# African Perspectives on Natural Language Processing Challenges and Opportunities

**Omotola Adekunle**[1,2]**, Femi Adesina**[3]**, Temitope Olayinka**[4,5]**, Hammed Olumide**[1,5]

[1] University of Maiduguri

[2] National Centre for Technology Management (NACETEM)

[3] Department of Cybersecurity, National Centre for Technology Management (NACETEM)

[4] University of Ilorin

[5] Nigerian Institute of Social and Economic Research (NISER)

**Author notes**

*Omotola Adekunle is affiliated with University of Maiduguri and focuses on Computer Science research in Africa.*
*Femi Adesina is affiliated with Department of Cybersecurity, National Centre for Technology Management (NACETEM) and focuses on Computer Science research in Africa.*
*Temitope Olayinka is affiliated with University of Ilorin and focuses on Computer Science research in Africa.*
*Hammed Olumide is affiliated with University of Maiduguri and focuses on Computer Science research in Africa.*

## Abstract

Natural Language Processing (NLP) is a field within Computer Science that focuses on enabling computers to understand and process human language. Despite its widespread applications in various domains, NLP has not fully addressed the challenges posed by African languages due to their unique linguistic features. A mixed-method approach was employed, including surveys of linguistic experts and practitioners, focus group discussions with language users, and an online survey targeting developers working on NLP projects. Data were analysed using thematic analysis and statistical software to quantify findings and identify patterns. The analysis revealed that the primary challenge in developing NLP systems for African languages is the limited availability of annotated datasets, which constitute more than 70% of the variance explained by the model ($R^2 = 0.74$). This finding underscores a critical gap in current research efforts. This study highlights the significant challenges faced in developing NLP technologies for African languages and identifies a clear need for greater investment in annotated datasets to support research and development in this field. Developers should prioritise the creation of comprehensive annotated datasets, while researchers must collaborate more closely with linguistic experts to ensure that these resources are culturally and linguistically appropriate. Policy makers could facilitate this by allocating funds towards building such datasets. Model estimation used $\{\hat{\theta}\} = argmin\{\theta\} sumiell(yi, f\theta(\xi)) + \lambda lVert\theta rVert 2^2$, with performance evaluated using out-of-sample error.

**Keywords:** *African, Bantu, Computational, Morphology, Transformational, Linguistics, Orthography*

# 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