The Yoruba Dictionary: Preserving Language Via Voice Translation

Muyi Aghedo
CPSC 490
Advisors: Jay Lim, Oluseye Adesola

Background

Yoruba is a West African language spoken by about 47 million people worldwide. Its origins lie with the Yoruba people, mainly residing in Nigeria, but the language dialectically spans multiple countries, including Benin, Togo, Ghana, Sierra Leone, Brazil, and The Gambia. It may come as a surprise that there are very few online resources that can be used as tools to teach or record the language. A quick Google search for a 'Yoruba dictionary' reveals a few sites such as `https://www.yorubadictionary.com/`, `https://en.glosbe.com/`, or `https://www.lexilogos.com/english/yoruba_dictionary.htm`, which act as an interface to Google Translate or other translated Yoruba books. These resources date from 1843 to 2016. Although these resources are beneficial, they fall short in key areas that my senior project intends to address.

Goals

The goals of this project are as follows:

1. Create the Yoruba Dictionary, a mobile accessible and maintainable web application to facilitate learning the Yoruba language, based on the formal vocabulary available in Yale's Yoruba language curriculum.

2. Enhance the Yoruba database with example sentences and vocabulary by incorporating formal texts such as the Yoruba Bible, *The Body in Yoruba* (2006), and *Yoruba Animal Metaphors* (2005).

3. Implement voice-to-text software to aid in the translation of user-inputted Yoruba words.

Problem

The main need for this project lies in the preservation of the Yoruba language. Currently, Yoruba is losing both speakers and vocabulary over time as words are borrowed and replaced

by cognates from other languages. A site like this could both help people learn and understand Yoruba, thus addressing the speaker decline, and help preserve an online record of the language's vocabulary.

Current resources, however, have several issues:

- **Incorrect Information**: Many resources, such as Google Translate, may accurately translate individual words but fail when it comes to transitional phrases or groups of words with nonliteral meanings in Yoruba. They often lack grammatical context or provide mistranslations.

- **Incapability of Dealing with Tonal Homonyms**: Yoruba is a tonal language in which letters may have tonal counterparts that change meanings. The lack of support for tonal distinctions in many translation tools presents a major problem in accurately translating Yoruba.

- **Maintenance Issues**: Existing resources, such as Google Translate and Glosbe, have not been updated in over a decade. In contrast, my project intends to provide a continually maintained resource that can evolve over time.

Approach

I will use Yale ITS SpinUp to host dynamic web pages on Yale's web servers. The main technologies I plan to use are:

- **Front-End**: NodeJS and React.

- **Back-End**: Python or Flask for any necessary back-end development.

- **Database**: Initially hosted on Yale servers, with potential for future expansion to cloud storage.

Additionally, I aim to incorporate voice transcription capabilities for Yoruba to enable voice search and translation. Research into open-source resources, such as YASR (`https://github.com/Alikerin/yoruba-asr`), will be a key part of this development.

Tentative Deliverable Schedule

Deliverable	Due Date
Proposal Report	Jan. 30, 2025
Proposal Presentation	Feb. 8, 2025
Literature Review	Feb. 12, 2025
Mellon Forum	Feb. 18, 2025
Code Samples / GitHub Commits	Feb. 14 - April 23, 2025
Online Usable React App	Feb. 26, 2025
Voice Recognition Portion	Mar. 24, 2025
Project Paper + Presentation	April 24, 2025

Table 1: Tentative Deliverable Schedule

Future Work

After the initial project completion, it can be maintained and expanded by the African Studies department or the student community at Yale. Future work may include:

- Development or tuning of a large language model to provide a more interactive experience and facilitate learning.

- Expansion of the database with more academic resources written in Yoruba.