A novel method for redefining language ecology and endangerment in Nigeria – towards a geospatial solution

Udoh, Imelda
University of Uyo
imeldaudoh@uniuyo.edu.ng
Ekpenyong, Moses
University of Uyo
mosesekpenyong@uniuyo.edu.ng
Urua, Eno-Abasi
University of Uyo
eno-abasiurua@uniuyo.edu.ng
Adeniyi, Harrison
Lagos State University
harrison.adeniyi@lasu.edu.ng
Obiamalu, Greg
Nnamdi Azikiwe University
go.obiamalu@unizik.edu.ng
Yusuff, Ayo
University of Lagos
yoyussuf@yahoo.co.uk
Anyanwu, Ogbonna
University of Uyo
ogbonnaanyanwu@uniuyo.edu.ng
Obikudo, Ebitare
University of Port Harcourt
ebitare.obikudo@uniport.edu.ng

Abstract

Being a multilingual and multicultural nation, Nigeria is blessed with over 525 languages (Blench, 2014) from four different language families. The sheer number of indigenous languages makes an interesting tapestry! Unfortunately, not much attention has been paid to the study of our indigenous languages, with all the abundant prospects. This paper is a work in progress and a recently funded research project by the Tertiary Education Trust Fund (TETFund) of Nigeria, to rescue language endangerment in Nigeria by redefining the ecology of languages through geospatial technologies; concentrating on those languages located in the southern part of Nigeria, and scalable to other regions of Nigeria and beyond. The main objective of the project is to implement a location-aware infrastructure or framework that functions proactively in real-time, for enhanced language ecology, and precise visualization of the language's vitality status for each language spoken in the region under study. The project draws a multidisciplinary team consisting of linguists and language experts, computational scientists, and geographical information system (GIS) specialists, to examine our language ecology through practical fieldwork and integrate same into a geospatial framework for the purpose of revitalizing our indigenous languages and making them readily available. This research project is therefore significant as it will not only provide a cooperative solution for advancing our understanding of the spatial pattern that describes language ecology, but also provide an early appraisal of the degree of vitality and endangerment in relation to factors such as social and cultural changes. The immediate impact is the prompt revitalization of our indigenous languages and the development of an effective language policy to strengthen our local heritage.

Keywords: Nigerian languages, endangerment status, geo-linguistic database, geospatial framework, language ecology app, location-based system.

1   Introduction

Language endangerment can be described as a process whereby the vitality of a language (i.e., the extent at which the language is applied as a means of communication in various social contexts and use in specific domains, or purpose) is no longer being pursued by the 'owners' of the language or language community, causing a threat of disuse or

abandonment of the language. Endangerment can be caused by natural factors (wars, earthquakes, etc.) or social factors (e.g., abandonment). Today several Nigerian languages are gradually going into extinction because many speakers of the language have either abandoned their languages for other languages or have neglected the core responsibility of developing their languages to support intergenerational transfer (i.e., inability to transfer the language from one generation to another),

The endangerment status of a language can be classified into five levels as follows (Udoh and Urua, 2015 after Connell 1994):

- Moribund – no longer in use, nor transmitted, or are threatened.

- Retreating – dying in an area but flourishing in another area: inter-country boundaries.

- Under-developed – without orthographies, written literature, meta-language.

- Developing – with fairly-developed orthographies, and which literature tradition and meta-language are in the process of development.

- Infiltrated or pidgin (Blench, Spriggs and Connell 1999) – mostly used in informal conversation.

The integrity of a language is further enhanced when its status significantly permeates new domains, develops into a language of the educated class, and adopts new societal values (Urua and Ekpenyong, 2018).

Various indices permit the determination of the endangerment status of a language. The most dominant ones include: the rate of population growth; dominance by a more powerful language or a language with greater economic influence; and/or, the lack of adequate descriptive evidence of the language, hence limiting the development or effective use that language (cf. Bamgbose 1976; Fakuade 1999). Whereas the degree of language use may influence the determination of language endangerment, the various variables work in varying ways to improve the vitality of the language (Adewale and Oshodi, 2013).

Interests in pursuing a spatial solution to language infrastructure development have greatly declined. While most studies concentrate on the use of GIS for literary studies (Kretzschmar, 2013), only three studies to the best of our knowledge are situated within the context of our study and are discussed as follows.

In 2009, the GIS-based Linguistic Geography of Thailand Project was initiated under the sponsorship of Chulalongkorn University with its key aim to promote the use of Geographic Information System (GIS) in linguistics. The project enabled scholars from different fields of knowledge to work together, in this case, geographers and linguists. A series of research work has been conducted since then. The first in the series was the research work of the Word Geography Maps of Thailand project, producing a geographic database of 170 Thai dialect vocabularies based on the data collection in 2002-2003 (Teerarojanarat and Tingsabadh, 2008). The second was the extension work of the first project – the creation of the boundary map of Central and Non-Central Thai Dialects by overlaying 170 map layers (Teerarojanarat and Tingsabadh, 2011).

A recent work was the Word Geography Maps of the Northeastern Thai Dialect. In this work, Thai dialect vocabularies in the northeastern region of Thailand based on data collection in 1979 was converted and transformed under a GIS environment to be available in a digital map.

In Podobnikar et al. (2009), a detailed determination of the local speech areas which are parts of the Slovenian Linguistic Atlas (SLA) was performed. The SLA is a geo-linguistic project designed in the 1930s and published in 2010. One of the goals of this project was to analyze the impact areas and to geographically allocate local speech data for a more precise determination of the isoglosses. Various evidence that influenced formation of the speech were also analyzed with GIS, resulting in intra-linguistic and extra-linguistic indicators (i.e., geographic, and historic).

While the foregoing related works concentrate on aspects of map production and analysis, our project introduces a novel integration of location-aware information plus details of data that determine the vitality and endangerment status of languages, to enable real-time spatial information storage, updates, and retrieval. The purpose of introducing a spatial information system is to enable prompt visualization of stored language attributes, and enhance policy decisions, instead of relying on belated information.

1.1 The Problem

Nigerian languages have not been adequately harnessed. Although a few of them are developing, several are still at risk of endangerment and death. The indigenous languages archive the rich cultural values of the people and valuable knowledge systems, which are important for development. The death of the languages leads to the loss of the rich cultural heritage of the people. Many of the languages are non-vital because they are used in rather restricted domains, and many have even lost the privilege of intergenerational transfer, a serious threat to the languages, their speakers and humanity. Besides, we lack correct information on the number of languages, not to talk of updates on them. There is no absolute one-to-one mapping between languages and the increasing amount of data on dominant languages on the Web, as there appears to be an unbalanced relationship between languages, as the level of endangerment further widens without notice. Also, no defined criteria/models exist to give immediate and precise picture (in real-time) of how these languages fare and when endangerment happens.

Traditional archiving methods (e.g., use of static media – texts, non-animated images, maps) could not offer real-time processing and visualization of resources through location-based (or spatial) services, as they are limited by the opportunities for language development and internalization. Furthermore, not giving back the products of research to the language community has caused loss of interest in preserving and propagating them into the future. Hence, the urgent need to characterize existing linguistic knowledge and resources, to crucially influence a possible revived use or abandonment of linguistic structures and varieties is pursued in this project.

1.2 Project Objectives

The objectives of the project include:

- To investigate the vitality and level of endangerment of languages in the 3 Zones of Southern Nigeria using the UNESCO 2003 LVE (language vitality and endangerment) Scale, for efficient knowledge extraction.

- To investigate the relation between the languages and their speakers and ground truth these details to the specific geolocations.

- To develop a spatial database system of the languages spoken in the 17 States in the 3 Geopolitical Zones of Southern Nigeria, with descriptive information about their location, point clouds, and high-resolution satellite images.

- To integrate extracted information on the spatial system, embedding or mapping important language attributes, for real-time visualization of the language's vitality and endangerment status.

2 Conceptual Framework

The proposed conceptual framework as presented in Figure 1, explains the key concepts or variables of the study and their relationships. It begins with the physical environment and connects technology, for precise information system development. Within the physical environment, the main variable of our study is the spoken language where sociolinguistic, ecological and location parameters are discoverable for the purpose of building:

- spatial linguistic database (for location-based queries and analysis)

- natural language processing (NLP)

- tool development (for linguistic and model development) and

- GIS tool and linguistic database (to establish natural patterns and trends that document and process available spatial datasets as well as model relationships that exists between them).

The outcomes include precise knowledge of linguistic diversity, endangerment threats projection and informed policy decision.

3 Methods

This project will be conducted in the 3 Zones of Southern Nigeria made up of 17 States. The project will be multidisciplinary and shall involve a tripartite collaboration between the academia (staff and students from various disciplines), professional association (Linguistic Association of Nigeria) and the speaking community. To allow for efficient monitoring of the project, supervisors and coordinators will be appointed to supervise/coordinate the primary data collection and training in the respective geopolitical zones.

The Research Population: The research population covers the number of people in a community that will be visited by the field assistants in each LGA, as provided by the 2006 census.

Sample and Sampling Procedure: a multi-stage sampling will be adopted. A systematic sampling method will be used to select houses, where every tenth house on the right side of the street will be selected. A stratified random sampling will then be used to select from a range of recipients within the houses – elders, youths, men, women, children from households in the community.

The Primary Data: The primary data are the languages spoken by the communities in the 355 local government areas (LGAs).

The Data Collection Instrument: The 2003 UNESCO LVE instrument will be adapted for use to collect the primary data. The questionnaire has three main components: the meta data, the real data, and a reliability index. A 5-point scale will be designed to get responses from the communities in the following areas:

- Inter-generational transmission from parents to children

- The absolute number of speakers per language.

- The proportion of speakers within the total population.

- Loss of existing language domains.

- Response to new domains and media.

- Existence materials for language education and literacy.

- The policies of government and institutions concerning the respective languages.

- The attitude of community members towards their own languages

- The amount and quality of documentation the languages have enjoyed.

The reliability index on each of these issues will cover:

- Evidence from field work and direct observation.

- Evidence from other reliable sources.

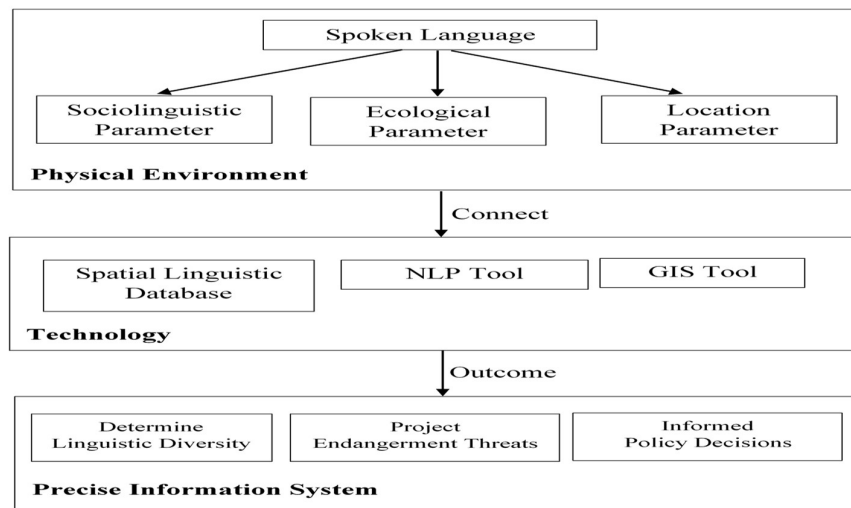- Very little evidence – a 'best guess'.

- No data available.

Figure 1. Proposed conceptual framework

Primary Data Analysis: The responses will be used to code the items in the questions in the questionnaire. These responses will be used to feed the GIS and statistical assessment of the linguistic ecology.

The methodology is divided into 5 major activities as summarized below:

Training of Coordinators/Research/Field Assistants: Training workshop will be organized for coordinators/research/field assistants and students involved in the project.

Fieldwork: Data for the study will be collected from 355 LGAs in the 17 States through fieldwork. Linguistic and location data including boundary information will be collected, processed, and documented as outlined above.

Spatial Database Design: A spatial linguistic database for storing linguistic and location information will be developed. The steps to accomplish this task include (1) satellite image (base map) acquisition and reconnaissance survey of study area to capture relevant location data and establishment of the start and end boundary, using GPS device; (2) Identification and abstraction of relevant features within the study environment.

Geo-Modelling, GIS Mapping and Test-bed Design: Mapping of the linguistic and location information will be achieved using ArcGIS. The steps to accomplish this process include (1) superimposing the GPS data on the extracted surface after image digitization. (2) geo-database modelling and integration; (3)

prototype test-bed system development for test datasets visualization.

Language Ecology App Design: Using the Python programming language, the user interface for communicating with the spatial database will be developed to enable real-time information storage and retrieval.

4 Expected Results

4.1 Outputs and Outcomes

The research outputs and outcomes of this project include:

- Open-Source Geo-Linguistic Database, for efficient community crowdsource and documentation of language resources in Nigeria.

- Language Data Model, for integration of extracted linguistic features into High-Definition maps.

- Geospatial (language ecology) App, for real-time visualization of available language resource, monitoring of the vitality and endangerment status of languages, efficient policy formulation, quick educational reference tool, and community-wide access.

- High quality publications and patent, for advancing linguistic research on language ecology, documentation, and endangerment

4.2 Dissemination

The project outcomes will be brought to the attention of key stakeholders through several avenues. There is a plan to host an international conference to do this. A project website will also be developed to share the project achievements, publications, and other resources, especially with the research community. An App will be developed for use by all. These efforts will contribute to redefining the linguistic ecology of our

communities as well as improve the vitality of these languages.

4.3 Expected Impact

The immediate impact of this project will be a redefinition of approach to preserving and revitalizing Nigerian languages, especially the endangered ones. The proposed approach will strengthen the vitality of our languages, improve their ecology, and stem the tide of their death. Technologically this project will:

(1) enhance the investigation and quantification of language change across speaking groups;

(2) provide detailed analysis and trace of language change and contact;

(3) encourage community contribution of language resources.

To achieve efficient community participation/contribution, the proposed application will be open source, with access rights for manipulating data. Furthermore, resources can be deposited to mapped locations and retrieved from same in real-time.

With a rich repository enabled by community contributions, a spatio-temporal analysis of the language is possible.

The long-term impact of the project is that it will help policy formulation and implementation, and contribute greatly to the language policy, as well as confirm the languages spoken in the LGAs.

5 Conclusion

GIS has found usefulness in documenting the languages of the world, but its application is still evolving. The multidisciplinary approach for redefining language ecology proposed in this project is innovative, as it integrates technology to remodel our perception towards proactive solution to the endangerment problem, which hitherto was achieved via

intensive fieldwork, some of which appear inconsistent. The solution this proposal will present will enable community efforts to develop language resources and could be upscaled to the other three zones of Northern Nigeria and beyond. The Ecology app will embed descriptions of culture, polity, and diversity of the language for effective visualization; hence, enabling cost-effective solution to the current language crises in Nigeria.

Acknowledgements

References

Adewale, R. K, Oshodi, B, 2013. Language Endagerment in Nigeria: Focus on Small Ethnolinguistic Communities in Niger State. Dialectologia, 11 (2013), 17-45.

Bambose, A (ed.), 1976. Mother Tongue Education: The West African Experience, London: Hodder and Stoughton.

Blench, R. M, Spriggs, M, Connell, B (eds.), 1999. The languages of Africa: Macrophyla proposals and implications for archaeological interpretation. Archaeology and Language 4. London: Routledge.

Connell, B., 1994. The Lower Cross Languages: A Pronegomena to the classification of the Cross River languages. Journal of West African Languages. XXIV, 1: 3-46.

Connell, B. 1998. Moribund languages of the Nigeria-Cameroon Borderland. Endangered languages in Africa. (Ed.) Brenzinger, M. Cologne: Koppe, 207-225.

Blench, R. M., 2014. An Atlas of Nigerian Languages. Oxford: Kay Williamson Educational Foundation.

Fakuade, G, 1999. Language Endangerment in the North-Eastern Part of Nigeria: Instances Language Empowerment: Theory and Reality, Aba: National Institute for Nigerian Languages, 50-66.

Haruna, A., 1998. Language death: The case of Bubbure in Southern Bauchi Area, Northern Nigeria.' Endangered languages in Africa. (Ed.) Brenzinger, M. Cologne: Koppe. Pp 227-251.

Kretzschmar, W.A., 2013. GIS for language and literary study. In: Siemens, R. and Price, K. (eds.) Literary Studies in the Digital Age: An Evolving Anthology. Modern Language Association: New York, NY, USA.

Podobnikar, T., Skofic, J. and Horvat, M., 2009. Mapping and analysing the local language areas for Slovenian linguistic atlas. In Cartography in Central and Eastern Europe (pp. 361-382). Springer, Berlin, Heidelberg.

Teerarojanarat, S. and Tingsabadh, K., 2011. Using GIS for linguistic study: a case of dialect change in the northeastern region of Thailand. Procedia-Social and Behavioral Sciences, 21, pp.362-371.

Udoh, I., Urua, E-A., 2015. Sustaining Nigerian Languages Through Use, Technology and Education Education in the 21st Century: A Festschrift in Honour of Professor Mrs. Comfort Ekpo, Urua, E., Udofot, I. M. & Uduk, H., University of Uyo, Uyo, pp. 62-73.

Urua, E-A., Ekpenyong, M, 2018, Strengthening Language Documentation in Africa through Effective Cooperative Research. In: Leonard, J. L. and Rialland, A. (eds). Linguistique Africaine - Perspectives Croisees. Journees scientifiques de la societe de linguistique de paris, 113-142.