

Odunayo Kareemat Buliaminu

+234 815 456 0286 buliaminuodunayo12@gmail.com odunayo-buliaminu-9a9037360
buliaminuodunayo12-eng

Professional Summary

Results-driven **Linguist** with First Class Honours in Linguistics and practical experience in **data labeling, linguistic annotation, and dataset curation** for African languages. Skilled in **Python, spaCy, and ELAN**, with solid understanding of **NLP pipelines, quality assurance, and AI dataset development**. Passionate about building inclusive, culturally grounded linguistic resources that power speech recognition and natural language understanding systems.

Professional Experience

Tonative AI Remote
Lead Validator (Volunteer, Nigerian Pidgin) 2025–Present

- Led a team of 20 validators ensuring linguistic accuracy, clarity, and cultural authenticity in Nigerian Pidgin datasets.
- Applied annotation guidelines to maintain dataset consistency and quality.
- Collaborated with engineers to refine dataset pipelines and improve AI model performance.

Elevvo Pathways Remote
NLP Intern Aug 2025–Sept 2025

- Conducted sentiment analysis and named entity recognition tasks on multilingual datasets.
- Supported the training of language models for underrepresented African languages.
- Utilized Python, Pandas, and spaCy for data cleaning, tokenization, and analysis.

Data Science Nigeria Remote
Audio Validator / Voicing Specialist Mar 2025–Aug 2025

- Validated 3,000+ Nigerian Pidgin recordings and created 100+ speech samples for AI datasets.
- Performed quality assurance checks on audio and transcription alignment.

Data Science Nigeria Remote
Data Annotator Feb 2025–Jul 2025

- Annotated 5,000+ Naija Pidgin sentences using ELAN for phonetic and orthographic precision.
- Contributed to culturally accurate dataset curation for multilingual speech technology.

Education

University of Benin Benin City, Nigeria
B.A. Linguistics (First Class Honours) 2024
Thesis: *Relativization in Standard Yoruba: A Government and Binding Approach.*

Technical & Linguistic Skills

Programming: Python, spaCy, NumPy, Pandas, Matplotlib, SQLite
Annotation Tools: ELAN, Label Studio, Prodigy, Smartcat, Trados, MemoQ
Data & NLP: Annotation, Labeling, Dataset Development, Tokenization, Speech Data Validation
Linguistics: Phonetic Transcription, Syntax & Phonology, Corpus Development, Language Documentation
Soft Skills: Leadership, Communication, Analytical Thinking, Quality Control

Publications & Conferences

- **2025:** Co-author — *7PERFECTION: Community-Based Development of Commonsense Dataset for Seven Nigerian Languages*. EMNLP 2025 (Japan)
- **2025:** Presenter — *A Corpus-Based Approach to Owe and Ofo in Jagun Jagun and King of Boys*. University of Ibadan Conference

Languages

Yoruba: Native **English:** Fluent **Nigerian Pidgin:** Fluent and Culturally Competent