Context-Aware Neural Machine Translation of English Numerical Expressions to Yorùbá: A Fine-Tuning Approach for Financial and General Domains

Anonymous authors
Paper under double-blind review

Abstract

The increasing prevalence of digital financial services across Africa and emerging markets has created an urgent demand for accurate, context-aware machine translation systems capable of handling specialized financial terminology and numerical expressions, including currencies, percentages, interest rates, and monetary values. This research presents a novel approach to financial machine translation that addresses the unique challenges of translating numerical content in low-resource African languages. We develop a specialized dataset focused on financial numerical expressions for English-Yorùbá translation and conduct comprehensive evaluations across multiple state-of-the-art machine translation models. Our methodology emphasizes context-sensitive translation of numbers within financial documents, recognizing that numerical accuracy is paramount for financial comprehension and decision-making. Experimental results demonstrate that the NLLB-3.3B parameter model achieves superior performance across all evaluation metrics, recording BLEU scores of 44.50, CHRF++ scores of 74.52, and Africomet scores of 0.75, while maintaining the lowest Translation Edit Rate (TER) of 34.58, indicating minimal post-editing requirements. These findings suggest significant improvements in translation quality for financial numerical content. This work contributes to the advancement of multilingual financial AI systems for African languages, supporting broader financial inclusion initiatives and enabling more effective cross-lingual financial communication. The research addresses a critical gap in specialized machine translation for financial domains in low-resource language settings.

1 Introduction

The increasing adoption of digital financial services across Africa and other emerging markets has created a critical need for accurate and context-sensitive machine translation systems capable of handling financial terminology, particularly numerical expressions such as currencies, percentages, interest rates, and large financial figures. While recent advancements in multilingual Natural Language Processing (NLP) — including models like No Language Left Behind (NLLB) (Team et al., 2022), mT5 (Xue et al., 2021), and AfriTeVa (Jude Ogundepo et al., 2022) — have significantly improved general-purpose translation across diverse languages, challenges remain in domain-specific translation tasks involving numerically intensive financial content, especially for low-resource languages.

Financial documents, transaction records, and customer communications often rely heavily on numerical expressions, which require accurate linguistic adaptation (Jørgensen et al., 2023). Unlike general translation tasks, financial language presents unique challenges due to its dependence on numerical precision, structured formatting, and culturally contextual representations of monetary values. In many African languages, including Yoruba, numbers follow distinct verbalization rules that are linguistically and contextually complex (Elesemoyo & Odejobi, 2022). These structures differ significantly from those in English and other high-resource languages, posing difficulties for existing machine translation systems.

However, most current translation models do not explicitly address the complexities of monetary formats, local verbalization rules, or culturally specific methods of expressing financial information. As noted by Elesemoyo & Odejobi (2022), the translation of numbers is not a straightforward word-to-word substitution; it requires nuanced understanding of syntax, morphology, and cultural context. Structural disparities in numerical systems, variations in currency representation, and context-driven articulation further complicate translation tasks in this domain.

To address these gaps, this paper introduces a financial translation methodology aimed at improving the translation of financial expressions in low-resource languages. Our key contributions are as follows: (1) the development of a manually curated financial numerical translation dataset specifically targeting Yoruba; (2) the fine-tuning of state-of-the-art multilingual translation models on this dataset; and (3) a comprehensive evaluation of model performance across diverse financial categories. This research not only advances machine translation for underrepresented languages but also contributes toward enhancing financial inclusion and multilingual accessibility across African markets.

2 Literature Review

Accurate translation of numerical expressions in financial texts is critical—not only for linguistic precision but also for financial reporting, regulatory compliance, and digital banking services (Tang et al., 2025). Despite the progress in neural machine translation (NMT) powered by large language models, challenges persist in adapting numerical formats, currency expressions, and domain-specific terminologies to different linguistic and cultural frameworks (Alghamdi et al., 2023). This section reviews the key issues in financial numerical translation, existing solutions for addressing these challenges, and domain adaptation strategies relevant to machine translation.

2.1 Numerical Translation Challenges in Financial Texts

Numerical translation presents distinct challenges from general text translation due to the requirement for semantic accuracy and mathematical consistency. In the financial domain, even minor mistranslations can result in significant miscommunication or financial loss (Tang et al., 2025). Prior studies have identified three major challenges:

1. Unit conversion errors: Differences in number representation across languages, such as the English ”one billion” versus the Chinese ”ten hundred million,” can lead to translation inconsistencies and severe numerical misinterpretation (Tang et al., 2025).

2. Contextual ambiguity: Financial terms are often embedded within dynamic contexts—interest rates (e.g., 5%), exchange rates (e.g., ₦500/USD), or debt notations (e.g., $1.5M). These require context-aware translation strategies (Alghamdi et al., 2023).

3. Linguistic divergence in numerical systems: Languages such as Yoruba follow subtractive numeral systems, in contrast to the additive systems used in English. This linguistic complexity influences how numbers and currency are verbalized and interpreted (Elesemoyo & Odejobi, 2022).

2.2 Addressing Numerical Translation Errors in Financial Texts

Several approaches have been proposed to mitigate numerical mistranslations in financial texts. Wang et al. (2021) introduce a framework that incorporates the following strategies:

- Separating numerical tokens from text, applying rule-based translation, and reinserting them post-processing.
- Leveraging financially annotated parallel corpora to train NMT systems for improved accuracy.
- Evaluating translation quality using custom test cases designed specifically for numerical content, rather than relying solely on general metrics like BLEU.

2.3 Multilingual Machine Translation for African Languages

The growing interest in building translation systems for African languages has led to the emergence of community-driven and large-scale efforts aimed at bridging the digital language divide. The Masakhane project (Adelani et al., 2021) introduced a grassroots approach to multilingual machine translation for over 30 African languages, highlighting the value of participatory research and dataset curation. While impactful, most Masakhane models are trained on general-domain corpora, which may not capture domain-specific language use, such as financial terminology.

Meta AI's No Language Left Behind (NLLB) project (Team et al., 2022) represents a significant step forward by training a single model capable of translating across 200+ languages, including several African ones. However, NLLB's zero-shot performance on domain-specific content, particularly numerically dense financial text, remains underexplored. Additionally, models like AfriTeVa (Jude Ogundepo et al., 2022) focus on task transfer and text generation for African languages, though their financial translation capabilities have not been evaluated.

Our work complements these efforts by introducing a domain-specific dataset and evaluation pipeline that targets financial numerical translation in Yoruba — a task that existing multilingual models struggle with due to lack of training data and linguistic nuance. This work builds on the momentum created by prior African MT efforts but distinguishes itself through its focus on structured numerical translation in real-world financial applications.

2.4 Summary of Gaps in the Literature

Despite growing interest in machine translation for low-resource settings, several critical gaps remain in the domain of financial numerical translation. First, state-of-the-art NMT systems continue to struggle with adapting to the specialized numeric structures and expressions that characterize financial discourse (Tang et al., 2025). These models typically lack mechanisms to distinguish between general numerical references and those that carry legal, monetary, or regulatory significance.

Secondly, there is a noticeable scarcity of multilingual datasets tailored to financial contexts in African languages. Resources for translating concepts such as interest rates, currency amounts, and transactional descriptions remain largely unavailable, limiting both research and practical applications in this space (Alghamdi et al., 2023).

Finally, current MT systems often yield translation errors when applied to regulatory and financial documents. Such errors, particularly in low-resource languages like Yoruba, can stem from a lack of culturally contextualized linguistic data and an inability to accurately represent subtractive or localized numeral formats (Elesemoyo & Odejobi, 2022).

To address these challenges, this paper introduces a domain-specific dataset focused on financial numerical translation from English to Yoruba. We investigate how fine-tuning multilingual NMT models—such as mT5, NLLB, mT0, and AfriTeVa—on this dataset can enhance translation accuracy in a low-resource setting. By benchmarking performance across different model families and financial expression types, we contribute new insights to the advancement of financial AI applications in underrepresented languages.

3 Methodology

This section provides an overview of the methods and algorithms for domain adaptation in Yoruba machine translation utilising transformer models. Initially, we provide information regarding the collected bilingual dataset. Subsequently, we present our approach, followed by a description of the metrics employed for evaluation.

3.1 Dataset and Translation

We manually translated 5,000 financial terms from English to Yoruba to create a comprehensive dataset for financial numerical translation. The sentences in the dataset comprises of examples such as:

1. Cardinal numerals (e.g., ”one thousand two hundred”).

Table 1: A table showing Models Explored

Sentences	Synonyms
The loan was approved for one million Naira	The bank sanctioned a credit of a million Naira.
The exchange rate for USD to NGN is 750 naira per dollar	The conversion rate for the US dollar to naira stands at 750 naira per USD.
The bank offers a simple interest rate of 5% per annum	The financial institution provides a flat interest rate of 5% yearly.

Table 2: A table showing Models Explored

Models	Description
mT5 (Xue et al., 2021)	A massively multilingual version of T5, trained on 101 languages, effective for low-resource translation tasks.
mT0 (Muennighoff et al., 2023)	A Multitask prompted finetuning (MTF) variant of mT5..
AfriTeva Large (Jude Ogundepo et al., 2022)	a family of sequence-to-sequence models derived from T5 that are pretrained on 10 African languages from scratch.
AfriTeva-v2 Large (Jude Ogundepo et al., 2022)	An improved version of AfriTeva, designed for better domain adaptation in financial and legal translations.
NLLB (Team et al., 2022)	A Facebook AI model trained on 200+ languages, optimized for low-resource language translation.

2. Monetary expressions (e.g., ”$1.5 million”).
3. Interest rates and percentages (e.g., ”5.25% interest rate
4. Large numerical values (e.g., ”three billion”).
5. Exchange rate representations (e.g., ”₦750 for each USD”).

However, due to constraints on the size of the data set, data set, we enhanced the dataset through synonym substitution and data duplication techniques.

3.2 Data Augmentation Strategy

To enhance the diversity and volume of the data set, we implemented automated data augmentation through WordNet-based synonym substitution (Sun et al., 2011)for English words. This approach preserves numerical precision while ensuring semantic variety in source sequences. We created a method to randomly substitute words in English text with their synonyms, with a 10% likelihood of substitution for each term. we made sure replacement are unique and semantically significant while terms without appropriate synonyms stay unchanged, as shown in Table[1]:

In order to have equal length for our parallel data, we created the duplicate of the 5000 Yoruba translated sentences to make up for the 10000 English sentences.

3.3 Model Traning Setup

We fine-tuned multiple state-of-the-art model machine translation models on our dataset to develop a domain-specific financial machine translation system. The selected models comprise general-purpose, multilingual, and Africa-centric language models.

3.4 Model Training Strategy

To check the performance of existing models in financial numerical translation, we fine-tuned five machine translation models (mT5, mT0, AfriTeva, AfriTeva-v2 Large, NLLB). The training process includes tokenization, hyperparameter-tuning, optimisation strategies, and early stopping techniques, etc.

3.4.1 Metrics

For evaluation, we implemented the BLEU score (PAPINENI, 2001), which makes use of a SentencePiece tokenizer, along with chrF (Popović, 2015) and TER (Snover et al., 2006). In addition, we calculate AfriCOMET (Wang et al., 2024), a recently introduced evaluation metric that contains evaluation data with human ratings for under-resourced languages.

4 Result and Discussion

This section compares the evaluation outcomes of existing baseline models and fine-tuned models for financial numerical translation. The models investigated include mT5, mT0, AfriTeva Large, AfriTeva-v2 Large, and NLLB. The evaluation process used BLEU, CHRF++, TER, and AfriCOMET scores to determine translation quality, numerical precision, and domain adaption.

4.1 Evaluation of Existing Models Before Fine-Tuning

Prior to fine-tuning, the baseline models (mT5 and mT0) performed poorly in all metrics, indicating that their translations required considerable manual modifications. Similarly, mT0 performed marginally better but still struggled with financial numerical translations. These findings indicate that general-purpose models are not well suited for financial numerical translation without domain-specific adaptations. While some models fared slightly higher.

4.2 Performance of Fine-Tuned Models

Fine-tuning improves translation performance, especially for low-resource language models. NLLB performed best among fine-tuned models, scoring highest in BLEU (44.50), CHRF++ (74.52), and AfriCOMET (0.75) and lowest in TER (34.58), indicating minimum post-editing.

AfriTeva-v2 Large had higher BLEU and CHRF++ ratings than AfriTeva Large, demonstrating that fine-tuning improved its financial numerical translation performance. However, both AfriTeva models had higher TER ratings than NLLB, requiring more human modifications.

Without fine-tuning, mT5 and mT0 struggle with financial words, currency values, and numerical formatting. These findings recommend changing translation models for domain-specific applications, especially in low-resource African languages.

Fine-tuning improved NLLB and AfriTeva models, proving domain-specific training works. To further reduce TER scores across all models, number tokenisation, decimal handling, and currency formatting must be improved.

Table 3: Model Performance Evaluation

Models	BLEU Score ↑	CHRF++ Score ↑	TER Score ↓	AfriCOMET Score ↑
mT5	0.00	0.09	100.00	0.14
mT0	23.50	60.91	52.99	0.70
AfriTeva Large	10.46	44.88	68.80	0.55
AfriTeva-v2 Large	18.78	53.99	55.87	0.68
NLLB-3.3b	**44.50**	**74.52**	**34.58**	**0.75**

4.3 Common Translation Errors and Error Analysis

Although NLLB performed well across all standard metrics, qualitative analysis of translations revealed recurring errors, particularly in currency formatting, large-number translations, and decimal placement. Some common errors include misrepresentation of exchange rates, omission of decimal values, and incorrect rounding of large numbers. These errors highlight areas where further fine-tuning or hybrid translation approaches could enhance accuracy in financial translation tasks.

5 Methodology

This section outlines the methods and algorithms employed to adapt machine translation models to the financial domain in a low-resource setting, focusing specifically on English-to-Yoruba translation. We begin by describing our dataset and data augmentation strategy, followed by the model fine-tuning procedure and evaluation framework.

5.1 Dataset and Translation

We developed a parallel dataset comprising 5,000 English-Yoruba sentence pairs, manually translated to cover a range of financial numerical expressions. These include cardinal numerals (e.g., “one thousand two hundred”), monetary expressions (e.g., “$1.5 million”), interest rates (e.g., “5.25% interest rate”), large numbers (e.g., “three billion”), and exchange rate formats (e.g., “₦750 for each USD”).

Due to the relatively limited dataset size, we implemented a data augmentation pipeline to expand coverage and improve model generalization.

5.2 Data Augmentation Strategy

To enhance the dataset’s diversity, we applied WordNet-based synonym substitution for English source sequences, inspired by prior work in low-resource NLP (Sun et al., 2011). In this process, each token in the English sentence had a 10% probability of being replaced with a semantically appropriate synonym. Only one unique synonym per word was allowed, and words without suitable substitutions were left unchanged to maintain semantic integrity and numerical precision. Table **??** provides examples of augmented sentences.

In addition, to balance the source and target lengths for model compatibility, each Yoruba translation was duplicated once, resulting in a total of 10,000 aligned sentence pairs.

5.3 Model Fine-tuning Setup

We fine-tuned five state-of-the-art transformer-based machine translation models to evaluate their effectiveness in translating financial text into Yoruba. These include:

- **mT5** – a multilingual text-to-text transformer trained on diverse languages.
- **mT0** – a zero-shot translation model built on T5 and prompted with task instructions.
- **AfriTeVa** and **AfriTeVa-v2 Large** – pretrained models for African languages with encoder-decoder architecture.
- **NLLB (600M variant)** – Meta AI’s No Language Left Behind model, trained on over 200 languages.

All models were trained on our parallel dataset using early stopping, appropriate batch sizes, and standard tokenization techniques. The SentencePiece tokenizer was employed for models requiring subword segmentation.

5.4 Evaluation Metrics

We used multiple automatic evaluation metrics to assess translation quality. These include BLEU (PAPINENI, 2001), which measures n-gram precision; chrF (Popović, 2015), which considers character-level F-scores; and TER (Snover et al., 2006), which evaluates edit distance from reference translations.

Additionally, we incorporated AfriCOMET (Wang et al., 2024), a recently introduced metric calibrated for African languages using human-aligned evaluation data. This allows us to assess not only surface-level accuracy but also semantic adequacy for under-resourced contexts.

6 Conclusion and Future Work

This study evaluated the impact of fine-tuning neural machine translation models for financial numerical translation in low-resource languages, using Yoruba as a case study. The results show that fine-tuned models significantly outperformed baseline models, particularly in handling financial terms, currency values, and numerical expressions.

Among the fine-tuned models, NLLB achieved the best overall performance, with the highest BLEU (44.50), CHRF++ (74.52), and AfriCOMET (0.75) scores, while also maintaining the lowest TER (34.58), indicating minimal post-editing requirements. AfriTeva-v2 Large showed notable improvements over AfriTeva Large, confirming that domain-specific adaptation enhances financial translation accuracy. Conversely, mT5 and mT0 struggled with numerical representation, reinforcing the need for specialized models in financial applications.

Future work should focus on improving numerical tokenization and formatting to reduce errors in large numbers, decimals, and currency expressions. Expanding domain-specific financial datasets, particularly with real-world banking and transaction data, will further enhance model performance.

Exploring hybrid neural-symbolic approaches can improve numerical consistency and reduce post-editing needs. Additionally, fine-tuning these models for other African languages will increase their applicability in financial services. Finally, integrating these models into real-world financial applications, such as multilingual banking systems and fintech solutions, will further support financial inclusion.

This research highlights the importance of AI-driven financial translation for low-resource languages and provides a foundation for future advancements in financial communication and automation.

References

David Ifeoluwa Adelani, Jesujoba Alabi, Alexis Tapo, and Others. Masakhane-mt: A grassroots approach to multilingual machine translation for african languages. In *Transactions of the Association for Computational Linguistics (TACL)*, volume 9, pp. 1116–1131, 2021.

Emad A. Alghamdi, Jezia Zakraoui, and Fares A. Abanmy. Domain adaptation for arabic machine translation: The case of financial texts, 2023. URL `https://arxiv.org/abs/2309.12863`.

Isaac O. Elesemoyo and Odetunji A. Odejobi. Machine translation system for numeral in english text to yorùbá language. *Ife Journal of Information and Communication Technology*, 6, 2022. ISSN 2645-2413.

Odunayo Jude Ogundepo, Akintunde Oladipo, Mofetoluwa Adeyemi, Kelechi Ogueji, and Jimmy Lin. AfriTeVA: Extending ?small data? pretraining approaches to sequence-to-sequence models. In Colin Cherry, Angela Fan, George Foster, Gholamreza (Reza) Haffari, Shahram Khadivi, Nanyun (Violet) Peng, Xiang Ren, Ehsan Shareghi, and Swabha Swayamdipta (eds.), *Proceedings of the Third Workshop on Deep Learning for Low-Resource Natural Language Processing*, pp. 126–135, Hybrid, July 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.deeplo-1.14. URL `https://aclanthology.org/2022.deeplo-1.14/`.

Rasmus Kær Jørgensen, Oliver Brandt, Mareike Hartmann, Xiang Dai, Christian Igel, and Desmond Elliott. MULTIFIN: A dataset for multilingual financial NLP. In *Findings of the Association for Computational Linguistics: EACL 2023*, pp. 894–909. Association for Computational Linguistics, May 2–6 2023. URL `https://aclanthology.org/2023.findings-eacl.66.pdf`.

Niklas Muennighoff, Thomas Wang, Lintang Sutawika, Adam Roberts, Stella Biderman, Teven Le Scao, M Saiful Bari, Sheng Shen, Zheng-Xin Yong, Hailey Schoelkopf, Xiangru Tang, Dragomir Radev, Alham Fikri Aji, Khalid Almubarak, Samuel Albanie, Zaid Alyafeai, Albert Webson, Edward Raff, and Colin Raffel. Crosslingual generalization through multitask finetuning, 2023. URL `https://arxiv.org/abs/2211.01786`.

K. PAPINENI. Bleu : a method for automatic evaluation of mt. *Research Report, Computer Science RC22176 (W0109-022)*, 2001. URL `https://cir.nii.ac.jp/crid/1573387449944584832`.

Maja Popović. chrF: character n-gram F-score for automatic MT evaluation. In Ondřej Bojar, Rajan Chatterjee, Christian Federmann, Barry Haddow, Chris Hokamp, Matthias Huck, Varvara Logacheva, and Pavel Pecina (eds.), *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pp. 392–395, Lisbon, Portugal, September 2015. Association for Computational Linguistics. doi: 10.18653/v1/W15-3049. URL `https://aclanthology.org/W15-3049/`.

Matthew Snover, Bonnie Dorr, Rich Schwartz, Linnea Micciulla, and John Makhoul. A study of translation edit rate with targeted human annotation. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers*, pp. 223–231, Cambridge, Massachusetts, USA, August 8-12 2006. Association for Machine Translation in the Americas. URL `https://aclanthology.org/2006.amta-papers.25/`.

Koun-Tem Sun, Yueh-Min Huang, and Ming-Chi Liu. A wordnet-based near-synonyms and similar-looking word learning system. *Educational Technology Society*, 14:121–134, 01 2011.

Wei Tang, Jiawei Yu, Yuang Li, Yanqing Zhao, Weidong Zhang, Wei Feng, Min Zhang, and Hao Yang. Investigating numerical translation with large language models, 2025. URL `https://arxiv.org/abs/2501.04927`.

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, John Hoffman, Semarley Jarrett, Kaushik Ram Sadagopan, Dirk Rowe, Shannon Spruit, Chau Tran, Pierre Andrews, Necip Fazil Ayan, Shruti Bhosale, Sergey Edunov, Angela Fan, Cynthia Gao, Vedanuj Goswami, Francisco Guzmán, Philipp Koehn, Alexandre Mourachko, Christophe Ropers, Safiyyah Saleem, Holger Schwenk, and Jeff Wang. No language left behind: Scaling human-centered machine translation, 2022. URL `https://arxiv.org/abs/2207.04672`.

Jiayi Wang, David Ifeoluwa Adelani, Sweta Agrawal, Marek Masiak, Ricardo Rei, Eleftheria Briakou, Marine Carpuat, Xuanli He, Sofia Bourhim, Andiswa Bukula, Muhidin Mohamed, Temitayo Olatoye, Tosin Adewumi, Hamam Mokayed, Christine Mwase, Wangui Kimotho, Foutse Yuehgoh, Anuoluwapo Aremu, Jessica Ojo, Shamsuddeen Hassan Muhammad, Salomey Osei, Abdul-Hakeem Omotayo, Chiamaka Chukwuneke, Perez Ogayo, Oumaima Hourrane, Salma El Anigri, Lolwethu Ndolela, Thabiso Mangwana, Shafie Abdi Mohamed, Ayinde Hassan, Oluwabusayo Olufunke Awoyomi, Lama Alkhaled, Sana Al-Azzawi, Naome A. Etori, Millicent Ochieng, Clemencia Siro, Samuel Njoroge, Eric Muchiri, Wangari Kimotho, Lyse Naomi Wamba Momo, Daud Abolade, Simbiat Ajao, Iyanuoluwa Shode, Ricky Macharm, Ruqayya Nasir Iro, Saheed S. Abdullahi, Stephen E. Moore, Bernard Opoku, Zainab Akinjobi, Abeeb Afolabi, Nnaemeka Obiefuna, Onyekachi Raphael Ogbu, Sam Brian, Verrah Akinyi Otiende, Chinedu Emmanuel Mbonu, Sakayo Toadoum Sari, Yao Lu, and Pontus Stenetorp. Afrimte and africomet: Enhancing comet to embrace under-resourced african languages, 2024. URL `https://arxiv.org/abs/2311.09828`.

Jun Wang, Chang Xu, Francisco Guzman, Ahmed El-Kishky, Benjamin I. P. Rubinstein, and Trevor Cohn. As easy as 1, 2, 3: Behavioural testing of nmt systems for numerical translation, 2021. URL `https://arxiv.org/abs/2107.08357`.

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. mt5: A massively multilingual pre-trained text-to-text transformer, 2021. URL `https://arxiv.org/abs/2010.11934`.