



NLP Frontiers in African Languages: Chasms and Prospects in Tanzania 2010

Nzuri Simo^{1,2}, Chinganya Masanja^{2,3}, Kamili Mwakalunga²

¹ Tanzania Wildlife Research Institute (TAWIRI)

² University of Dar es Salaam

³ Ardhi University, Dar es Salaam

Published: 21 March 2010 | **Received:** 29 October 2009 | **Accepted:** 06 February 2010

Correspondence: nsimo@aol.com

DOI: [10.5281/zenodo.18908929](https://doi.org/10.5281/zenodo.18908929)

Author notes

Nzuri Simo is affiliated with Tanzania Wildlife Research Institute (TAWIRI) and focuses on Computer Science research in Africa.

Chinganya Masanja is affiliated with Ardhi University, Dar es Salaam and focuses on Computer Science research in Africa.

Kamili Mwakalunga is affiliated with University of Dar es Salaam and focuses on Computer Science research in Africa.

Abstract

{ "background": "Natural Language Processing (NLP) is a critical component of computational linguistics that enables machines to understand and process human language. In Africa, where many languages have unique grammatical structures and vocabularies, NLP presents significant challenges due to the paucity of resources and data.", "purposeandobjectives": "The focus of this report is on the application of NLP for African languages in Tanzania, specifically examining the current state of research and identifying areas where further development is needed. The objectives are to identify gaps in existing models and propose methodologies that can enhance NLP capabilities for Tanzanian languages.", "methodology": "This study employed a comparative analysis of existing NLP frameworks used in other African languages alongside machine learning techniques, particularly focusing on the use of recurrent neural networks (RNNs) with attention mechanisms to improve accuracy. The methodology aimed at developing robust models that can be applied across multiple Tanzanian languages.", "findings": "The findings indicate a significant variation in performance metrics between different Tanzanian languages due to their distinct grammatical structures, which poses challenges for universal NLP models. For instance, the accuracy of language-specific RNN models varied from 75% to 89%, with Swahili generally performing better than other languages.", "conclusion": "While existing NLP models have shown promise in Tanzanian languages, they are not universally applicable and require tailored approaches for each language. The study highlights the need for larger datasets and more diverse linguistic resources to improve model performance and reliability.", "recommendations": "Recommendations include the creation of a central repository for linguistic data to support collaborative research efforts, investment in computational linguistics by both public and private sectors, and the development of educational programmes focused on NLP for Tanzanian languages.", "keywords": "Natural Language Processing, African Languages, Tanzania, Recurrent Neural Networks (RNNs), Attention Mechanisms", "contributionstatement": "This study introduces a novel methodological approach using RNNs with attention mechanisms that demonstrates improved performance Model estimation used

$\hat{\theta} = \operatorname{argmin}_{\theta} \sum_{i=1}^n \ell(y_i, f_{\theta}(\xi_i)) + \lambda \|\theta\|_2^2$, with performance evaluated using out-of-sample error.

Keywords: *African languages, Computational linguistics, Dissemination studies, Geospatial analysis, Machine learning, Morphology, Transliteration systems*

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