Workshop TALAI @ CRI'2025

Transcribing African Languages for Artificial Intelligence Applications

Date : December 15, 2025
Venue : University of Yaounde I

Event collocated with CRI'2025 (http://cri-info.cm/)

CALL FOR SHORT PAPERS

SCOPE

Creating robust corpora, standardizing writing rules, and fostering interdisciplinary collaboration are critical first steps for leveraging African languages in the digital age. It's essential to create large, annotated textual corpora because these datasets are the foundational "food" needed to train Natural Language Processing (NLP) and Artificial Intelligence (AI) models, enabling crucial applications like speech recognition and machine translation for local languages. Furthermore, standardizing orthography and grammar is necessary to ensure that the AI models can process input consistently, overcoming the current challenge of high variability and regional differences in writing, which otherwise makes the data unusable for computational tasks. This systematic process simultaneously serves the preservation of cultural and indigenous knowledge by providing a stable, documented form of these often-oral languages for future generations and academic study. Finally, collaboration among linguists (for rules and transcription), computer scientists (for modeling), and community leaders (for validation and ethical context) is the only practical path to ensure that the resulting AI applications are both linguistically accurate and culturally relevant. Contributions are welcome from disciplines like data mining, linguistics, machine learning, AI, and data science. The first focus is on the technical challenges of putting Cameroonian languages into a consistent, machine-readable written format : alphabet and orthography standardization, modeling language specificities : tonal marking, morphology, syntax and grammar. The second focus is data creation and transcription : transcription protocols, data Annotation and tagging, data Curation and sharing. The third focus is AI application and industrialization : building Language Models (LLMs) from transcripts, speech-to-Text (STT), translation and accessibility tools, roadmap and collaboration.

TOPICS OF INTEREST

The workshop will be structured around three main objectives: Standardization, Data Generation, and AI Application.

The main topics of interest include:

- Data collection and curation
- Speech Recognition
- Text to Speech
- Text Summarization
- Machine Translation
- Computational linguistics
- Dialect and Language Documentation
- Conversation Analysis
- Phonetics and Phonology
- Ethic and Responsible AI
- Creating Linguistic Resources

All innovative research addressing these themes is welcomed, irrespective of the field of application.

SUBMISSIONS

Researchers, academics, and students working on the field of Natural Language Processing, Artificial Intelligence, Linguistic, Anthropology with a focus on Cameroon as target county are invited to submit extended abstracts or short papers for oral presentations or posters. Submitted papers must be in English and will be reviewed by the workshop committees for suitability and interest to the TALAI audience. The authors can submit papers of unpublished work reporting original and early results, introducing new ideas or describing prototypes.

Every accepted submission must have at least one author registered for the workshop. All submitted extended abstracts or short paper must follow the LNCS format with a page limit of up to 2 and 6 pages respectively including the title page, figures, references, and an optional appendix. The abstracts should be submitted electronically in PDF format via EasyChair via the following link: https://easychair.org/conferences?conf=talai2025

LNCS template :

Latex : https://resource-cms.springernature.com/springer-cms/rest/v1/content/19238648/data/v8

Word : https://resource-cms.springernature.com/springer-cms/rest/v1/content/19238706/data/v5

IMPORTANT DATES :

- Submission deadline: 17 November, 2025

- Notification to authors: 01 December , 2025

- Workshop date: 15 December , 2025

PARTICIPATION

The workshop will be held in Yaounde, as an event affiliated to CRI'2025. This workshop will be in-person event.

WORKSHOP CHAIRS :

- Gabriel Mba, University of Yaounde I, Cameroun

- Jean Romain Kouesso, University of Dschang, Cameroun

- Florence Tabe Ako Enoh, University of Yaounde I, Cameroun

- Maxime Yves Julien Manifi Abouh, University of Yaounde I, Cameroon

- Paulin Melatagia, University of Yaounde I, Cameroon

- Justin Moskolai, University of Douala, Cameroon