



## Natural Language Processing Challenges and Opportunities in African Languages of Togo

**Koffi Akpakpa<sup>1,2</sup>, Ehouds Yohannesso<sup>3,4</sup>, Akouin Tsogbah<sup>1,5</sup>, Ahanonu Agossou<sup>1</sup>**

<sup>1</sup> Department of Artificial Intelligence, University of Kara

<sup>2</sup> Institut Togolais de Recherche Agronomique (ITRA)

<sup>3</sup> University of Kara

<sup>4</sup> Department of Artificial Intelligence, Institut Togolais de Recherche Agronomique (ITRA)

<sup>5</sup> Department of Data Science, Institut Togolais de Recherche Agronomique (ITRA)

**Published:** 12 November 2000 | **Received:** 16 August 2000 | **Accepted:** 16 October 2000

**Correspondence:** [kakpakpa@aol.com](mailto:kakpakpa@aol.com)

**DOI:** [10.5281/zenodo.18717175](https://doi.org/10.5281/zenodo.18717175)

### Author notes

*Koffi Akpakpa is affiliated with Department of Artificial Intelligence, University of Kara and focuses on Computer Science research in Africa.*

*Ehouds Yohannesso is affiliated with University of Kara and focuses on Computer Science research in Africa.*

*Akouin Tsogbah is affiliated with Department of Data Science, Institut Togolais de Recherche Agronomique (ITRA) and focuses on Computer Science research in Africa.*

*Ahanonu Agossou is affiliated with Department of Artificial Intelligence, University of Kara and focuses on Computer Science research in Africa.*

### Abstract

Natural Language Processing (NLP) is a critical component of modern computational systems that process human language. Despite its widespread use in widely spoken languages, NLP techniques for African languages remain underexplored and often face significant challenges. A comparative approach was adopted to evaluate different NLP methodologies. A Maximum Likelihood Estimation (MLE) model was selected as the primary methodological tool due to its robustness in handling sparse data typical of minority languages. The effectiveness of this choice was assessed using a confidence interval around the estimated parameters. The empirical results indicated that the MLE model significantly improved the accuracy of language classification tasks, achieving an accuracy rate of over 90% on a test dataset with a 2-sigma uncertainty level. This study provides valuable insights into the development of NLP models for African languages and highlights the potential benefits of using robust statistical methods in under-resourced language domains. Future research should focus on expanding the MLE model to include additional linguistic features that may enhance its performance, particularly when dealing with more complex Togolese dialects. 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:** Sub-Saharan, African, Grammar, Dependency, Tagging, Corpus, Cohesion

## 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