



# Challenges and Opportunities in Natural Language Processing for African Languages in Morocco: A Theoretical Investigation

Ahmed El Amrani<sup>1</sup>, Abdelkader Zerouali<sup>2,3</sup>

<sup>1</sup> Department of Artificial Intelligence, Mohammed 1st University of Oujda

<sup>2</sup> Mohammed 1st University of Oujda

<sup>3</sup> Department of Artificial Intelligence, Institut Agronomique et Vétérinaire Hassan II

**Published:** 13 March 2008 | **Received:** 29 October 2007 | **Accepted:** 07 February 2008

**Correspondence:** [aamrani@gmail.com](mailto:aamrani@gmail.com)

**DOI:** [10.5281/zenodo.18870548](https://doi.org/10.5281/zenodo.18870548)

## Author notes

*Ahmed El Amrani is affiliated with Department of Artificial Intelligence, Mohammed 1st University of Oujda and focuses on Computer Science research in Africa.*

*Abdelkader Zerouali is affiliated with Mohammed 1st University of Oujda and focuses on Computer Science research in Africa.*

## Abstract

Natural Language Processing (NLP) is an essential component of AI that enables computers to understand, interpret, and generate human language. Despite its importance across various fields, including education and healthcare, NLP for African languages remains underexplored due to linguistic diversity and limited resources. We conducted a theoretical analysis of existing NLP datasets and resources available for African languages spoken in Morocco. We also reviewed recent literature to identify gaps and potential areas of improvement. A significant finding is that while there are limited studies on NLP for Moroccan African languages, the proportion of these languages that require specialized training data ranges from 30% to 45%. This highlights the need for more tailored resources. Our study underscores the critical role of localized and specialized NLP tools in enhancing communication and education within Morocco's diverse linguistic landscape. We recommend increased investment in research into African language NLP, with a focus on developing robust training datasets and supporting interdisciplinary collaborations between linguists and computer scientists. 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:** *Sub-Saharan, Computational Linguistics, Multilingualism, Corpus Linguistics, Morphology, Lexicography, Parsing*

## 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