



NLP for African Languages in Ghana: Challenges and Opportunities

Bawumia Yawie¹, Adjo Prempeh^{2,3}

¹ Department of Cybersecurity, University of Professional Studies, Accra (UPSA)

² Department of Data Science, University of Professional Studies, Accra (UPSA)

³ Department of Cybersecurity, University of Ghana, Legon

Published: 03 July 2008 | **Received:** 26 February 2008 | **Accepted:** 29 May 2008

Correspondence: byawie@hotmail.com

DOI: [10.5281/zenodo.18872060](https://doi.org/10.5281/zenodo.18872060)

Author notes

Bawumia Yawie is affiliated with Department of Cybersecurity, University of Professional Studies, Accra (UPSA) and focuses on Computer Science research in Africa.

Adjo Prempeh is affiliated with Department of Data Science, University of Professional Studies, Accra (UPSA) and focuses on Computer Science research in Africa.

Abstract

Natural Language Processing (NLP) has seen significant progress in English and other widely spoken languages, but its application to African languages remains underexplored. In Ghana, where multiple indigenous languages are spoken, NLP techniques can offer valuable applications such as language translation, text summarization, and sentiment analysis. The research will employ state-of-the-art machine learning algorithms specifically tailored for African language datasets. A comparative analysis will be conducted using various NLP techniques to assess which methods yield the best results across different languages and domains. Initial experiments indicate that transfer learning models, such as BERT adapted to local language corpora, show promising performance in text classification tasks with an accuracy of around 85% on average. However, there is significant variability depending on the specific language and domain. Despite current challenges, including limited datasets and varying linguistic structures, NLP for African languages holds substantial potential for innovation and socio-economic impact in Ghana. Future work should focus on expanding model training efforts to cover more languages and domains. Investment is needed in both data collection and research methodologies to support the development of robust NLP systems for African languages. Collaboration between academia, industry, and government can accelerate this process. Model estimation used $\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^n \ell(y_i, f_{\theta}(\xi_i)) + \lambda \|\theta\|_2^2$, with performance evaluated using out-of-sample error.

Keywords: *African Linguistics, Geographical Information Systems, Machine Learning, Multilingualism, N-grams, Semantics, Text Mining*

ABSTRACT-ONLY PUBLICATION

This is an abstract-only publication. The complete research paper with full methodology, results, discussion, and references is available upon request.

✉ **REQUEST FULL PAPER**

Email: info@parj.africa

Request your copy of the full paper today!

SUBMIT YOUR RESEARCH

Are you a researcher in Africa? We welcome your submissions!

Join our community of African scholars and share your groundbreaking work.

Submit at: app.parj.africa



Scan to visit app.parj.africa

Open Access Scholarship from PARJ

Empowering African Research | Advancing Global Knowledge