

# Nlp systems overview: deriving the meaning from human language

[Science](#)



Today's world is now a global village where technological solutions have become more efficient and enhanced the ways in which we as humans communicate. Slowly but surely most of our African native local languages are being buried by the powerful 'international languages' building a bridge between the literate and illiterate societies [1]. This social difference has lowered the rate of embracing technology in most parts of Africa. The concept of analyzing, understanding and deriving the meaning of a human language has become the future through the NLP techniques governed by the art of Artificial Intelligence. As such, part of African culture is to preserve a rich history and heritage it possesses, and this can be achieved through using the current technologies to maintain and preserve native spoken languages.

It is evident that many NLP systems that focus on preserving IK are not yet realized, for example an NLP system that was developed by Bobrow in 1964 focused on solving high school algebra problems. A group of multidisciplinary (MDT) researchers collaborated on an NLP system that focused on tackling problems related to medicine. This MDT worked together for 2 years and noted down the main challenges that they experienced with AI based technologies in the medical field. Furthermore, some of the challenges included creating the appropriate data sets, collaboration with professionals of different backgrounds and many more. In Namibia the CARACAL system was developed, and it uses a mobile based technology (Android smartphones) that captures and annotates data of real life objects to produce information related to the captured object. Although much NLP concentration is not on preserving IK, some study has shown some existence

of such systems. “ Breaking the Unwritten Language Barrier” project was focused on creating automated processing tools of unwritten languages. The project was focused on three main Bantu languages which include; Bassa, Myene and Embosi but all these were translated to French.

With many NLP systems being developed to solve other problems, there is an additional sense of urgency in trying to preserve IK. Statistical information that come from Africa’s endangered languages recorded that there are approximately 308 highly endangered languages spoken in Africa. This is about 12% of all African languages and at least 201 languages that have gone extinct. The project done by Bobrow in the 60’s was one of the first human intelligent systems to be developed but now it has long been outdated and this study seeks to address a pressing generational issue that seek to avoid IK going extinct. Researchers in where mainly focusing on a medical point of view of the use of NLP systems and outlined some significant challenges that the process brings forth, for example developing NLPs requires the collaboration of professionals with different backgrounds in respect to their area of study. This is also enlightens this study which will also have a multidisciplinary team that includes Shona language linguists, computer scientist among others to work together in order to successfully develop the proposed NLP system.

The CARACAL system developed in Namibia focused mainly on annotating objects, people and places in order to get information and knowledge. Their approach to preserving IK using AI was based on augmented technology and they also outlined that this approach has not yet been fully incorporated into

the daily IK preservation processes because the indigenous people were not fully part of the research, as such, the solution was not fully appreciated.

With this great revelation, this study seeks to provide a more technologically advanced approach to preserve neglected spoken languages that will encourage linguists and Africanists alike to work to protect them through using AI.

The preservation of historical data, particularly indigenous knowledge, has widely evolved and has also progressively improved from oral tradition to manuscripts, audio recordings, documentary artifacts among many others. However, these techniques have taken us this far but can no longer keep up with time, and this has resulted in data loss through the dying oral traditions due to urbanization, eroded manuscripts due to their longevity in dusty archival store rooms to mention just a few. AI has revealed its utmost potential for the future where machines can learn a concept and be able to analyze, process and preserve. As evidently highlighted by that a considerable number of African native languages have already gone extinct, this knowledge shows that indeed preserving IK is a problem in Africa as such this study seeks to utilize the most advanced technological approach to address this problem through the use of NLP.