

## 3L Summer School 2008

### Language Documentation and Language Description

Peter K Austin

ELAP, Department of Linguistics

SOAS, University of London

23<sup>rd</sup> June 2008

#### 1. Outline

- ❖ Documentary linguistics and language documentation
- ❖ Components and skills for documentation
- ❖ Language documentation projects
- ❖ Current and future concerns
- ❖ Conclusions

#### 2. Documentary linguistics

- ❖ 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)
- ❖ has developed over the last decade in large part in response to the urgent need to make an enduring record of the world’s many endangered languages and to support speakers of these languages in their desire to maintain them, fuelled also by developments in information and communication technologies
- ❖ essentially concerned with role of language speakers and their rights and needs

#### 3. Features of documentary linguistics

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.

#### 4. A contrast

- ❖ **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

#### 5. Uses of documentation

Documentation outputs are multifunctional for:

- ❖ linguistic research - phonology, grammar, discourse, sociolinguistics, typology,

historical reconstruction

- ❖ folklore - oral literature and folklore
- ❖ poetics - metrical and music aspect of oral literature
- ❖ anthropology - cultural aspects, kinship, interaction styles, ritual
- ❖ oral history, and
- ❖ education - applications in teaching
- ❖ language revitalisation

## 6. Users of documentation

Collection, analysis and presentation of data

- useful not only for linguistics but also for research into the socio-cultural life of the community
- analysed and processed so it can be understood by researchers of other disciplines and does not require any prior knowledge of the language in question
- usable by members of the speaker community
- respects intellectual property rights, moral rights, individual and cultural sensitivities about access and use and is done in most ethical manner possible

## 7. The documentation record

- ❖ core of a documentation is 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*
- ❖ as a result documentation is increasingly done by teams rather than 'lone wolf linguists'
- ❖ need to see grammatical analysis and description as a *tertiary-level activity* contingent on and emergent from the documentation corpus

## 8. Phases in documentation project

- ❖ Project conceptualisation and design
- ❖ Establishment of field site and permissions
- ❖ Funding application
- ❖ Data collecting and processing
- ❖ Creation of outputs
- ❖ Monitoring, evaluation and reporting

## 9. Phases in data collection and analysis

- ❖ *Recording* – of media and text (including metadata)
- ❖ *Capture* – analogue to digital transfer
- ❖ *Analysis* – transcription, translation, annotation, notation of metadata
- ❖ *Archiving* – creating archival objects, assigning access and usage rights
- ❖ *Mobilisation* – publication and distribution of materials

## 10. Skills for language documentation

- ❖ *Project conception, design and management* - familiarity with documentation theory, applied ethics, IPR, moral rights and socio-cultural issues, stakeholder communication, project management

- ❖ *Grant application writing*
- ❖ *Fieldwork outside the classroom* – interpersonal skills to establish and develop relationships on site, select and manage research methods, observation skills, personal comfort, power and equipment management, health and safety management
- ❖ *Media management* – audio and video recording and editing techniques, data transfer, backup
- ❖ *Data and metadata management* – data and metadata representation (XML, relational database models), transcription, linguistic analysis (phonetics, phonology, morphology, syntax, semantics, discourse) and annotation, use of linguistic software tools (eg. Transcriber, Shoebox/Toolbox, ELAN, IMDI), data integrity and sustainability (a la Bird and Simons), workflow design and management
- ❖ *Mobilisation* – respond to community needs, familiarity with applied linguistics concepts (orthography design, lexicography, curriculum development, policy formation, revitalisation), publication skills, multimedia design and implementation
- ❖ *Team-based research*, skills sharing and transfer, capacity development
- ❖ *Monitoring, evaluation and reporting* – presentation, writing and communication skills
- ❖ *Advocacy* – political and communication skills, willingness to put oneself out there

## 11. Is it just too hard?

- ❖ No single individual can have **all** these skills – we’re not Superman or Wonder Woman
- ❖ Projects must be *appropriate* to context with *realistic* goals – there is no generic “record everything for posterity” project
- ❖ Walcott (1999:73) “if you ... really did have all the skills [listed above], why would you invest your time plodding along with [language documentation]?”

## 12. What skills do communities respect?

<http://languagespeak.wordpress.com/2007/05/31/what-are-linguists-good-for/>

“Two weeks ago our entire group attended the Workshop for American Indigenous Languages (WAIL) in Santa Barbara. There are 8 linguists on our team and 4 community language activists, making ‘our entire group’ a rather overwhelming, but nonetheless easy-going crew. We gave a group presentation on collaborative linguistics. What our presentation stressed was the necessity of forming a collaborative partnership between academics and communities in efforts to maintain and revitalize endangered languages. Our talk was the last one of the session on the last day of the conference. Now of course, the audience was hardly impressed with the linguists on the team, but the community language activists were literally pummeled with questions after the talk was over. One woman asked the language activists something like, “What one aspect of linguistics has been crucial to the development of your project?” She said she wanted to know because she was interested in teaching linguistics to community activists and would like to know where to start. (I know that she was looking for an answer like, “Oh it was morphology! Once I understood the morphology and how to break words apart into meaningful units everything else made sense!” I know that she really wanted to hear what part of linguistics was actually useful to people doing language work.) However, the answer she got from our community language activists was not like this at all. Instead they responded by mentioning how enthusiastic the linguists always were about doing language work (they said something like, “they keep showing up”), and how much they enjoyed meeting with us, and ultimately how much they trusted us. Later on at the party I heard someone fondly summarize their answer as “Trust and love. What are linguists really good for? Trust and love.” At first, after hearing this, the academic in me was disappointed. There has to be something from my discipline which is more useful to language revitalization, right? I mean, I’ve been studying linguistics for over 5 years ... was it all a waste of time? But then I got to thinking about how many negative things linguists have done throughout history ... when it comes down to it, I ought to be overjoyed that there is a community that likes me and thinks I’m a trustworthy academic. In fact, in the

end, maybe it's not so bad to be known for that.”

Dobrin 2004, 2007 – understanding the values of oneself and the community members, and analysing and managing potential value conflicts, plays a **key role** in language documentation projects

### 13. Documentation projects

- HRELP
- Volkswagen DoBeS
- NSF-NEH DEL
- Others

### 14. Some current issues and challenges

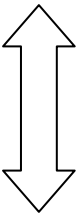
- Documentation versus description
- The ‘comprehensive’ record
- Quality of language documentation
- Interdisciplinarity
- Training for language documentation

### 15. Documentation vs description

Himmelman and others have tried to distinguish *language documentation* from *language description*, but it is unclear whether such a separation is truly meaningful, and even if it is where the boundaries between the two might lie. Documentation projects must rely on application of theoretical and descriptive linguistic techniques, if only to ensure that they are usable (i.e. have accessible entry points via transcription, translation and annotation) as well as to ensure that they are comprehensive. It is only through linguistic analysis that we can discover that some crucial speech genre, lexical form, grammatical paradigm or sentence construction is missing or under-represented in the documentary record. Without good analysis, recorded audio and video materials do not serve as data for any community of potential users. Similarly, linguistic description without documentary support is sterile, opaque and untestable.

### 16. The “comprehensive” record

On a theoretical level, one can define “comprehensive” documentation as the collection of representative texts of *all* discourse types, *all* registers and genres, from speakers representing *all* ages, generations, socioeconomic classes, and so on. On a practical level, however, there are concrete limitations to the range and number of texts which can be collected. Most linguists cannot devote their entire careers to time in the field, which would be required for a truly thorough collection and analysis of data. It is clear that the success of a documentation project rests on intimate collaboration with community members. In the ideal, they can be trained to be engaged in data collection themselves, thereby expediting the process (eg. Florey 2004). Even if this is not possible, community members can direct (external) linguists to varying discourse types and to differing speech patterns. Himmelman (2006:66) identifies five major types of communicative events ranged along a continuum from unplanned to planned (see below) however it is not clear that this typology is applicable to all languages and all speech communities – just what is a comprehensive account of language in use remains unclear

Parameter	Major Types	Examples	
	unplanned	exclamative	'ouch!', 'fire!'
		directive	'scalpel!'
		conversational	greetings, small talk, chat, discussion, interview
		monological	narrative, description, speech, formal address
	planned	ritual	litany

## 17. Quality of documentation

There is a tendency among some researchers to equate documentation outcomes with archival objects (part of what David Nathan has termed 'archivism'), that is, the number and volume of recorded digital audio and/or video files and their related transcription, annotation, translation and metadata. Mere quantity of objects is not a proxy for quality of research. Equally, some would argue that outcomes which contribute to language maintenance and revitalization are the true measure of the quality of a documentation project (what better success of an endangered language project than that the language continues to be used?).

## 18. Interdisciplinarity

Himmelmann and others have pointed to the importance of taking a multidisciplinary perspective in language documentation and drawing in researchers, theories and methods from a wide range of areas, including anthropology, musicology, psychology, ecology, applied linguistics etc (see Harrison 2005, Coelho 2005, Eisenbeiss 2005). True interdisciplinary research, is difficult to achieve, both because of theoretically different orientations, and practical differences in approach (ranging from differences in linguists' and anthropologists' practices concerning payments for consultants traditionally have differed, to more significant differences in academic paradigm that make communication and understanding fraught). Mainstream linguistics has tended to turn away from other disciplines and to emphasise its 'independence' by concentrating on theoretical concerns that are of internal interest to linguists only (minimalism, OT phonology – see Libermann 2007). Documentary linguistics opens new doors to interdisciplinary collaboration but we need to work out **how** to achieve it.

## 19. Conclusions

- ❖ Documentary linguistics (theory) and language documentation (practice) is an exciting new development in linguistics
- ❖ Potential to revolutionise field of Linguistics through concern for data, concern for representation, openness to alternative analyses, brings together theory (including linguistic typology) and fieldwork data collection, applying new ethical and collaborative approaches
- ❖ A range of challenges and opportunities for research, practice and training remain to be met
- ❖ So join in!

## 27. References

- Austin, Peter K. and Lenore A. Grenoble 2007. Current trends in language documentation. In Peter K. Austin (ed.) *Language Documentation and Description*, Volume 4. London: SOAS.
- Bird, Steven and Gary Simons. 2003 Seven Dimensions of Portability for Language Documentation and Description *Language* 79.
- Coelho, Gail 2005. Language documentation and ecology: areas of interaction. In Peter K. Austin (ed.) *Language Documentation and Description*, Vol 3. London: School of Oriental and African Studies.
- Dobrin, Lise 2004. When their values conflict with ours: linguists and community empowerment in Melanesia. In Peter K. Austin (ed.) *Language Documentation and Description* volume 3: 42-52. London: SOAS.
- Dobrin, Lise 2007. Eliciting the linguist. To appear in *Language*.
- Eisebeiss, Sonja 2005. Psycholinguistic contributions to language documentation. In Peter K. Austin (ed.) *Language Documentation and Description*, Vol. 3, 00-00. London: School of Oriental and African Studies.
- Florey, Margaret 2004. Countering purism: confronting the emergence of new varieties in a training program for community language workers. In Peter K. Austin (ed.) *Language Documentation and Description* Vol. 2, 00-00. London: School of Oriental and African Studies.
- Gippert, Jost, Nikolaus Himmelmann and Ulrike Mosel (eds.) *Fundamentals of Language Documentation*. Berlin: Mouton.
- Harrison, K. David 2005. Ethnographically informed language documentation. In Peter K. Austin (ed.) *Language Documentation and Description*, Vol 3. London: School of Oriental and African Studies.
- Himmelmann, Nikolaus 1998 Documentary and descriptive linguistics *Linguistics* 36:161-195.
- Himmelmann, Nikolaus P. 1998. Documentary and descriptive linguistics. *Linguistics* 36:161-95.
- Himmelmann, Nikolaus P. 2006. Language documentation: What is it and what is it good for? In Jost Gippert, Nikolaus P. Himmelmann and Ulrike Mosel (eds.) 2006. *Essentials of Language Documentation* (Trends in Linguistics. Studies and Monographs, 178), 1-30. Berlin: Mouton de Gruyter.
- Libermann, Mark 2007. The future of Linguistics. Invited talk, LSA meeting Los Angeles. January 2007.
- Walcott, Fred. 1999 *The Art of Fieldwork*. Los Angeles: Alta Mira.
- Woodbury, Anthony 2003 'Defining Language documentation' in Peter K. Austin (ed.) *Language Documentation and Description*, Vol 1, 35-51. SOAS.