Cureus Journal of Computer Science
https://doi.org/10.7759/s44389-026-00202-y

Original Article

Low-Resource Machine Translation of Yoruba Text to Nigerian Pidgin Using NLLB-200 and Parameter-Efficient Fine-Tuning

Oyebola Akande [1], Goodness Opateye [1,]

1. *Computer Science, Babcock University, Ilishan-Remo, NGA*

Received: May 26, 2026 | Review began: June 30, 2026 | Review ended: July 23, 2026 | Published: August 03, 2026

Abstract

Machine translation for low-resource and structurally divergent languages remains a significant challenge, particularly when mapping highly tonal languages to contact languages with fluid orthographies. This study presents the development of a translation system for Yoruba to Nigerian Pidgin, which addresses a critical gap in African natural language processing. Using a custom-curated parallel corpus of 1,846 pairs, the NLLB-200 (600M) foundational model was adapted for this specific language pair. To reduce the computational cost of full-model fine-tuning, Low-Rank Adaptation (LoRA), a Parameter-Efficient Fine-Tuning (PEFT) technique, was applied to the transformer's attention layers. Applying systematic hyperparameter ablation, the model was evaluated across various LoRA ranks and beam search decoding sizes. The results from the evaluation demonstrate that the optimal configuration (LoRA rank 16, Beam Size 4) significantly outperformed the zero-shot baseline (paired bootstrap resampling, $p < 0.001$), improving the BLEU (Bilingual Evaluation Understudy) score from 0.74 to 30.97 and the character-level score from 9.59 to 50.63. Further qualitative evaluations confirmed the model's ability to generate semantically adequate and colloquially natural translations despite the lack of a standardized dictionary for Nigerian Pidgin. These findings demonstrate the efficacy of parameter-efficient adaptation strategies in democratizing translation technologies for underrepresented, non-standardized linguistic environments.

Categories: AI applications, Deep Learning, Natural Language Processing (NLP)

Keywords: machine translation, low-resource language, nigerian pidgin, nllb-200, lora, parameter-efficient fine-tuning

Introduction

Machine translation is the automatic conversion of text from a source language to another (target language) while preserving meaning, grammatical correctness, and fluency [1]. It is a central task in Natural Language Processing (NLP) as it plays a critical role in multilingual communication, information access, and cross-lingual transfer. It has evolved through several paradigms over decades. The early approaches were rule-based systems using dictionaries and hand-crafted rules, which were labor-intensive and struggled with fluency. Then statistical machine translation followed, in which translation probabilities were modeled using phrase-based and hierarchical alignments trained on parallel corpora. This achieved major improvements but still required intensive feature engineering [2].

A major paradigm shift occurred with the introduction of neural machine translation, which employs recurrent neural networks, particularly long short-term memory networks, with attention mechanisms in sequence-to-sequence architectures [3,4]. The introduction of the transformer architecture in 2017 was a breakthrough as it replaced recurrence with self-attention mechanisms, enabling parallel processing, long-range dependency modeling, and scalability. Transformers have now become the dominant paradigm powering state-of-the-art systems [5].

How to cite this article:
Akande O, Opateye G (August 03, 2026) Low-Resource Machine Translation of Yoruba Text to Nigerian Pidgin Using NLLB-200 and Parameter-Efficient Fine-Tuning. Cureus J Comput Sci 3 : es44389-026-00202-y. DOI
https://doi.org/10.7759/s44389-026-00202-y

Despite the remarkable success of neural machine translation in high-resource language pairs, translation performance remains highly uneven across the globe. For low-resource languages, the lack of large-scale parallel corpora still remains a major bottleneck, resulting in limited linguistic coverage, higher error rates, and poor generalization [6,7]. Furthermore, high vocabulary sparsity and flexible word order in morphologically rich languages pose additional challenges, which complicate alignment during training and decoding processes [8]. Consequently, there is a growing body of research focused on mitigation strategies for low-resource languages, including transfer learning, multilingual training, and parameter-efficient fine-tuning (PEFT).

For many African languages, which include Yoruba and Nigerian Pidgin, dialectical variations, diacritics, non-standardized orthographies, and frequent code-switching are all major challenges [9]. Despite a recent spike in NLP studies across the continent, this growth remains disproportionately spread across different dialects and functions, with linguistic translation initiatives for specific Nigerian dialects remaining notably limited [10,11]. Yoruba, spoken by about 47 million people in West Africa, poses unique phonetic and tonal characteristics that introduce linguistic complexity [12-14]. Its lexical structure relies heavily on tone marks, which are called diacritics (high, mid, and low tones). A single alphabetic sequence can yield entirely different semantic meanings depending on the applied diacritics. For example, the word "igba" can mean time, calabash, or two hundred, strictly dictated by its tonal representation. This makes precise character encoding during translation critical. A recent systematic review spanning a decade of Yorùbá NLP research reveals that annotated parallel corpora and pretrained models remain limited, while tonal complexity and reliance on diacritics continue to represent the field's most pressing unresolved challenges [15]. Similarly, Nigerian Pidgin, despite acting as a major lingua franca spoken by over 100 million people across Nigeria and parts of West Africa, lacks standardized orthography and exhibits highly fluid grammar and extreme lexical variations [9]. Its vocabulary is heavily influenced by indigenous local languages, Portuguese, and English. Consequently, a single word may possess multiple acceptable spellings depending on regional dialects (e.g., "pikin" vs. "pekin" for child). Developing a translation system between a strict, tonal language like Yoruba and a fluid, non-standardized language like Nigerian Pidgin requires a highly adaptable subword tokenization strategy. Addressing these specific lexical and orthographic inconsistencies is critical to developing effective translation tools for multilingual African societies.

PEFT has emerged as the dominant strategy for adapting large pretrained models without the computational cost of full fine-tuning. Early approaches inserted lightweight Adapter modules between transformer layers, training only these new parameters while freezing the backbone [16]. Prompt-based alternatives followed a different strategy: Prefix-Tuning prepends trainable continuous vectors to each transformer layer [17], while Prompt Tuning restricts trainable parameters to a small set of soft tokens at the input layer alone, achieving competitive performance at larger model scales [18]. Low-Rank Adaptation (LoRA) took a different approach, freezing pretrained weights entirely and instead injecting trainable low-rank decomposition matrices into attention layers, substantially reducing trainable parameter counts while matching full fine-tuning performance on numerous benchmarks [19]. Building on this, QLoRA combined LoRA with quantized backbone weights to further reduce memory requirements, enabling fine-tuning of significantly larger models on consumer-grade hardware [20]. Among these approaches, LoRA has become particularly widely adopted for low-resource and multilingual adaptation tasks due to its favorable trade-off between adaptation quality, computational cost, and the absence of added inference latency once adapters are merged back into the base model. Building on these general-purpose PEFT strategies, a growing body of work has applied transfer learning and multilingual adaptation specifically to Nigerian and other African languages.

In the context of Nigerian languages, early neural approaches demonstrated the feasibility of translation despite data scarcity. For instance, Ahia and Ogueji [21] performed supervised and unsupervised neural machine translation between English and Nigerian Pidgin using a transformer architecture trained on the JW300 corpus. Their supervised model achieved a BLEU (Bilingual Evaluation Understudy) score of 24.29 for English-to-Pidgin translation and a BLEU score of 24.67 for Pidgin-to-English. Similarly, Ekle and Das [22] developed a neural machine translation system for the low-resource English-to-Igbo language pair utilizing recurrent architecture and transfer learning via MarianMT to achieve approximately 70% semantic accuracy.

How to cite this article:
Akande O, Opateye G (August 03, 2026) Low-Resource Machine Translation of Yoruba Text to Nigerian Pidgin Using NLLB-200 and Parameter-Efficient Fine-Tuning. Cureus J Comput Sci 3 : es44389-026-00202-y. DOI
https://doi.org/10.7759/s44389-026-00202-y

Recent advances have shifted towards leveraging massively multilingual pretrained foundational models. Adelani et al. [23] leveraged pretrained models such as mT5, ByT5, mBART-50, and M2M-100 to create translation systems for 16 low-resource African languages, demonstrating that continual pre-training on monolingual data significantly boosts translation scores. Building on this, Moslem et al. [24] developed a series of lightweight multilingual machine translation models for 15 African language pairs (30 directions), including Swahili, Hausa, Yoruba, Amharic, Somali, Zulu, Lingala, Afrikaans, Wolof, Egyptian Arabic, and official AU languages like English. The study compressed the NLLB-200 baseline via iterative layer pruning and fine-tuned it using knowledge distillation, achieving competitive spBLEU scores while significantly reducing computational overhead.

The effectiveness of these multilingual transformer models extends beyond African languages, proving robust for low-resource scenarios globally. For example, Oni and Prama [25] fine-tuned mBART-50, MarianMT, and NLLB-200 for the Bengali-Sylheti language pair using advanced data augmentation techniques. They observed that NLLB-200 excelled in semantic adequacy (BLEU), while MarianMT performed best in character-level fidelity (Character n-gram F-score [chrF]). Likewise, Man et al. [26] fine-tuned mT5 for translation between Myanmar and the zero-resource Tedim Chin language, outperforming standard baselines. Furthermore, Khatri et al. [27] demonstrated that transferring learned weights from a related language pair (German to Upper Sorbian) to a target pair (German to Lower Sorbian) is a highly effective mitigation strategy for parallel data scarcity. Most recently, language-specific LoRA fine-tuning strategies have been shown to improve translation quality for other underrepresented language pairs by concentrating adaptation capacity on attention and feed-forward layers most relevant to the target language's structure [28], and similar PEFT-based approaches have demonstrated feasibility even for very low-resource African language pairs with limited parallel data [29].

While NLLB-200 and related multilingual foundation models have substantially advanced low-resource machine translation, three specific gaps remain unaddressed for the Yoruba-Nigerian Pidgin pair. First, to our knowledge, no prior study has directly adapted NLLB-200 for Yoruba-to-Pidgin translation; existing work has focused on English-centric pairs (English-Pidgin [21], English-Igbo [22]) or on other African language combinations that do not involve Nigerian Pidgin as a target. Second, this is the first study to apply PEFT specifically to this language pair, rather than the full fine-tuning or continual pretraining approaches used elsewhere [23,24]. Third, this work presents the first systematic LoRA-rank ablation, jointly varying rank and beam search decoding size, for a Yoruba-to-Pidgin translation task. Together, these contributions position this study as a targeted response to the underexplored problem of direct translation between culturally adjacent, non-English-centric African language pairs.

This study makes the following contributions:

• A curated parallel corpus: 1,846 manually annotated and deduplicated Yoruba-Nigerian Pidgin sentence pairs, including retained multi-reference translations to capture natural Pidgin variation.

• First direct LoRA adaptation of NLLB-200 for this language pair: demonstrating that PEFT can substantially close the gap between zero-shot and fine-tuned performance for a previously unstudied direct translation direction.

• A systematic LoRA-rank and beam-size ablation: jointly evaluating ranks 8, 16, and 32 across beam sizes 1, 2, 4, and 8 to identify optimal configurations for this task.

• A reproducible, resource-efficient pipeline: built on an openly available foundational model and standard PEFT tooling, requiring substantially less compute than full fine-tuning.

• Qualitative error analysis: a manual inspection of model outputs to characterize the model's semantic and lexical adaptation behavior.

This study is guided by the following research questions:

RQ1: Can LoRA effectively adapt the NLLB-200 foundational model for direct Yoruba-to-Nigerian Pidgin translation?

RQ2: Which LoRA rank (8, 16, or 32) produces the optimal translation performance for this language pair, as measured by BLEU and chrF?

How to cite this article:
Akande O, Opateye G (August 03, 2026) Low-Resource Machine Translation of Yoruba Text to Nigerian Pidgin Using NLLB-200 and Parameter-Efficient Fine-Tuning. Cureus J Comput Sci 3 : es44389-026-00202-y. DOI
https://doi.org/10.7759/s44389-026-00202-y

RQ3: What beam search decoding size maximizes translation quality without incurring diminishing returns in computational cost?

Materials And Methods

Data collection and preprocessing

A parallel dataset of Yoruba sentences and their Nigerian Pidgin translations was utilized for this study. This dataset was compiled from custom annotations, which involved manually translating the Yoruba source text into Nigerian Pidgin. The researcher, a native speaker of both Yoruba and Nigerian Pidgin, translated each Yoruba sentence directly into Pidgin while preserving idiomatic and colloquial phrasing. Sentences spanned everyday conversational registers (family, food, religion, and social interaction) rather than domain-specific or formal text, reflecting the intended use case of the resulting system. Where a Yoruba sentence admitted more than one natural Pidgin rendering, for example, due to regional lexical variation, the rendering used most frequently in the researcher's own linguistic community was recorded, while retaining a secondary reference where a genuinely distinct but equally valid alternative existed. As a single-annotator dataset, no formal inter-annotator agreement statistic could be computed.

Initially, the raw annotated dataset comprised 2,487 sentence pairs. Exact duplicate removal was applied at the sentence-pair level using case-sensitive string matching prior to normalization, ensuring that the reported duplication rate reflects genuinely repeated content rather than normalization artifacts; a total of 641 exact duplicate pairs (approximately 25.8% of the raw data) were removed on this basis to avoid data leakage during training. The final clean dataset consists of 1,846 highly unique parallel sentence pairs. Of these, 136 pairs were found to have multiple valid Pidgin translations arising from the regional variation described above; these were retained as multi-reference pairs to help the model learn the natural variance of Pidgin text. A comprehensive breakdown of the dataset is presented in Table *1*.

Initial Raw Sentence Pairs	**2,487**
Exact Duplicates Removed	641
Total Unique Parallel Pairs	1,846
Multi-Reference Yoruba Sentences	136
Training Split	90%
Testing Split	10%

TABLE 1: Parallel Corpus Statistics

Text preprocessing was meticulously handled due to the contrasting linguistic structures of the two languages. For the Yoruba text, the primary challenge here is tone-marking with diacritics, which denote pitch and dictate semantic meaning. To address this, Unicode Normalization Form C (NFC) was applied across the corpus to ensure that decomposed diacritics were consistently combined into single characters. Extraneous non-linguistic symbols, ASCII, and Unicode punctuation were filtered out to improve sequence alignment. Furthermore, the texts were standardized by converting all characters to lowercase and collapsing redundant whitespaces to reduce token sparsity.

How to cite this article:
Akande O, Opateye G (August 03, 2026) Low-Resource Machine Translation of Yoruba Text to Nigerian Pidgin Using NLLB-200 and Parameter-Efficient Fine-Tuning. Cureus J Comput Sci 3 : es44389-026-00202-y. DOI
https://doi.org/10.7759/s44389-026-00202-y

Conversely, for Nigerian Pidgin, aggressive linguistic normalization was deliberately avoided. Spelling variants, loan words, and contractions (for example, "wan," "want to," etc.) were retained as they appear in the data. This allows the NLLB model to learn appropriate mappings through its subword-based tokenization. Basic cleaning for Nigerian Pidgin was limited to Unicode normalization, lowercasing, and whitespace corrections.

Tokenization

The massively multilingual SentencePiece-based Byte-Pair Encoding tokenizer integrated within Meta's NLLB-200 architecture was used, as it was trained on over 200 languages, providing a deeply shared subword vocabulary. This naturally accommodates the heavy use of tonal diacritics in Yorùbá alongside the extreme spelling variability and colloquialisms of Nigerian Pidgin. During tokenization, the tokenizer was explicitly initialized with specialized language codes: yor_Latn for the source inputs and pcm_Latn for the Nigerian Pidgin target output. To guarantee consistency during fine-tuning and evaluation, the sequence boundaries were controlled by applying a maximum token length of 128, with truncation enabled.

Model architecture and fine-tuning

The distilled 600M-parameter variant of NLLB-200 (Meta AI) was used as the foundational model due to its reduced computational requirements while still maintaining strong translation performance, making it appropriate for resource-constrained environments. Given that Yoruba and Nigerian Pidgin are structurally divergent languages, the foundational model required fine-tuning to properly learn the specific lexical mapping between them. To handle this efficiently, the study applied LoRA, a recognized PEFT technique. Rather than updating the entire network, LoRA inserts trainable low-rank matrices directly into the attention layers (q_proj, k_proj, v_proj, and out_proj) and freezes the original pre-trained weights. This approach heavily optimizes memory usage while ensuring the model retains its foundational multilingual knowledge. The set of hyperparameters and configurations is shown in Table *2*.

Parameter	Value
LoRA Target Modules	Attention Layers (q_proj, k_proj, v_proj, out_proj)
LoRA ranks tested	8, 16, 32
Dropout rate	0.05
Optimizer	AdamW
Learning Rate	2×10^{-4}
Weight Decay	0.01
Per-Device Batch Size	8
Training Epoch	20

TABLE 2: NLLB-200 LoRA Fine-Tuning Hyperparameters

LoRA, Low-Rank Adaptation

How to cite this article:
Akande O, Opateye G (August 03, 2026) Low-Resource Machine Translation of Yoruba Text to Nigerian Pidgin Using NLLB-200 and Parameter-Efficient Fine-Tuning. Cureus J Comput Sci 3 : es44389-026-00202-y. DOI https://doi.org/10.7759/s44389-026-00202-y

Model evaluation

The performance of the fine-tuned model was evaluated using two standard metrics: BLEU and chrF. BLEU measures n-gram overlap between system translations and human reference Pidgin translations. Translations generated for the held-out Yoruba test set were compared against reference Pidgin sentences. However, because BLEU primarily captures word-level overlap, it may not fully reflect translation quality when acceptable paraphrasing or multiple valid renderings occur. To complement BLEU, the chrF metric was also computed. chrF operates at the character level, evaluating n-gram precision and recall to produce an F-score that is more sensitive to morphological variations and minor spelling differences. This makes chrF particularly suitable for evaluating Nigerian Pidgin, which exhibits flexible spelling and variable orthography. Beam search was utilized during inference, with beam sizes of 1, 2, 4, and 8 tested for each LoRA rank to determine the optimal balance between translation fluency and computational latency. To assess whether observed differences between configurations were statistically robust, paired bootstrap resampling was applied to corpus-level BLEU and chrF scores across key configuration pairs.

Ethical considerations

Yoruba and Nigerian Pidgin are linguistically and culturally significant languages in Nigeria. Yoruba relies on tone and diacritic markings to distinguish lexical and grammatical meaning. Therefore, these features were preserved during preprocessing through Unicode NFC normalization rather than being removed. Nigerian Pidgin exhibits considerable orthographic variation due to the absence of a universally accepted writing standard. In this study, such variation was treated as a legitimate linguistic characteristic, and potential downstream applications of this system should avoid favouring any particular regional orthographic convention. The parallel corpus was manually translated by the researcher rather than scraped from third-party sources, so no personal or community-generated text was used without consent, and the dataset contains no identifiable information. Cultural sensitivity is maintained by respecting linguistic norms and documenting standardization decisions. This research work was conducted in accordance with the ethical standards set by the Babcock University Health Research Committee (BUHREC).

Results And Discussion

Experimental setup

All computational experiments were conducted using a single NVIDIA A100 GPU on a Google Colab runtime, with T4 as a fallback configuration. PyTorch 2.2.2, Hugging Face Transformers 4.40.2, Accelerate 0.30.1, PEFT 0.11.1, and Datasets 2.19.1 were used. Mixed precision training (bf16 where supported, otherwise fp16) was used throughout. For reproducibility, a fixed random seed of 42 was applied across Python, NumPy, and PyTorch. The cleaned corpus was split into training and testing sets using a 90/10 split. Each LoRA configuration was trained for up to 20 epochs with early stopping (patience of 5 epochs, monitored on BLEU) using the Hugging Face Seq2SeqTrainer.

Data preprocessing outcomes

To improve alignment consistency and reduce noise, language-specific normalization was carried out. As illustrated in Table *3*, all texts were converted to lowercase, extraneous quotations were filtered out, and excessive whitespaces were collapsed. The removal of quotation marks, punctuation, and excessive whitespaces was implemented.

How to cite this article:
Akande O, Opateye G (August 03, 2026) Low-Resource Machine Translation of Yoruba Text to Nigerian Pidgin Using NLLB-200 and Parameter-Efficient Fine-Tuning. Cureus J Comput Sci 3 : es44389-026-00202-y. DOI
https://doi.org/10.7759/s44389-026-00202-y

Before Preprocessing	After Preprocessing
Dem don lock your shop	dem don lock your shop
Lafihan na him win the king's award	lafihan na him win the kings award
Who carry the water wey I fetch?	who carry the water wey I fetch

TABLE 3: Pidgin Text Preprocessing

As shown in Table *4*, comparing the text samples before and after preprocessing demonstrates that preprocessing standardizes transcription while maintaining semantic and phonological integrity for the Yoruba text.

Before Preprocessing	After Preprocessing
Ifálékè ti j'Ọba o.	ifálékè ti joba o
Ibo ni ọkọ̀ yí ń lọ?	ibo ni ọkọ̀ yí ń lọ
Èyàn mẹ́fà ló ń jẹ́ Máíkẹ̀lì ní kílàsì wa.	eyàn mẹ́fà ló ń jẹ́ Máíkẹ̀lì ní kílàsì wa
Wón ní kí a wá gba ẹran ọdún.	wón ní kí a wá gba ẹran ọdún

TABLE 4: Yoruba Text Preprocessing

Quantitative performance and ablation

Before applying any fine-tuning, the base NLLB-200 model was tested directly on the Yoruba-Pidgin translation task to establish a zero-shot baseline. As expected, the base NLLB model struggled significantly due to localized vocabulary and unique syntactical structures of the language, returning a BLEU score of 0.74 and a chrF score of 9.59. This clearly indicates that the model does not perform well without fine-tuning.

As observed in Table *5*, rank 16 emerged as the best-performing model, outperforming rank 8 and rank 32. Going from greedy decoding (beam size 1) to a beam size of 4 with rank 16 improved the BLEU score to an optimal 30.97 and the chrF score to 50.63. This shows a BLEU improvement of +30.23 over the zero-shot baseline and a chrF improvement of +41.04 over the baseline. The performance plateaued at a beam size of 8, which makes a beam size of 4 the optimal decoding strategy.

How to cite this article:
Akande O, Opateye G (August 03, 2026) Low-Resource Machine Translation of Yoruba Text to Nigerian Pidgin Using NLLB-200 and Parameter-Efficient Fine-Tuning. Cureus J Comput Sci 3 : es44389-026-00202-y. DOI https://doi.org/10.7759/s44389-026-00202-y

Rank	Metric	Beam 1	Beam 2	Beam 4	Beam 8
8	BLEU	28.0062	27.8037	28.3036	28.1943
	chrF	48.4113	48.0543	48.3848	48.4493
16	BLEU	29.7568	30.1957	30.9740	30.8457
	chrF	49.6599	50.2519	50.6341	50.6301
32	BLEU	28.3567	28.5172	29.7106	29.4307
	chrF	47.6526	48.1753	49.1459	48.7602

TABLE 5: Ablation Summary: BLEU and chrF Scores by LoRA Rank and Beam Size

BLEU, Bilingual Evaluation Understudy; chrF, Character n-gram F-score; LoRA, Low-Rank Adaptation

To better illustrate these performance variances, the results are visually compared in Figure *1* (BLEU scores) and Figure *2* (chrF scores). As shown in Figure *1* and Figure *2*, while LoRA rank 16 achieved the highest point estimates across all beam sizes, the 95% confidence intervals overlap substantially with those of rank 8 and rank 32.

How to cite this article:
Akande O, Opateye G (August 03, 2026) Low-Resource Machine Translation of Yoruba Text to Nigerian Pidgin Using NLLB-200 and Parameter-Efficient Fine-Tuning. Cureus J Comput Sci 3 : es44389-026-00202-y. DOI https://doi.org/10.7759/s44389-026-00202-y

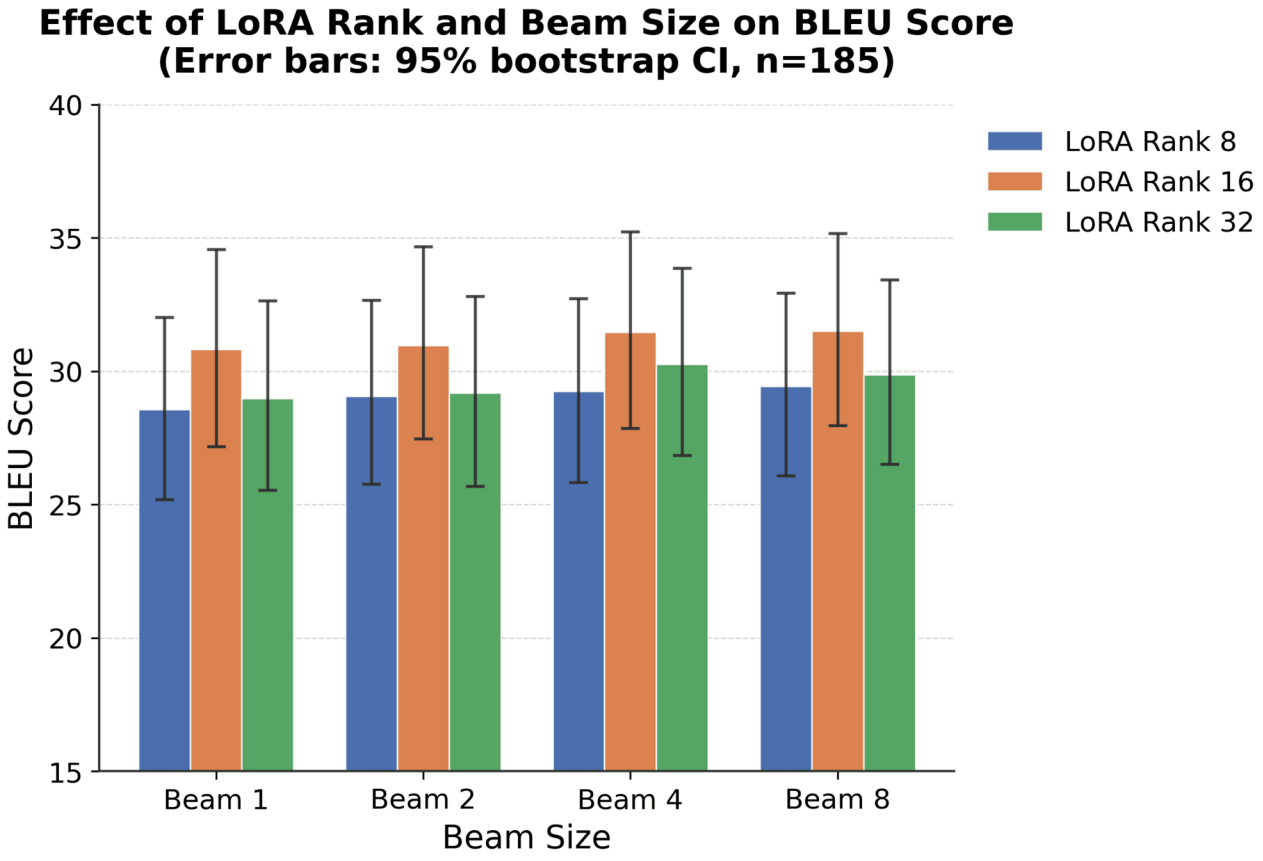

FIGURE 1: Effect of LoRA Rank and Beam Size on BLEU Score

BLEU, Bilingual Evaluation Understudy; LoRA, Low-Rank Adaptation

How to cite this article:
Akande O, Opateye G (August 03, 2026) Low-Resource Machine Translation of Yoruba Text to Nigerian Pidgin Using NLLB-200 and Parameter-Efficient Fine-Tuning. Cureus J Comput Sci 3 : es44389-026-00202-y. DOI https://doi.org/10.7759/s44389-026-00202-y

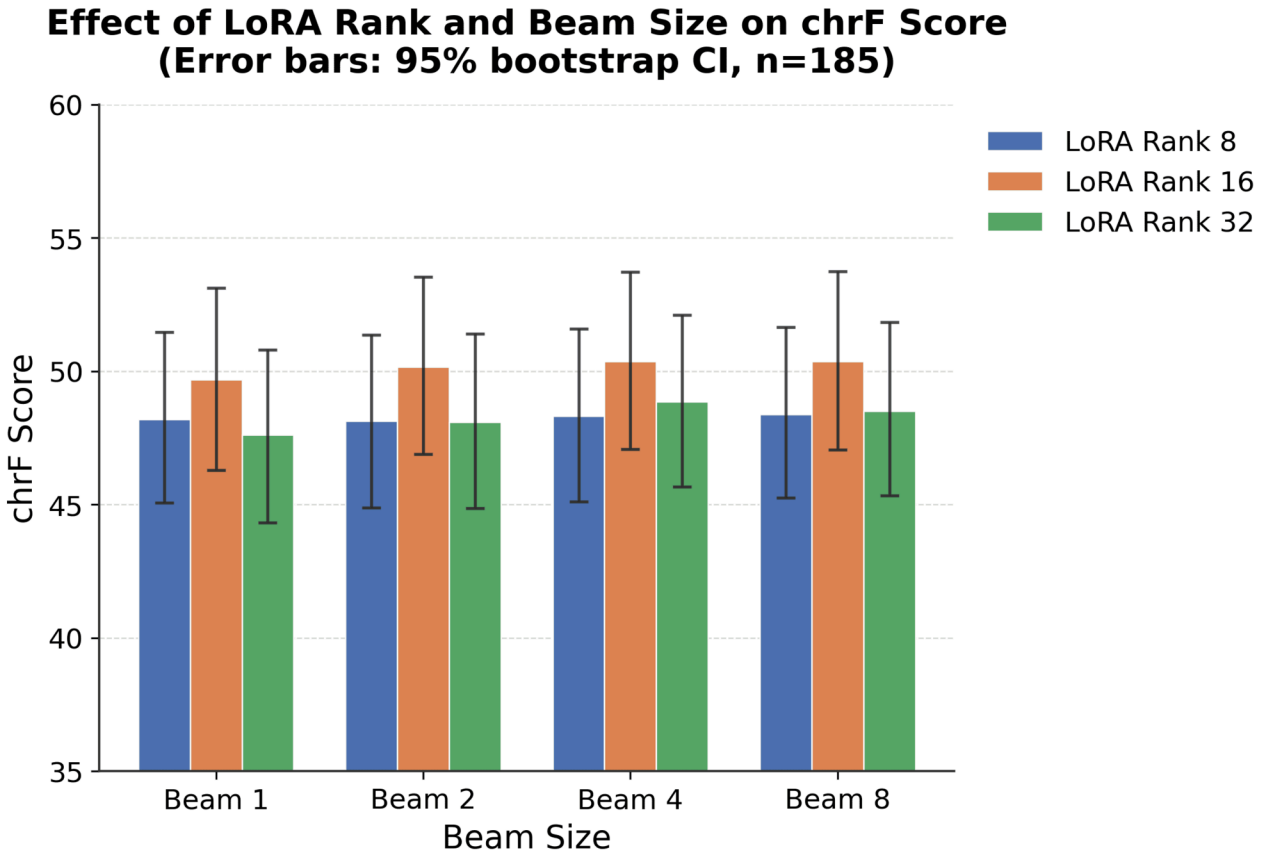

FIGURE 2: Effect of LoRA Rank and Beam Size on chrF Score

chrF, Character n-gram F-score; LoRA, Low-Rank Adaptation

To determine whether these differences reflect genuine improvements rather than sampling variation, paired bootstrap resampling was conducted on corpus-level BLEU and chrF for five key configuration pairs. The results are summarized in Table *6*.

How to cite this article:
Akande O, Opateye G (August 03, 2026) Low-Resource Machine Translation of Yoruba Text to Nigerian Pidgin Using NLLB-200 and Parameter-Efficient Fine-Tuning. Cureus J Comput Sci 3 : es44389-026-00202-y. DOI https://doi.org/10.7759/s44389-026-00202-y

Comparison	ΔBLEU	p (BLEU)	ΔchrF	p (chrF)
Fine-tuned (Rank 16, Beam 4) vs. zero-shot baseline	+30.23	<0.001 ***	+41.04	<0.001 ***
Rank 16 vs. Rank 8 (Beam 4)	+2.67	0.003 **	+2.25	0.009 **
Rank 16 vs. Rank 32 (Beam 4)	+1.26	0.163 (NS)	+1.49	0.045 *
Beam 4 vs. Beam 1 (Rank 16)	+1.22	0.037 *	+0.97	0.052 (NS)
Beam 4 vs. Beam 8 (Rank 16)	+0.13	0.411 (NS)	+0.00	0.494 (NS)

TABLE 6: Paired Bootstrap Significance Testing

*p < 0.05; **p < 0.01; ***p < 0.001
BLEU, Bilingual Evaluation Understudy; chrF, Character n-gram F-score; NS, Not Significant

Rank 16's improvement over rank 8 is significant on both metrics, but its advantage over rank 32 is significant only for chrF, not BLEU, indicating that the two higher-capacity configurations are statistically distinguishable at the character level but not at the word-overlap level. Similarly, Beam 4 offers a significant BLEU improvement over greedy decoding (Beam 1) but shows no significant difference from Beam 8 on either metric, confirming that decoding quality plateaus beyond Beam 4.

The NLLB-200 model demonstrated exceptionally stable and consistent learning. The validation loss decreased steadily from 2.73 at Epoch 1 to an optimal 1.87 at Epoch 14, moving in tandem with the training loss before stabilizing. This proves that the fine-tuning method facilitates stable optimization, allowing the model to generalize effectively.

To validate the quantitative metrics, Table *7* compares the source Yoruba text, the ground-truth Pidgin reference, and the model's prediction at LoRA rank 16 and Beam Size 4. This serves as the qualitative manual inspection of the generated translations.

S/N	Source	Reference	Prediction
1	ọmọ náà fẹ́ràn láti ma wò mí	the pikin like to dey look me	dat pikin like to dey look me
2	ìmọ́tótó ṣe pàtàkì fún àti dénà ààrùn	cleanliness dey important to prevent sickness	cleaning important for sickness cure
3	adébánjọ fẹ́ran ẹja tútù láti jẹ	adebajo like to dey chop cold fish	adebànjọ like fresh fish to chop
4	ìyàwó ba oko rẹ̀ nínú jẹ́	wife spoil the mind of her husband	wife spoil her farm inside
5	mo járọ́ọ́ láídé	i spoil laide lie	i talk nonsense for ade
6	ó gbèjà àbúrò rẹ̀ gan	e defend im younger well well	e protect im younger brother well
7	ìmọ̀ràn náà dára a ó múu lò	di advice good we go use am	that advice good we go use am
8	oúnjẹ kékeré ló sè lọ fún un	na small food e prepare go for am	na small food e cook go give am
9	wọn fi ẹ̀mí ìmoore hàn	dem show spirit of gratitude	dem show spirit of gratitude
10	ibà ló ń ṣe é látàárọ̀	fever dey catch am since morning	na fever dey do am all morning
11	ọ̀pọ̀lọpọ̀ ọmọdé fẹ́ràn patí	many children like party	plenty pikin like paddle
12	ọgbà àjàrà ni ó yí ilé náà ká	vineyard dey surround di house	na vineyard dey surround the house
13	ìjọ tí mo lọ lánàá wọ́n ń pariwo	the church wey i go yesterday dem dey shout	di church wey i go yesterday dem dey shout
14	alágídí lọmọ náà ó fi jọ ẹ̀gbọ́n rẹ̀ ni	that pikin stubborn e resemble im senior brother	that pikin mistake na im resemble im brother
15	àwọn ọ̀rẹ́ mi ni wọ́n kọ́ mi bí wọ́n ṣe máa ń se oúnjẹ	na my friends teach me how to cook	na my friends teach me how dem dey cook food

TABLE 7: Source Text vs. Reference vs. Prediction

How to cite this article:
Akande O, Opateye G (August 03, 2026) Low-Resource Machine Translation of Yoruba Text to Nigerian Pidgin Using NLLB-200 and Parameter-Efficient Fine-Tuning. Cureus J Comput Sci 3 : es44389-026-00202-y. DOI https://doi.org/10.7759/s44389-026-00202-y

The qualitative analysis indicates that the fine-tuned model is capable of producing semantically adequate translations in a number of cases. Among the fully correct predictions, the model frequently substitutes valid orthographic and lexical variants without altering intended meanings. In sample 1, the model produces "dat pikin" instead of "the pikin," which is still a valid variation in Nigerian Pidgin and preserves the intended meaning. Sample 13 also shows the same, accepted "di" versus "the" spelling variance. In samples 6, 8, and 15, the model introduces synonymous or slightly more explicit phrasing, predicting "defend" as "protect" in sample 6 and adding the object food in sample 15 without compromising semantic accuracy. This suggests that the model captures the natural paraphrastic variability characteristics of fluent Pidgin usage. Furthermore, in sample 3, "ẹja tútù" is translated as "fresh fish" rather than the literal "cold fish." This may indicate an attempt to produce a more natural expression. In sample 10, the model paraphrased by replacing the verb "catch" with "do" and changed the temporal expression from "since morning" to "all morning." Although these lexical and temporal substitutions modify the emphasis of the utterance, the core semantic content remains largely preserved. However, not all outputs maintain full semantic fidelity. In Sample 2, the model shifts from "prevent sickness" in the reference to "sickness cure," altering the meaning from prevention to treatment. This demonstrates a loss of semantic precision despite partial preservation of the overall topic.

Beyond its methodological contributions, a working Yoruba-to-Pidgin system has clear practical value. Bilingual educational materials, public health messages, and official announcements could reach Pidgin-dominant communities more effectively, particularly in southern and coastal Nigeria, where more people speak Pidgin than formal English. In healthcare, accurate Yoruba-to-Pidgin translation could improve patient-provider communication in multilingual clinical settings. In agricultural extension advisories or emergency alerts, rapid and accurate translation into Pidgin could improve the timely dissemination of important information to rural and peri-urban communities. In digital inclusion, the work contributes towards closing the language technology gap for Nigerian Pidgin. It could serve as a component within larger future chatbot or voice assistant systems designed to serve underrepresented African language communities.

Limitations to the study

The parallel corpus, while carefully curated, is relatively small (1,846 sentence pairs) and was manually translated and annotated by a single researcher without independent inter-annotator validation. Nigerian Pidgin's non-standardized orthography further means the study's reference translations reflect one set of regional spelling conventions among several valid options. While 95% bootstrap confidence intervals and paired significance tests (Table *6*) were computed across ablation configurations, these were derived from a single train-test split rather than cross-validated folds, and the ablation itself varied only LoRA rank and beam size; dropout, learning rate, epochs, and target module selection remain unexplored. mBART-50 and M2M-100 were considered as comparison baselines but excluded, as neither supports Nigerian Pidgin as a target language in their released vocabularies; broader benchmarking against larger NLLB variants, other multilingual systems, and alternative fine-tuning approaches (e.g., full fine-tuning, Adapters, QLoRA) remains an important direction for future work with greater computational resources. Finally, human evaluation was not carried out due to resource constraints; a future evaluation with independent bilingual annotators rating adequacy and fluency, with inter-annotator agreement reported, would strengthen these findings.

Conclusions

This study demonstrates that a parameter-efficient fine-tuning technique can effectively adapt a multilingual foundational model for direct Yoruba-to-Nigerian Pidgin translation, with LoRA rank 16 and beam size 4 substantially outperforming the zero-shot baseline. This study answers all three research questions, confirming that targeted, low-cost adaptation is effective for this language pair. The study contributes methodologically by providing a reproducible, resource-efficient pipeline that combines a curated parallel corpus, a systematic rank and beam-size ablation validated with bootstrap confidence intervals, and a qualitative error analysis. This work proposes a template for adapting foundation models to other culturally adjacent low-resource African language pairs. Practically, the study moves towards closing the language technology gap for an estimated 100 million Nigerian Pidgin speakers who are currently underserved by mainstream translation systems. The proposed approach has potential applications in education, public

How to cite this article:
Akande O, Opateye G (August 03, 2026) Low-Resource Machine Translation of Yoruba Text to Nigerian Pidgin Using NLLB-200 and Parameter-Efficient Fine-Tuning. Cureus J Comput Sci 3 : es44389-026-00202-y. DOI
https://doi.org/10.7759/s44389-026-00202-y

health communication, and digital inclusion. Future work should prioritize extending the hyperparameter search beyond rank and beam size to include dropout, learning rate, and target module selection; benchmarking LoRA against alternative fine-tuning strategies and, where computational resources permit, larger or differently pretrained multilingual systems; and incorporating structured human evaluation with independent bilingual annotators.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Goodness Opateye, Oyebola Akande

Acquisition, analysis, or interpretation of data: Goodness Opateye

Drafting of the manuscript: Goodness Opateye

Critical review of the manuscript for important intellectual content: Goodness Opateye, Oyebola Akande

Supervision: Oyebola Akande

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

The datasets (and/or code) supporting this study are available from the corresponding author upon reasonable request.

References

1. Stahlberg F: Neural machine translation: a review. Journal of Artificial Intelligence Research. 2020, 69:343-418. 10.1613/jair.1.12007
2. Koehn P: History. Neural Machine Translation. Cambridge University Press, Cambridge; 2020. 29-40. 10.1017/9781108608480.004
3. Sutskever I, Vinyals O, Le QV: Sequence to sequence learning with neural networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. 2014, 2:3104-3112.
4. Bahdanau D, Cho K, Bengio Y: Neural machine translation by jointly learning to align and translate [PREPRINT]. arXiv. 2016, 10.48550/arXiv.1409.0473
5. Vaswani A, Shazeer N, Parmar N, et al.: Attention is all you need. 31st Conference on Neural Information Processing Systems. 2017, 30:5998-6008.
6. Magueresse A, Carles V, Heetderks E: Low-resource languages: a review of past work and future challenges [PREPRINT]. arXiv. 2020, 10.48550/arXiv.2006.07264
7. NLLB Team: Scaling neural machine translation to 200 languages. Nature. 2024, 630:841-846. 10.1038/s41586-024-07335-x
8. Alabi JO, Adelani DI, Mosbach M, Klakow D: Adapting pre-trained language models to African languages via multilingual adaptive fine-tuning. Proceedings of the 29th International Conference on Computational Linguistics. 2022, 4336-4349.
9. Lin PJ, Scholman M, Saeed M, Demberg V: Modeling orthographic variation improves NLP performance for Nigerian Pidgin. Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation. 2024,

11510-11522.

10. Alabi JO, Hedderich MA, Adelani DI, Klakow D: Charting the landscape of African NLP: mapping progress and shaping the road ahead. Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing. 2025, 27807-27841. 10.18653/v1/2025.emnlp-main.1414

11. Inuwa-Dutse I: NaijaNLP: a survey of Nigerian low-resource languages [PREPRINT]. arXiv. 2025, 10.48550/arXiv.2502.19784

12. Ahia O, Aremu A, Abagyan D, et al.: Voices unheard: NLP resources and models for Yorùbá regional dialects. Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. 2024, 4392-4409. 10.18653/v1/2024.emnlp-main.251

13. Akinwonmi AE: Development of a prosodic read speech syllabic corpus of the Yoruba language. Communications on Applied Electronics. 2021, 7:13-32. 10.5120/cae2021652884

14. Alabi JO, Amponsah-Kaakyire K, Adelani DI, España-Bonet C: Massive vs. curated embeddings for low-resourced languages: the case of Yorùbá and Twi. Proceedings of the 12th Language Resources and Evaluation Conference (LREC). 2020, 2754-2762.

15. Jimoh TA, De Wille T, Nikolov NS: Bridging gaps in natural language processing for Yorùbá: A systematic review of a decade of progress and prospects. Natural Language Processing Journal. 2025, 13:100194. 10.1016/j.nlp.2025.100194

16. Houlsby N, Giurgiu A, Jastrzebski S, et al.: Parameter-efficient transfer learning for NLP [PREPRINT]. arXiv. 2019, 10.48550/arXiv.1902.00751

17. Li XL, Liang P: Prefix-tuning: optimizing continuous prompts for generation. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. 2021, 1:4582-4597. 10.18653/v1/2021.acl-long.353

18. Lester B, Al-Rfou R, Constant N: The power of scale for parameter-efficient prompt tuning [PREPRINT]. arXiv. 2021, 10.48550/arXiv.2104.08691

19. Hu EJ, Shen Y, Wallis P, et al.: LoRA: low-rank adaptation of large language models [PREPRINT]. arXiv. 2021, 10.48550/arXiv.2106.09685

20. Dettmers T, Pagnoni A, Holtzman A, Zettlemoyer L: QLoRA: efficient finetuning of quantized LLMs [PREPRINT]. arXiv. 2023, 10.48550/arXiv.2305.14314

21. Ahia O, Ogueji K: Towards supervised and unsupervised neural machine translation baselines for Nigerian Pidgin [PREPRINT]. arXiv. 2020, 10.48550/arXiv.2003.12660

22. Ekle OA, Das B: Low-resource neural machine translation using recurrent neural networks and transfer learning: a case study on English-to-Igbo [PREPRINT]. arXiv. 2025, 10.48550/arXiv.2504.17252

23. Adelani DI, Alabi JO, Fan A, et al.: A few thousand translations go a long way! Leveraging pre-trained models for African news translation. Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2022, 3053-3070. 10.18653/v1/2022.naacl-main.223

24. Moslem Y, Wassie AK, Abebe AG: AfriNLLB: efficient translation models for African languages [PREPRINT]. arXiv. 2026, 10.48550/arXiv.2602.09373

25. Oni MK, Prama TT: Transformer-based low-resource language translation: a study on standard Bengali to Sylheti [PREPRINT]. arXiv. 2025, 10.48550/arXiv.2510.18898

26. Man CZ, Win SSM, Khine KLL: Fine-tuning the mT5 model on bidirectional Myanmar and Tedim Chin machine translation system. IAENG International Journal of Computer Science. 2025, 52:4589-4599.

27. Khatri J, Murthy RV, Bhattacharyya P: Language model pre-training and transfer learning for very low resource languages. Proceedings of the Sixth Conference on Machine Translation (WMT). 2021, 995-998.

28. Liang X, Khaw YMJ, Liew SY, Tan TP, Qin D: Toward low-resource languages machine translation: A language-specific fine-tuning with LoRA for specialized large language models. IEEE Access. 2025, 13:46616-46626. 10.1109/access.2025.3549795

29. Kariuki MW, Muguro J, Maina CW, Wanzare LDA: Bridging the language gap: fine-tuning Llama for machine translation in low-resource African languages. Proceedings of the AI for African Languages Conference 2025. 2026, 314:37-40.

How to cite this article:
Akande O, Opateye G (August 03, 2026) Low-Resource Machine Translation of Yoruba Text to Nigerian Pidgin Using NLLB-200 and Parameter-Efficient Fine-Tuning. Cureus J Comput Sci 3 : es44389-026-00202-y. DOI https://doi.org/10.7759/s44389-026-00202-y