Neural Machine Translation for French–Mooré: Adapting Large Language Models to Low-Resource Languages

Walker Stanislas R. Compaoré[1,2], Maimouna Ouattara[1,4], Rodrique Kafando[1], Abdoul Kader Kaboré[1,4], Aminata Sabané[1,3] and Tegawendé F. Bissyandé[1,4]

[1] Centre d'Excellence Interdisciplinaire en Intelligence Artificielle pour le Developpement (CITADEL)
[2] Ecole Polytechnique de Ouagadougou [3] Université Joseph KI-ZERBO
[4] University of Luxembourg

Correspondence: walk.compaore@gmail.com, maimouna.ouattara@uni.lu, rodrique.kafando@citadel.bf, abdoulkader.kabore@uni.lu, aminata.sabane@ujkz.bf, tegawende.bissyande@uni.lu

Abstract

This work focuses on neural machine translation between French and Mooré, leveraging the capabilities of Large Language Models (LLMs) in a low-resource language context. Mooré is a local language widely spoken in Burkina Faso but remains underrepresented in digital resources. Alongside Mooré, French, now a working language, remains widely used in administration, education, justice, etc. The coexistence of these two languages creates a growing demand for effective translation tools. However, Mooré, like many low-resource languages, poses significant challenges for machine translation due to the scarcity of parallel corpora and its complex morphology.

The main objective of this work is to adapt LLMs for French–Mooré translation. Three pre-trained models were selected: No Language Left Behind (NLLB-200), mBART50, and AfroLM. A corpus of approximately 83,000 validated sentence pairs was compiled from an initial collection of 97,060 pairs through pre-processing, semantic filtering, and human evaluation. Specific adaptations to tokenizers and model architectures were applied to improve translation quality.

The results show that the fine-tuned NLLB model outperforms the others, highlighting the importance of native language support. mBART50 achieves comparable performance after fine-tuning, while AfroLM remains less effective. Despite existing limitations, this study demonstrates the potential of fine-tuned LLMs for African low-resource languages.

1 Introduction

Africa is characterized by high linguistic diversity, with over 2,000 languages spoken across the continent (Smith and Levasseur (2009)). Today, these native languages coexist with languages introduced during colonization, such as French, English, and Portuguese, which often serve as official or administrative languages (Diki-Kidiri, 2008). This diversity reflects the continent's rich indigenous linguistic heritage and presents both opportunities and challenges for natural language processing research. Indeed, most African languages are considered as low-resource languages due to the scarcity of digital corpora and language processing tools (Ouattara et al., 2025).

This situation is exemplified in Burkina Faso, where several national languages coexist. Among them, Mooré occupies a central position. According to the 5th General Population and Housing Census, it is spoken by approximately 52.9% of the population [1] . As the identity language of the Mossi ethnic group, Mooré is understood by many other communities.

At the same time, French, formerly the sole official language and now a working language, continues to play an important role in administrative, educational, judicial, and institutional domains. The coexistence of French and local languages, particularly Mooré, generates a strong demand for translation, notably for administrative documents, media content, educational materials, and digital communication.

In this context, neural machine translation emerges as a promising solution. Previous studies have explored French–Mooré translation. Notably, Ouily et al. (2024) introduced a first neural translation model for this language pair, achieving encouraging yet limited results. These limitations are mainly due to the scarcity of parallel corpora and restricted access to high-performance models.

The advent of large language models (LLMs) has significantly transformed natural language processing. Trained on large-scale multilingual corpora, these models demonstrate strong capabilities in machine translation. Among them, the *No Lan-*

[1] `https://www.insd.bf/fr/resultats`

Proceedings of the Second Workshop on Language Models for Low-Resource Languages (LoResLM 2026), pages 615–622
March 29, 2026

guage Left Behind (NLLB) model (NLLB Team et al., 2022), which covers more than 200 languages including Mooré and is released as open source, offers a particularly relevant foundation. In this context, this work explores the adaptation of LLMs for bidirectional French–Mooré machine translation using parallel data and targeted fine-tuning strategies.

This work contributes a novel French–Mooré corpus of approximately 83,000 validated sentence pairs (filtered from an initial collection of 97,060 pairs), including a held-out test set of approximately 8,000 pairs for benchmarking. We provide tokenizers adapted to Mooré morphological segmentation, two high-performing fine-tuned models accessible via a web-based translation platform. Finally, we provide a French–Mooré text translation API for others use cases.

2 Background

This section presents the characteristics of Mooré, reviews prior work on machine translation for low-resource and African languages, with a particular focus on approaches relevant to the French–Mooré language pair. We then discuss multilingual pre-trained models that motivate the design choices explored in this study.

2.1 Linguistic characteristics of Mooré

Mooré belongs to the Gur branch of the Niger-Congo language family (COCOON Database, 2024) and is the most widely spoken language in Burkina Faso. Its linguistic structure differs substantially from French, presenting specific challenges for machine translation.

Mooré exhibits a rich agglutinative morphology, particularly in its nominal and verbal systems (Ouoba Kabore, 2025). Nouns are organized into noun classes marked by suffixes that indicate number and semantic categories. For example, the singular suffix *-a* becomes *-ã* in the plural for certain classes. Verbs carry multiple affixes encoding tense, aspect, mood, and negation, often resulting in complex verbal forms.

Mooré follows a Subject-Verb-Object (SVO) word order, similar to French (Ouoba Kabore, 2025). However, modifiers typically follow the noun, and the language employs postpositions rather than prepositions. Tense and aspect are often expressed through auxiliary constructions and particles.

2.2 Machine translation for low-resource and African languages

Neural machine translation has achieved major advances since the introduction of the Transformer architecture (Vaswani et al., 2017). However, these improvements have primarily benefited resource-rich languages. For low-resource languages, which suffer from a severe scarcity of parallel and monolingual data, transfer learning and multilingual modeling have become the dominant strategies (Neubig and Hu, 2018; Aharoni et al., 2019).

Massively multilingual pre-trained models such as mBERT (Devlin et al., 2019) and XLM-R (Conneau et al., 2020) demonstrate the effectiveness of cross-lingual transfer, but their coverage of African languages remains limited. Pre-training followed by fine-tuning on small parallel corpora has nonetheless emerged as an effective paradigm, significantly reducing data requirements compared to training from scratch (Zoph et al., 2016).

Despite more than 2,000 languages spoken across Africa (Smith and Levasseur, 2009), the vast majority remain unsupported by machine translation systems. While commercial platforms such as Google Translate and Microsoft Translator cover a few dozen African languages, this represents less than 3% of the continent's linguistic diversity. Early research focused on a small subset of high-resource African languages such as Swahili, Amharic, and Hausa (Martinus and Abbott, 2020). Standard benchmarks such as FLORES-200 (NLLB Team et al., 2022) include 55 African languages, yet many remain absent, including Mooré.

Available corpora for African languages are predominantly derived from religious sources, notably JW300 (Agić and Vulić, 2019), which severely limits its domain diversity. Community-driven initiatives such as Masakhane (Nekoto et al., 2020) aim to address this gap, but resources remain insufficient for most languages.

For Mooré specifically, research on machine translation remains extremely limited, with only a handful of studies addressing this language pair. The most recent and notable contribution is from Ouily et al. (2024), who developed the first neural machine translation system for the French–Mooré language pair. Their Transformer-based model, trained from scratch on 36,178 sentence pairs drawn primarily from biblical texts, achieves a BLEU score of 44.82 on the full corpus and 65.76

on the biblical domain alone. This performance gap highlights the impact of domain homogeneity and the limited generalization capacity of models trained from scratch on small and thematically narrow corpora.

2.3 Pre-trained multilingual models

The emergence of large multilingual language models has significantly reshaped machine translation. These models can be categorized according to their size and training strategy (Hussen et al., 2024). **Large Language Models (LLMs)**, typically comprising more than 7 billion parameters, offer strong generalization capabilities at the cost of high computational requirements. **Small Language Models (SLMs)**, with fewer than 7 billion parameters, prioritize computational efficiency. **Specialized Small Language Models (SSLMs)** adopt an intermediate strategy by focusing on specific domains, language families, or tasks, thereby maximizing effectiveness in targeted contexts. This distinction is particularly relevant for low-resource languages, where the trade-off between broad coverage and targeted specialization directly influences model design choices.

Our study compares three models representing these distinct strategies.

NLLB-200 (NLLB Team et al., 2022), developed by Meta AI, exemplifies a strategy of *massive generalization*. With 600 million parameters (distilled version), it supports over 200 languages, including 38 African languages, natively incorporates Mooré, and leverages back-translation to enhance learning for under-resourced languages. Released as open source, NLLB-200 serves as a strong baseline for evaluating the impact of native language support.

mBART50 (Liu et al., 2020) is an encoder–decoder model pre-trained via denoising autoencoding on 50 languages. It includes only three African languages and does not cover Mooré. Its comparable size and architecture allow for a controlled comparison with NLLB-200, particularly with respect to the effect of native language inclusion and transfer from typologically related languages.

AfroLM (Dossou et al., 2023) represents a *regionally specialized* SSLM, focusing exclusively on 23 African languages, including Mooré and Bambara. This encoder-only RoBERTa-based model employs active learning and synthetic data augmentation to improve representations under limited data conditions, thereby enhancing typological transfer.

Table 1 summarizes the coverage of African languages across these models.

Model	Params	African lan.	Mooré
mBERT	110M	6	No
XLM-R	270M	8	No
mBART50	610M	3	No
NLLB-200	600M	38	Yes
AfroLM	1B	23	Yes

Table 1: African language coverage of multilingual models

2.4 Positioning of our contribution

Our work differs from prior approaches along several key dimensions. Unlike Ouily et al. (2024), who train a model from scratch on a homogeneous biblical corpus, we leverage transfer learning through fine-tuning of pre-trained multilingual models. Our final dataset comprises 82,934 validated parallel sentence pairs, more than twice the size of that used by Ouily et al. It exhibits substantial domain diversity, including educational, administrative, religious, legal, and literary texts, with the novel inclusion of institutional government data.

We conduct a systematic comparative evaluation of three models representing distinct design strategies: NLLB-200, mBART50, and AfroLM. This setup enables an explicit investigation of two research hypotheses: the impact of native Mooré support during pre-training (H1) and the effectiveness of typological transfer from closely related languages (H2).

Finally, we adopt a multi-metric evaluation framework, including BLEU, SP-BLEU, chrF++, and ROUGE-L, tailored to the morphological characteristics of Mooré. This quantitative evaluation is complemented by qualitative translation analysis and the development of a web-based platform to support continuous user-driven evaluation.

3 Methodology

This section presents the methodological framework adopted in this work. Figure 1 illustrates the overall workflow, from corpus construction to model evaluation.

3.1 Parallel corpus construction

The French–Mooré parallel corpus was built by aggregating seven complementary sources, totaling

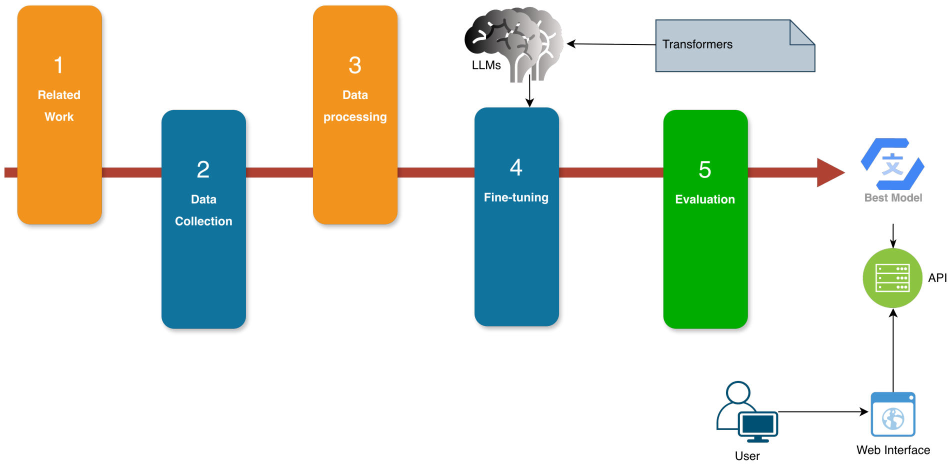

Figure 1: Overall methodological workflow, covering parallel corpus construction, pre-trained model adaptation, fine-tuning, and evaluation.

97,060 sentence pairs. Table 2 summarizes the sources and their respective contributions.

Source	Domain	Pairs	%
Fr–Mos Dictionary	Lexical	39,922	41.1
JW.org (Bible)	Religious	30,725	31.7
CITADEL Corpus	Educational	16,973	17.5
Masakhane	Mixed	5,559	5.7
Sig.gov.bf	Administrative	3,089	3.2
UDHR + Tales	Legal / Literary	752	0.8
Proverbs	Idiomatic	40	0.04
Total	**Multi-domain**	**97,060**	**100**

Table 2: Sources of the French–Mooré parallel corpus

This diversity contrasts sharply with previous work (Ouily et al., 2024), which relied on biblical texts for more than 90% of the data. The main contribution lies in the integration of institutional government data, which is rarely available for African languages.

Data collection relied on heterogeneous methods. Biblical texts were extracted using the JW.org API. Government documents were aligned position-wise. PDF files were processed using Gemini 2.5 Flash, with support from a French–English–Mooré dictionary. Existing corpora (CITADEL, Masakhane, UDHR) were integrated directly.

The raw corpus was processed using a five-step pipeline: (1) data cleaning and Unicode NFC normalization of Mooré diacritics; (2) semantic similarity computation using LaBSE; (3) filtering with a similarity threshold of 0.55; (4) human review of low-similarity pairs; (5) splitting into training, validation, and test sets (80/10/10).

The final corpus contains **82,934 sentence pairs** with high semantic consistency. The average sentence length is 16.8 words in French and 18.3 words in Mooré.

3.2 Research hypotheses

This study investigates two research hypotheses concerning the adaptation of multilingual models in low-resource context:

H1: Impact of native language support. We hypothesize that models with native support for Mooré during pre-training will achieve superior translation performance compared to models without such support. Specifically, NLLB-200, which includes Mooré in its pre-training data, is expected to outperform mBART50, which does not cover Mooré.

H2: Effectiveness of typological transfer. We hypothesize that models specialized in African languages will benefit from typological transfer, leveraging structural similarities between related languages. AfroLM, trained exclusively on 23 African languages including Mooré and typologically related languages such as Bambara, is expected to demonstrate effective transfer despite its smaller scale.

These hypotheses guide our experimental design and model selection, enabling a systematic comparison of distinct adaptation strategies for low-resource machine translation.

3.3 Model selection and adaptation

Three pre-trained models, representing distinct adaptation strategies, were selected for French–Mooré translation.

- **NLLB-200**: With native support for Mooré, the distilled version (`nllb-200-distilled-600M`) was chosen as the first model to be fine-tuned. This choice allows us to test the impact of native language support.

Preliminary analysis indicates satisfactory segmentation, with token-to-word ratios of 1.26 for French and 1.19 for Mooré. The model was therefore fine-tuned directly without modifying the tokenizer.

- **mBART50**: The mBART50 model does not support Mooré (Liu et al., 2020). This choice allows us to test the impact of native support by comparing fine-tuned mBART50 and fine-tuned NLLB.

  The tokenizer was extended by training a SentencePiece model on the Mooré corpus and merging it with the original vocabulary. This added 9,149 Mooré tokens, along with a dedicated language token (`mos_Latn`). The embedding matrix was resized, and the new embeddings were learned during fine-tuning. This adaptation substantially reduced Mooré over-segmentation.

- **AfroLM**: The AfroLM model, originally encoder-only (Dossou et al., 2023), was chosen to test the impact of typological transfer from specialized African language models.

  Two major adaptations were required. First, a bilingual SentencePiece tokenizer was trained and merged with the existing vocabulary. Second, an encoder–decoder architecture was constructed by pairing the AfroLM encoder with an XLM-RoBERTa decoder equipped with cross-attention. This design enables the exploitation of African-specific representations while supporting neural machine translation.

3.4 Fine-tuning protocol

Each model was fine-tuned in both translation directions (French→Mooré and Mooré→French) using a single bidirectional training setup. The translation direction is specified via a language token prepended to the input sequence.

All models were trained using AdamW, a linear learning rate scheduler with 500 warmup steps, an effective batch size of 32, and a maximum sequence length of 256 tokens. Early stopping with a patience of three epochs was applied. Fine-tuning was conducted on an NVIDIA A100 GPU.

3.5 Evaluation protocol

Evaluation was performed on a held-out test set of 8,294 sentence pairs, strictly reserved for final assessment. Sampling was random and stratified to preserve source distribution.

Four automatic metrics were used:

- **BLEU**: n-gram-based lexical precision;
- **SP-BLEU**: BLEU computed at the subword level;
- **chrF++**: character-level F-score;
- **ROUGE-L**: longest common subsequence.

4 Experimental Results

This section reports the performance of the models evaluated on our test corpus of 8,294 French–Mooré sentence pairs.

4.1 Overall performance

Table 3 reports the results for both translation directions across all evaluation metrics.

Fine-tuned NLLB achieves the highest scores across all metrics and both directions. Fine-tuned mBART50 obtains competitive results, with performance close to that of fine-tuned NLLB. Fine-tuned AfroLM yields lower scores across all metrics.

4.2 Directional asymmetry

Translation performance varies depending on the direction. The NLLB baseline exhibits a strong asymmetry between FR→MOS and MOS→FR. This gap is substantially reduced after fine-tuning. Fine-tuned mBART50 shows a smaller difference between the two directions. Fine-tuned AfroLM achieves comparable performance in both directions. A pronounced discrepancy between BLEU and SP-BLEU is observed for all models particularly the NLLB baseline.

4.3 Qualitative examples

Table 4 presents qualitative examples of translations produced by the three fine-tuned models for both translation directions. Each example shows the source sentence, the human reference translation, and the outputs from each model. The rightmost column indicates translation quality :

V (checkmark) denotes an exact match with the reference, ***** indicates minor variations or added information, and **X** marks outputs containing factual errors or significant divergences.

Model	Direction	BLEU	SP-BLEU	chrF++	ROUGE-L
NLLB Baseline	FR→MOS	9,37	18,85	28,18	26,72
	MOS→FR	2,75	3,75	18,10	11,42
NLLB Fine-tuned	FR→MOS	**34,70**	**47,78**	**53,78**	**46,84**
	MOS→FR	**31,76**	**33,45**	**52,35**	**41,25**
mBART50 Fine-tuned	FR→MOS	29,64	42,74	48,69	46,33
	MOS→FR	28,16	32,15	48,52	42,44
AfroLM Fine-tuned	FR→MOS	13,10	31,64	32,53	30,18
	MOS→FR	13,46	28,16	33,99	32,31

Table 3: Model performance on French ↔ Mooré translation. The best scores for each metric and direction are shown in **bold**.

Example 1: FR→MOS		
Source (FR)	*Mes clients continuent d'acheter mes poulets sans faire de problème.*	–
Reference	*Mam raadensã ket n daada m noosã ti koeeg ka be ye.*	–
NLLB fine-tuned	*Mam raadensã ket n daada m noosã ti koeeg ka be ye.*	V
mBART50 fine-tuned	*Mam raadensã ket n daada mam noosã ti yell ka be ye.*	*
AfroLM fine-tuned	*m raadengua ket n daada mam noosã ti koeeg ka ye.*	X
Example 2: MOS→FR		
Source (MOS)	*Y lamus daar yoom tus-a yi la piig la yiib sig-noy kiuug rasem-a tãabo.*	–
Reference	*C'était le jeudi 3 mai 2012.*	–
NLLB fine-tuned	*C'était le jeudi 3 mai 2012.*	V
mBART50 fine-tuned	*C'était le jeudi 3 mai 2012 à Ouagadougou.*	*
AfroLM fine-tuned	*C'était le vendredi 3 mai 2012 à Ouagadougou.*	X
V = exact match; * = added information; X = factual error		

Table 4: Qualitative examples of translations produced by the models.

For the FR→MOS example, fine-tuned NLLB exactly matches the reference. Fine-tuned mBART50 shows a minor lexical variation. Fine-tuned AfroLM exhibits several formal divergences. For the MOS→FR example, fine-tuned NLLB exactly matches the reference. Fine-tuned mBART50 introduces additional information not present in the source. Fine-tuned AfroLM produces a factual error.

4.4 Comparison with prior work

We compare our approach to the related work of Ouily et al. (2024) by evaluating our fine-tuned NLLB model on their test set. While their model achieves a BLEU score of 44.82 for French-to-Mooré translation, our fine-tuned NLLB model attains 46.91 BLEU, 55.64 SP-BLEU, and 69.86 ROUGE-L (FR→MOS), demonstrating consistent improvements over Ouily et al. (2024)'s results on their evaluation domain.

Beyond this single domain, our models also demonstrate strong performance on educational and administrative texts, confirming their ability to generalize across content types. When evaluated on our full multi-domain test set (our test set), which spans eight distinct domains, the fine-tuned NLLB model achieves a BLEU score of 34.70. These results demonstrate that our approach enables effective multi-domain generalization for French–Mooré translation.

5 Analysis and discussion

This section interprets the results and relates them to our research hypotheses.

5.1 Impact of native Mooré support

The results support hypothesis (H1). Fine-tuned NLLB consistently outperforms fine-tuned mBART50 in both translation directions and across all metrics.

However, the performance gap remains moderate. Fine-tuned mBART50 reaches approximately 85% of NLLB's BLEU score despite the absence of Mooré in its pre-training. This indicates that fine-tuning on a high-quality parallel corpus can largely compensate for the lack of native language support.

These results highlight the effectiveness of large multilingual models for low-resource machine translation, even when the target language is not explicitly covered during pre-training.

5.2 Typological transfer and African specialization

AfroLM, despite its specialization in African languages, obtains the lowest performance among the evaluated models.

This underperformance can be attributed to several factors. First, effective coverage of Mooré during pre-training remains limited, as reflected by tokenizer segmentation issues. Second, AfroLM relies on an encoder-only architecture adapted to sequence-to-sequence translation, which may be less suitable than native encoder–decoder models. Finally, the scale and diversity of its pre-training corpus are substantially smaller than those of NLLB and mBART50.

These results suggest that typological proximity alone is insufficient to ensure effective transfer without adequate model capacity and large-scale pre-training.

5.3 Directional asymmetry

The observed directional asymmetry reflects tokenization challenges for Mooré. Standard subword tokenizers, tend to over-segment words into multiple tokens, which particularly affects subword-level metrics such as SP-BLEU. This effect is most pronounced in the NLLB baseline and is substantially reduced after fine-tuning.

The reduction of asymmetry highlights the importance of task-specific adaptation for low-resource languages. In contrast, mBART50 exhibits more balanced performance across directions, likely due to symmetric bidirectional fine-tuning.

6 Conclusion

This paper addressed French–Mooré neural machine translation in a low-resource setting. We introduced a novel multi-domain parallel corpus, including institutional data rarely available for African languages.

Three modeling strategies were evaluated: a multilingual model with native support (NLLB-200), a multilingual model without native support (mBART50), and a specialized African model (AfroLM). Results show that fine-tuning large pre-trained models is the most effective approach. Fine-tuned NLLB achieves the best performance, while mBART50 reaches comparable results despite lacking native Mooré support.

These findings highlight that corpus quality and diversity are often more decisive than native language coverage. In contrast, linguistic specialization alone does not guarantee strong performance without sufficient model capacity and large-scale pre-training.

Directional analysis further shows that fine-tuning substantially reduces translation asymmetry. Overall, this work confirms the effectiveness of adapting existing multilingual models for low-resource African languages.

Future work will focus on human evaluation, testing on unseen domains, and lightweight fine-tuning methods. The web-based translation platform developed in this study also enables continuous, community-driven improvement.

Limitations

This work has a few limitations. Firstly, automatic evaluation metrics do not fully reflect translation quality. Human evaluation is required to assess fluency, adequacy, and cultural acceptability.

Secondly, generalization to unseen domains remains an open question. Although the test set is multi-domain, it follows the same distribution as the training data.

Finally, computational constraints limited extensive hyperparameter exploration and prevented experiments with very large proprietary LLMs such as GPT-4 or Claude.

Acknowledgements

We thank Centre d'Excellence Interdisciplinaire en Intelligence Artificielle pour le Developpement (CITADEL) for supporting this work.

References

Željko Agić and Ivan Vulić. 2019. JW300: A wide-coverage parallel corpus for low-resource languages. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 3204–3210, Florence, Italy. Association for Computational Linguistics.

Roee Aharoni, Melvin Johnson, and Orhan Firat. 2019. Massively multilingual neural machine translation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 3874–3884, Minneapolis, Minnesota. Association for Computational Linguistics.

COCOON Database. 2024. Cocoon: Informations linguistiques sur la langue mossi.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 8440–8451, Online. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Marcel Diki-Kidiri. 2008. *Langues africaines et sociétés*. Éditions Harmattan, Paris, France.

Bonaventure F. P. Dossou, Atnafu Lambebo Tonja, Chris Chinenye Emezue, Miehleketo Sibanda, Kelechi Ogueji, Jonathan Munkoh-Buabeng, David Ifeoluwa Adelani, Shruti Rijhwani, Jesujoba O. Palen-Michel, Catherine Leong, and Eleanor Chodroff. 2023. AfroLM: A self-active learning-based multilingual pretrained language model for 23 african languages. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 15529–15543, Singapore. Association for Computational Linguistics.

Kedir Yassin Hussen, Michael Woldeyohannis, Seid Muhie Jabesa, Amanuel Mersha Gebremichael, Ibrahim Said Mohammed, Abinew Ali Yousuf, and Seid Muhie Yimam. 2024. The state of large language models for african languages: Progress and challenges. *arXiv preprint arXiv:2412.12485*.

Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. 2020. Multilingual denoising pre-training for neural machine translation. *Transactions of the Association for Computational Linguistics*, 8:726–742.

Laura Martinus and Jade Z. Abbott. 2020. Vukuzenzele and IsiZulu: Introducing two new castilian-South African machine translation datasets. In *Proceedings of the Fourth Widening NLP Workshop*, pages 146–148, Seattle, USA. Association for Computational Linguistics.

Wilhelmina Nekoto, Vukosi Marivate, Tshinondiwa Matsila, Timi Fasubaa, Tajudeen Kolawole, Taiwo Fagbohungbe, Solomon Oluwole Akinola, Shamsuddeen Hassan Muhammad, Salomey Kabongo Kabenamualu, Salomey Osei, and 1 others. 2020. Participatory research for low-resourced machine translation: A case study in African languages. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2144–2160, Online. Association for Computational Linguistics.

Graham Neubig and Junjie Hu. 2018. Rapid adaptation of neural machine translation to new languages. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 875–880, Brussels, Belgium. Association for Computational Linguistics.

NLLB Team, Marta R. Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, Anna Sun, Skyler Wang, Guillaume Wenzek, Al Youngblood, Bapi Akula, Loic Barrault, Gabriel Mejia Gonzalez, Prangthip Hansanti, and 20 others. 2022. No language left behind: Scaling human-centered machine translation. *arXiv preprint arXiv:2207.04672*.

Maimouna Ouattara, Abdoul Kader Kaboré, Jacques Klein, and Tegawendé F. Bissyandé. 2025. Bridging literacy gaps in African informal business management with low-resource conversational agents. In *Proceedings of the First Workshop on Language Models for Low-Resource Languages*, pages 193–203, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Hamed Joseph Ouily, Aminata Sabane, Delwende Eliane Birba, Rodrique Kafando, Abdoul Kader Kabore, and Tégawendé F. Bissyandé. 2024. Neural machine translation for mooré, a low-resource language. In *Proceedings of JRI 2023*, Ouagadougou, Burkina Faso.

Fatoumata Ouoba Kabore. 2025. Development of a rule-based morphological analyzer for mooré using apertium lttoolbox. Master's thesis in language technology, Uppsala University, Department of Linguistics and Philology, Uppsala, Sweden, June.

Stephen Smith and Claire Levasseur. 2009. *La Diversité Linguistique en Afrique*. Éditions Universitaires, Paris, France.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 30, pages 5998–6008.

Barret Zoph, Deniz Yuret, Jonathan May, and Kevin Knight. 2016. Transfer learning for low-resource neural machine translation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1568–1575, Austin, Texas. Association for Computational Linguistics.