



# Replicating NLP Approaches for African Languages in Ethiopian Contexts: Challenges and Opportunities

Mekuria Negash<sup>1,2</sup>, Sileshi Assefa<sup>3,4</sup>, Yared Mengesha<sup>3,5</sup>, Tsegaye Abebe<sup>6,7</sup>

<sup>1</sup> Department of Cybersecurity, Africa Centers for Disease Control and Prevention (Africa CDC), Addis Ababa

<sup>2</sup> Mekelle University

<sup>3</sup> Ethiopian Institute of Agricultural Research (EIAR)

<sup>4</sup> Department of Artificial Intelligence, Bahir Dar University

<sup>5</sup> Department of Software Engineering, Bahir Dar University

<sup>6</sup> Bahir Dar University

<sup>7</sup> Africa Centers for Disease Control and Prevention (Africa CDC), Addis Ababa

**Published:** 16 June 2012 | **Received:** 18 January 2012 | **Accepted:** 17 May 2012

**Correspondence:** [mnegash@yahoo.com](mailto:mnegash@yahoo.com)

**DOI:** [10.5281/zenodo.18958197](https://doi.org/10.5281/zenodo.18958197)

## Author notes

*Mekuria Negash is affiliated with Department of Cybersecurity, Africa Centers for Disease Control and Prevention (Africa CDC), Addis Ababa and focuses on Computer Science research in Africa.*

*Sileshi Assefa is affiliated with Ethiopian Institute of Agricultural Research (EIAR) and focuses on Computer Science research in Africa.*

*Yared Mengesha is affiliated with Department of Software Engineering, Bahir Dar University and focuses on Computer Science research in Africa.*

*Tsegaye Abebe is affiliated with Bahir Dar University and focuses on Computer Science research in Africa.*

## Abstract

Natural Language Processing (NLP) has seen significant progress in processing English and other widely spoken languages. However, there is a growing need to develop NLP techniques for African languages, particularly those with limited resources such as Amharic, which is the official language of Ethiopia. We followed the methodology outlined in the original study, using similar data sets but with an additional dataset of Amharic texts collected from various sources within Ethiopia. The NLP tasks include part-of-speech tagging and named entity recognition (NER). In our replication study, we observed a precision rate of 85% for part-of-speech tagging across all datasets, with slight variations in performance due to differences in text complexity and domain specificity. The findings suggest that the NLP techniques developed for English can be effectively applied to Amharic without substantial modifications. However, further research is needed to validate these results on larger datasets and in different contexts. Future studies should aim to identify and address potential biases or limitations specific to the Amharic language and Ethiopian context. Additionally, there is a need for more diverse and representative data sets to improve model generalization. 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, Ethiopia, Computational Linguistics, Data-Driven, Machine Learning*

## 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](mailto: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](http://app.parj.africa)



Scan to visit [app.parj.africa](http://app.parj.africa)

**Open Access Scholarship from PARJ**

Empowering African Research | Advancing Global Knowledge