Editor's Preface and List of Contributors (LDD 5)

PETER K. AUSTIN

_____

_____

Editor's Preface

Peter K. Austin

The papers in this volume arise from various sources. Those by Ruth Finnegan, Maurizio Gnerre and Frank Seifart are written up versions of talks given at a workshop entitled *What counts (and what doesn't)? Data and methodology in language documentation* held at the School of Oriental and African Studies (SOAS) on 4th-5th November 2006 and co-organised by Leora Bar-el and Friederike Luepke. Louise Ashmore's contribution grows out of her MA dissertation submitted at SOAS in 2006, while David Nathan's is an elaborated version of a paper presented at the *International Conference on Austronesian Endangered Language Documentation* held at Providence University, Taiwan, 5th-7th June 2007. Margaret Florey's paper was presented orally at the *Conference on Language Documentation and Linguistic Theory* held at SOAS on 7th-8th December 2007 and appears in written form here for the first time. All the papers were anonymously refereed by colleagues at SOAS; I am particularly grateful for assistance with refereeing from Oliver Bond and David Nathan. Tom Castle did all the layout and formatting and looked after printing and binding, as well as designing the new cover that was established with LDD 4.

As mentioned, the first three papers were presented at the December 2006 workshop, the goals of which were to explore what we mean when we refer to 'data' within language documentation and what methodologies we use when collecting, presenting, and analyzing it in an interdisciplinary context involving linguists and other scholars. Among the issues considered by workshop participants were:

1. What counts (and what *does not* count) as data?
2. Can we make different assumptions about data and still do comparative work?
3. How do our various areas and disciplines reflect on their own approaches to data and methodology?
4. Are there different, perhaps conflicting, data traditions within our areas and disciplines?
5. Do our consultants share our views on what data is? What is *their* view?

Ruth Finnegan's contribution explores the questions 'what is a language?' and 'what is the data by which we study a language?'. She begins by looking at traditional conceptions of language as a set of written texts, a view that was

shaken for her when she began fieldwork in Africa. She came to realise the importance of *oral performance* and 'the subtle characterisations, the drama, the way the tellers used volume, pitch, tempo, repetition, emphasis, dynamics, silence, timbre, onomatopoeia, and a whole plethora of non-verbal indications to convey humour, pathos, irony, atmosphere' which could not be captured by unilinear text. She argues, indeed, that we lack the vocabulary to describe and analyse these performance dimensions of oral language as they have been excluded from linguistics, and this leads her to re-examine the essentially cognitive view of language that has permeated Western thought. She argues that 'once we go beyond models of language as centred on written text or on abstract or cognitive systems, and consider practice and performance, it also becomes inescapable that human communicating is commonly multisensory'. In addition, since language performers are essentially social beings, '*all* data wherever it originates has to be treated critically, with full awareness of the providers' social situatedness whether outside or within 'the field'. Looking for neutral informants channelling neutral data is unrealistic'. She challenges the distinction current in language documentation discourse between *data* and *metadata* (see, for example, Austin 2006:93), and argues for the need to see both speakers and researchers as people in the world and to explore what the consequences of that are for documenting endangered languages.

Maurizio Gnerre also addresses the social situatedness of linguistic research especially the relevance of non-linguistic aspects of field research as they relate to the quality of the material that researchers collect and label as 'linguistic data'. He argues for paying attention to the processes of data collection as part 'of the 'natural history' of the data themselves, and … particularly relevant for the genesis of the artefacts that are typical products of linguists' work: grammars, dictionaries, articles focussing (from many possible theoretical angles) on a specific aspect of a particular language'. He makes a case for attending closely and introspectively to the circumstances of data collection, seeing it as involving 'fragments' of interactions. Thus, each word or sentence we collect (or quote in our publications) has its own autobiography, both in terms of the context of its collection but also in terms of the biography of the researcher who collected it. He argues that language data ranges along a continuum from 'given' (ie. volunteered by some speaker) through to 'taken' (ie. observed, or extracted via elicitation) and that is is not possible to describe language 'in its own terms' (cf. Mithun 2001) because of the importance of this 'natural history' of the data collection process. He presents two case studies from his own research to illustrate these points: from work on Shuar and Achuar (Upper Amazon) ways of speaking, and secondly from research (in collaboration with Flavia Cuturi) on Huave, spoken in Oaxaca, Mexico. The Shuar and Achuar case concerns their ceremonial visiting dialogues and narrative discourse dialogues that have mythological content, and the ways they have changed over 40 years in terms of structure

and function. There are important issues about *how* the relevant material was collected that raise questions about the nature of the discourses as linguistic data. The second case concerns the expression of 'concomitance relations' in Huave, a linguistic isolate from Mexico. Gnerre shows the impact of different speakers and different contexts of taking/giving on the example sentences that appear in a paper on this topic he published with Cuturi (Cuturi and Gnerre 2005), what he calls the 'speaker's polyphony'. Only by careful attention to the biography (or 'natural history') of each sentence can a full understanding of their significance be gained. He concludes by arguing that 'any linguistic output of field research should include at least an extended note, if not a whole chapter … making clear that it is not the patchwork somebody could assume it is, but rather a piece of scholarship reflecting some of (not all!) the dimensions of linguistic variation and polyphony found in a speech community'.

Frank Seifart's paper looks at the question of how a documentary corpus can be properly representative of the language being documented, in accordance with Himmelman's conception of language documentation as (Himmelmann 2006:1):

> 'a field of linguistic inquiry and practice in its own right which is primarily concerned with the compilation and preservation of linguistic primary data and interfaces between primary data and various types of analyses based on these data'

Seifart describes different possible approaches to sampling of this primary data, including *convenience sampling*, *externally-motivated sampling*, and *systematic sampling*. For the last of these he goes on to explore criteria that might be used rigorously to select which events should be recorded, most particularly those 'based on a systematic classification of communicative event types' which can therefore be theoretically grounded. After presenting a case study from his research with the Bora and neighbouring groups in South America he concludes that 'representativeness must be based on a careful analysis of culture-specific event types and that it is not possible to achieve representativeness for all kinds of communicative events to the same degree'. In addition, he notes that ethical and ideological considerations may limit the range of data it is possible to include in the corpus, since speaker communities will often have ideas about what it is appropriate or acceptable to document.

Louise Ashmore's paper takes up the issue of the place of digital video in language documentation. As she notes, 'video has been promoted as "the ideal recording device" for language documentation', and is often recommended to researchers as a means to fully record 'linguistic data'. Its value as a way of capturing visual and spatial aspects of interaction, as assistance for transcription, and as a preference of speech communities is well established.

Yet important issues for discussion remain, including ethical and methodological concerns about the impact of video recording and distribution on communities, the enormous cost of archiving unedited video 'data', and the need to develop 'goals, methods and evaluative criteria appropriate for the different ways video is used in language documentation work'. Based on empirical research, including questionnaires completed by field linguists and case studies, Ashmore explores how language documenters use video in their research projects, how native speakers self-document, and what goals are set for video recordings. She then explores selection and types of video records, wondering whether the 'recordability' of certain speech events and their attractiveness 'on film' may lead to biases in the nature and quantity of interactions that end up in the corpus. In addition, video recording introduces problems for the idea that a corpus should be 'multifunctional' since the kind of material and production values that serve for research purposes may well be rather different from that which is good for publication and distribution (often involving editing to 'tell a story'). Finally, evaluation of quality of video records is a problematic issue. Ashmore identifies three dimenstions here: (1) the technical elements of the record, (2) the contextual information and analysis that transforms video data into a video record, and (3) video data in relation to broader documentation aims. She concludes that 'digital video can be useful and effective for language documentation. However its utility is dependant on individual language situations and necessitates some level of planning, consultation and commitment to provide contextual information'. We can expect that there will be on-going discussion of these issues as equipment costs decrease and more and more researchers decide to make video recordings, and hence, hopefully, will need to grapple with the impact that video has on their work.

Building on his experience of setting up the Endangered Languages Archive (ELAR) at SOAS, David Nathan argues that archives have become an essential component of language documentation research from a number of perspectives. These include interactions between documenters and the archive (which are multifarious and which Nathan specifies in detail), training, fieldwork, preparation of materials for archiving, including the exchange of samples and feedback on them, and the propagation of format guidelines. He presents a detailed analysis of 150 queries he received from about 50 depositors (or potential depositors) at various stages of their projects over a period of two years, categorising the nature of the questions and showing through case studies how the archivist and the language documenter can work productively and creatively together. Archives also have an important role in developing, disseminating and supporting format guidelines 'aimed at encouraging the production of resources that are 'portable', as described in 'seven dimensions' by Bird and Simons 2003'. How this is achieved, and how depositors can be assisted in coming to grips with such guidelines (and

transforming their data for use in different contexts) is discussed in some detail. As Nathan concludes, the main issue here for the relations between archives and researchers is how to maximise the amount and quality of language documentation while taking into account real-world issues of skills, divisions of labour, and resource allocation. His paper ends with some deeper philosophical challenges about the future roles of archives in terms of setting documentation agendas, ensuring preservation, and involvement as agents of 'mobilisation', ensuring that archived data meets the needs of language community members, educators and those engaged in language revitalisation.

The final paper by Margaret Florey discusses how language activism through training can play a role in raising the quality and quantity of language documentation and support for revitalisation activities, especially in linguistically diverse but resource poor countries such as Indonesia. She argues that the dual concerns for loss of linguistic diversity and the promotion of indigenous rights internationally have led to a 'new linguistics' in which capacity building and mentoring are core activities through which language activists outside language communities can support those within them. She presents a case study of this 'new linguistics' applied to language training workshops held in Indonesia in 2006 and 2007. These workshops were aimed at developing skills for local language activists, but Florey shows that subsequently there has been an impact on the kinds of language documentation work that the trainees are now doing across the country. In addition, the initial training has fostered further training and skill-sharing with their colleagues by those who took part in the training workshops. Application of these ideas and experiences to other locations around the world where languages are endangered and documentation is sparse would be welcome.

As usual, readers are encouraged to send comments and feedback on the papers presented here, directed to the address on the inside front cover.

Bloomsbury, London

May 2008

References

Austin, Peter K. 2006. Data and language documentation. In Jost Gippert, Nikolaus P. Himmelmann and Ulrike Mosel (eds.) *Essentials of Language Documentation*, 87-112. Berlin:Mouton de Gruyter.

Bird, Steven and Simons, Gary. 2003. Seven Dimensions of Portability for Language Documentation and Description. *Language* 79, 557-582.

Cuturi, Flavia and Maurizio Gnerre. 2005. Concomitance in Huave. In
          Rosemary Beam de Azcona and Mary Paster (eds.) *Conference on
          Otomanguean and Oaxacan Languages* (COOL), 51-86. Berkeley:
          Department of Linguistics. Survey of California and Other Indian
          Languages. Report 13.

Himmelmann, Nikolaus P. 2006. What is language documentation and what is
          it good for? In Jost Gippert, Nikolaus P. Himmelmann and Ulrike
          Mosel (eds.) *Essentials of Language Documentation*, 1-30.
          Berlin:Mouton de Gruyter.

Mithun, Marianne. 2001. Who shapes the record: the speaker and the linguist.
          In Paul Newman and Martha Ratliff (eds.) *Linguistic Fieldwork*, 34-54.
          Cambridge: Cambridge University Press.

List of Contributors

Louise Ashmore
Department of Linguistics,
SOAS
Thornhaugh Street, Russell Square
London WC1H 0XG
UK
email: louise.ashmore@soas.ac.uk

Peter K. Austin
Department of Linguistics,
SOAS
Thornhaugh Street, Russell Square
London WC1H 0XG
UK
email: pa2@soas.ac.uk

Ruth Finnegan
Faculty of Social Sciences
The Open University
Walton Hall
Milton Keynes
MK7 6AA
UK
email: r.h.finnegan@open.ac.uk

Magaret Florey
Linguistics Program
School of Languages, Cultures and
Linguistics
Monash University, Victoria 3800
Australia
email:
Margaret.Florey@arts.monash.edu.au

Maurizio Gnerre
Università Orientale
Naples
Italy
email:
maurizio_gnerre@hotmail.com

David Nathan
Endangered Languages Archive
Department of Linguistics, SOAS
Thornhaugh Street, Russell Square
London WC1H 0XG
UK
email: djn@soas.ac.uk

Frank Seifart
Allgemeine und Vergelichende
Sprachwissenschaft
Universität Regensburg
D-93040 Regensburg
Germany
email: seifart@berlin.de