

**Old data and new technologies: the
seamless integration of linguistics, literacy
and translation for Nigerian minority
languages**

Jos Linguistic Circle

Jos, 13th March, 2019

Roger Blench

McDonald Institute for Archaeological Research, Cambridge

Department of History, University of Jos

Kay Williamson Educational Foundation

Visiting Professor, University of New England, Armidale

Academic Visitor, Museu Emilio Goeldi, Belem, Brazil

Where are we?

- ❖ Nigeria has over 500 languages; so it is one of the countries in the world with most languages in absolute numbers, but also in terms of diversity
- ❖ Indonesia has more languages in absolute terms, but they are mostly Austronesian and thus much more similar
- ❖ Expanding aspirations of minorities have had the consequence that far from waiting for them to die, there is active agitation to develop even quite small languages
- ❖ As a consequence, there is far more demand for language development than there is local or international expertise
- ❖ Why should this be?
- ❖ Well, Nigeria is not such an easy country to work in for outsiders
- ❖ And, more disappointingly, the universities, despite a plethora of departments of linguistics, have contributed little

How to get things done?

- ❖ One of the major drivers of a push for literacy has been the translation of scripture
- ❖ This started as a rather stately SIL enterprise, but there are now a large number of players, often making improbable promises about the speed of achievement and the quality of output
- ❖ This sounds good at first, but in reality it has meant an explosion of amateur work
- ❖ And the complexity of Nigerian languages is such that this isn't really a job for amateurs
- ❖ The consequence has been a major waste of human and financial resources, literacy programmes moribund and translations unread.

Does this need to be the case?

- ❖ Obviously not. A confluence of new technology, greater energy and enthusiasm from minorities and a generation of trained students makes possible a more streamlined approach
- ❖ Which would produce;
 - ❖ A comprehensive sociolinguistic analysis prior to project initiation
 - ❖ A sophisticated linguistic analysis which should pass international scrutiny
 - ❖ Literacy and aural materials combined with an outline strategy of sustainability in a broader media environment
 - ❖ Translation and preparation of written materials
- ❖ This presentation will outline how this might be achieved

Bringing new life to old data

- ❖ An initial step would be to bring back to life the rich heritage of material from the paper era, much of which is out of print, or circulates in Xerox or pdfs
- ❖ The dictionaries of major languages, such as Hausa, Yoruba and Igbo as well as forgotten masterpieces such as the 2-volume Nembe dictionary by Kaliai
- ❖ Could be made accessible through new versions for Android phones, with sound files, imagery, modernised orthography
- ❖ This is important, because the next generation of Nigerian speakers definitely want to use phones and tablets for these purposes
- ❖ Paper is so last century..

The first step I

- ❖ There's no way round it; these have to be typed. OCR is simply not practical because of the abundance of special and non-Unicode characters
- ❖ But anyway, retyping allows you to reformat the text with tab delimiters and other useful markers which will in turn make possible the transition to a database
- ❖ Unfortunately, early retyping mimicked the format of the printed page, making this type of repurposing far more difficult
- ❖ I've still never managed to move Kay Williamson's Igbo dictionary to a useful format

The first step II

- ❖ However, over the last thirty years I have been creating a variety of dictionaries of Central Nigerian languages
- ❖ Some on a large scale such as the Berom dictionary
- ❖ There are now about twenty of such dictionaries mostly in Word format
- ❖ So, with careful preparation you can then;
 - ❖ Move them into Flex
 - ❖ Create orthographic columns and IPA columns
 - ❖ Add a reverse
 - ❖ Add pictures
 - ❖ And sound files
- ❖ And then they can be exported to Dictionary App Builder and used on Android devices

What are the advantages?

- ❖ First and most important is that this is a way of stimulating youth to explore and enrich their speech
- ❖ Nigerian minority languages are fortunately not disappearing at as fast a rate as once feared, but they are being creolised with heavy borrowing
- ❖ Sound files help ensure correct pronunciation
- ❖ Pictures help recall increasingly rare cultural items
- ❖ And the orthographic columns can reflect changing research but are there for writers and translators to consult easily and conveniently

- ❖ It is for this reason the IPA is essential; orthographies are not stable, nor probably should they be

- ❖ Android dictionaries can be made available on various App stores for nominal payment and regularly updated

OK, how would 'seamless integration' work?

- ❖ A major cause of faulty procedures is a confusion about goals and the route by which they can be reached
- ❖ If your only goal is, for example, a printed New Testament, and linguistics/literacy is some annoying detour that is required, then it is extremely likely the project will fail, even if you have a printed volume to show the sponsors
- ❖ If you actually don't care if it is read, and don't worry about piles of material gathering dust, there's nothing to say
- ❖ But you can go about things another way, and the result will be far more satisfying

I: The sociolinguistic survey, indispensable 1

- ❖ Nigeria remains a complex country which is poorly known linguistically
- ❖ And frequently, what we think we know is wrong
- ❖ You might think, therefore, that it would be useful to gather and make use of background information on ethnolinguistic communities before spending considerable human and financial resources on a project
- ❖ You can find out;
 - ❖ Numbers and distribution of the population
 - ❖ Status of the language and the role of LWCs
 - ❖ Religious affiliation of the population
 - ❖ Existing levels of literacy and language attitudes (do people want to write their language?)
 - ❖ Use of language in broadcast and social media
 - ❖ Distribution and prestige of dialects

I: The sociolinguistic survey, indispensable 2

- ❖ Useful? So you might think
- ❖ But in fact most projects go ahead in blithe ignorance of such basic facts
- ❖ Often they are prey to 'enthusiastic pastor' syndrome, where an individual (often from a marginal dialect) convinces the organisation to undertake a project, regardless of any likelihood of success
- ❖ Unless you redefine success
- ❖ Resources are finite, they have to be allocated
- ❖ Use surveys to allocate resources efficiently

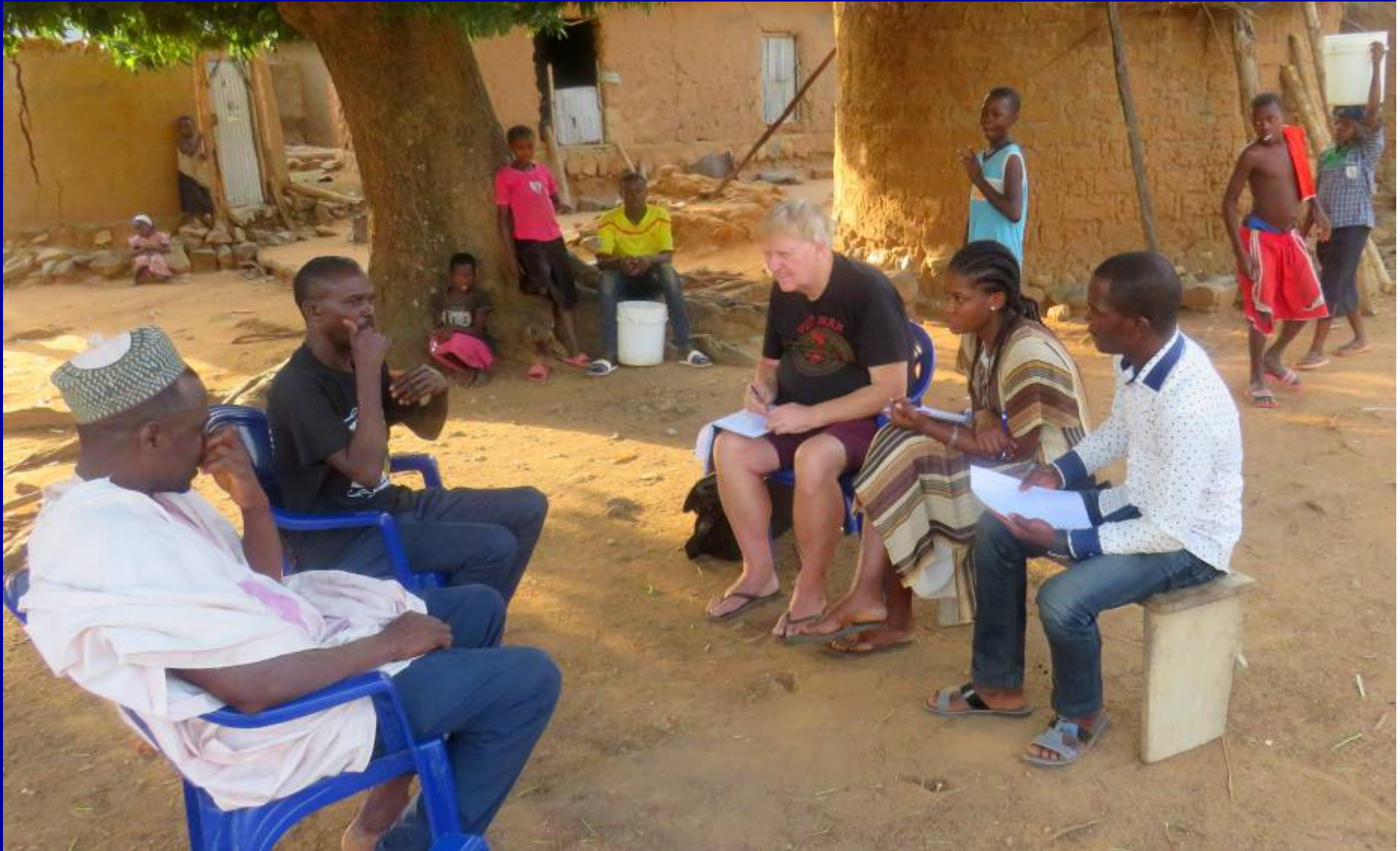
II: Read the literature

- ❖ You might think this one was fairly obvious as well
- ❖ But in fact, project after project (both literacy and bible translation and academic projects) go ahead without consulting prior publications
- ❖ They re-invent the wheel and sometimes rather badly
- ❖ Build on the findings of the past, don't repeat them
- ❖ Through the magic of the pdf it has never been easier to build up a library of existing publications
- ❖ It will save you time and perhaps you might learn something

III: Do the linguistics 1

- ❖ One of the more bizarre assumptions in circulation is that all languages can be 'done' according to some schedule
- ❖ Apart from the unfortunate fact that Nigeria shelters some of the most complex languages in the world
- ❖ Languages are extraordinarily different from one another and simply take different amounts of time to analyse and document
- ❖ Another assumption frequently made is that what you need is phonology and discourse and everything else will somehow take care of itself
- ❖ Nope. The phonology may be relatively simple, but the morphology almost never is
- ❖ You simply can't do word division consistently unless you have understood the morphology

Tal work in Balong



III: Do the linguistics 2

- ❖ In Mwaghavul, written for nearly a century, we are still arguing over word division
- ❖ Similarly, there are areas of phonology which still frighten people so they ignore them and hope they go away
- ❖ Tone...strikes fear into the heart of many including professional linguists
- ❖ But with the exception of Fulfulde, all Nigerian languages are tonal and some constitute global records, such as the five-level languages in the Jukun area
- ❖ Crucial lexical and morphosyntactic distinctions are made by tone
- ❖ And these *must be understood* before reaching orthographic solutions

III: Linguistics is open science

- ❖ Linguistic analyses need to be reviewed, criticisms taken on board and revised
- ❖ This is standard procedure in science, and this is science
- ❖ All too often it is said the phonology has been 'done' although there is no public document to this effect and no evidence it has been reviewed
- ❖ There is *no reason at all* not to publish drafts on the web and invite comments
- ❖ Similarly, moving from the underlying phonology to orthography should be transparent
- ❖ As well as the process of testing the orthography and the basis for community acceptance
- ❖ Leave an (electronic) paper trail

Multiple languages at work



IV: Reading and writing 1

- ❖ The next phase on from the orthography is to consider how to make reading and writing widespread and most importantly sustainable
- ❖ The first stage is the preparation of a reading and writing book
- ❖ This translates the linguistic insights into a practical writing system in a way which can be used by textbook preparers and teachers
- ❖ It is remarkable how many far advanced literacy and translation projects do not have such a guide.
- ❖ With this is hand, it is possible to start preparing teaching materials, in particular primers and wallcharts
- ❖ And these have to be tested and reprinted at regular intervals

IV: Reading and writing 2

- ❖ Literacy projects are another graveyard. There are many examples of literacy materials which were used in the preliminary phase of a project but which now lie gathering dust
- ❖ There's a strong tendency to imagine that if people turn up at church on Sunday with a printed Bible under their arm this shows reading is widespread
- ❖ This simply isn't true. And the failure of many churches to move to the vernacular (.e.g. Hausa), *even in monoglot areas*, shows the process is faltering
- ❖ The most important way vernacular literacy can be embedded is with the school system
- ❖ Attempts to do this go back to Hans Wolff (1954)
- ❖ Strangely, the situation was better thirty years ago than today.

IV: Reading and writing 3

- ❖ However, there is evidence for new initiatives to re-introduce vernaculars into the school system in Plateau, at least
- ❖ The workshops held by CONAECDA are pressing for pilot projects in six (?) languages
- ❖ Obviously government is not really going to be proactive in preparing school materials. But if you hand it to them this gives them the opportunity to print and allocate teachers
- ❖ So it is also worth talking to potential teachers about their skills and availability
- ❖ If reading and writing is embedded in the school system, then this is worth multiple informal classes taught by volunteers

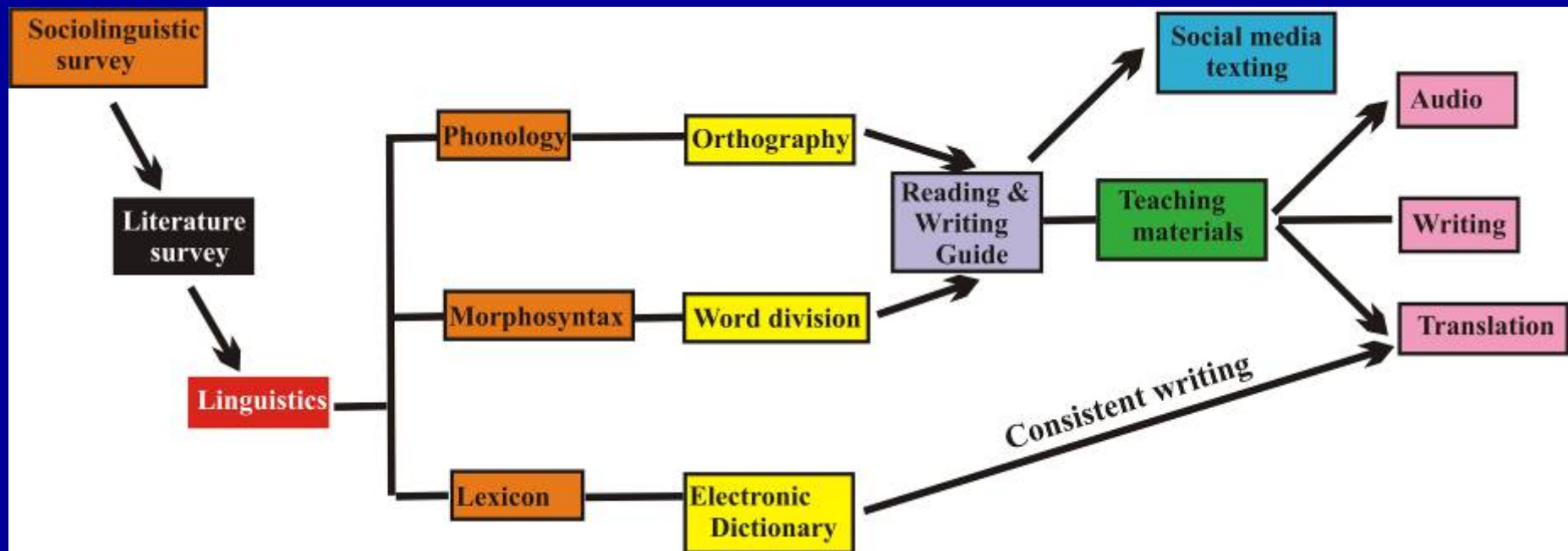
V: Keeping the community informed

- ❖ One of the disadvantages of earlier non-connected eras was the long time it took for any output to reach the community
- ❖ Now there is a younger generation with high levels of enthusiasm for linguistic and cultural development
- ❖ Technology can enable us to return results quickly, for example by printing a low-run physical dictionary or out putting Android phone dictionaries
- ❖ These also enable rapid feedback from the community

VI: Translation and general writing

- ❖ There is a strong argument that unless these steps are gone through systematically it is really not worth beginning scriptural work or indeed the production of other reading materials
- ❖ Nigeria is rich in meetings of worthies discussing and even preparing translations with no relevant expertise to hand
- ❖ This doesn't mean there is no product, but it does mean it is unlikely to be truly read as opposed to being an iconic product
- ❖ Part of this is pressure from funders who really don't understand all these preliminary processes and perhaps wouldn't care if they did
- ❖ But even those mainly motivated in translation perhaps see these other aspects as detours on the road to the main goal

Seamless integration flowchart



They do things slightly differently elsewhere



THANKS

- To a wide range of individual linguists who have made unpublished material available
- To my many collaborators over the years including quite a few in the room
- To SIL for friendly relations and software assistance especially Robert Hedinger
- To KWEF, for funding

