

Dialect-Aware Neural Models for Low-Resource African Languages: A Case Study on Igbo

Low-resource languages often exhibit rich dialectal variation, yet most natural language systems treat them as monolithic, resulting in reduced accuracy and inclusivity. Igbo, spoken by over 30 million people across southeastern Nigeria, presents a diverse dialect landscape spanning multiple states and regions. Existing NLP and speech systems, primarily trained on Standard Igbo, often fail to recognize dialect-specific lexical and phonological patterns, leading to degraded performance for speakers of non-standard dialects.

We propose a dialect-aware framework that identifies Igbo dialects in both text and speech and adapts system behavior accordingly. For text inputs, transformer-based classifiers (AfroXLM-R) capture lexical and orthographic patterns across dialects. For speech inputs, wav2vec/XLS-R embeddings are processed by lightweight neural classifiers to detect dialect-specific phonological cues. The system also generates interpretable explanations, highlighting the lexical and phonetic features driving its predictions.

An adaptive routing mechanism dynamically selects dialect-specific models for ASR, NLU, and TTS components, improving accuracy and user experience. Experiments on a multi-dialect Igbo dataset demonstrate effective dialect discrimination and transparent explanation generation.

This work applies deep learning methods to language and speech applications while addressing inclusivity and interpretability for speakers of diverse dialects. It contributes to inclusive NLP for low-resource languages, establishing a novel approach for dialect identification, explanation, and adaptive system behavior, and provides a foundation for extending dialect-aware machine learning methods to other linguistically diverse, low-resource languages.