Critical Perspective. arXiv:2503.20783.

Lucas, M.; Burgueño, A.; Carazas, M.; Buenadicha Sánchez, C.; Ramirez Rufino, S.; Rosales Torres, C. S.; Korn, D.; Tesfaye, H.; and Deo, G. 2025. The Performance of Artificial Intelligence in the Use of Indigenous American Languages.

Merx, R.; Mahmudi, A.; Langford, K.; de Araujo, L. A.; and Vylomova, E. 2024. Low-Resource Machine Translation through Retrieval-Augmented LLM Prompting: A Study on the Mambai Language. arXiv:2404.04809.

Morris, A.; Maier, V.; and Green, P. 2004. From WER and RIL to MER and WIL: improved evaluation measures for connected speech recognition.

NLLBTeam. 2022. No Language Left Behind: Scaling Human-Centered Machine Translation. arXiv:2207.04672.

Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C. L.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K.; Ray, A.; Schulman, J.; Hilton, J.; Kelton, F.; Miller, L.; Simens, M.; Askell, A.; Welinder, P.; Christiano, P.; Leike, J.; and Lowe, R. 2022. Training language models to follow instructions with human feedback. arXiv:2203.02155.

Papineni, K.; Roukos, S.; Ward, T.; and Zhu, W.-J. 2002. BLEU: A Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02. USA: Association for Computational Linguistics.

Prieto, J.; Martinez, C.; Robles, M.; Moreno, A.; Palacios, S.; and Manrique, R. 2024. Translation systems for low-resource Colombian Indigenous languages, a first step towards cultural preservation. In Mager, M.; Ebrahimi, A.; Rijhwani, S.; Oncevay, A.; Chiruzzo, L.; Pugh, R.; and von der Wense, K., eds., *Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2024)*, 7–14. Mexico City, Mexico: Association for Computational Linguistics.

Qwen; :; Yang, A.; Yang, B.; Zhang, B.; Hui, B.; Zheng, B.; Yu, B.; Li, C.; Liu, D.; Huang, F.; Wei, H.; Lin, H.; Yang, J.; Tu, J.; Zhang, J.; Yang, J.; Yang, J.; Zhou, J.; Lin, J.; Dang, K.; Lu, K.; Bao, K.; Yang, K.; Yu, L.; Li, M.; Xue, M.; Zhang, P.; Zhu, Q.; Men, R.; Lin, R.; Li, T.; Tang, T.; Xia, T.; Ren, X.; Ren, X.; Fan, Y.; Su, Y.; Zhang, Y.; Wan, Y.; Liu, Y.; Cui, Z.; Zhang, Z.; and Qiu, Z. 2025. Qwen2.5 Technical Report. arXiv:2412.15115.

Rasley, J.; Rajbhandari, S.; Ruwase, O.; and He, Y. 2020. DeepSpeed: System Optimizations Enable Training Deep Learning Models with Over 100 Billion Parameters. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '20, 3505–3506. New York, NY, USA: Association for Computing Machinery. ISBN 9781450379984.

Robles, M.; Martínez, C. A.; Prieto, J. C.; Palacios, S.; and Manrique, R. 2024. Preserving Heritage: Developing a Translation Tool for Indigenous Dialects. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, WSDM '24, 1200–1203. New York, NY, USA: Association for Computing Machinery. ISBN 9798400703713.

Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; and Klimov, O. 2017. Proximal Policy Optimization Algorithms. arXiv:1707.06347.

Shao, Z.; Wang, P.; Zhu, Q.; Xu, R.; Song, J.; Bi, X.; Zhang, H.; Zhang, M.; Li, Y. K.; Wu, Y.; and Guo, D. 2024. DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models. arXiv:2402.03300.

Shu, P.; Chen, J.; Liu, Z.; Wang, H.; Wu, Z.; Zhong, T.; Li, Y.; Zhao, H.; Jiang, H.; Pan, Y.; Zhou, Y.; Owl, C.; Zhai, X.; Liu, N.; Saunt, C.; and Liu, T. 2024. Transcending Language Boundaries: Harnessing LLMs for Low-Resource Language Translation. arXiv:2411.11295.

Verne, J. 1864. *Journey to the Center of the Earth*. Voyages extraordinaires. Paris: Pierre-Jules Hetzel. Originally published as *Voyage au centre de la Terre*.

Yu, Q.; Zhang, Z.; Zhu, R.; Yuan, Y.; Zuo, X.; Yue, Y.; Dai, W.; Fan, T.; Liu, G.; Liu, L.; Liu, X.; Lin, H.; Lin, Z.; Ma, B.; Sheng, G.; Tong, Y.; Zhang, C.; Zhang, M.; Zhang, W.; Zhu, H.; Zhu, J.; Chen, J.; Chen, J.; Wang, C.; Yu, H.; Song, Y.; Wei, X.; Zhou, H.; Liu, J.; Ma, W.-Y.; Zhang, Y.-Q.; Yan, L.; Qiao, M.; Wu, Y.; and Wang, M. 2025. DAPO: An Open-Source LLM Reinforcement Learning System at Scale. arXiv:2503.14476.

Zhang, H.; Liu, M.; Li, C.; Chen, Y.; Xu, J.; and Zhou, M. 2024. A Reinforcement Learning Approach to Improve Low-Resource Machine Translation Leveraging Domain Monolingual Data. In Calzolari, N.; Kan, M.-Y.; Hoste, V.; Lenci, A.; Sakti, S.; and Xue, N., eds., *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, 1486–1497. Torino, Italia: ELRA and ICCL.