

Training in documentation

Felix K. Ameka



Universiteit Leiden

Universiteit Leiden. Universiteit om te ontdekken.

Attributed to Roman Jakobson

Linguista sum, nihil
linguistici a me
alienum puto

Outline

- Introductory remarks
- Some historical background reflecting on the context of the emergence of documentary linguistics
- Refreshing our memories: what is language documentation
- Skills for language documentation
- Interdisciplinarity
- Documentary products and resources
- Orthographies, standardization, literacy

HISTORICAL REFLECTIONS

Linguistics curricula pre 1992

- The cry for field methods courses in linguistics programs and for more descriptive work in graduate schools leading to the production of the trilogy for many languages happened around the same time as the call to arms for linguists to care about endangered languages.
- (Newman 1991 (reprinted 2009 in LDC);
- David P. Wilkins 1992

Linguistics curricula 2

- Linguistics training had shifted from producing people describing languages to people whose goal was to apply the latest model to a part of a language (or their idiolect) to show whether a particular principle of UG holds or the model needs some modification here and there.

Data, empiricism, speakers

- A second problem at the time was the empirical bases of the claims being made:
 - Representativity of data
 - Accountability
 - Recognising variation
 - Sociolinguistic variationist studies do not speak to parametric variation theorists

A view on data

- Bickerton (1996: 157) quoted in Lefebvre (1999: 383):
- “It is a foundation stone of our discipline that native speakers of a language do not and in the nature of things CANNOT ‘correct’, ‘modify’, ‘rectify’, or ‘repair’ the judgments of other native speakers: they just make different judgments” (emphasis in original)

Saddock 1996 on PIFL

Principle of Information Free Linguistics

- Covert scrambling at LF in Tophar
- (i) a Pictures of themselves seem to him to bother each other
- b *Pictures of him seem to them to resemble each other
- (ii) a Whose self do picture of each other bother him?
- b *Where is the nearest men's room?

PIFL - 2

- “I am a native speaker of Tophar [pronounced [dopHAR] – Oops!] ... and find your (ia) and (iia) horribly ungrammatical and your (ib) and (iib) just fine. Doesn't that cast some doubt on your principle of LF scrambling?”

PIFL – 3 Linguist's response

- “Uhhh, Noo. To the theoretically naïve in the audience, the dispute over the accuracy of your data might seem pretty serious, but given the gigantic gulf between data and theoretically relevant fact and the further chasm between fact and principle, it is not really of any consequence. LF scrambling could well exist **regardless** of any data from Tophar or any other language, for that matter.

PIFL –4 Linguist's response

- Given what we know of UG, LF scrambling is a clear possibility, a virtual conceptual necessity, as a matter of fact. So in this case, simply smile patronizingly and say “My data are different from yours”, for indeed they are.” (Sadock 1996: 136)

What is the “best record”?

- 1995: Best records workshop at the MPI for Psycholinguistics, Nijmegen, organised by David P. Wilkins.
- Himmelman’s famous 1998 paper was presented there
- It was evident at the time for revitalization projects that grammatical descriptions, dictionaries and monologue narrative texts were not enough

Best reord?

- One could not find in the grammars how the speakers at the time for example, greeted one another?
- One could find questions and one could find the structure of answers but one could not find how these were used in everyday interaction.

Holes in the record

- A problem with the descriptive trilogy is that gaps and holes tend to occur.
- They do not capture specific ways of expressing oneself idiomatically in a given variety
- For example: Formulaic language used in everyday interaction is still very hard to describe and there is no methodology or theoretical model for incorporating them
- Into descriptions . (Himmelman)

Summer School on Fieldwork

- 1993: A summer school on fieldwork funded by the VW Stiftung was held at the University of Cologne- somehow the precursor to many of the training programmes today in field linguistics and language documentation.
- Colette Grinevald was one of the distinguished lecturers at that school



DEVIL'S
MEAN'S
EXIT-RED

THE DOCUMENTARY LINGUISTICS TURN

Language loss

- Language loss leads to an irrevocable loss of human heritage (David Crystal 2000)
- Each language that vanishes without being documented leaves an enormous gap in our understanding of some of the many complex structures the human mind is capable of producing (David Harrison, quoted in Financial Times March 26, 2004)

Noam Chomsky on the loss of language

- The loss of a language “is much more of a tragedy for linguists whose interests are mostly theoretical like me [Chomsky] than for descriptive linguists who focus on specific languages since it means the permanent loss of the most relevant data for general theoretical work”
- Chronicle of Higher Education (1 June 2009)

- However not all practitioners feel this way about data or the empiricism that characterises documentary linguistics.

Data and linguistic methodology

- Robert Friedin 2009 : A note on methodology in linguistics (Commentary on Evans & Levinson p. 355 BBS)

Evans and Levinson say:

- “Each constraint in UG needs to be taken as no more than a working hypothesis, hopefully sufficiently clearly articulated that it could be falsified by cross-linguistic data”

(sect 3 para 6)



Data and linguistic methodology-2

- The first part is a truism. However, the assumption in the second part that data alone can falsify any working hypothesis (especially one that works for a significant range of data) is clearly false.
- Data alone cannot speak to the validity of explicit proposals about the content of UG.

Data and linguistic methodology-3

- What is required is an explicit analysis of data that follows from a precisely formulated fragment of a grammar.
- In science there is no alternative to providing explicit analysis of data.

Himmelman's view

Documentary linguistics is a new 'turn' in linguistics that is "concerned with the methods, tools, and theoretical underpinnings for compiling a representative and lasting multipurpose record of a natural language or one of its varieties" (Himmelman 2006: v)

Woodbury's view

- Documentary linguistics is “the creation, annotation, preservation and dissemination of transparent records of a language.”

(Woodbury 2010, see also 2003)

Dimmendal's view (2010)

- Describing languages in the sense understood in traditional descriptive linguistics means collecting, transcribing, translating, and analyzing linguistic data. But the experience with the description of endangered languages of course had shown

Dimmendaal's view (2010) -2

- that this undertaking was not quite enough to document an endangered language, e.g., for those interested in revitalization. Grammars as abstracted and often idealized analyses of structures do not quite reflect language (“parole”) as a living medium.

Dimmendaal's view (2010) - 3

- Whereas in the descriptive tradition statements may be illustrated by one or two examples, the result of language documentation in principle should be a comprehensive record of the essential linguistic practices and traditions characteristic of a given speech community.

Dimmendaal's view (2010) -4

- By definition, this kind of archiving requires various kinds of data. Data collected with the aim of documenting (rather than describing) a language also requires the collecting of information of potential relevance for other disciplines, including sociology, anthropology, and the study of oral history.

Dimmendaal's view (2010) -5

- As pointed out by Himmelmann (1998:163), this presupposes that the data set contains data and information amenable to the research methodologies of these disciplines. Given the fact that one is often dealing with endangered languages,

Dimmendaal's view (2010) - 6

- the speech community itself might also be interested in a record of these linguistic practices.

Key features of language documentation

- Collection and analysis of a varied array of data made available for a wide range of users
- Explicit concern for accountability- makes evaluation of linguistic analyses possible

Key features of language documentation -2

- Concern for long term storage and preservation of primary data- a focus on archiving ensuring that documentary materials are available to users
- Work in interdisciplinary teams- requires input from other disciplines and not restricted to linguistics

Key features of language documentation -3

- Close cooperation and collaboration with and direct involvement of the speech community (Himmelman 2006)

Austin 2010 adds:

- Diversity- documentation projects show a diversity of approaches, techniques, methodologies, skills and responses

Components of documentation research

Recording – of media and text (including metadata) in context

- Transfer – to a data management environment
- Adding value - the transcription, translation, annotation and linking of metadata to recordings
- Archiving – creating archival objects
- Mobilisation – creation, publication and distribution of objects

Documentary corpus

- The documentary corpus will ideally cover a *diverse* range of genres and contexts and be *large, expandable, opportunistic, portable, transparent, ethical and preservable* (Woodbury 2010)

Multipurpose corpus

Documentary corpus should be able
then to lead to - Revitalisation

- Preservation
 - Conservation
 - (re) valorization of languages
- to be used for
- Interdisciplinary research
 - Production of pedagogical material

WHAT SKILLS TO TRAIN

The formidable enemies

- “Those of us who are concerned about the endangered languages question and the problem of language extinction are up against a formidable enemy – and that enemy is our own discipline of linguistics and the individuals who make it up.”
(Newman 2003:11)

Skills for language documentation

- (see Austin 2007; Jukes 2011)
- Basic linguistic analytic skills
- Field methods
- Field skills - a knowledge of ethics, health and safety, hygiene, culture shock and dislocation, how to handle gender and sexuality in the community, and other capabilities that contribute to a fieldworker's well being

Skills for LD - 2

- Skills in the application of information, communication and media technologies
- Knowledge and skills typically associated with applied linguistics areas such as
 - Orthography development
 - Lexicography
 - Translation
 - Pedagogy and curriculum design

Skills for LD -3

- Multimedia
- Language policy
- Needs assessment and advocacy

- These arise from community desires as well as the interdisciplinary nature of documentary linguistics

Specific skills for LD projects

- Austin (2007: 28-9) Jukes (2011:426-7)
- **Project conception, design and management**
 - Familiarity with documentation theory
 - Applied ethics, intellectual property rights etc.
 - **Grant application writing**

Specific skills - LD projects -2

- **Media Management**

- Recording techniques, field methods, data transfer, back-up

- **Data and metadata management**

- Data representation, transcription, linguistic analysis, annotation, use of software tools, workflow design and management

Specific skills - LD projects -3

- **Mobilization**

- Applied aspects

- **Team based research**

- skills sharing, transfer, capacity development

- **Reporting**

- Presentation, writing and communication skills

Intellectual capacity building

The kind of training received impacts on the type of documentation work done

Documentation requires a certain mind-set

Any training for documentation and in documentation should give opportunities to the documenters

- to think in interdisciplinary terms
- to reflect on the pitfalls in documentation
- to question and examine documentation practices

Some dangers and pitfalls in documentation work

- “hyper-traditionalisation”
- “museumisation”
- “hypervalorisation”
- divide and rule
- creating uniformity ignoring variation
- linguistic purism – ignoring contact and multilingualism

Contact in documentation

- Language contact is important in ... language documentation. However, fieldwork focused on individual speech varieties has tended to see multilingualism as a problem rather than an opportunity, as a source of contamination of the data under consideration rather than something to study in its own right. ... excluding contact means excluding potentially relevant data. (Bower)

Some questions to reflect on

- Is writing integral to revitalisation?
- Are the documentary products for the communities produced in the right way, do they have the right effect?
- Do some of these processes kill the very diversity we are trying to document?

INTERDISCIPLINARITY

Facets of interdisciplinarity

- Language documentation is by its nature **multidisciplinary** (Austin 2010: 12)
- It draws on “**concepts and techniques from linguistics, ethnography, psychology, computer science, recording arts and more.**” (Woodbury 2010)

INTER- vs MULTI- disciplinarity

- Documentation work is best carried out in interdisciplinary teams.
- HOWEVER For me there is a difference between multidisciplinary and interdisciplinary approaches
- Multidisciplinary approach involves the juxtaposition of the different disciplines – each person or each discipline doing their own thing, the only common thing is that we are working together on the same subject; no synergies are created

Interdisciplinarity-advantages

- Interdisciplinary approach means asking research questions and developing methodologies that are inspired by the different disciplinary perspectives (cf. Widlok 2004, 2005)
- Such an approach:
 - Facilitates the acquisition of funds
 - Allows researchers to cover more ground and creates synergies (more not in terms of size but in terms of representation in terms of text genres)

Interdisciplinarity-advantages

-2

- Facilitates multi-media and multi-modal representation (one medium is not enough)
- Enrichment of the components of the trilogy- vocabulary list is not enough but augmented by different kinds of information not traditionally included in vocabulary lists or dictionaries

Interdisciplinarity

- Helps in determining how to organise data in the corpus-cross disciplinary work facilitates the identification of genres, and situations as well as how the speakers define them

Difficult but not impossible

- In our experience, true interdisciplinary research, especially in teams carrying out fieldwork in remote locations, is difficult to achieve, both because of theoretically different orientations and practical differences in approach (Austin and Grenoble 2007: 22)
- Language documentation opens new doors to multi-disciplinary collaborations but we need to work out how to achieve it (Austin 2010: 27)

How can it be achieved?

- (i) Train language documenters to think in interdisciplinary rather than disciplinary terms
- (ii) Ask questions that are informed from an INTER rather than a MULTI or MONO disciplinary perspective
- (iii) develop methods for addressing the questions holistically

Some research areas for interdisciplinary work

- Non-verbal interaction (gestures and how to study them)
- Language contact effects – multilingualism
- Variation within communities
- Change – In most contexts languages don't disappear en bloc, it is various domains of use and segments of knowledge that are lost gradually. How should documentary linguistics handle that

- Context and genres
- How to describe and model findings from interdisciplinary research

DOCUMENTARY PRODUCTS AND RESOURCES

Audio-visual corpus

- For the community we generate multi-media outputs such as CDs, DVDs, etc. and we make the material available online through the world wide web.
- We tend to think that there are big infrastructural requirements and problems associated with these; e.g. communities may not have the means to play the CDs; no access to internet etc.

Written materials

- Grammars, story books, spelling books , dictionaries (based on and extracted from the corpus).
- We tend to think these do not have any infrastructure problems
- We forget that wrting is a technology
- Associated with certain practices, especially literacy practices, and the process of creating these written materials leads to norm and standard creation,

Written materials 2

- leveling variation and introducing linguistic purism or correctness ideologies at least in one medium

The “sins” of language documentation practitioners

ORTHOGRAPHIES, STANDARDIZATION, LITERACY

Diversity - a feature of humans

- In the language sciences, there is slow realisation that diversity is the most striking feature of human beings- there is no other animal on the planet, as far as we know, which has such myriad variants of form and meaning at every level in its communication system. Levinson 2012: 2

“The original sin of the Cognitive Sciences”

This is the original sin of the cognitive sciences- the denial of variation and diversity in human cognition ... we know that brains of literate persons are substantially rewired compared with their illiterate siblings.
(Levinson 2012: 2)

The “sins” of documentation

Perhaps one of the sins of language documenters is that some of their practices tend to kill the very diversity that they seek to document. This is one of the aspects that must be addressed in any documentation program in and for documentation.

One can also say that revitalisation programs which also base themselves on schooling and literacy materials are also guilty of this sin as they ignore the context of the languages.

Some reminders

If these communities do not have a written tradition, should writing be the medium of teaching the languages?

We should not forget that writing is technology.

We should not forget that literacy comes with particular skills

We should not forget that cultures and communities differ in how they acquire knowledge:

-in literate societies it is getting knowledge from the written word



Some reminders 2

- In other cultures the mode is that of observation, memorisation, imitation/ repetition
- Communities differ in their socialisation patterns: master-apprentice; independence and interdependence etc.
- What do such differences tell us about documentation and revitalisation practices and processes

Challenges

- An example from Totela (Zambia, Crane 2009).
- Mismatch between sentence and intonation units
- “A large challenge in the revision of the stories was the lack of a written tradition in Totela, which made the delineation of sentences quite problematic; some sentences in the revised versions do not match with the intonation patterns of the original recordings.”

Totela -2

- “In addition, Totela and the regional lingua franca Lozi have influenced each other heavily for many years in the area, and there was sometimes disagreement about which words were Lozi borrowings. In some cases, the younger revisers attempted to introduce constructions from Lozi.”

Totela -3

- “The lack of written Totela also had some advantages: although some general prescriptive language rules were known to the speakers through their schooling, work with the texts can be considered largely oral to oral revisions, an effect that might be difficult to reproduce in a language with a long-standing written tradition.”

Totela -Types of changes

- “Many of the changes seen between the originals and the revised versions served to clarify or ‘polish’ the texts. Such changes included the removal of extraneous material ((1a) vs. (1b)) and code switching, the removal of interjections ((2a) vs. (2b)), overt character specification ((2a) vs. (2b)), direct/indirect object specification, and the amendment of plot details.”

Totela - 4

- (2) (a) *Mm! Bà-yá*
 - INTERJ 3PL.PST-go
 - *kù-mpìlì kù-kà-limà-limà.*
 - to-field **INF-DIST-cultivate~REDUP**
- ‘Mm! They went to the field and busied themselves cultivating there.’
- Nyawi-Nyawi transcription, line 8

Totela - 5

- (b) *Abanakazi noku-ya ku-mpili.*
- women COM.INF-go to-field
- *Ku-ka-sika-ko*
- INF-DIST-arrive-there
- *noku-ka-lima-ima*
- COM.INF-DIST-cultivate-REDUP
- 'Then the women went to the field. Upon arriving there, they busied themselves cultivating there.'
- Nyawi-Nyawi revision, line 8

Restricted genres? Ideophones in Spoken vs. written

- “... ideophones are part of spoken language ... rather than written language.” (Voeltz and Kilian-Hatz 2001:3)
- “... existing written materials (school textbooks and Bible translations) have few ideophones.” (Amha 2001: 49)
- “Ideophones in Dagaare ... are very much used in oral descriptions” (Bodomo 2004: 7)

Standardization -1

- Standardization kills diversity and bad orthographic practices may stand as obstacles to literacy.
- Some would say books and literacy materials and literacy practices as a whole add value and prestige to the language
- BUT at what cost?

Standardization -2

- It is also possible to learn an important lesson from looking at the history of highly standardised languages like English—namely that standardization is not a prerequisite for literacy. We ... know that a very impressive body of literature was written in English long before the language achieved the level of standardization that exhibits today. (Rheg 2004)

Some concluding issues

- Bad documentation practices can be an obstacle to the maintenance and recording of linguistic diversity
- Language documentation is too important to serve linguistics alone!
- Time for a sea change in linguistics (Levinson and Evans 2010)

Some concluding issues -2

- Time for documentary linguistics to have greater impact and shape the goals of the language sciences.
- We need more practitioners of language documentation to influence the mindset of linguists.
- They can only do this if we set our house in order by thinking about the challenges, developing ways of addressing them and above all developing a critical and inquiring mind to question our own practices

Gratias

- Akpe kakakakakakakaka!!!!
- Merci beaucoup
- Thank you very much
- Dank u wel!!!!

Ghana-Togo Mountain area



Ghana-Togo Mountain area

