

Name: Lorraine Paula Arinaitwe

Access no. B20729

Reg. No. M23B38/004

### Insights Derived from the Review Article Titled “Building Natural Language Processing Tools for Runyakitara”

The article values insight into the intersection of linguistics and computational modeling for Uganda’s indigenous languages. The authors document the first large-scale effort to develop Runyakitara-specific NLP resources, such as the RunyaCorpus, RunyaMorph morphological analyzer, and RunyaSpeller. These tools were designed to process four closely related dialects—Runyankore, Rukiga, Runyoro, and Rutooro—spoken by more than five million people across Western Uganda.

From the article, one major insight is the critical shortage of digital linguistic resources for African languages. Despite the sizeable population of Runyakitara speakers, there is limited presence of written, digitized, or annotated texts that could be used to train NLP models. The researchers therefore adopted a resource-constrained methodology, combining manual corpus creation with semi-automatic text collection from student essays, religious publications, and news articles. This highlights an important principle in low-resource NLP: that linguistic expertise and community participation can substitute, to some degree, for data abundance.

A second insight concerns morphological complexity. Runyakitara exhibits rich inflectional and derivational morphology, meaning that a single verb root can yield thousands of surface forms. The authors demonstrate how rule-based morphological analyzers such as RunyaMorph can complement statistical language models, improving both spelling correction and word-form recognition. This integration of linguistics and computation is particularly instructive for students of data science working in multilingual contexts.

Thirdly, the article underscores language endangerment in the digital age. While Runyakitara is not critically endangered in terms of speakers, its digital invisibility threatens its future use in education and technology. The creation of RunyaSpeller and LearnRunya prototypes reveals how NLP can actively support literacy and language learning. The study therefore shifts the conversation from documentation toward technological empowerment, demonstrating that data science can serve cultural sustainability.

Finally, the article provides methodological insight into evaluation and iterative improvement. The researchers relied on human linguistic feedback to refine the morphological analyzer and spell-checker, emphasizing a participatory approach to NLP model development. This shows that sustainable NLP for African languages must blend computational rigor with social engagement.

## 2. Recommendations to Ethnic Leaders of Runyakitara Speakers.

Building on the insights above, several recommendations can guide ethnic and cultural leaders in promoting language preservation through NLP:

- Institutional Support for Digital Language Resources

Local leaders should collaborate with universities and cultural institutions to fund the expansion of RunyaCorpus and similar datasets. Making community texts, folk stories, and oral histories available in digital form will sustain ongoing NLP research and literacy applications.

- Community Participation in Data Collection

Elders, teachers, and youth should be mobilized to contribute authentic Runyakitara texts and speech recordings. Participatory data collection ensures dialectal balance and improves model accuracy, while fostering community ownership of the technology.

- Integration of NLP Tools into Education

Leaders can advocate for the inclusion of RunyaSpeller and LearnRunya applications in local schools. Using such digital aids for spelling, reading, and vocabulary will encourage written fluency among learners and raise the prestige of the language in formal settings.

- Partnerships with Tech and Policy Stakeholders

Collaborations between cultural councils, the Uganda National Council for Science and Technology, and private tech firms could scale up existing prototypes.