EDUCATION IN NIGERIA IN THE 21ST CENTURY:
FOCUS & IMPERATIVES

A Festschrift in Honour of
PROFESSOR MBONG A. UDOFOT

Etuk N. Etuk
Inyang M. Udofot
Alice E. Udosen

EDUCATION IN NIGERIA IN THE 21ST CENTURY:
FOCUS AND IMPERATIVES

CHAPTER 24

LINGUISTIC GLOBALIZATION AND ITS EFFECTS ON LANGUAGE DOCUMENTATION IN NIGERIA

By

Imelda Icheji Udoh

Abstract
Globalization, an interaction of different aspects of human life in the traditional space of place, and the space of flows, implies sharing of the interacting elements. Linguistic globalization therefore involves the interaction and sharing of Linguistic activities, particularly language data. For Linguistic data to be shared in space, they need to be in electronic forms. This has changed the nature of Language documentation, as it has had to conform to globalization trends. For Linguistic data to be shared, they have to be stored and accessed. Such data are today stored in electronic archives. To be able to produce data that can be processed at this level, language documentation has had to radically shift its form and modus operandi. This chapter explores linguistic globalization and some of the effects it has had on language documentation.

Introduction
The world is gradually becoming a global village. This implies that a lot of interaction goes on mostly in space, such that at the click of a mouse one has access to huge amounts of data from any part of the world. There is a current technological revolution going on in information technology, closely associated with globalization. There are several challenges associated with this.

Linguistic data have had to interact at this level thus leading to linguistic globalization.

Globalization, is the interaction of different aspects of human life in the traditional space of place, and the current revolutionary space of flows. Globalization has been used to describe many things from different perspectives, ranging from a process, a condition, a force and an age. it is a 'set of processes that manifests in cultural, economic, political, technological, ecological and many other dimensions' (Adegbola, 2004). But basically, it involves some kind of interaction. Steger (2003), cited in Adegbola (2004) presents four features of globalization:

- Creation and multiplication of social network activities that go beyond the traditional political, economic, cultural and geographical boundaries.
- Reflection in the expansion and stretching of social relations, activities and interdependencies.
- Intensification and acceleration of social activities and exchanges.
- Involvement of the subjective plane of human consciousness.

Linguistics as a scientific study of language deals with different aspects of language. Traditionally, it deals with analyses and description of languages with the aim mostly to develop the languages, and have some form of record. A consequence of this is an array of descriptive works, dictionaries, grammars, orthographies and use of such materials in developing a literary tradition. The linguistic enterprise therefore has several branches, all of which begin with time consuming elicitations, analyses and descriptions.

Linguistic globalization involves the interaction and sharing of the Linguistic enterprise, particularly language data in both the

space of place and the space of flows. For linguistic data to be stored, shared and accessed in space, they need to be in electronic forms. This has changed the nature of language documentation and language documentation has had to conform to globalization tools. Naturally with this medium of operation from place to space, there has also arisen the need to extend the types of media used, particularly as multi-data are involved. Language documentation is done using several media like video, audio, text and meta.

Language Description and Documentation (Himmelman, 1998)

Linguists have always done some form of documentation, as a way of recording linguistic data. Traditionally, from elicitation of data from different aspects of a language, and subsequent analysis, grammars, dictionaries, etc. emerge. These constitute some form of documentation. But in the light of current developments, a new area of linguistics is gradually emerging, whose aims are slightly different from the traditional aims of language description. Himmelmann (1998) has tried to differentiate between description and documentation. The aims of description are different. A record of the language with emphasis on abstract elements, constructions and underlying rules, form the focus of the elicitation, and these precede the analysis. On the other hand, the aims of documentation are basically to record the linguistic behaviour as exhibited in discourse. This shift in aims, methodology and tools has evolved a new area of Linguistics called Documentary Linguistics.

Documentary Linguistics

Documentary Linguistics is a new area of Linguistics. It was motivated by the need to record endangered Amerindian

languages. Its methods are therefore slightly different from previous linguistic analyses. Its primary concern is having a direct record and representation of naturally occurring discourse. Discourse is therefore primary, while description and analysis are secondary by products. Descriptive and explanatory materials like grammars, and dictionaries, annotate the documentary corpus (Woodbury, 2003).

Documentary Linguistics deals with huge multi-media data, which are ultimately stored in archives. It encourages researchers to collect a wide range of linguistic phenomena in genuine communicative situations, and it uses high quality sound and video recording to make sure that the results are the best possible record of the language.

Theoretical Considerations

Documentary Linguistics is based on American Anthropological concerns about the disappearance of Amerindian languages. It has as a theoretical basis the importance of discourse. According to Boas (1911: 60),

> …much information can be gained by listening to conversations of natives and by taking part in their daily life, which, to the observer who has no command of the language, will remain entirely inaccessible.

Documentary Linguistics is therefore based on Boas 'Ethnography of speaking' (Gumperz and Hymes, 1964), which was later expanded to 'Discourse-centered approach' (Sherzer, 1990, Urban, 1991). Emphasis is placed on how natural discourse data can be represented, transcribed, preserved,

disseminated, and made accessible worldwide, and this is done in connection with digital archiving.

The tradition for this new discipline is being set, and its specifications for analysis and what can be considered as best Linguistic practices are currently being worked out in the areas of data annotation and markup, electronic and card catalogue for linguistic data in digital archives, general ontologies, and the curation of linguistic data. Being a new area, the modalities and detailed specifications for its concepts need to be worked out.

There are many language documentation projects all over the world and some of them are currently working on different aspects of formalizing Documentary Linguistics. For instance, the E-MELD has had about four workshops on different aspects, the latest being in July, 2005, which critiqued the General Ontology for Linguistic Description, with the theme, 'Linguistic Ontologies and Data Categories for Language Resources'.

Language documentation is done in connection with digital archiving. Emphasis is placed on how natural discourse data can be represented, transcribed, preserved, disseminated and made accessible worldwide. And such access is only possible in the space of flows. The focus at this stage therefore, is setting the right traditions for its formalism in the areas of data annotation and mark-up, as well as electronic and card cataloguing for linguistic data, especially multi-media data in digital archives.

Since a language documentation project aims to collect and create data in different media (audio, video, text) covering use of language in different social and cultural contexts, its priorities are basically:

- To create a range of high quality materials to support description.
- To enable the recovery of knowledge of the language even if all other resources are lost.
- To generate resources in support of language research.

Language Documentation projects create materials in the following media, each of which has both strengths and weaknesses.

- Video-data in video form have immediate and rich authenticity. They are multi-dimensional in content and they can be produced by the community without the assistance of the researcher. On the other hand, they are more difficult to process and store, and they need a lot of annotation which is time consuming and difficult to preserve.
- Audio-data in audio form contain less information, but they are simple. Audio files are easier to work with and there are several software for editing and presenting sound.
- Text-data in text form is the traditional way. Text is compact, stable and easy to store, access, index and revise. But when text is used to document language, it requires a lot of analyses and abstraction. Such analyses can sometimes lose information that were in the original event or record.
- Meta data – This is data about data. It is structured information describing characters of events, recordings and other data files. It is in the form of text, but it is an independent media. Its collection is dependent on the type of materials described. There are several kinds of metadata, like: catalogue (speakers, collectors, time, place of recording, etc.), description (content, relationship to other resources, etc.), technical (performance and preservation information), administrative (responsibilities and access statement).

Basically, the methods, terminology and media all aim at making knowledge about a language accessible to a wide range of audience like: academics, community members, learners and teachers.

Given the nature of this new enterprise, special tools are needed. Of course, the basic fieldwork materials are still useful, like writing materials, recording (audio/video) equipment, etc., but in addition to this, computational resources are also needed for creating, describing, analyzing, viewing, querying or generally using electronic language data. Tools of application programmes, components, fonts, style sheets, converter programmes, etc., are therefore very important.

Some Effects of Globalization on Documentation

The interaction in space has affected several aspects of language documentation, especially in its aims, tools and its products.

- Aims of documentation: Emphasis has shifted from elicitation, analysis and description to discourse. Since discourse is primary, the emphasis is on audio and video recording of such discourse and linguistic behaviour of a people. Description and analysis are secondary, as these are gotten from the record of the discourse.

- Tools of documentation: These have gone beyond the basic fieldwork tools: The tools are those that can be used as computational resources for creating, describing, analyzing, viewing, querying, or otherwise using language data. Such tools are those that can be used as application programs, components, fonts, style sheets, converter programs for transferring data between formats and programs, etc. and they are such that can be used to handle electronic data.

- Products of documentation: The products are mostly in electronic form. Some of these include annotated primary language data, archived as searchable corpora, audio, video, photography, metadata, links and references between the items of data, talking picture dictionaries, audio CDs, CD-Roms, web pages, etc. However, some traditional hard copy products can still be developed from language documentation. For instance, linguistic/ native speakers' grammars, text collections and story books, primers and articles are still products of documentation.

Some Documentation Projects

A lot of work is being done on different aspects of language documentation. Most of the projects are in Europe and America. Some of these include among others: E-MELD, HRELP, SIL, DOBES, OLAC, ALT-I and WALA. The ALT-I, African Language Technology Initiative is based in Ibadan. This is important to us. You may visit its website for more information. The West African Language Archive (WALA) is also close. This is one of the archives hosting West African data. Currently, the WALA which is a result of a cooperation between the Universities of Bielefeld, Cocody and Uyo, has multi-media data on Ega and Ibibio. Data on more languages are needed. More details on these projects and their activities can be found in their websites, some of which are given below:

*   E-MELD (http://emeld.org)
    -     GOLD
    -     SCHOOL OF LINGUISTIC BEST PRACTICES IN DIGITAL LANGUAGE DOCUMENTATION (http://emeld.org/school/).
    -     QUERY ROOMS
    -     THE LINGUIST LIST
*   HRELP (http://hrelp.org/) Hosted by SOAS, London.
*   THE ROSETTA PROJECT (htt:/www.rosettaproject.org/).
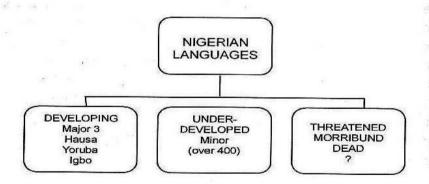
* SIL (http://www.sil.org/).
* DOBES (http://www.sil.org/).
* OLAC (http://www.language-archives.org/)
* ALT-I (http://www.alt-i.org/projects.htm/)
* WALA

The Effects of Linguistic Globalization of Language Documentation in Nigeria

Linguistic globalization has had an effect on language documentation in Nigeria. Of course, we have always documented Nigerian languages in the traditional sense of describing languages. Before we explore the direction of some of these effects, let us look very briefly at the status of the languages of Nigeria.

The Status of Nigerian Languages

The linguistic situation in Nigeria is rather complicated. Within a hierarchy of a geo-political structure and ethnic groupings, there are over 500 languages and dialects. There is an official dichotomy between major languages (i.e. Hausa, Yoruba and Igbo) and the minor languages (i.e. all others). The three major languages have some kind of a 'national' status, and they enjoy

a lot more government patronage from both the Federal and State governments where they are spoken. Perhaps, due to this, an awareness has been created in communities where these languages are spoken and the communities are more involved in their development. On the other hand, the minor languages are at different levels of gross underdevelopment. The huge number may not be the only problem responsible for this lack of attention or interest. It may be more a problem of lack of advocacy and inadequate knowledge of the importance of the language of a people to both its speakers and humanity.

On the basis of the dichotomy that exists between major and minor languages in Nigeria, Nigerian languages can be classified into three groups: developing languages made up of the major three languages, underdeveloped languages made up of about 500 minor languages at different stages of under-development, and a small number that is highly endangered made up of moribund and dead languages. Grimes (2000) records eight in this last group.

Documenting Nigerian Languages for Development

In the light of these, what should we be doing with regard to language documentation in Nigeria in this 21st Century? The area to cover is both vast and fruitful. We need to harness all our resources so that we can document Nigerian languages both for development and archiving.

The vibrant living and developing languages should be documented for development. Such development should aim at different things including producing databases to be converted to dictionaries, grammars, orthographies, etc. It should also aid the communities that speak these languages in preservation of the languages in developing teaching materials, in planning (both

corpus and status) and revival of some languages that may be heading to a moribund stage.

Several benefits accrue to us for the development of Nigerian languages. These include: linguistic analyses, database of primary language data to support this, orthographies, dictionaries, text collections, journal articles, primers and literary works.

Documenting Nigerian Languages for Archiving

For those languages that are threatened, they should be documented with the sole aim of archiving. There should be a plan for regular conversion of data into an active format using XML files, or even tab-delimited text.

Several benefits also accrue to us for the archiving of Nigerian languages. Primarily, we will have huge databases of transcribed and annotated languages data, searchable corpora that can be accessed and shared worldwide in the space of flows. These data can be in multi-media forms, metadata, talking picture dictionaries, audio CDs, web pages and search engines.

In addition to documenting the endangered languages for archiving, they can also be documented for revival and development. It is true that some languages have become moribund, but such language can be revived and then developed. For instance, the Kiong languages spoken by the Akpab Okoyong people of Cross River State, has been moribund for several years, such that it is known (not spoken) by less than 20 people. A documentation of discourse on several aspects of the languages will provide a data base which will form a basis for several things like the production of teaching and literary materials.

Conclusion

There are many products of documentation and most of these are related to the benefits already mentioned for both development and archiving. Some of these include annotated primary and searchable data, in both audio and video forms, talking dictionaries, traditional dictionaries and encyclopedia, grammars, audio Ds, video CDs, primers, web pages, search engines on language matters, especially as they relate to Nigerian languages.

While a new area of linguistics is evolving in the area of Languages Documentation, we need to contribute to that development, especially as we have the data resources from which we can tap. We need to join in the formalization of the subject. But if we are not researching into it, then, we will not have much to offer. And while the specifications are being worked out, many interesting areas of Nigerian languages will be left out of the ontologies that are being fine-tuned now if we do not take part in what is going on globally.

Several workshops testing out some of these specifications by the E-MELD workshop series on Digital Language Documentation are run every summer. The last one held in July, 2005 at Harvard tested out the General Ontology for Linguistic Description (GOLD) specification on some language groups: Native American, Sino-Tibethan, Australian and Austronesian, Caucasian, Turkic and Mongolic. There was no working group on African languages due to almost zero representation. After the details are worked out there may be interesting and important features that could be left out of the formalism which GOLD is drawing. For instance, in the area of property terms, there was nothing on 'ideophones', and when this was brought up, we had problems getting references from the internet to support it like

wąs done for other features that were African specific. The question one may ask at this stage is, are we not doing enough research? But we are. Our research reports should be processed to a level that they can be shared globally.

We should begin to document Nigerian languages in the light of these new developments in the field, both for development and archiving. That way, we will be contributing to humanity and the development of a linguistic formalism for modern languages documentation. NINLAN has an important role to play in all these by way of coordinating the funding, manpower training and the actual research on these languages.

For linguistic data to be shared in the globalization revolution, they have to be stored, accessed and shared in electronic archives. To be able to produce data that can be processed at this level therefore, languages documentation has had to radically shift its form and modus operandi to conform to globalization, especially in this 21st Century Nigeria.

References

Adegbola, Tunde (2004), 'Globalization: Colonizing the Space Flows', Plenary paper Presented at the West African Languages Congress, University of Ibadan.

Boas, Franz (1911), *Handbook of American Indian Languages*, vol 1, Washington D. C. Bureau of American Ethnography, 40.

Grimes, Babara (2000), Ethnologue, Dallas: SIL.

Gumperz, John and Dell Hymes (1964). 'The Ethnography of Communication', American Anthropologist, 66:6.

Himmelmann, N. (1998), Documentary and Descriptive Linguistics, *Linguistics* 36. 161-195.

Sherzer, Joel (1990), *Verbal Art in San Blass: Kuna Culture through its Discourse*, Cambridge, UK: C. U. P.

Urban, Greg (1991), *A Discourse-centered Approach to Culture: Native South American Myths and Rituals,* Austin: University of Texas Press.

Woodbury, Tony (2003).Defining Documentary Linguistics, *Language Documentation and Description,* Vols. 1. Tony Woodbury, (Ed), London: The Hans Rausing Endangered languages Project.