Harnessing NLP for Indigenous Language Education: Fine-Tuning Large Language Models for Sentence Transformation

Mahshar Yahan, Dr. Mohammad Amanul Islam
Department of Computer Science and Engineering
Uttara University, Bangladesh
mahshar@uttara.ac.bd, amanul.islam@uttarauniversity.edu.bd

Abstract

Indigenous languages face significant challenges due to their endangered status and limited resources which makes their integration into NLP systems difficult. This study investigates the use of Large Language Models (LLMs) for sentence transformation tasks in Indigenous languages, focusing on Bribri, Guarani, and Maya. Here, the dataset from the AmericasNLP 2025 Shared Task 2 is used to explore sentence transformations in Indigenous languages. The goal is to create educational tools by modifying sentences based on linguistic instructions, such as changes in tense, aspect, voice, person, and other grammatical features. The methodology involves preprocessing data, simplifying transformation tags, and designing zero-shot and few-shot prompts to guide LLMs in sentence rewriting. Fine-tuning techniques like LoRA and Bits-and-Bytes quantization were employed to optimize model performance while reducing computational costs. Among the tested models, Llama 3.2(3B-Instruct) demonstrated superior performance across all languages with high BLEU and ChrF++ scores, particularly excelling in few-shot settings. The Llama 3.2 model achieved BLEU scores of 19.51 for Bribri, 13.67 for Guarani, and 55.86 for Maya in test settings. Additionally, ChrF++ scores reached 50.29 for Bribri, 58.55 for Guarani, and 80.12 for Maya, showcasing its effectiveness in handling sentence transformation. These results highlight the potential of LLMs that can improve NLP tools for indigenous languages and help preserve linguistic diversity.

1 Introduction

Indigenous languages are an important part of human history and culture, but many are on the verge of disappearing. These languages hold unique knowledge and traditions that should be preserved for future generations. Thankfully, advancements in Natural Language Processing (NLP) offer new ways to protect and revitalize them.

For example, in New Zealand, technology is playing a key role in revitalizing the Maori language. Apps like 'Kōrerorero' are making it easier for people to learn and practice the language in their daily lives[1]. Similarly, in Canada, the FirstVoices app offers resources to support learning for more than 60 indigenous languages, helping to preserve and promote these rich cultural traditions[2].

The AmericasNLP 2025 Shared Task 2 (de Gibert et al., 2025) focuses on creating educational tools for Indigenous languages in the Americas, including Bribri, Guarani, Maya, and Nahuatl Omitlan. The initiative leverages NLP techniques to develop systems that can generate language learning exercises by transforming sentences based on grammatical changes, such as tense or type.

In this shared task, the provided dataset contains a source sentence and instruction that need to be applied to achieve the target sentence. The goal is to train a system capable of modifying the source sentences according to specified grammatical transformations. For instance, an example of sentence transformation in the Maya language,

Source: Táan u bin tu kool *(He is going to the field)*

Change(Instruction): TYPE:NEG

Target(Transformed): Ma' táan u bin ich kooli' *(He is not going to the field)*

Each of these languages presents unique linguistic characteristics. For example, Bribri is a tonal language with SOV(Subject-Object-Verb) word order supported by tools like morphological analyzers and electronic dictionaries (Coto-Solano et al., 2021). Guarani is a highly agglutinative language, where prefixes and suffixes are used to express grammatical information (Lucas et al., 2024).

[1]https://linguisticsnews.com/insight/case-study-the-evolution-of-indigenous-languages/

[2]https://autogpt.net/the-impact-of-ai-in-languages-preservation/

Proceedings of the Fifth Workshop on NLP for Indigenous Languages of the Americas (AmericasNLP), pages 119–125
May 4, 2025 ©2025 Association for Computational Linguistics

Maya languages exhibit fascinating linguistic features, such as aspectual marking instead of tense conjugation to express time-related information (Pugh et al., 2023).

The task of sentence transformation for Indigenous languages creates a unique challenge due to it's complex linguistic structures. To address these, the dataset was carefully preprocessed by cleaning text, standardizing formatting, and simplifying transformation tags into actionable instructions. Prompt design played a critical role, with zero-shot and few-shot prompts guiding models effectively in rewriting sentences based on linguistic instructions. Few-shot prompts consistently outperformed zero-shot prompts by providing examples for better learning. Large Language Models (LLMs) were fine-tuned using techniques like LoRA (Low-Rank Adaptation) and Bits-and-Bytes (BNB) quantization to optimize performance while reducing computational costs. Post-processing ensured concise outputs by extracting only the relevant transformed sentences. The results showed that Llama 3.2 achieved the best performance across Bribri, Guarani, and Maya languages, with high BLEU and ChrF++ scores on development and test datasets. Few-shot prompting proved particularly effective for low-resource languages, highlighting its advantage in multilingual NLP tasks.

The major contributions of our research work are as follows-

- We proposed an innovative sentence transformation system for Indigenous languages, utilizing LLMs for effective results.
- We executed a range of experiments on the dataset and presented a comprehensive analysis of their performance.

The experimentation details have been provided in the GitHub repository. [3]

2 Related Work

Indigenous languages are often low-resource, making them challenging for NLP systems that rely on extensive annotated data. Previous studies have showed the potential of NLP in preserving these languages by creating tools like machine translation systems and educational resources.

Leveraging pre-trained models like mBERT and XLM-R for cross-lingual knowledge transfer can help adapt high-resource language models to low-resource settings, enabling better sentence transformations (Pakray et al., 2025). In prior work organized by AmericasNLP, researchers demonstrated that GPT-4 and other large language models perform effectively in few-shot learning for low-resource languages (Ginn et al., 2024). Additionally, they highlighted data augmentation strategies that can address data scarcity and enhance model generalization in low-resource settings.

In a study, the author of the paper (Hammond, 2024) implemented a multilingual transformer-based model(mBERT) and an edit tree method to address the sentence transformation task, which performed poorly. Then, they applied a morphosyntactic similarity approach, which significantly improved performance by utilizing linguistic features. In another research work, the authors (Niklaus et al., 2019) introduced the idea of changing complex sentences into simple ones using recursive sentence simplification and a semantic hierarchy. In a separate study (Silfverberg et al., 2017), researchers proposed an efficient data augmentation technique by modifying morphological patterns, which helps with low-resource language with limited data.

The paper (Su et al., 2024) explores fine-tuning transformer models like NLLB-200, Claude 3 Opus and demonstrates their effectiveness in capturing sentence-level morphological inflections. For Maya, fine-tuning with data augmentation (using StemCorrupt) yielded the best performance. Another shared task paper by AmericansNLP explores sentence transformation using Pointer-Generator LSTM, Mixtral 8x7B (SICL[4] with LoRA), and GPT-4 (ICL[5]) (Bui and Von Der Wense, 2024). Also, they have proposed an ensemble method that outperforms single models by boosting accuracy by almost 4%.

An innovative approach by the authors combines rule-based NLP techniques with prompt-based methods leveraging large language models (LLMs) and POS[6] tagging(Vasselli et al., 2024). This approach balances general processing with language-specific customization for grammatical sentence transformation. Another study demon-

[3] https://github.com/mahshar-yahan/AmericansNLP-2025/tree/main/Shared%20Task-2

[4] SICL: Supervised In-Context Learning

[5] ICL: In-Context Learning

[6] POS: Parts of Speech

strates that minimal CSV-style prompting using large language models (LLMs) like GPT-4 and GPT-3.5 can achieve competitive performance in low-resource morphological tasks (Haley, 2024).

3 Dataset

We have utilized a dataset created for AmericasNLP 2025 Shared Task 2 (de Gibert et al., 2025), which aims to develop educational tools for Indigenous languages. The dataset includes four low-resource languages: Bribri, Guarani, Maya, and Nahuatl Omitlan. It is designed for sentence transformation tasks, where sentences are modified based on linguistic instructions such as changes in tense, aspect, polarity and so on. It includes 16 major categories with a total of 68 unique values across these categories. The dataset is divided into Train, Development (Dev), and Test sets, as shown in Table 1:

Split	Bribri	Guarani	Maya	Nahuatl Omitlan
Train	391	178	594	392
Dev	176	179	149	177
Test	120	364	310	121

Table 1: Language-wise distribution in the dataset

The relatively small size of the training data, particularly for some languages, presents a challenge for robust model training. The dataset also features a mix of simple and complex instructions, allowing for a wide array of sentence transformations to be applied. Table 2 offers some excellent examples to illustrate these transformations and their English equivalents.

4 Methodology

In this section, we have provided an overview of the methods and techniques applied to the dataset described earlier. Initially, the data was preprocessed, and the transformation tags were transferred to instruction. Subsequently, various LLMs were utilized to enhance performance. These models were fine-tuned and evaluated to optimize their effectiveness, as illustrated in Figure 1.

4.1 Data Preprocessing

Several preprocessing steps have been implemented on the given dataset of different language to achieve optimal outcomes. These steps include removing of unnecessary changes, standardizing text formatting and addressing inconsistencies in the data. Each step is designed to enhance the model's ability to process linguistic transformations effectively.

4.1.1 Removal of Unnecessary Changes

The dataset contains entries where the change column is tagged as NA, indicating that no modifications are required for the source sentence. These entries are removed to obtain meaningful transformation. This ensures that only actionable instructions are remained in the dataset. For example,
Before Removal: VOICE:MID, PERSON:NA
After Removal: VOICE:MID

4.1.2 Text Standardization

To ensure uniform formatting in the dataset, we have cleaned the text by removing punctuation, special characters, and unnecessary whitespace. This step helps reduce noise and improves the model's ability to focus on meaningful linguistic patterns. For instance in Bribri,
Before Removal: Ye' tö i kít
After Removal: Ye tö i kít *(Apostrophe removed)*

In the given example for cases English translation is *And here it is*. But in some cases, this preprocessing step may have impacted results, where apostrophes represent glottalization and differentiate minimal pairs.

4.2 Tag Simplification

This process simplifies complex tag combinations into clear, actionable instructions that are easy to understand. It helps the model interpret and apply transformations more effectively. For example, in the dataset, the Change field may contain multiple complex tags like TYPE:NEG, TENSE:PRF_REC.
Input Tags: TYPE:NEG, TENSE:PRF_REC
Simplified Instructions: Make the sentence negative and change to recent perfect tense.
This makes the dataset easier to work with and helps the model learn better. This ensures the model is trained on instructions it can accurately process and apply, resulting in more precise transformations.

4.3 Prompt Design

In this task, zero-shot-prompt and few-shot-prompt are utilized to rewrite sentences according to instructions. These prompts are structured to provide

Source *(Original Sentence)*	**Change** *(Instruction)*	**Target** *(Transformed Sentence)*
Kin in suut koonol merkaado *(I am returning to the market)*	ASPECT: INS	Je'el in suut koonol merkaado *(I will return to the market)*
Kin in suut koonol merkaado*(I am returning to the market)*	ASPECT:TER, TENSE: PAS_SIM	Ts'o'ok in suut koonol merkaado*(I have returned to the market)*

Table 2: Illustrative examples of single and multi-instruction sentence transformations in the Maya language

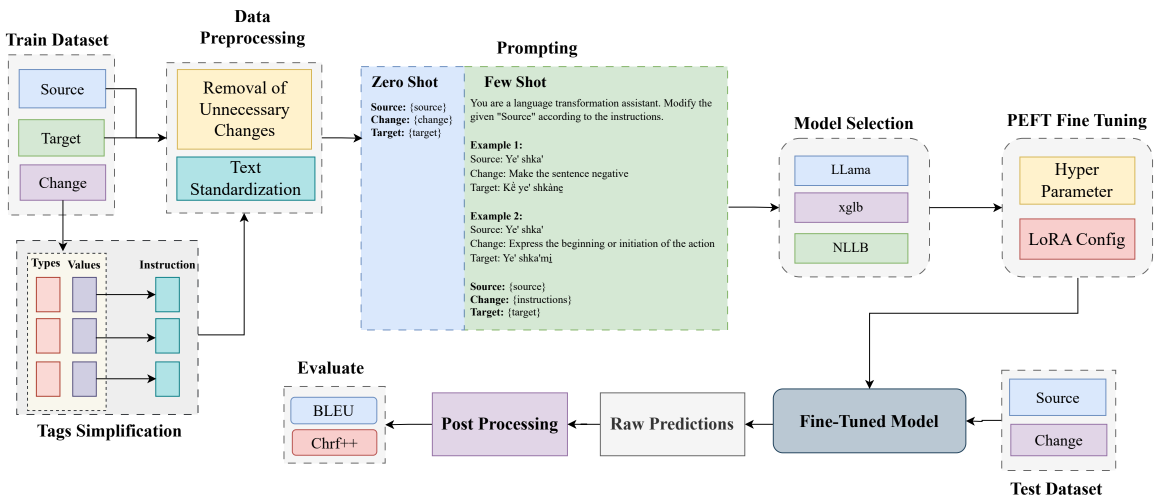

Figure 1: Methodological Workflow for Sentence Transformation in Indigenous Language Education Using Large Language Models

clear guidance to the model while ensuring consistency in the transformation process.

4.3.1 Zero Shot Prompt Design

These prompts are designed to evaluate the model's ability to perform translations and linguistic transformations without relying on specific training examples. The model is expected to independently generate the correct output based solely on the provided instruction. For instance, consider the following training prompt in Bribri:

Zero Shot Prompt
Source: Ye' shka'
Instruction: MODE:ADVERS
Target: Ye' shka'

In handling test data, a slightly modified version of the prompt is used to isolate the predicted sentence in the output: "Provide only the Target sentence, nothing else". This ensures that the generated output is concise and aligned with the task output.

4.3.2 Few Shot Prompt Design

Few-shot prompts incorporate multiple examples of source sentences paired with instructions and their corresponding target sentences. These examples act as references, helping the model learn transformation patterns and apply them accurately. For instance,

Few Shot Prompt:
Language: Bribri, Rewrite and change the Source sentence to the Target sentence according to the given instruction.
Example 1:
Instruction: Change to recent perfect tense. (*TENSE:PRF_REC*)
Source: Ye' shka'
Target: Ye' shké
Example 2:
Instruction: Make the sentence negative and change to recent perfect tense. (*TYPE:NEG, TENSE:PRF_REC*)
Source: Ye' shka'
Target: Ye' kë shkàne
Rewrite following sentence using instruction:
Instruction: Change to potential future tense and change to imperfective aspect. (*TENSE:FUT_POT, ASPECT:IPFV*)
Source: Ye' shka'
Target: Ye' shkömi

Language	Model	Prompt Type	Acc	BLEU	Chrf++
Bribri	Llama 3.1(8B-Instruct)	Zero Shot	1.79	2.74	11.28
	Llama 3.1(8B-Instruct)	Few Shot	5.11	12.73	50.22
	Llama 3.2(3B-Instruct)	Zero Shot	5.59	4.94	33.50
	Llama 3.2(3B-Instruct)	Few Shot	**6.21**	**22.36**	**50.46**
	Xglm 1.7B	Zero Shot	0.89	1.16	30.29
	Xglm 1.7B	Few Shot	2.51	13.56	29.19
Gurarani	Llama 3.1(8B-Instruct)	Zero Shot	6.47	3.16	29.56
	Llama 3.1(8B-Instruct)	Few Shot	9.01	18.34	28.15
	Llama 3.2(3B-Instruct)	Zero Shot	7.57	**24.14**	41.10
	Llama 3.2(3B-Instruct)	Few Shot	**10.53**	22.99	**58.30**
	Xglm 1.7B	Zero Shot	2.55	8.34	52.17
	Xglm 1.7B	Few Shot	4.19	6.24	48.17
Maya	Llama 3.1(8B-Instruct)	Zero Shot	8.29	17.11	56.56
	Llama 3.1(8B-Instruct)	Few Shot	10.11	19.56	68.15
	Llama 3.2(3B-Instruct)	Zero Shot	17.39	43.45	70.23
	Llama 3.2(3B-Instruct)	Few Shot	**21.31**	**57.16**	**82.48**
	Xglm 1.7B	Zero Shot	13.51	43.45	70.53
	Xglm 1.7B	Few Shot	1.16	11.23	40.44

Table 3: Performance Evaluation of Different Models and Zero-Shot and Few-Shot Prompt on the Dev Dataset for Bribri, Guarani, and Maya Languages using Accuracy, BLEU and ChrF++ Metrics

This approach bridges the gap between zero-shot learning and fully supervised training, making it highly effective for multilingual sentence transformation.

4.4 Train

The training process for sentence transformation task involves fine-tuning large language models (LLMs) such as Llama (Touvron et al., 2023), XGLM (Lin et al., 2021) and NLLB (Costa-Jussà et al., 2022) to accurately rewrite sentences based on instructions provided in the dataset. This task focuses on transforming sentences across different dimensions, such as tense, mood, aspect, voice, and negation.

To adapt pre-trained LLMs to the task-specific requirements, we have employed efficient fine-tuning techniques using LoRA (Low-Rank Adaptation) and quantization with Bits and Bytes (BNB). These methods allow us to optimize memory usage and computational efficiency while maintaining the model's performance. LoRA modifies only a subset of the model's parameters, making it ideal for tasks requiring domain-specific adjustments without retraining the entire model. BNB enables 4-bit quantization of model weights, significantly reducing memory consumption during training.

4.5 Post Processing

When using a casual language model for sentence transformation tasks, the generated output may include extra information beyond the desired target sentence. To address this, we have employed a simple linear search on the output to locate the keyword **"Target"**. Once the keyword is identified, everything following it is extracted as the final transformed sentence. This method ensures that only the relevant portion of the model's output is retained.

5 Results and Analysis

In this section, we have provided a comprehensive comparison of the performance across different approaches to large language models (LLMs) for different languages.

5.1 Parameter Setting

Table 4 shows parameter settings for different models.

In Table 4, *lr*, *optim*,*la* and *l4* represents *learning_rate*,*optimizer*, *lora_alpha* and *load_in_4bit* and respectively.

5.2 Evaluation Metrics

The performance of various models has been evaluated using the Bilingual Evaluation Under-

Model	lr	optim	la	l4
Llama 3.1 (8B-Instruct)	$3e^{-4}$	Paged Adamw	4	8
Llama 3.2 (3B-Instruct)	$2e^{-4}$	Paged Adamw	4	8
XGLM 1.7B	$2e^{-3}$	Adam	4	8

Table 4: Parameter settings for different models

study (BLEU) score, the Character-level F-score++ (ChrF++), and Accuracy metrics on the development and test dataset.

5.3 Comparative Analysis

From Table 3, we observed that Llama 3.2 (8B-Instruct) demonstrated the best performance for sentence transformation tasks in Bribri, Guarani, and Maya languages. For Bribri, it achieved the highest BLEU score of 22.36 and ChrF++ of 50.46 in few-shot settings on the development set. Similarly, for Guarani, it secured a BLEU score of 22.99 and ChrF++ of 58.35, while for Maya, it excelled with a BLEU score of 57.16 and ChrF++ of 82.48. In contrast, XGLM 1.7B performed poorly with significantly lower scores across all languages and settings. Few-shot prompting consistently outperformed zero-shot prompting for all models, demonstrating its advantage in low-resource language tasks. The submitted system using Llama 3.2 (8B-Instruct) performed well on the test sets, as shown in Table 5. It achieved competitive BLEU and ChrF++ scores, particularly for Maya, and secured 9th place on the leaderboard.

Language	Type	Evaluation Metrics		
		Acc	BLEU	Chrf++
Bribri	dev	6.21	22.36	50.46
	test	0.4167	19.51	50.29
	base	5.66	20.35	45.56
Gurarani	dev	10.53	22.99	58.30
	test	1.92	13.67	58.55
	base	22.78	34.99	78.72
Maya	dev	21.31	57.16	82.48
	test	13.55	55.86	80.12
	base	26.17	52.38	78.72

Table 5: The results of the submitted system on the development and test sets using Llama 3.2(3B-Instruct) Model

6 Conclusion

The research demonstrates the feasibility of using LLMs for sentence transformation tasks in Indigenous languages. The performance of the models, particularly when compared to a simple edit tree baseline, fell short across all tested languages. Factors such as excessive preprocessing, overly complex prompts, a small dataset size, and high out-of-vocabulary (<unk>) token rates in the model tokenizer may cause these challenges. Among the experimented models, Llama 3.2 is the most effective system. Few-shot prompting proved particularly advantageous for low-resource languages. However, this work provides valuable insights into the obstacles faced when applying LLMs to low-resource languages.

Limitations

Several limitations were identified in this study. First, the provided dataset is quite small, which impacted model generalization. The limited availability of annotated data particularly affected Guarani, where language-specific adaptations were not implemented due to time constraints. Computational constraints also restricted broader experimentation with larger-scale models or ensemble techniques. Addressing these limitations will be crucial for future advancements in Indigenous language processing.

Future Work

Future research should focus on advancing dataset quality and diversity through innovative data augmentation techniques, such as back-translation, contextual embedding-based augmentation, and syntax tree manipulation. Using methods like ensemble learning or hybrid modeling could also boost performance in sentence transformation tasks. Additionally, integrating neural morphology extensions to handle complex linguistic structures would improve sentence transformation tasks. Expanding this work to include more endangered languages could help preserve cultural heritage through NLP.

References

Minh Duc Bui and Katharina Von Der Wense. 2024. Jgu mainz's submission to the americasnlp 2024 shared task on the creation of educational materials for indigenous languages. In *Proceedings of the 4th Workshop on Natural Language Processing for Indigenous*

Languages of the Americas (AmericasNLP 2024), pages 195–200.

Marta R Costa-Jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, et al. 2022. No language left behind: Scaling human-centered machine translation. *arXiv preprint arXiv:2207.04672*.

Rolando Coto-Solano, Sharid Loáiciga, and Sofía Flores-Solórzano. 2021. Towards universal dependencies for bribri. In *Proceedings of the Fifth Workshop on Universal Dependencies (UDW, SyntaxFest 2021)*, pages 16–29.

Ona de Gibert, Raul Vazquez, Robert Pugh, Abteen Ebrahimi, Pavel Denisov, Ali Marashian, Enora Rice, Edward Gow-Smith, Juan C. Prieto, Melissa Robles, Rubén Manrique, Oscar Moreno Veliz, Ángel Lino Campos, Rolando Coto-Solano, Aldo Alvarez, Marvin Agüero-Torales, John E. Ortega, Luis Chiruzzo, Arturo Oncevay, Shruti Rijhwani, Katharina von der Wense, and Manuel Mager. 2025. Findings of the AmericasNLP shared tasks on machine translation, creation of educational material, and translation metrics for indigenous languages. In *Proceedings of the 5th Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2025)*, Albuquerque, New Mexico. Association for Computational Linguistics.

Michael Ginn, Ali Marashian, Bhargav Shandilya, Claire Post, Enora Rice, Juan Vásquez, Marie Mcgregor, Matthew Buchholz, Mans Hulden, and Alexis Palmer. 2024. On the robustness of neural models for full sentence transformation. In *Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2024)*, pages 159–173.

Coleman Haley. 2024. The unreasonable effectiveness of large language models for low-resource clause-level morphology: In-context generalization or prior exposure? In *The 4th Workshop on NLP for Indigenous Languages of the Americas*, pages 174–178. Association for Computational Linguistics (ACL).

Michael Hammond. 2024. The role of morphosyntactic similarity in generating related sentences. In *Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2024)*, pages 221–223.

Xi Victoria Lin, Todor Mihaylov, Mikel Artetxe, Tianlu Wang, Shuohui Chen, Daniel Simig, Myle Ott, Naman Goyal, Shruti Bhosale, Jingfei Du, et al. 2021. Few-shot learning with multilingual language models. *arXiv preprint arXiv:2112.10668*.

Agustín Lucas, Alexis Baladón, Victoria Pardiñas, Marvin Agüero-Torales, Santiago Góngora, and Luis Chiruzzo. 2024. Grammar-based data augmentation for low-resource languages: The case of guarani-spanish neural machine translation. 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 6385–6397.

Christina Niklaus, Matthias Cetto, André Freitas, and Siegfried Handschuh. 2019. Transforming complex sentences into a semantic hierarchy. *arXiv preprint arXiv:1906.01038*.

Partha Pakray, Alexander Gelbukh, and Sivaji Bandyopadhyay. 2025. Natural language processing applications for low-resource languages. *Natural Language Processing*, 31(2):183–197.

Robert Pugh, Francis Tyers, and Quetzil Castañeda. 2023. Developing finite-state language technology for maya. In *Proceedings of the Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP)*, pages 30–39.

Miikka Silfverberg, Adam Wiemerslage, Ling Liu, and Lingshuang Jack Mao. 2017. Data augmentation for morphological reinflection. In *Proceedings of the CoNLL SIGMORPHON 2017 Shared Task: Universal Morphological Reinflection*, pages 90–99.

Jim Su, Justin Ho, George Broadwell, Sarah Moeller, and Bonnie Dorr. 2024. A comparison of fine-tuning and in-context learning for clause-level morphosyntactic alternation. In *Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2024)*, pages 179–187.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Justin Vasselli, Arturo Martínez Peguero, Junehwan Sung, and Taro Watanabe. 2024. Applying linguistic expertise to llms for educational material development in indigenous languages. In *Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2024)*, pages 201–208.