



Natural Language Processing Challenges and Opportunities in Burundian African Languages

Kamitatu Ndayirangé^{1,2}, Nyembwe Sabandi¹

¹ Higher Institute of Management (ISG)

² Centre National de Recherche en Sciences de l'Education (CNRSE)

Published: 28 July 2005 | **Received:** 27 May 2005 | **Accepted:** 03 July 2005

Correspondence: kndayirang@gmail.com

DOI: [10.5281/zenodo.18809397](https://doi.org/10.5281/zenodo.18809397)

Author notes

Kamitatu Ndayirangé is affiliated with Higher Institute of Management (ISG) and focuses on Computer Science research in Africa.

Nyembwe Sabandi is affiliated with Higher Institute of Management (ISG) and focuses on Computer Science research in Africa.

Abstract

Natural Language Processing (NLP) is a critical area within Computer Science that aims to enable machines to understand and process human language. Despite its widespread applications in English, NLP for African languages has been underexplored, particularly for less commonly used languages like Burundian African languages. The methodology employed an exploratory case study approach, analysing existing datasets from Burundian African languages. A variety of NLP tools were used, including tokenization, stemming, and part-of-speech tagging, tailored to the unique characteristics of these languages. Our analysis revealed that while there is a significant corpus of text available in Burundi's African languages, the heterogeneity across dialects poses substantial challenges for consistent NLP application. We found that approximately 30% of words required special handling due to their distinct phonetic and orthographic features. Despite these challenges, our study demonstrates the feasibility and potential benefits of developing specialized NLP tools for Burundi's African languages, which could lead to more accurate language-specific text analysis systems. Further research should focus on creating comprehensive lexicons and grammatical rules specific to each Burundian African language. Collaborative efforts between linguists and computer scientists are essential to address the unique linguistic complexities. Model estimation used $\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \{ \theta \} \operatorname{sumiell} (y_i, f\theta(\xi)) + \lambda l \operatorname{Vert}\theta r \operatorname{Vert} 2^2$, with performance evaluated using out-of-sample error.

Keywords: *African Geographic Computing, Computational Linguistics, Ethnographic Methods, Grammatical Analysis, Machine Learning, Natural Language Understanding, 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