Yoruba Language Code-Mixing Project

Duration: 5th April – 14th June 2024

Overview:

This project delves into the fascinating phenomenon of code-mixing, particularly between the Yoruba and English languages, exploring how bilingual speakers seamlessly blend words from both languages to create meaningful and contextually appropriate sentences.

Objectives:

- To understand the dynamics of code-mixing in a multilingual society.
- To analyze the cultural and linguistic implications of code-mixing in bilingual communication.
- To create a robust dataset that can be used for further research and development in language processing and AI applications.

Key Contributions:

- **Dataset Creation:** Developed a comprehensive dataset consisting of sentences where Yoruba and English words were artfully mixed. This dataset serves as a valuable resource for studying the syntax, semantics, and pragmatics of code-mixing in bilingual communication.
- **Linguistic Analysis:** Conducted an in-depth analysis of the linguistic patterns observed in the dataset, focusing on how speakers navigate between the two languages. The analysis provided insights into the syntactic structures, word choices, and cultural references that are prevalent in code-mixed speech.
- **Cultural Implications:** Explored the cultural significance of code-mixing in Yoruba-English bilingual communities, highlighting how language serves as a bridge between different cultural identities. The project underscored the adaptability of language in reflecting the lived experiences of multilingual speakers.
- **AI and Language Learning:** The findings from this project have potential applications in the development of AI-driven language processing tools that can better understand and respond to code-mixed language. Additionally, the insights gained can be leveraged to enhance language learning methodologies, making them more relevant to bilingual or multilingual learners.
- **Collaboration:** Collaborated closely with Native Yoruba Language Instructors to ensure the accuracy and cultural relevance of the code-mixed sentences. Their expertise was instrumental in validating the linguistic patterns and ensuring that the project outcomes were reflective of real-world language use.

Outcomes:

This project not only contributed to the academic understanding of code-mixing but also provided practical insights that can be applied in technology, education, and communication. The dataset and findings are available for public access and can serve as a foundation for future research in this area.