

3L Lyon Summer School 2008 Course and Lecture Descriptions

1. Plenary Lectures

What is language documentation? How does it differ from language description?

Peter K. Austin,
Monday 23rd June 10:30 am

Documentary Linguistics is a new field of linguistics “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 1998, 2006). Himmelman (2006:15) identifies important new features of documentary linguistics:

- Focus on primary data – language documentation concerns the collection and analysis of an array of primary language data to be made available for a wide range of users;
- Explicit concern for accountability – access to primary data and representations of it makes evaluation of linguistic analyses possible and expected;
- Concern for long-term storage and preservation of primary data – language documentation includes a focus on archiving in order to ensure that documentary materials are made available to potential users into the distant future;
- Work in interdisciplinary teams – documentation requires input and expertise from a range of disciplines and is not restricted to linguistics alone;
- Close cooperation with and direct involvement of the speech community – language documentation requires active and collaborative work with community members both as producers of language materials and as co-researchers.

Some people draw a distinction between:

- *language documentation*: activity of systematic recording, transcription, translation and analysis of the broadest possible variety of spoken (and written) language samples collected within their appropriate social and cultural context
- *language description*: activity of writing grammar, dictionary, text collection, typically for linguists

In this lecture we will explore the nature of documentary linguistics, and similarities and differences with language description. We will finish by looking at some current issues in language documentation.

Data management

Peter K. Austin
Tuesday 24th June 9:00 am

The core of a language documentation is generally understood to be a *corpus* of audio and/or video materials with transcription, multi-tier annotation, translation into a language of wider communication, and relevant metadata on context and use of the materials. The corpus will

ideally be *large*, cover a *diverse* range of genres and contexts, be *expandable*, *opportunistic*, *portable*, *transparent*, *ethical* and *preservable*. In this lecture we will look at:

- What the corpus should contain
- How the data in the corpus should be represented
- Good practices in data representation and data management
- The nature and role of metadata
- Corpus management and work flows
- Overview of tools for language documentation

We will look at some data samples to explore each of these concepts.

Archiving

David Nathan

Wednesday 25th June 9:00 am

Although some people see archiving as the end point of language documentation, archives and archiving principles are involved in various phases of the documentation data lifecycle, starting from the very beginning of projects, and archives should not be considered the sole final goal for documentation materials.

This session describes various kinds of archives and their functions, including storage and security for long term data preservation, and providing controlled access to data. Today's digital archives work with depositors at many points throughout their projects, from providing training and advice to documenters who are starting out, to evaluating samples, to working with depositors to convert materials to archive formats, and providing facilities for depositors to continue managing their data and metadata to keep it current. Modern digital language archives place equal emphasis on both preservation and access, so that archives and their collections should cater for the needs of the eventual users of data.

The session will describe what types of data can be archived in a modern digital archive, how to select and prepare materials for deposit, with details of formats, packaging and transmission, and types and preparation of metadata.

Linguistic Anthropology: Linguistic Practices and their Cultural Context

Lolke van der Veen

Friday 27th of June, 9:00 am

One way to analyze a language is by looking at it as a complex formal device comprising a set of cognitive lexical and grammatical forms as well as a set of rules or constraints capable of producing the well-formed utterances of the language. In this case, the linguist examines the system of the language, i.e. its structural and functional features, in a fairly abstract way.

Another important way to examine a language is by looking at the different ways it is actually used by its speakers within a specific cultural setting. Following this option, the linguist analyzes the actual linguistic practices of a particular speech community. These practices widely differ from one community to another. Cultures impose different norms for the use and form of language in similar social situations, and linguistic practices that are more or less specific to the corresponding languages accordingly. Anthropological linguists study

linguistic productions (texts or parts of them) in the light of these specific social and cultural conventions. How do (oral or written) texts relate to *genre*, *discourse* and *social practice*? Identical linguistic constructions or entities may take different values according to the genre they occur in. In order to demonstrate that cultural values and belief are constitutive of linguistic reality and that social activities bring into play specific relationships between three different semiotic spheres (i.e. semiotic, cognitive and physical), examples will be drawn from an array of languages of the world and in particular from a case study of Galwa (Bantu, Myene) greetings from Gabon.

Fieldwork Methodology And Ethics

Colette Grinevald

Monday 30th June 9:00 am

This conference will consider the multidimensional and complex intertwining of issues of fieldwork and ethics, approaching this complexity for various angles. It will first consider questions such as who is accountable to whom, when, for what (as in the VW-DOBES flow chart), and then the evolution of fieldwork frameworks over the last decades (à la Cameron, as in fieldwork *on/for/with*, and now *by* speakers), while contemplating the inherent differences across continents. It will then consider the whole process in its cyclic nature, with the three phases of fieldwork of *before/during/after* times, each phase bringing up different ethical issues (from choosing language and field site and planning fieldwork, to entering the field, gathering data, and leaving the field, to returning to academia and producing (analysis, archives, materials). The conference will recommend in particular the 2000 AIATSIS ‘Guidelines for Ethical Research in Indigenous Studies’ from Australia, with its 11 ‘principles’, explained and illustrated by practical applications, and organized in three clusters of considerations:

- A. consultation, negotiation, mutual understanding
- B. respect, recognition and involvement
- C. benefits, outcomes and agreements.

Most of all, the conference wants to emphasize how besides, or beyond, the legalistic, technological and socio-political aspects of ‘ethics’, is the sense that to be ‘ethical’ is to adopt an attitude or a posture, recognizing ethics is an ongoing process, to be considered at every turn of a project, in each specific field situation of the encounter of a linguist with a language of a particular community, living in a particular country.

Data collection methods in field-based language documentation

Friederike Lüpke

Tuesday 1st of July, 9:00 am

Language documentation – in the sense of the creation of extensive corpora of annotated speech data in audio and video format – is a newly emerging field of empirical linguistics. Since it owes its existence to rapidly evolving digital technologies for the recording, processing, and archiving of these data, it is not surprising that the main methodological focus so far has been on the form a documentation should take. Thus, we find explicit recommendations on suitable data formats, form and content of metadata descriptions, minimally required levels of annotation, access to and portability of data, etc. (cf. Austin 2006, Bird & Simons 2003, Nathan 2004, Nathan & Austin 2004, Thieberger 2004, Wittenburg 2003, inter alia). This lecture focuses on another central question for good practice in language documentation and description (henceforth LDD): what linguistic goals

it can have and how methods of data collection are related to achieving these goals. Since language documentation has data at its very core, it is only consequent to extend this focus on data on the ways in which they are collected, if language documentation aims at fulfilling the demands put upon it. These are generally understood to be the creation of a record of a language showing the actual linguistic practices of a speech community designated for a broad audience ranging from linguists and researchers from neighboring disciplines to members of the speech communities whose languages are documented (see for instance Austin 2006, Austin & Grenoble 2007, Himmelmann 1998, 2006, Woodbury 2003). Already, the goal to observe and document language use distinguishes LDD from linguistic frameworks concentrating on I-language, competence, or langue as a non-observable phenomenon and bears on the nature of necessary data as well as their interpretation.

Yet, LDD corpora crucially need to integrate analysis and hence cannot be limited to observed and observable linguistic behavior but have to complement the data reflecting it with information on negative evidence, metalinguistic awareness, felicity conditions for utterances, etc. Thus, they need to contain data resulting from a battery of methods aimed at elucidating different aspects of linguistic structure and knowledge. A documentary corpus consequently needs to be as representative and varied as possible, not only in terms of what social variables are covered through the choice of speakers of different ages, genders, etc. but also in terms of the linguistic parameters included. This is what this lecture attempts: to provide an overview of different methods of linguistic data collection in field-based documentary research; to illustrate how different ways of gathering data have an influence on the insights researchers can obtain from them; and to explore how different kinds of data contribute jointly to providing a representative insight into the structure of a language, the linguistic intuitions of its speakers, and their repertoires.

Documenting concepts in contact (conceptional transfer vs. conceptional shift)

Matthias Brenzinger

Thursday 3rd of July 9:00 am

“Documenting concepts in contact” requires the detection, analysis and description of non-conformist, genuine concepts underlying languages spoken by non-western societies. It can be done through the study of intergenerational variations in concepts on space, numbers, colours, biotaxonomies, etc. The observations presented in the following are based on fieldwork with African hunter-gatherer communities, mainly from the Khwe of southern Africa.

All languages spoken some 12,000 years ago were based on a hunter-gatherer view of the world. New concepts of the environment emerged with the rise of agricultural and pastoral economies, and vocabularies and categories developed to accommodate new communication needs. In rural Africa, the mode of production is still among the most essential components of culture and, with that, it constitutes the foundation for conceptualisation.

Concepts of the world with individuals and among communities transform constantly with altering fashions of explaining and understanding an also continuously changing social and physical environments. These evolutionary processes follow unidirectional conceptual transfer patterns that lead "from concrete, or less abstract, meanings to more abstract meanings" (Heine 1997:7).

Such internal processes of conceptional transfers contrast to others, contact phenomena, which may be referred to as conceptional shifts. Concepts that reflect autonomous traditional views of marginalized cultures are rapidly replaced by globally dominant thoughts in that members of such communities either copy conceptual strategies or even entire target concepts. "Globally dominant" concepts are spread through rather uniform government-run educational systems (with concepts mostly from scientific frameworks), or through religious mission work (with concepts from various believe systems). Speakers copy abstract ideas, such as the concepts of numerals, cardinal directions, left-right, sets of colour terms and sets of personal names, etc. Starting from these otherwise targets of conceptual transfers, they search for items in their lexicon to accommodate the new concepts, i.e. go back to the sources. Concepts might already exist, for which no or at least no conventionalized terminology has yet been invented: targets without sources.

As we struggle to learn more about such nonconformist conceptual structures in the languages of non-Western societies, we realize that these habitual thoughts, the ones diverging from the world culture's concepts - are the most vulnerable assets of these languages and documenting them should be given high priority.

2. Fieldwork Workshops

Phonology

Presenter: Lolke van der Veen (DDL, University of Lyon)

Timetable: 23rd and 24th of June, 2 - 3:30 pm (French)
25th and 26th of June, 2 - 3:30 pm (English)

Part 1 Introduction.

A short overview of the most basic issues such as how to choose your informants, how to choose the type of data (words, utterances, oral texts), how to conduct an elicitation session, how to stay in good terms with your informants, can one start doing phonology without knowing anything about the morphosyntax of the language, etc.

Part 2 Practical work with a Bantu-speaking informant from Gabon.

- Eliciting word list material (core vocabulary).
- Hearing and transcribing new sounds (segments and suprasegmental features).
- Aiming for narrow transcriptions.
- Handling transcription errors.
- Handling variation (individual variation, variation between informants).
- Learning as much as possible about the morphosyntax of the language.
- Using language games.
- Keeping your notebook and diary up to date.
- Recording limited sets of well-chosen words containing unusual or problematic sounds for experimental phonetics (e.g. PRAAT).

Tonology

Presenter: Sophie Manus (DDL, University of Lyon)

Timetable: 23rd and 24th of June, 2 - 3:30 pm (English)
25th and 26th of June, 2 - 3:30 pm (French)

This workshop is a general presentation of issues a descriptive linguist has to deal with in tonology, from initial fieldwork to the final analysis.

It does not require any previous knowledge of tonology and will thus introduce briefly the notion of "tone" (the world's tonal languages; tone, pitch accent & stress; different types of tone; lexical & grammatical tone; various transcription systems commonly used for Amerindian, African and Asian languages; and phonetics & acoustics).

Various fieldwork related issues will be addressed such as research protocols, the choice of informants, and recording, transcribing and analysing first-hand data.

Numerous transcribed and recorded examples taken from various language families from all over the world will be used throughout the workshop.

Morphology

Presenter: Denis Creissels (DDL, University of Lyon)

Timetable: 30th of June and 1st of July, 2 - 3:30 pm (French)
2nd and 3rd of July, 2 - 3:30 pm (English)

The practice of eliciting data for morphosyntactic analysis will be carried out through work with a native speaker of Q'anjob'al, a Mayan language of the highlands of Guatemala. The workshop will focus on the structure of the noun phrase, in particular its possessive constructions. The classes will alternate between elicitation demonstration and analysis of corpora of data provided to the students to advance the discussion. Appropriate readings in the morphosyntax of nominal phrases will be provided in time.

Specialized lexicon, cultural references

Presenter: Lolke van der Veen (DDL, University of Lyon)

Timetable: 30th of June and 1st of July, 2 - 3:30 pm (English)
2nd and 3rd of July, 2 - 3:30 pm (French)

Part 1 Introduction.

A short overview of crucial issues such as how to choose the informants (age, sex, in-group status, etc.), how to prepare the elicitation of specialized vocabulary (training, preparation of corpus of stimuli, approach), how to use well-chosen stimuli, and how to choose the most appropriate site and season for this kind of elicitation activities.

Part 2 Practical work with a Bantu-speaking informant from Gabon.

- Eliciting specialized vocabulary (illness and diseases).

Eliciting cultural references related to this specialized vocabulary (principles governing the process of naming, role of taboo, indigenous perception, indigenous categorization, parallel vocabularies (initiate vs. non-initiate; opaque vs. transparent; degrees of specialization), etc.).

3. Technical Workshops

Field Linguist's Toolbox

Presenter: Antoine Guillaume (DDL, University of Lyon)

Timetable: 23rd and 24th of June, 4 - 5:30 pm (French)
25th and 26th of June, 4 - 5:30 pm (English)

Toolbox/Shoebox is a data management and analysis tool for field linguists, available for free download at <http://www.sil.org/computing/toolbox>. It is especially useful for (1) parsing and interlinearizing text and (2) maintaining lexical data. The purpose of this workshop is an introduction and hands-on training for beginners to these two major functions of the software, working with data from the Bolivian Amazonian language Reyesano, and data from the students own field languages if available. We will learn how to create a project, enter words, sentences and full texts, parse and glose them automatically, set up a lexicon of the morphemes encountered in the corpus, elaborate on this lexicon in order to prepare for a real dictionary. We will also practice on the various options available for sorting, searching, browsing, filtering, etc.

Audio Recording Techniques

Presenter: David Nathan (SOAS, University of London)

Timetable: 23rd and 24th of June, 4 - 5:30 pm (French)
25th and 26th of June, 4 - 5:30 pm (English)

Audio recording is at the heart of language documentation. We must make the best possible recordings while we can, since we may not have the same opportunity to record again. Creating good recordings depends on many factors: having clear goals, having a range of good equipment and knowing how to use it, and a basic understanding of acoustics. The two factors which will most influence the quality of your recordings are the choice and use of microphones, and monitoring the recorded audio to ensure it is what you expect.

This session will discuss managing the acoustic environment, psychoacoustics, microphones, recorders, and the basic properties of audio signals and digital audio.

Praat

Presenter: Emmanuel Ferragne (DDL, University of Lyon)

Timetable: 30th of June and 1st of July, 4 - 5:30 pm (French)
2nd and 3rd of July, 4 - 5:30 pm (English)

Over the past ten-odd years, the Praat program has become the most widely used software package for the study of speech sounds. It offers a comprehensive range of functions including annotation and manipulation of the speech signal, and, of course, acoustic measurements. In addition to this, it is freely available, frequently updated, and highly customizable thanks to its programming language. Anyone working with oral material should therefore seriously consider using it for their research.

The Praat tutorials will come in two flavours:

1. Getting to grips with basic manipulations;
2. Exploring advanced functions and programming.

We will do our utmost to tailor the tutorials to the needs of the participants, so this outline is just a rough sketch. The "getting to grips" course will include the following:

- Opening, closing, displaying and browsing through speech data;
- Editing and annotating a waveform;
- Visualizing the speech signal (spectrograms, spectra, pitch curves, etc.);
- Making reliable acoustic measurements, etc.

The specific acoustic features that we will be analyzing have yet to be determined. The "advanced" tutorial will tackle basic programming with Praat (no programming skills required). It is primarily aimed at people whose research involves tedious repetitive tasks (e.g. opening and listening to several sound files in a row, measuring the duration of a sound in hundreds of different sound files, etc.). At the end of the tutorial, the participants will be

able to create their own programs, and will therefore be able to have the computer carry out a number of time-consuming processes for them.

The use of stimuli in field-based language documentation

Presenter: Friederike Lüpke (SOAS, University of London)

Timetable: 30th of June and 1st of July, 4 - 5:30 pm (English)
2nd and 3rd of July, 4 - 5:30 pm (French)

Language documentation corpora ideally contain representative samples of the language use of a given speech community complemented by data providing negative evidence and linguistic intuitions of that community. It is useful to have a way of categorizing these different types of data based on the manner in which they were collected. Himmelmann (1998, 2002) presents a three-way distinction of “communicative events”, as he calls them, which are ideally included in a field-based documentary corpus. He distinguishes the following event types:

- Observed communicative events, or OCVs in shorthand, are those events in which the only influence of the researchers is (ideally) their presence.
- Staged communicative events or SCVs occupy the middle ground between elicitations and OCVs: they are prompted or “staged” for linguistic purposes, but often use non-linguistic prompts such as pictures, video clips that consultants are asked to describe or sort or card games they are invited to play. While they owe their existence to the research goals and hence do not constitute speech events in the sense of Humes, their linguistic structure is less likely to be influenced directly by the researcher than that of elicitations.
- Elicitations comprise all types of communicative events heavily linguistically influenced by and only created for the sake of the researcher, such as word lists, verb paradigms or acceptability judgments.

This course introduces a particular type of SCVs: those collected with the aid of nonverbal stimuli, i.e. pictures, videos, animations, etc. Providing examples of non-verbal stimuli that have been used in the past, and drawing on data collected with them, it illustrates the strengths and weaknesses of the different kinds of resulting data.

Doing fieldwork with a camera: producing videos for linguistic analysis

Presenter: Lorenza Mondada (ICAR research lab, CNRS & University of Lyon)

Timetable: 30th of June, 5:30 - 6:30 pm (English)
2nd of July, 5:30 - 6:30 pm (French)

More and more linguistic data are collected through video recordings. This increasing use of video is related both to technological advances - such as the miniaturization of video equipments and their sophistication - and to theoretical advances - such as the increasing focus of attention of linguists for the multimodal dimensions of language. This lecture will offer some landmarks about the history of use of video in the social sciences for the documentation and study of human linguistic, social and cultural activities. On the basis of good (and bad) examples of video shots, the lecture will present technical problems related to difficulties of doing fieldwork with a camera. It will also show that these problems are always

connected to epistemological and analytical issues. Thus, the course will aim at discussing technical questions of image shooting, framing, zooming and editing in the light of the analytical use of these images.

4. Areal Courses

Topics in African (Cushitic) languages

Presenter: Maarten Mous (University of Leiden)

Timetable: Monday - Thursday 23rd to 26th June, and Monday - Thursday 30th of June to 3rd July, 2 - 3:30 pm (8 classes)

I discuss a number of topics that are important for many African languages and beyond. For each of the topics I present an overview of the situation in Cushitic including the analytical issues involved. This I contrast to the facts in other African languages. The aim is two-fold: to learn about the topics involved and to get an overview of Cushitic and to some extent of the linguistic diversity of Africa.

Overview of topics:

In the first lecture I present an introduction to African linguistic landscape: classification, typology, sociolinguistic profiles, research traditions and developments.

Tone or pitch accent: There is a wide variety of systems within Cushitic from clearly tonal to accent realised by pitch and this enables us to go deeper into the essential difference between tone and accent and contrast this to reduced tone systems in other African languages.

Gender and number: Gender and number are interrelated in Cushitic with plural as a value for gender. This brings us to category definition and nominal classification in Niger-Congo.

Verbal conjugation: Comparing verb conjugations crucially involves a double approach from to function and function to form. On the basis of Cushitic we discuss the issues in dealing with verb conjugation typologically and descriptively.

Verbal derivation and argument structure: Cushitic has passives of intransitive, impersonal constructions, indirect causatives, middle with a wide semantic range. This serves as the basis for discussing valence changing verbal derivations in Africa and brings us to a discussion of verbal diathesis and argument structure including languages using labile verbs and serialisation.

Case and adpositions: Cushitic languages have a marked nominative case system which has properties of both nominative-accusative and ergative-absolutive. The comparative discussion focuses on the coding of other case relations such as instrumental which is in Cushitic either by clitics or by multipurpose adpositions (sometimes stacked).

Topic and focus: Topic and focus drive Cushitic syntax and this has attracted considerable attention which I position into the vast literature on other African languages.

Typological Features of Mayan Languages

Presenters: Prof. Colette Grinevald, and Marc Peake (DDL, University of Lyon)

Timetable: Monday - Thursday 23rd to 26th June, and Monday - Thursday 30th of June to 3rd July, 4 - 5:30 pm (8 classes)

This course will cover the major typological features of the Mayan family of languages. This language family, spoken in Guatemala and parts of Southern Mexico, has been extensively studied since the 70ies, and is rather well known today for its many contributions to general and typologically oriented linguistics. After an introduction to the language family and a brief history of the development of Mayan linguistics, we will discuss the major phonological, morphological, and syntactic features of Mayan languages. The course will first quickly present phonological inventories; major word classes; noun morphology (possession, classifier systems); relational nouns vs. prepositions, verb-initial word order with rigid vs. flexible systems. It will then give an overview of the very rich verbal morphosyntax (TAM, and the multiple markings of transitivity), and treat in turn well known themes of this family, such as ergativity and various alignments, multiple voice systems (several types of passives and antipassives, applicative and agentive/inverse constructions). While still rare for most families of languages of the Americas, the study of the basic syntax of complex sentence formation is rather extensive today for Mayan languages; topics such as relativisation, subordination, coordination, and considerations of ergative syntax will be considered in turn. Time permitting, the last session(s) will consider nominal classification systems (particularly the rare noun classifier system of Q'anjob'alan languages, and the expression of space (through categories of positionals and directionals, in the Tzeltalan and Q'anjob'alan branches).