FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 10 No. 11, June, 2026, pp 299 – 305
DOI: https://doi.org/10.33003/fjs-2026-1011-5311

Development of a Transformer-Based Neural Machine Translation System for English to Ebira Language

***[1]Tijani Musari Abdulmusawir, [2]Amina Hassan Abubakar,**
[2]Armand Florentin Donfack Kana and [2]Mohammed Abdullahi

[1]Department of Computer Science, School of Technology, Federal Polytechnic Idah, Kogi state, Nigeria.
[2]Department of Computer Science, Faculty of Physical Sciences, Ahmadu Bello University, Zaria, Kaduna state, Nigeria.

*Corresponding authors' email: musaritijani@gmail.com

Received: 22 Jan. 2026
Accepted: 12 Feb. 2026
Published: 30 May 2026

ABSTRACT
Transformer-based neural machine translation (NMT) models have boosted translation accuracy for high-resource languages; however, research has largely bypassed unwritten and low-resource languages, particularly African languages such as Ebira. Ebira is an unwritten, low-resource language spoken by approximately 2.5 million people predominantly in Kogi State, Nigeria. Existing Ebira machine translation (MT) systems suffer from poor fluency, accuracy, and missed nuances, constrained by small datasets and rule-based methods. This study presents the development of a neural machine translation (NMT) system for English-to-Ebira translation using Google's T5-base transformer model. A bilingual parallel corpus of 32,322 English-Ebira sentence pairs was compiled and used to fine-tune the model. The system achieved a corpus-level BLEU score of 40.95%, with 87% of evaluated sentences scoring 0.5 BLEU or higher, surpassing the prior rule-based system's threshold result of 81.50%, corresponding to 6.75% relative improvement. Human evaluation by ten native Ebira speakers yielded a mean rating of 8.33/10 for fluency, accuracy, and cultural relevance. This research demonstrated that the application of transfer learning on transformer NMT model significantly improves the quality of (MT) systems; and also provides a foundational step for the development of computational resources for Ebira and supports the broader goal of linguistic inclusivity in artificial intelligence.

Keywords: BLEU Score, Fine-Tuning, Low-Resource Languages, Neural Machine Translation, Transformer Architecture

INTRODUCTION

Language is central to communication, education, and cultural identity. In Nigeria, over 200 million people speak more than 500 indigenous languages, yet most remain underrepresented in digital technologies (Oluwatoki et al., 2021). The dominance of English in ICT limits access to online resources and services for speakers of low-resource languages.

Machine Translation (MT) automates text conversion between languages and is a subfield of computational linguistics. Neural Machine Translation (NMT) using transformer-based encoder-decoder models with cross-attention has improved accuracy and fluency for high-resource languages (Vaswani et al., 2017). These advances are yet to reach most African languages due to data scarcity and limited tools.

Ebira is a Nupoid language spoken by about 2.5 million people in Kogi, Nasarawa, Ondo, Edo, and Abuja (Joshua Project, 2024). Though vital for cultural heritage, Ebira is unwritten and lacks a standardized orthography. Younger speakers are shifting to English and dominant languages due to schooling and socioeconomic pressure, threatening its long-term vitality (Nekoto et al., 2020). Ebira is an unwritten language — According to (U.S. Election Assistance Commission, 2022) a language that lacks a standardised written system or a widely adopted orthographic convention is an unwritten language.

Existing English-Ebira translation systems by (Etuk et al., 2020; Abdulmusawir et al., 2021) are rule-based and suffer from poor fluency and inability to handle complex syntax. While NMT models exist for related Nigerian languages like Igbo (Ekle and Das, 2025), Igala (Makoji and Sani, 2025), and Nigerian Pidgin (Lin et al., 2023), none has been developed for Ebira.

This study develops an English-to-Ebira NMT model to address this gap. The aim is to improve access to global knowledge for Ebira speakers and contribute to the computational documentation and preservation of an endangered African language. The study developed an automatic English-Ebira translation system by fine-tuning Google's T5-base transformer model. The specific objectives of the research were: (i) to construct a large-scale English-Ebira parallel corpus through manual method; (ii) to develop a neural machine translation model for English-to-Ebira translation with improved fluency, accuracy, and contextual fidelity; and (iii) to evaluate the system using automatic metrics and structured human assessment by native Ebira speakers.

Related Literature

Recent advancements in NMT have demonstrated the efficacy of encoder-decoder architectures, particularly when combined with attention mechanisms (Vaswani et al., 2017). However, most of these advancements have been applied to high-resource languages. For low-resource languages, researchers have adopted techniques such as transfer learning, subword modelling, and data augmentation (Sennrich, Haddow, and Birch, 2016).

Within the Nigerian language context, earlier MT efforts employed rule-based approaches. Etuk et al. (2020) and Abdulmusawir et al. (2021) developed rule-based English-to-Ebira translation systems; the latter achieved a BLEU

threshold of 81.5% for the majority of 300 test sentences but suffered from poor fluency and inability to handle complex sentence structures. Neural systems have been developed for related low-resource Nigerian languages: Orife (2020) achieved BLEU scores of approximately 37% for four Edoid languages; Lin et al. (2023) fine-tuned T5-base for Nigerian Pidgin, achieving 36.35% BLEU on 29,700 pairs; Ekle and Das (2025) applied T5-base to English-Igbo translation on 30,000 pairs, achieving 35-36% BLEU after 80 training epochs; and Makoji and Sani (2025) developed an English-to-Igala NMT system using RNN-GRU with Bahdanau attention, with above 81.0% of sentences scoring 0.5 BLEU or higher. Anazia et al.(2024) further contributes to this landscape by designing a secured real-time speech-to-text platform that translates spoken input into English and Yoruba, demonstrating encoder-decoder-based MT in a low-resource Nigerian context and highlighting the practical potential of integrating speech and translation for low-resource languages. These studies highlight that NMT substantially improves over rule-based approaches but that no prior neural system existed for Ebira. The T5 (Text-to-Text Transfer Transformer) model (Raffel et al., 2020) unifies all NLP tasks into a sequence-to-sequence framework, enabling effective transfer learning for low-resource translation. Comparative studies (Lin et al., 2023) demonstrate that English-pretrained T5-base consistently outperforms multilingual mT5 on low-resource pairs where the target language is absent from pre-training data, providing the methodological justification for T5-base selection in this study.

Previous efforts to develop English-Ebira translation systems laid important groundwork but remained limited. The two existing rule-based systems (Etuk et al., 2020; Abdulmusawir et al., 2021) produced poor fluency, failed on complex sentence structures, and suffered significant loss of linguistic nuance. No neural machine translation system existed for English-to-Ebira prior to this work, representing a fundamental gap this study addresses.

MATERIAL AND METHODS

This section presents a two-phase approach: first, the construction of a 32,322-sentence-pair parallel corpus through manual translation and data augmentation; second, the application of transfer learning to adapt the pre-trained T5-base transformer model to English-to-Ebira translation.

Problem Formulation

This subsection formally defines the English-to-Ebira translation task as a computational problem, establishing the mathematical notation used throughout the methodology and clarifying the connection between the corpus construction and model training stages.

Let E denote the set of all well-formed English sentences and B denote the set of all well-formed Ebira sentences. Let E denote the set of all well-formed English sentences and B denote the set of all well-formed Ebira sentences. The English-to-Ebira translation task is defined (by equation 1) as learning a conditional mapping:

$$f : E \rightarrow B$$

Such that,

$$f(e) = argmax\ P(b \mid e;\ \theta)\ over\ b\ \in B \quad (1)$$

Where, $e\ \in E$ is a source English sentence, $b\ \in B$ is a target Ebira sentence, and θ denotes the parameters of the translation model. The optimal translation $b^* = f(e)$ maximises the conditional probability $P(b|e;\ \theta)$, estimated through neural sequence-to-sequence learning.

This research addresses two structural extensions of the standard formulation arising from the specific properties of the English-Ebira translation problem. First, the low-resource condition requires that the parameters θ be estimated from a parallel corpus $D = \{(e_i, b_i)\}N_{i=1},\ of\ size\ N\ \ll 10^6$, placing the problem firmly in the low-resource regime where standard maximum-likelihood estimation over a large corpus is infeasible. This constraint motivates the transfer learning strategy in Section 3.3, in which θ is initialised from a pre-trained T5-base checkpoint rather than from random initialisation. Second, in Stage 1 (Section 3.2), the parallel corpus is constructed and quality-verified, then augmented to D_final of size 32,322; in Stage 2 (Sections 3.3–3.4), D_final is used to estimate θ* through gradient-based optimisation of the cross-entropy objective using the pre-trained T5-base checkpoint as initialisation.

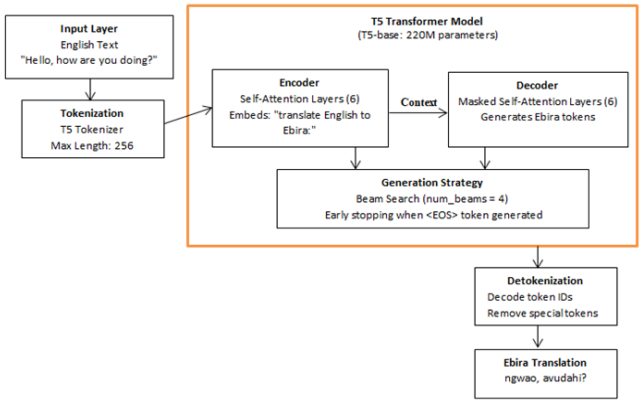

Figure 1: English-Ebira NMT System Architectural Diagram

English-Ebira Neural Machine Translation System Architecture

The neural translation component is built upon Google's T5-base (Text-to-Text Transfer Transformer) model (Raffel et al., 2020), implemented via the HuggingFace Transformers library. T5-base comprises 220 million parameters in an encoder-decoder architecture. Figure 1 illustrates the architectural design of the English-Ebira NMT system. The multi-head self-attention mechanism that underpins both the encoder and decoder is defined in equation 2 as:

$$Attention\ (Q, K, V) = \text{softmax}\ \left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

where Q, K, and V denote the query, key, and value projection matrices, and d_k is the dimensionality of each attention head's

key vectors (Vaswani, 2017). T5-base employs 12 parallel attention heads, enabling simultaneous capture of different linguistic relationships across the sequence. Three model variants were considered: T5-small (60M parameters), T5-base (220M parameters), and T5-large (770M parameters); T5-base was selected as the optimal balance between translation quality and computational feasibility. The pipeline processes English input through five main stages: (i) the Input Layer receives raw English text; (ii) the Tokenisation stage converts text into numerical token representations using the T5 tokeniser with a maximum sequence length of 256 tokens; (iii) the Encoder processes the tokenised input through six self-attention layers to capture contextual relationships; (iv) the Decoder generates the Ebira translation using six masked self-attention layers with cross-attention to the encoder states; and (v) the Detokenisation stage converts the generated token IDs back into readable Ebira text.

Corpus Development

The development of a high-quality English-Ebira parallel corpus is the primary contribution of this work. Given the absence of any existing corpus and Ebira's unwritten nature with no standardized orthography, a systematic multi-stage methodology was implemented to ensure translation quality, consistency, and orthographic standardization.

All English sentences were collected from four source channels and later translated to Ebira by expert native speakers. The sources included: 1) Religious texts: English segments from Quran and Bible translations, yielding 7,000 sentences digitized via OCR for formal register and spiritual vocabulary. 2) Lexical resources: dictionary applications and web-based PDF resources providing 3,114 sentences of basic vocabulary and everyday expressions. 3) Broadcast media: 1,886 contemporary English sentences from Jatto FM and Tao FM radio stations representing modern usage and news terminology. 4) Corpus adaptation: 13,000 English sentences from Arabic-English and Dutch-English parallel corpora in the Tatoeba database to provide diverse domains and varied linguistic complexity. This produced 25,000 English-Ebira sentence pairs.

Following this, data augmentation was applied to increase size and diversity while reducing overfitting. Both techniques were applied only to the English source side, with all expert-validated Ebira targets unchanged. Character-level noise injection added 5,110 pairs using RandomCharAug with 0.10 perturbation probability to simulate real-world typographic variations. Word substitution added 2,212 pairs by replacing up to 20% of content words with WordNet synonyms to expand lexical diversity. Together, augmentation expanded the corpus from 25,000 to 32,322 sentence pairs, a 29.3% increase. Table 1 presents the final corpus composition.

Table 1: Final Corpus Composition After Augmentation

Source	Pairs	% of Total	Characteristics
Religious texts (Quran and Bible)	7,000	21.66%	Formal register, spiritual vocabulary
Lexical resources	3,114	9.63%	Basic vocabulary, everyday expressions
Broadcast media (Jatto FM, Tao FM)	1,886	5.83%	Contemporary usage, news terminology
Adapted corpora (Tatoeba)	13,000	40.23%	Diverse domains, varied complexity
Total initial corpus	25,000	77.35%	Before augmentation
Character-level noise augmentation	5,110	15.81%	Source-side only; Ebira targets unchanged
Word substitution augmentation	2,212	6.84%	Source-side only; Ebira targets unchanged
Total corpus after augmentation	32,322	100%	Comprehensive coverage

Training and Evaluation

The T5-base model was fine-tuned using the HuggingFace Transformers Trainer API with the AdamW optimiser (learning rate 1e-4, linear warmup over 500 steps). FP16 mixed-precision training and gradient accumulation over two mini-batches (effective batch size 8) were employed to improve training efficiency on a single NVIDIA T4 GPU. Early stopping with patience of three evaluation cycles was applied. Training was completed in eight epochs (271 minutes; 25,800 training steps).

System performance was assessed using both automatic metrics and structured human evaluation. BLEU (Papineni et al., 2002) served as the primary benchmark metric, computed at corpus level over the full 3,232-sentence test set and at sentence level over a 200-sentence expert-evaluated baseline set. METEOR (Banerjee and Lavie, 2005) was applied to assess semantic correspondence, and ChrF (Popovic, 2015) to capture character-level morphological accuracy. For human evaluation, ten native Ebira speakers were recruited from the Ebira-speaking community. Each evaluator rated 20 translation pairs from the 200-sentence baseline set on a 10-point Likert scale (1 = very poor, 10 = excellent) across three independently scored criteria: fluency (the degree to which the Ebira translation reads naturally and flows as a native speaker would express the idea); accuracy (the degree to which the Ebira translation preserves the complete meaning of the English source without information added, omitted, or distorted); and cultural relevance (the degree to which the translation employs culturally appropriate Ebira expression for the given communicative function). Written annotation guidelines with three anchor examples at scale points 2, 5, and 9 were provided to all evaluators before the exercise. Evaluators completed the rating task independently without access to reference translations or other evaluators' judgements. The 200-sentence evaluation pool exceeds the minimum statistically valid sample of 196 derived using Cochran's (1977) finite-population-corrected formula at 95% confidence with a ±7% margin of error for the human evaluation sub-population. Mean ratings were computed across all ten evaluators.

RESULTS AND DISCUSSION

Training Convergence and Loss Trajectory

The fine-tuned T5-base model completed training in eight epochs (271 minutes; 25,800 training steps) on a CUDA-enabled NVIDIA T4 GPU. Both training and validation losses decreased consistently throughout, confirming stable convergence without overfitting. Training loss decreased from 2.754 at epoch 1 to 0.532 at epoch 8, a reduction of 80.7%. Validation loss decreased from 2.559 to 0.487 (81.0%). Throughout all 25,800 steps, the validation loss tracked within 0.10 to 0.20 points of training loss, confirming the absence of overfitting and successful generalisation. Early stopping was triggered at epoch 8 with patience = 3. Table 2 and Figure 2 present the epoch-level loss values.

Table 2: Epoch-Level Training and Validation Loss Values Over Eight Fine-Tuning Epochs

Epoch	Training Loss	Validation Loss	Val. Loss Reduction	Notes
1	2.754	2.559	(baseline)	Initial rapid adaptation
2	2.201	2.034	−0.525	Steep early convergence
3	1.687	1.543	−0.491	Warmup steps complete
4	1.243	1.118	−0.425	Stable learning phase
5	0.981	0.872	−0.246	Consistent convergence
6	0.789	0.701	−0.171	Fine-tuning of patterns
7	0.641	0.574	−0.127	Near-optimal configuration
8	0.532	0.487	−0.087	Best checkpoint; early stopping

Note. Total Training Duration: 271 Minutes on a Single NVIDIA T4 GPU (16 GB VRAM). Early Stopping Triggered at Epoch 8 with Patience = 3

Figure 2: Training and Validation Loss Convergence

Automatic Evaluation Results

Following training, the fine-tuned model was evaluated on both test sets. All sentences were translated successfully, corresponding to a 100% translation completion rate with no generation failures. The average translation speed was 2.85 sentences per second. Table 3 presents the full automatic metric results.

Table 3: Automatic Evaluation Metrics for the T5-Base Model

Metric	Score	Interpretation
Corpus-level BLEU (%)	40.95	Evaluated on 3,232-sentence held-out test set
Sentences ≥ 0.5 BLEU (%)	87.00	Evaluated on 200-sentence expert baseline set; same threshold as Abdulmusawir et al. (2021)
METEOR	0.73	High semantic correspondence; rewards synonym and stem matches beyond BLEU
ChrF	0.82	Strong character-level overlap; captures morphological accuracy for agglutinative Ebira
Human rating (/10)	8.33	Mean score from ten native Ebira speakers across 200 sentence-translation pairs
Translation success rate (%)	100.0	No generation failures across 3,232 test sentences
Average translation speed	2.85 sent/sec	Suitable for real-time applications

Note. Corpus-Level BLEU Computed on the 3,232-Sentence Held-out Test set. METEOR, ChrF, and BLEU Threshold Statistics Computed on the 200-Sentence Expert-Evaluated Baseline Set. Human Rating is the Mean of ten Evaluators' Scores Across 200 Sentence-Translation Pairs on a 10-Point Likert Scale Assessing Fluency, Accuracy, and Cultural Relevance

The threshold distribution on the 200-sentence baseline set shows 87.0% of sentences at or above the 0.5 BLEU threshold, with 55.0% achieving a perfect score of 1.0, indicating that more than half of the test sentences were translated identically to the reference. The METEOR of 0.73 and ChrF of 0.82, substantially higher relative to what the BLEU value alone would suggest, confirm that the model reproduces Ebira semantic content and morphological structure with considerably greater fidelity than word-level n-gram overlap captures — a critical consideration for an agglutinative tonal language with no standardised orthography.

Comparative Evaluation against Related Systems

Table 4 provides the systematic comparison of the proposed system against all reviewed Nigerian low-resource NMT systems. Corpus-level BLEU is computed using the standard multi-sentence algorithm over the full 3,232-sentence held-out test set, enabling direct comparison with other systems' published corpus-level BLEU scores (Ekle and Das, 2025; Lin et al., 2023). The BLEU threshold statistic (percentage of sentences at or above 0.5 BLEU) is computed over the 200-sentence expert-evaluated baseline set using the same test set, tokenisation, and threshold definition as Abdulmusawir et al. (2021), enabling a valid direct comparison with the prior rule-based system on this measure. These two metrics are not equivalent and no comparison presents one against the other as if they were interchangeable.

Table 4: Comparative Evaluation against Related Low-Resource Nigerian Language NMT Systems

Parameter	Ekle & Das (2025) Igbo	Lin et al. (2023) Pidgin	Makoji & Sani (2025) Igala	Proposed System (Ebira)
Architecture	T5-base	T5-base	RNN-GRU + Bahdanau	T5-base
Dataset size	30,000	29,700	1,000	32,322
Training epochs	80	NR	25	8
Corpus-level BLEU (%)	36.0	36.35	NR	40.95
Threshold ≥ 0.5 BLEU (%)	NR	NR	81.00	87.00
METEOR	NR	NR	NR	0.73
ChrF	NR	NR	NR	0.82
Prior rule-based (Ebira)	N/A	N/A	N/A	81.50% ≥ 0.5 BLEU

Note. NR = Not Reported. N/A = Not Applicable. Corpus-Level BLEU Comparisons (Ekle and Das, 2025; Lin et al., 2023) are on Different Test Sets and Different Language Pairs; they Provide Architectural Benchmarks, not Direct Equivalence. Threshold Comparisons are on the Same 200-Sentence Baseline set with the Same Tokenisation and Threshold. Prior Rule-Based System Result from Abdulmusawir et al. (2021) is Provided for Contextual Reference only Given the Architectural Paradigm Difference

The proposed model achieves the highest corpus-level BLEU of 40.95% among all T5-based comparators, representing a relative improvement of 13.75% over Ekle and Das (2025) at 36.0% and 12.65% over Lin et al. (2023) at 36.35%, while requiring 90% fewer training epochs (8 versus 80). Against Makoji and Sani (2025) at 81.0% on the threshold metric, the proposed system achieves a relative improvement of 7.41% as shown in Figure 3.

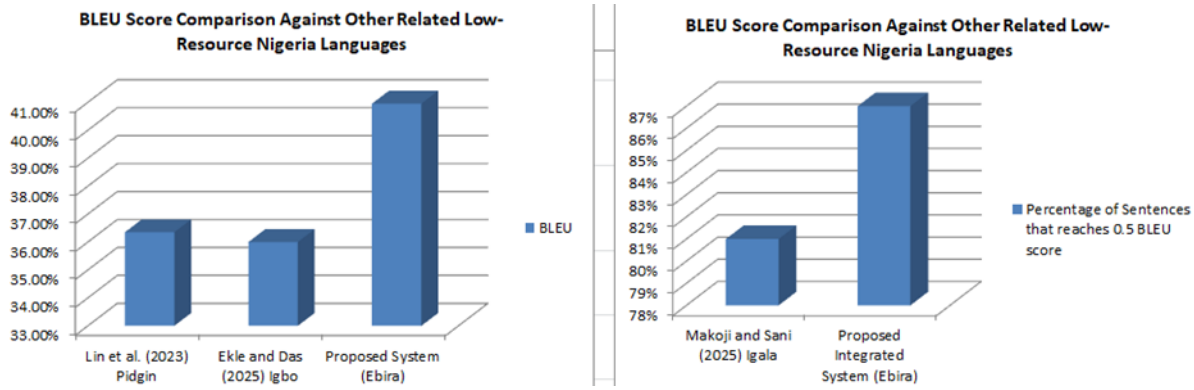

Figure 3: Comparative BLEU Threshold Performance (≥ 0.5) of the Proposed Integrated System Against Related Low-Resource Nigerian Language NMT Systems

Sample Translation Analysis

Table 5 presents ten English-to-Ebira translation outputs with professional human references and sentence-level accuracy scores. Sentence accuracy was computed as the percentage of correctly translated content words relative to the reference.

Table 5: Sample English-to-Ebira Translation Outputs with Accuracy and Metric Scores

S/N	English Input	Reference Translation	Model Output	Acc. (%)	BLEU	METEOR	ChrF
1	My name is Abu	Irehami ori Abu	irehami vi Abu	95	0.85	0.88	89.3
2	Hello, how are you doing?	ngwao, avudahi?	ngwao, avudahi?	100	1.00	0.94	100.0
3	What is your name?	irehawu ri	irehawu ri	100	1.00	0.98	100.0
4	Which state are you from?	state izuhureve?	state izuhureve?	100	1.00	0.98	100.0
5	Good morning.	nyene.	nyene.	100	1.00	0.98	100.0
6	The man gave him a book.	omuha ono osi uwe yo.	omuha ono osi uwe.	75	0.75	0.74	75.2
7	I finished the work.	mamukoro ono ta.	mamukoro ono ta.	100	1.00	0.98	100.0
8	I love you.	manyi oyisi awu.	manyi oyisi awu.	100	1.00	0.98	100.0
9	The little child died.	ozi inuvo ono owusu.	ozi ono owusu.	90	0.82	0.84	91.0
10	Where is the market?	Izohu ya?	Izi ohu ya?	75	0.75	0.80	82.0

Note. Ebira Translations are in italics in the Running Text. Sentence Accuracy Computed as the Percentage of Correctly Translated Content Words. METEOR Scores Range from 0 to 1. ChrF Scores are Reported as Percentages. Six of the Ten Sentences Achieve 100% Accuracy with Model Outputs Identical to the Professional Human Reference

Six of the ten sentences achieve 100% accuracy with model outputs identical to the professional human reference. Sentence 1 scores 95%, with the only deviation being the substitution of possessive marker vi for the expected ori, a minor morphological variant that preserves meaning. Sentences 6 and 9 score 75% and 90% respectively, arising from partial omissions of object or modifier words rather than mistranslation of core content. These results are consistent with the general finding that performance degrades slightly on longer, more morphologically complex Ebira target sentences. Figure 4 shows the system's sample interfaces.

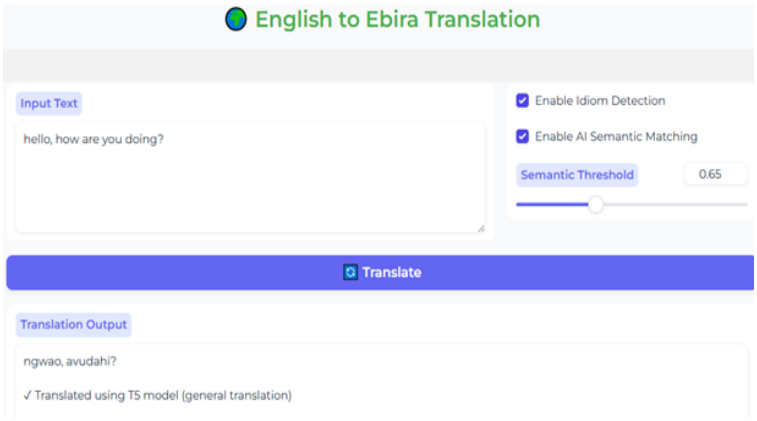

Figure 4: Sample Translation Interface

Discussion

The results affirm the capability of transformer-based NMT models in addressing translation for low-resource, unwritten languages. The fine-tuned T5-base model achieves the highest corpus-level BLEU of 40.95% among all T5-based comparators for related Nigerian low-resource languages, representing a relative improvement of approximately 13% over Ekle and Das (2025) at 36.0% and over Lin et al. (2023) at 36.35%, while requiring 90% fewer training epochs (8 versus 80). Against Makoji and Sani (2025) at 81.0% on the BLEU threshold metric, the proposed system achieves a relative improvement of 7.4%.

The BLEU score of 40.95% should be interpreted in context. Three structural features of Ebira constrain the BLEU ceiling: its tonal nature (surface n-gram BLEU cannot reward correct tonal representation); its agglutinative morphology (candidates may differ from the reference at the word level while conveying identical meaning); and single-reference evaluation for an orthographically unstandardised language. The complementary metrics — METEOR of 0.73 and ChrF of 0.82 — capture semantic and morphological accuracy more faithfully, and together with the human rating of 8.33/10 provide strong, multi-dimensional validation of the system's quality.

Some performance degradation was observed on longer sentences (exceeding approximately 50 tokens) and on idiomatic or culturally embedded expressions, consistent with findings across all reviewed NMT systems for Nigerian languages. These limitations suggest the need for larger training corpora and specialised mechanisms for idiomatic translation.

The successful implementation demonstrates the potential of transfer learning with T5-base for severely under-resourced, unwritten African languages.

CONCLUSIONS

The results affirm the capability of transformer-based NMT models in addressing translation for low-resource languages, even where the target language is entirely absent from the pre-training data. The T5-base fine-tuning strategy, combined with a large-scale purpose-built parallel corpus, successfully produced the first neural machine translation system for English-to-Ebira, achieving performance competitive with state-of-the-art T5-based systems for related Nigerian low-resource languages while requiring significantly fewer computational resources.

The 32,322-pair English-Ebira parallel corpus constitutes the first publicly available large-scale aligned parallel dataset for this language pair, filling the fundamental data scarcity that has prevented data-driven NMT development for Ebira. Beyond the technical contributions, this research carries cultural and societal significance by supporting the digital visibility and revitalisation of Ebira as a living language in the knowledge economy.

RECOMMENDATIONS

Given the promising results of this research, the following recommendations are proposed:

Corpus expansion beyond 32,322 pairs to 100,000 or more sentence pairs, incorporating additional domains such as healthcare, legal, and agricultural texts, to improve model generalisation.

Extension to bidirectional Ebira-to-English translation using the T5 text-to-text framework with a reverse-direction task prefix

Integration of speech recognition and text-to-speech synthesis to expand accessibility for community members with limited literacy, though developing speech components for a tonal language represents a significant independent research challenge

Fine-tuning of larger models (T5-large, 770M parameters) or multilingual variants once NMT resources for related Nupoid languages become available

Development of specialised mechanisms for detecting and translating idiomatic and culturally embedded expressions, which the current compositional neural decoding cannot reliably handle

ACKNOWLEDGMENT

The authors are grateful to the native communities, journal and Ahmadu Bello University and their staff.

DISCLOSURE

The author(s) report no conflicts of interest in this work

REFERENCES

Abdulmusawir, T. M., Ayegba, S. F., Kayode, Y. M., & Eze, C. C. (2021). A system for machine translation from English to Ebira using the rule-based approach. Journal of Scientific Research and Reports, 27(11), 137–148. https://doi.org/10.9734/jsrr/2021/v27i1130465

Anazia, E. K., Eti, E. F., Ovili, P.H., & Ogbimi, O. F. (2024). Speech-to-Text: A secured real-time language translation platform for students. *FUDMA Journal of Sciences*, *8*(6), 329–338. https://doi.org/10.33003/fjs-2024-0806-2890

Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. International Conference on Learning Representations (ICLR).

Banerjee, S., & Lavie, A. (2005). METEOR: An automatic metric for MT evaluation with improved correlation with human judgements. ACL Workshop on Evaluation Measures for MT, 65–72.

Ekle, O. A., & Das, B. (2025). Low-resource neural machine translation using recurrent neural networks and transfer learning: A case study on English-to-Igbo. arXiv:2504.17252.

Etuk, S. O., Oyefolahan, I. O., Zubairu, H. A., Babakano, F. J., and Momoh, L. O. (2020). Development of an English-to-Ebira and Ebira-to-English machine translation system. International Journal of Information Processing and Communication, 6(1), 1–12.

Joshua Project. (2024). Ebira in Nigeria people group profile. https://joshuaproject.net/people_groups/12188

Lin, P.-J., Saeed, M., Chang, E., & Scholman, M. (2023). Low-resource cross-lingual adaptive training for Nigerian Pidgin. Interspeech 2023, 3950–3954.

Makoji, E., & Sani, F. (2025). Development and evaluation of an English-to-Igala NMT system using deep learning. International Journal of Innovative Science and Research Technology, 10(5). https://doi.org/10.38124/ijisrt/25may556

Nekoto, W., Marivate, V., Matsila, T., Fasubaa, T., Kolawole, T., Akinadan, F., Adewole, D., Yamahata, K., & Hsu, C.-J. (2020). Participatory research for low-resourced machine translation: A case study in African languages. Findings of the Association for Computational Linguistics: EMNLP 2020, 2144–2160. https://doi.org/10.18653/v1/2020.findings-emnlp.195

Oluwatoki, T. G., Adetunmbi, A. O., Boyinbode, O. K., Badeji-Ajisafe, B., Abiola, O. B. and Popoọla, O. S. (2021). Machine translation systems for Nigerian indigenous languages: A statistical overview. *Nigerian Journal of Technological Research*, 16(2), 41–52.

Orife, I. (2020). Neural machine translation for Edoid languages. In AfricaNLP Workshop, ICLR 2020.

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). BLEU: A method for automatic evaluation of machine translation. In Proceedings of ACL 2002, 311–318.

Popovic, M. (2015). ChrF: Character n-gram F-score for automatic MT evaluation. Workshop on Statistical MT, 392–395.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21 (140), 1–67. https://doi.org/10.48550/arXiv.1910.10683

Sennrich, R., Haddow, B., & Birch, A. (2016). Improving neural machine translation models with monolingual data. ACL 2016, 86–96.

Tijani, M. A., Abubakar, A. H., & Donfack Kana, A. F. (2025). Leveraging data triangulation technique for the development of unwritten language parallel corpora for NLP applications. In Proceedings of ICCAIT2025, 234–241.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention is all you need. In Advances in Neural Information Processing Systems 30* (NeurIPS 2017) (pp. 5998–6008). https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Xue, L., Constant, N., Roberts, A., Kale, M., Al-Rfou, R., Siddhant, A., Barua, A., & Raffel, C. (2021). mT5: A massively multilingual pre-trained text-to-text transformer. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 483–498. https://doi.org/10.18653/v1/2021.naacl-main.41.