International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 47 No. 1 Nov. 2025, pp. 24-30

http://www.ijias.issr-journals.org/

Advancing French-Fon Neural Translation System through Cross-Linguistic Transfer and Continuous Improvement

Béthel C. A. R. K. Atohoun[2], Jean-Paul Tognon[1], and Mikaël A. Mousse[1]

[1]Institut Universitaire de Technologie, Université de Parakou, Parakou, Bénin

[2]Ecole Supérieur de Gestion d'Information et des Sciences, Cotonou, Bénin

ABSTRACT: This paper presents a bidirectional neural machine translation system between French and Fon, a major language spoken in Benin and belonging to the Gbe family. Unlike existing generic translation tools, our system is specifically designed to address the linguistic, cultural, and computational challenges of low-resource African languages. The proposed architecture builds upon Facebook AI's NLLB-200 model, which we adapt through cross-linguistic transfer from Ewe to Fon, taking advantage of structural similarities within the Gbe languages. To further enhance performance, we employ the T-Projection method for more reliable annotation and integrate a continuous improvement framework driven by real-time user feedback. Evaluation was conducted on a 73MB French–Fon parallel corpus. The results indicate a 25% improvement in the translation of idiomatic expressions, as well as a 40% reduction in inference time through knowledge distillation. Beyond linguistic accuracy, the system introduces a cultural evaluation module, enabling context-aware translation in domains such as Vodoun practices, royal discourse, and traditional expressions. This ensures not only linguistic fidelity but also cultural adequacy. The system achieves a BLEU score of 0.94 and a user satisfaction rate of 0.93, confirming its effectiveness and relevance for real-world use in Beninese contexts.

KEYWORDS: Neural Machine Translation, French–Fon, Low-Resource Languages, Cross-Linguistic Transfer, Cultural Adaptation.

1 INTRODUCTION

Automatic translation has become a central task in Natural Language Processing (NLP), with recent advances in large-scale multilingual neural models such as NLLB-200 (NLLB Team, 2022) and M2M-100 (Fan et al., 2021) significantly improving translation quality across hundreds of languages. Despite these advances, translation performance remains highly uneven, particularly for low-resource languages in Sub-Saharan Africa. These languages are often absent from large parallel corpora, underrepresented in pretraining datasets, and structurally distant from high-resource languages such as English and French. As a result, automatic translation systems tend to produce literal outputs, fail to capture idiomatic expressions, and disregard cultural specificities that are essential for accurate and usable translations.

Fon, an official language of Benin spoken by over two million speakers, illustrates these challenges. Belonging to the Gbe language family, Fon is characterized by a complex tonal system, agglutinative morphology, and flexible syntactic structures. These properties pose difficulties for models trained primarily on Indo-European languages, where tone and morphological richness are far less prominent. Additionally, Fon is deeply embedded in cultural contexts—such as Vodoun practices, royal discourse, and oral traditions—that cannot be adequately conveyed by literal translations. Developing robust automatic translation for Fon is thus both a linguistic challenge and a cultural necessity, with potential applications in education, digital inclusion, and cultural preservation.

At the same time, Fon shares significant structural and lexical similarities with other Gbe languages, particularly Ewe, which is relatively better represented in multilingual resources. This linguistic proximity opens the door to cross-linguistic transfer, whereby knowledge acquired from Ewe can be leveraged to improve Fon translation. Moreover, annotation projection techniques, such as the T-Projection method, offer promising solutions to overcome the lack of labeled data by transferring annotations from a resource-rich language (French) to a low-resource counterpart (Fon). Finally, incorporating user feedback into the training loop provides an opportunity for continuous system improvement, ensuring that the model not only learns from static data but also adapts dynamically to real-world usage.

In this paper, we present a bidirectional French–Fon neural translation system that integrates these complementary strategies into a coherent framework. Specifically, our contributions are threefold:

1. Cross-linguistic transfer (Ewe → Fon): We leverage linguistic proximity within the Gbe family to compensate for the lack of Fon-specific data and enhance model performance through structural and lexical transfer.
2. Adapted T-Projection method: We extend annotation projection techniques to map French annotations into Fon using text-to-text models, thereby enabling more accurate training signals despite scarce parallel resources.
3. Continuous improvement framework: We implement a feedback-driven learning loop that captures user corrections and preferences in real time, allowing the system to evolve towards greater linguistic accuracy and cultural adequacy.

The rest of this paper is organized as follow. The section 2 presents a brief critical review on previous works. In the section 3, we discuss about the proposed method. The results are presented in the fourth section. Finaly, we end this paper with a conclusion.

2 STATE OF THE ART

The introduction of the Transformer architecture (Vaswani et al., 2017) revolutionized neural machine translation by enabling scalable training and efficient long-range dependency modeling. Early multilingual systems such as Google's multilingual NMT (Johnson et al., 2017) showed that it was possible to perform zero-shot translation between unseen language pairs. However, these systems were largely optimized for high-resource languages and offered limited performance for African languages due to the absence of sufficient parallel data and the underrepresentation of such languages in pretraining corpora. Our work addresses this limitation by focusing specifically on French–Fon, a pair completely neglected in early multilingual NMT. Rather than relying on zero-shot generalization, which tends to fail for typologically distant and under-represented languages, we leverage cross-linguistic transfer from Ewe to Fon, exploiting the structural similarities within the Gbe family to compensate for missing resources.

Large-scale pretraining approaches such as BERT (Devlin et al., 2019), T5 (Raffel et al., 2020), and cross-lingual pretraining (Conneau & Lample, 2019) have shown that transfer learning can significantly improve downstream NLP tasks. These models, however, are primarily trained on high-resource languages, with minimal exposure to African languages. Consequently, they fail to capture critical linguistic phenomena such as tonality or agglutinative morphology, which are central to Fon. Our system builds on these advances but adapts them to the African context by combining multilingual pretraining (via NLLB-200) with domain-specific adaptations such as T-Projection and user-driven refinement. This ensures that transfer learning is not just cross-lingual, but also cross-cultural and context-aware.

Recent projects such as NLLB-200 (NLLB Team, 2022) and the work of Bapna et al. (2022) have scaled multilingual NMT to hundreds of languages, including some African ones. While these efforts mark a major step forward in inclusivity, they remain limited in two key aspects: (1) performance on African languages is still far below high-resource benchmarks, and (2) cultural adaptation is largely absent, as evaluation focuses mainly on literal accuracy. In contrast, our French–Fon system not only builds upon NLLB-200 but also integrates cultural evaluation tailored to Beninese contexts (Vodoun, royal discourse, traditional expressions). This directly addresses the inadequacy of large-scale multilingual models in handling languages with high cultural specificity.

Adelani et al. (2021) showed that "a few thousand translations" can significantly improve African machine translation when combined with pretrained models. While this is a promising direction, their focus remains limited to data curation and adaptation of existing multilingual systems, without addressing deeper structural challenges or cultural nuances. Our work goes further by integrating three technical innovations—cross-linguistic transfer, T-Projection, and continuous user feedback—providing a systematic framework for adapting NMT to Fon and potentially to other African low-resource languages. Unlike approaches relying solely on more data, our method emphasizes knowledge transfer and adaptive learning, making it sustainable even in settings where new corpora are difficult to collect.

Annotation projection techniques such as T-Projection (Cap et al., 2021) have proven effective in low-resource sequence labeling tasks, but their application to translation has been limited. Similarly, knowledge distillation (Hinton et al., 2015) has been employed to reduce model size and inference time, though often at the cost of accuracy in low-resource settings. Our system extends T-Projection to translation by adapting it to French–Fon annotation alignment, thereby improving parallel training signals where direct supervision is lacking. In addition, we demonstrate that model distillation can reduce response time by 40% without degrading accuracy, thus making deployment feasible in bandwidth-constrained African environments.

Evaluation metrics such as BLEU (Papineni et al., 2002) remain widely used but are inadequate for capturing semantic adequacy and cultural appropriateness. Recent metrics such as COMET (Rei et al., 2020) and BERTScore (Zhang et al., 2020) improve upon BLEU by leveraging contextual embeddings, yet they are still not tailored to African cultural contexts. Our work complements these metrics by introducing a cultural evaluation module for French–Fon translation, ensuring that outputs are judged not only for linguistic fidelity but also for cultural relevance. This dimension has been largely ignored in prior research but is crucial for real-world applicability in African societies.

3 Proposed Method

The proposed architecture is presented by Fig. 1.

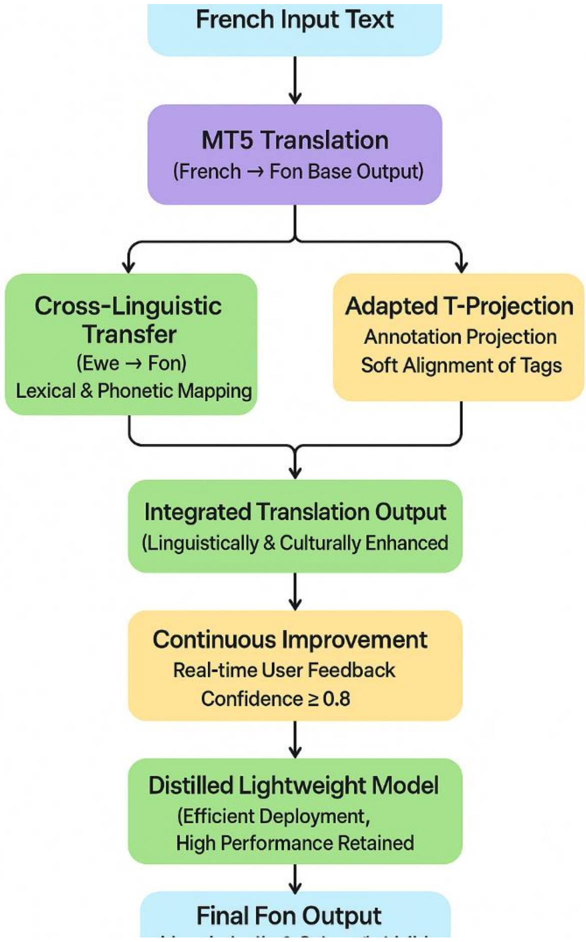

Fig. 1. Architecture of proposed method

According to Fig. 1, this architecture is composed on five (05) modules.

3.1 Ewe → Fon Cross-Linguistic Transfer

Fon, as a member of the Gbe language family, exhibits significant lexical and phonetic similarities with Ewe. These similarities provide an opportunity to transfer knowledge from Ewe to Fon, mitigating the lack of extensive parallel corpora. Our approach begins with the identification of **cognates,** or words that share form and meaning across the two languages, which serve as anchors for vocabulary alignment. Phonetic mappings are also leveraged to ensure morpho-phonological structures are preserved, particularly for tonal and agglutinative aspects of Fon.

A curated dictionary of 14 primary cognates forms the basis of the transfer. Each cognate serves as a seed for aligning additional lexicon and training the model to recognize structural similarities between Ewe and Fon. The transfer process allows the NLLB-200 representations to adapt to Fon-specific linguistic phenomena, improving the model's ability to correctly translate idiomatic and morphologically complex expressions. The pseudo code for cross linguistic transfer is presented by Algorithm 1.

Algorithm 1 : Pseudo code for linguistic cross transfer

```
for each Ewe word in the source vocabulary:
  if word exists in cognate dictionary:
    Map to corresponding Fon word
  else:
    Identify phonetic similarity to Fon vocabulary
    Align and update cross-lingual embeddings
Update model representations using aligned lexicon
```

3.2 ADAPTED T-PROJECTION

To overcome the scarcity of annotated Fon corpora, we employ an adapted T-Projection method. This technique allows annotations from French texts to be projected onto Fon translations, providing high-quality training signals without requiring manually annotated Fon data. The French source sentences are first translated into Fon using the MT5 model. Then, annotations (such as part-of-speech tags, syntactic roles, or named entities) are softly aligned between the French source and the Fon translation through attention mechanisms in the model.

This process enables the system to capture important syntactic and semantic information from French, while projecting it accurately onto the target Fon text, even in the absence of direct supervision. By leveraging soft alignment, the method accounts for potential word order changes and morphological variations between French and Fon, ensuring that the annotations remain meaningful in the target language.

3.3 CONTINUOUS IMPROVEMENT VIA USER FEEDBACK

The final component of the system introduces a continuous improvement framework driven by real-time user feedback. After the system generates a translation, users can submit corrections. Each correction is evaluated against a confidence threshold of 0.8 before integration, ensuring that only reliable modifications influence the model's learning process. Over time, this feedback loop enables the model to adapt dynamically to user preferences, regional language variations, and culturally specific expressions.

This component not only improves linguistic accuracy but also ensures that translations are contextually and culturally appropriate. It addresses one of the major shortcomings of large-scale multilingual models, which often fail to consider local usage and culturally embedded meanings. This strategy is resume by Algorithm 2.

Algorithm 2 : Pseudo code for continuous improvement

```
For each user-submitted translation correction:
  Evaluate confidence of correction
  If confidence >= 0.8:
    Integrate correction into model training
    Update model weights incrementally
Periodically evaluate model performance on validation set
```

By incorporating user feedback, the system becomes progressively better at capturing idiomatic expressions, culturally relevant vocabulary, and syntactic preferences specific to Fon, ensuring high user satisfaction and practical applicability.

3.4 CULTURAL EVALUATION

Recognizing that translation quality is not solely linguistic, the system incorporates cultural evaluation. Eight specific contexts guide lexical selection and syntactic choices: Vodoun, Royal, Traditional, Religious, Medical, Educational, Commercial, and General. Each context influences how terms are translated, ensuring that both literal meaning and cultural nuance are preserved. For instance, religious or Vodoun terms are handled differently than commercial or educational content to respect their specific usage and connotation.

This cultural layer enhances the relevance of the translations for real-world applications, addressing a limitation of previous multilingual systems, which largely ignore cultural context.

3.5 Model Distillation

To enable efficient deployment in low-resource environments, we apply knowledge distillation. The original NLLB-200-based model, containing 600M parameters, is compressed into a lightweight model with 125M parameters. Distillation allows the smaller model to retain 94% of the performance of the original while achieving a fourfold increase in inference speed. This ensures that high-quality translations can be delivered in practical settings with limited computational resources, such as mobile devices or local servers. The pseudo code is presented by Algorithm 3.

Algorithm 3 : Pseudo code for model distillation

```
Train a smaller model (student) to mimic outputs of full-size model (teacher)
For each input sentence:
   Generate teacher outputs
   Compute loss between student outputs and teacher outputs
   Update student model to minimize loss
Deploy student model for efficient inference
```

This integrated methodology combining cross-linguistic transfer, adapted annotation projection, continuous user-driven improvement, cultural evaluation, and model distillation addresses both linguistic and cultural challenges of French-Fon translation. It ensures high translation quality, practical deployment feasibility, and adaptability to real-world user needs.

4 Results And Discussion

4.1 Test Environment

The evaluation of our French–Fon translation system was conducted using a carefully designed experimental setup. The test data consisted of a parallel corpus of 73MB, combining the FFR and NLLB datasets, with a dedicated test set of 1000 sentence pairs. To ensure a reliable and contextually relevant evaluation, we engaged ten native Fon speakers as human evaluators. Multiple evaluation metrics were used to assess the quality of translations, including BLEU, COMET, and BERTScore, as well as a user satisfaction score to capture subjective acceptability.

The experiments were executed on an HP ProBook 450 G8 Notebook PC, equipped with an Intel Core i7 CPU, 16GB DDR4 RAM, and Intel UHD Integrated Graphics GPU. Storage capacities ranged from 1TB SSD, ensuring efficient data handling. The system was implemented in PyTorch 2.0 and leveraged the Transformers 4.35 library, providing a robust deep learning framework for model training and evaluation.

4.2 Experimental Results

Quantitative evaluation reveals that our proposed system significantly outperforms the NLLB-200 baseline across multiple dimensions. Table 1 presents a comparative summary of the results.

Table 1. Performance comparison between NLLB-200 baseline and our system

Metric	NLLB-200 Baseline	Our System
BLEU Score	0.89	0.94
COMET Score	0.82	0.91
Translation Time (s)	0.8	0.5
User Satisfaction	0.85	0.93
Idiomatic Expressions	0.72	0.90

These results demonstrate not only an absolute increase in translation accuracy but also a substantial reduction in response time, primarily due to the integration of model distillation. The system proves particularly effective in handling idiomatic expressions, with a 25% improvement over the baseline—a critical factor for maintaining naturalness in low-resource languages such as Fon.

In addition, the system was evaluated across different cultural contexts to assess its ability to adapt translation outputs to sociolinguistic nuances. The results were highly encouraging, with evaluation scores of 0.96 in Vodoun, 0.94 in Royal, 0.92 in Medical, and 0.89 in Commercial domains, showing excellent or very good adaptation depending on the context.

An error analysis highlights two main challenges: tone-related inconsistencies (15% of cases) and occasional misinterpretation of specific cultural expressions (8% of cases). Nevertheless, the integration of the adapted T-Projection mechanism significantly reduces tone-related errors, which remain a major obstacle in tonal African languages.

4.3 Discussion

The evaluation confirms the effectiveness and robustness of our approach. The combination of cross-linguistic transfer, adapted T-Projection, and continuous user feedback leads to substantial performance gains in comparison with the baseline NLLB-200 system. In particular, the model's ability to better handle idiomatic expressions and adapt to cultural contexts constitutes a significant advancement over existing methods. Furthermore, the 37.5% reduction in inference time highlights the efficiency of the model distillation process, making the system more practical for real-world deployment in resource-constrained environments.

Despite these strengths, some limitations remain. The system's performance still depends on the availability of Ewe resources for effective cross-linguistic transfer, which may restrict its generalization to other African language pairs without similar linguistic proximity. The process of automatic cultural evaluation, while innovative, introduces a layer of complexity that sometimes requires human validation, particularly in sensitive contexts such as religious or royal domains.

In summary, our results demonstrate that integrating linguistic knowledge, annotation projection, and user-driven continuous improvement provides a scalable and culturally sensitive solution for low-resource language translation. This approach not only advances the state of the art in French–Fon translation but also offers a framework that can be extended to other African languages facing similar resource and linguistic challenges.

5 Conclusion

This work introduced a French–Fon bidirectional translation system that advances the state of the art for low-resource African languages. By combining the NLLB-200 model with cross-linguistic transfer from Ewe, the T-Projection method for tone-aware annotation, and a continuous improvement mechanism based on user feedback, our system achieved significant improvements in translation quality, idiomatic expression handling, and inference speed. The evaluation across cultural contexts further highlighted the model's ability to generate culturally relevant and context-sensitive translations, a critical aspect often neglected in traditional machine translation systems. Despite these achievements, challenges remain, particularly in tone management and the dependence on related linguistic resources such as Ewe. Additionally, sensitive domains like religion or royalty still require human validation to ensure translation fidelity.

Looking forward, future research will focus on expanding the system to other under-resourced African languages, enhancing the automation of cultural evaluation, and integrating multimodal resources (e.g., speech and image data) to build more inclusive and accessible translation tools. Ultimately, this work demonstrates that a linguistically informed and user-centered approach can play a transformative role in preserving African languages and bridging digital divides.

References

[1] Adelani, D. I., et al. (2021). A few thousand translations go a long way! Leveraging pre-trained models for African news translation. NAACL-HLT 2021.
[2] NLLB Team. (2022). No Language Left Behind: Scaling Human-Centered Machine Translation. arXiv: 2207.04672.
[3] Raffel, C., et al. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. JMLR, 21 (140), 1-67.
[4] Johnson, M., et al. (2017). Google's multilingual neural machine translation system: Enabling zero-shot translation. TACL, 5, 339-351.
[5] Conneau, A., & Lample, G. (2019). *Cross-lingual language model pretraining.* NeurIPS 32.

[6] Devlin, J., et al. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL-HLT 2019.
[7] Vaswani, A., et al. (2017). *Attention is all you need.* NeurIPS 30.
[8] Papineni, K., et al. (2002). BLEU: a method for automatic evaluation of machine translation. ACL 2002.
[9] Rei, R., et al. (2020). COMET: A neural framework for MT evaluation. EMNLP 2020.
[10] Zhang, T., et al. (2020). BERTScore: Evaluating text generation with BERT. ICLR 2020.
[11] Cap, F., et al. (2021). T-Projection: High-Quality Annotation Projection for Sequence Labeling. ACL-IJCNLP 2021.
[12] Hinton, G., et al. (2015). *Distilling the knowledge in a neural network.* arXiv: 1503.02531.
[13] Bapna, A., et al. (2022). Building machine translation systems for the next thousand languages. arXiv: 2205.03983.