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

The accurate translation of numerical expressions poses significant challenges in machine translation, particularly for low-resource languages with complex numeral systems like Yorùbá. Existing models often struggle with context-dependent number translation, where the same English numeral requires different Yorùbá forms based on grammatical and semantic context. This limitation is especially pronounced in financial domains where numerical accuracy is critical for accessibility and comprehension.

This research addresses the contextual translation of English numerical expressions to Yorùbá through fine-tuning of state-of-the-art multilingual models including No Language Left Behind (NLLB), mT5, Afriteva-small, etc. Unlike previous rule-based approaches that handle numbers in isolation, our neural approach processes complete sentences to determine appropriate Yorùbá number forms (adjectival, adverbial, ordinal, and all-forms) based on surrounding context.

We leverage a curated parallel corpus of 5,000+ English-Yorùbá sentence pairs containing numerical expressions across financial and general domains. Our methodology employs multi-task learning to simultaneously perform sentence translation, numerical context classification, and appropriate Yorùbá form selection. Data augmentation techniques including template-based generation and back-translation address the data scarcity challenges inherent in low-resource language processing.

Preliminary evaluation on validation datasets demonstrates improvements over rule-based baselines, with particular strength in handling quantity expressions, temporal references, and financial values. The neural models show better generalization to unseen number-context combinations, addressing key limitations of existing approaches that achieve only 53% accuracy on complex numerical contexts.

This work contributes to bridging the digital language gap in financial literacy and numerical comprehension for Yorùbá-speaking communities. By providing contextually accurate numerical translations, our research supports the development of inclusive financial systems and educational tools while advancing the broader field of neural machine translation for African languages.

Keywords: Machine Translation, Numerical Processing, Yorùbá Language, Context-Aware Translation, Fine-Tuning, Low-Resource Languages, Financial Domain