

Title: Advancing Language Technology for Jola

Papa Ousmane Diedhiou, MA Student in African Studies University of Warsaw, Poland

Collaborator: Talent Mudenda, PHD Student University of Warsaw, Poland

Introduction

The Jola Lexical Database (JLD) is going to be a significant resource in the domain of computational linguistics, specifically focused on the Jola language, a member of the Atlantic branch of the Niger-Congo language family spoken in Senegal and Gambia. This database is going to serve as a repository of lexical information, offering a structured compilation of Jola words, their meanings, grammatical properties, and phonetic representations. Being developed with the aim of preserving and studying the Jola language, the JLD will provide valuable insights into linguistic diversity and contributes to the broader field of language documentation and revitalization.

Purpose

The development of the Jola Lexical Database emerged from the need to document and analyze endangered languages, particularly those with rich oral traditions but limited written resources. The project is initiated by myself as a linguist, native Jola speaker and a researcher specializing in African languages, recognizing the linguistic value of preserving and analyzing the Jola language. The database's construction will involve extensive fieldwork, collaborating with native Jola speakers to collect lexical data and document the language's intricacies.

Structure and Components

The Jola Lexical Database is structured to facilitate comprehensive linguistic analysis and computational processing. It comprises several key components:

Lexical Entries: Each entry in the database represents a unique Jola word or phrase, accompanied by detailed linguistic information, including phonetic transcriptions, semantic definitions, grammatical properties, and morphological analyses.

Phonetic Inventory: The JLD includes a phonetic inventory that outlines the sound system of the Jola language. This inventory assists in phonetic transcription and pronunciation analysis.

Semantic Relations: Semantic relations between words are encoded in the database, illustrating lexical fields, synonyms, antonyms, and other semantic connections within the Jola lexicon.

Grammatical Annotations: The database annotates each lexical entry with grammatical information, such as part of speech (e.g., noun, verb, adjective), inflectional patterns, and syntactic roles.

Morphological Analysis: Morphological information is incorporated to analyze word structure, including prefixes, suffixes, and root forms, enabling computational processing of Jola morphology.

Example Usage: The JLD includes illustrative examples of word usage in context, aiding in the understanding of Jola syntax and discourse structure.

Methodology and Data Collection

The construction of the Jola Lexical Database is going to involve meticulous data collection methodologies to ensure accuracy and comprehensiveness. Linguists and field researchers collaborated with native Jola speakers through ethnographic fieldwork, eliciting lexical data through structured

interviews, narratives, and linguistic elicitation tasks. The data collection process focused on capturing various aspects of the Jola language, including phonology, morphology, syntax, and semantics. Specialized tools and techniques, such as audio recordings, field notes, and multimedia documentation, are going to be employed to capture the nuances of spoken Jola and preserve cultural context alongside linguistic data.

Applications and Significance

The Jola Lexical Database is going to play a significant implication across academic, cultural, and technological domains:

Language Preservation: The JLD will play a crucial role in preserving and revitalizing the Jola language by providing a comprehensive linguistic resource for future generations of speakers and researchers.

Linguistic Research: The database will facilitate linguistic research on the Jola language, fostering investigations into language structure, historical linguistics, and language contact phenomena.

Computational Linguistics: The JLD will serve as a valuable dataset for computational linguistics applications, including natural language processing (NLP), machine translation, and speech recognition systems.

Cultural Heritage: By documenting the Jola language, the database will contribute to the preservation of intangible cultural heritage, fostering cultural appreciation and understanding.

Benefits of attending the summer training school

Participation in a training school can significantly benefit projects focused on language documentation, computational linguistics, and the development of lexical databases such as the Jola Lexical Database (JLD). Firstly, these schools offer specialized courses in linguistic methodologies like data collection techniques, phonetic transcription, and morphological analysis. Additionally, training schools teach computational tools and techniques used in language processing and database management, optimizing data processing and analysis for the JLD's development. Methodologically, attendees learn best practices for thorough and culturally sensitive data collection, crucial for maintaining the JLD's integrity. Quality control methods taught in these schools minimize errors in collected data. Interdisciplinary collaboration is facilitated through networking with experts from linguistics, anthropology, and computer science, leading to innovative approaches for database development. Participants also gain project management skills essential for coordinating efforts and meeting milestones efficiently. Lastly, participation in reputable training schools enhances career prospects in language documentation and computational linguistics while providing access to funding sources and grant-writing skills vital for sustaining and expanding projects like the JLD. In summary, attending a training school equips individuals with the specialized knowledge, methodological skills, ethical considerations, interdisciplinary networks, technological insights, and career opportunities necessary for advancing language documentation and lexical database development.

Conclusion

In summary, the Jola Lexical Database (JLD) will represent a significant contribution to language documentation, linguistic research, and cultural preservation. This comprehensive resource encapsulates the richness of the Jola language, offering valuable insights into its phonological, morphological, and semantic characteristics.