

IN THIS NUMBER:

Editorial

New Editorial Board	2
HRELP and the ELAR Archive at SOAS, University of London <i>David Nathan</i>	2
The Centre for Languages and Literature at Lund University <i>Sven Strömqvist</i>	2

Reports from DoBeS Research

Inventing Communicative Events: Conflicts Arising from the Aims of Language Documentation <i>Ulrike Mosel</i>	3
Ethnography in Language Documentation <i>Thomas Widlok</i>	4

Technical Reports

Sound Recording: Microphones <i>David Nathan</i>	6
Safeguarding the Documentary Heritage of Cultural and Linguistic Diversity <i>Dietrich Schüller</i>	9

Endangered Languages Events

Endangered Languages Conference: Endangered Languages Research in the Netherlands <i>Marc Linsse</i>	10
News in Brief from HRELP and ELAR <i>David Nathan</i>	11
DOBES: Next Application Deadline 1 November 2004 <i>Vera Szoelloesi-Brenig</i>	11
DOBES Presentation in the Nobel Museum, Stockholm <i>Vera Szoelloesi-Brenig</i>	11
Endangered Languages Public Event <i>Marloes Telle</i>	11
International Expert Meeting on Access Management for Distributed Language Archives <i>Peter Wittenburg</i>	12

News in Brief

ELAN Native Media Handling <i>Albert Russel & Peter Wittenburg</i>	12
Report from DoBeS training week, 2004 <i>Paul Trilsbeek</i>	12
LREC conference 2004 <i>Peter Wittenburg</i>	12
LinguaPax Forum on Language Diversity, Sustainability, and Peace <i>Peter Wittenburg</i>	13
ELAN 2.2 Now Available <i>Hennie Brugman</i>	13

Editorial

New Editorial Board

In order to expand the scope of the LAN with respect to topics and to the public, the MPI has decided to extend the group of editors such that other language archives — in particular those who deal with endangered and minority languages material — take an active role. Discussions with the linguistics department of Lund university (Sven Strömquist) and the HRELP programme at London university (Peter Austin) resulted in an agreement that all three institutions will take care of editing the coming LAN issues as a joint enterprise.

The person in charge at HRELP is David Nathan who is the archivist. He contributed three articles to the present issue. The person in charge at the Linguistics Department in Lund is Marcus Uneson who is responsible for archive organization matters. He created the new layout for this issue. Romuald Skiba is the corpus manager for the DoBeS and MPI corpora in Nijmegen.

Therefore, in this issue we include a brief introduction to the two collaborating institutions (see below).

Yours sincerely,

Hennie Brugman, Romuald Skiba, Peter Wittenburg

HRELP and the ELAR Archive at SOAS, University of London

David Nathan
SOAS, London

The Hans Rausing Endangered Languages Project (HRELP) was established with a commitment of 20 million pounds from the Lisbet Rausing Charitable Fund to document as many endangered languages as possible and to encourage the development of relevant skills. It shares many common interests and functions with the DoBeS Project. HRELP has three parts:

- Endangered Languages Documentation Programme (ELDP). ELDP manages the awarding of research grants. Overseen by an international selection panel, it currently awards about 20 grants per year for a total of 1 million pounds. The grants are administered by the School of Oriental and African Studies (SOAS). See the news article about the latest grants.

- Endangered Languages Academic Programme (ELAP). ELAP, headed by Professor Peter Austin, offers MA and PhD courses aimed at training the next generation of scholars in language documentation to create a sustainable field. It also offers occasional lectures, seminars, and training courses.
- Endangered Languages Archive (ELAR). ELAR is currently being established, headed by David Nathan. It will provide digital archive facilities for ELDP grantees and others, collaborate in digital materials development, and assist ELAR with training programs. By later in 2004 we will be accepting and archiving digital data and will be setting up facilities for supporting the development of materials assisting community efforts to strengthen languages.

The main ways that ELAR/HRELP differs from its distinguished “cousin” at DoBeS/MPI, are its integration with postgraduate degree programmes, and its twin focus on preservation of data and mobilisation of materials in support of language efforts in communities. We hope to make, like DoBeS/MPI, significant achievements in support of endangered languages worldwide.

To find out more about ELAR and HRELP, please visit our website at <http://www.hrelp.org>.

The Centre for Languages and Literature at Lund University

Sven Strömquist
Lund University

The Centre for Languages and Literature at Lund University is a new cross-disciplinary centre housing linguistic and phonetic sciences, language technology, language departments, and literary science. It also houses a new research library and a new research laboratory for psycholinguistics and cognitive science. Visit our website <http://www.sol.lu.se> (English translations yet to appear in some cases).

The Lund centre sets three main goals for its development over the next few years: to advance new language corpora combined with metadata and partly new analysis tools; to promote e-science for the Humanities, using the new centre in Lund as a testbed; and to expand the local lab environment so that it can serve as a data repository for multimedia data.

The Lund environment has a strong tradition of corpus linguistics with, among other things, the London–Lund corpus as a paradigmatic exemplar. Recently, Lund headed a state-of-art review in the project ECHO (European Cultural Heritage

Online). See http://www.ling.lu.se/projects/echo/state_of_art/index.html. In cooperation with MPI for Psycholinguistics Lund helped advance metadata and language corpora within the same project. For the Language Archiving Network, Lund will contribute data from, among other things, phonetic research on Swedish dialects; field research on South East Asian languages, for example Kammu; and cognitive and psycholinguistic research on reading and writing. The latter type of data include eye-tracking data and a growing crosslinguistic archive of textwriting data collected through key-stroke logging (using the local tool ScriptLog, <http://www.scriptlog.net>). Our rapidly growing archive of key-logged textwriting is at the foundation of a new type of research using frequency and production rate variables to make crosslinguistic comparisons ranging from spelling to rhetorical patterns and habits.

The Language Archiving Network has a great potential for the university world, where archives can be explored for both research and education. The new centre in Lund is keen to promote and test the LAN concept as an extension to the modern research library – in relation to both researchers and students. The Centre is currently creating a new organization with a technical group including a data manager. We also plan to build a server environment for a multimedia data repository.

Reports from DoBeS Research

Inventing Communicative Events: Conflicts Arising from the Aims of Language Documentation

Ulrike Mosel

University of Kiel, Teop Language Documentation Project

As linguists, our primary aim of language documentation is to

- i) provide primary data, in particular annotated audio and video recordings, not only for linguistics, but also for other disciplines of the humanities and social sciences; and
- ii) present them in such a way that they can be understood without prior knowledge of the documented languages even after the language has died.

The corpus of recordings should comprise a large variety of communicative events that are representative for the culture specific ways of speaking in the respective speech community (Himmelmann 1998, Foley 2003, Woodbury 2003). Since language documentation is necessarily based on intensive cooperation between the linguists and the speech community, and in view of the Linguistic Human Rights, the third aim of the DOBES program is to

- iii) produce language documentations that can be used by the speech community for language maintenance and revitalisation.

These three aims give rise to conflict.

The linguist's documentation consists of recordings, transcriptions and translations with comments on content and linguistic phenomena, a grammatical sketch and various kinds of additional materials to warrant understandability. But transcriptions, even when they are done in a practical orthography, are hard to read for non-linguists simply because spoken language is not meant to be written and read. Transcriptions are a tool for linguists and not an enjoyable read for researchers of other disciplines. People engaged in natural discourse repeat themselves, stutter, break up utterances, do not care about background information, mix names and make mistakes so that the recording and transcription may be incomprehensible for outsiders, unless it is accompanied by numerous footnotes, which, of course, make the text even less reader-friendly.

And the speech community? If they are literate in their own or a dominant language, they may not want their recordings be published in the form of transcriptions (as we researchers would not publish linguistically accurate transcriptions of our lectures). In addition, transcriptions cannot be directly used for the production of written materials that the community may want for maintenance and revitalisation measures.

In sum, what is linguistically justified and desirable, may be inconvenient for other researchers and unacceptable for the primary contributors to the language documentation.

Inventing communicative events

To solve the conflict arising from the heterogeneous DOBES aims in our Teop Language Documentation Project, we resorted to inventing communicative events without feeling guilty. All texts that seem suitable for unrestricted publication and interesting for the speech community and non-linguists are edited by Teop native speakers, i.e. rewritten to become a readable text. Since these texts might be used for educational materials later, we are well aware of the dangers of westernizing the Teops' ways of expressing themselves and encourage the editors to keep the specific Teop flavour

of these texts as it manifests itself in their phraseology and discourse structure. These edited texts are then also translated for those who do not understand Teop and are read by good readers for recordings that can be used in schools or for broadcastings. As was agreed upon in the DOBES guidelines, all original recordings are presented with an accurate transcription and a translation that renders the recording utterance by utterance, but they will presumably only be accessible with special permissions in the archive.

Thus the editions of recordings lead to new communicative events such as written legends, personal histories, or cooking recipes. Other inventions of communicative events follow from our dictionary work as dictionary workers write example sentences and monolingual definitions for words that are difficult to directly translate into English.

When previously preliterate speech communities want their language to become a written language and the means of instruction in primary schools, it certainly belongs to the responsibilities of linguists to help them create it, taking care of the uniqueness of their language, but also avoiding a rigid purism that would not be accepted by the younger generations. Linguistically these new communicative events are interesting because they allow us to observe the process of putting a previously unwritten language into writing. Being carefully planned, the written texts may also contain highly complex constructions that only rarely occur in spoken language data and thus may give us new insights into the language's expressive potential.

References

- Foley, William A. 2003. Genre, register and documentation. In Peter K. Austin (ed.), *Language documentation and description*. Vol. 1. Hans Rausing Endangered Languages Project, School of Oriental and African Studies, London, pp. 85–98.
- Himmelman, Nikolaus. 1998. Documentary and descriptive linguistics. *Linguistics* 36, pp. 161–195.
- Woodbury, Anthony. 2003. Defining documentary linguistics. In Peter K. Austin (ed.), *Language documentation and description*. Vol. 1. Hans Rausing Endangered Languages Project, School of Oriental and African Studies, London, pp.35–51.

linguists. This raises a number of methodological issues including that of the status of ethnography in linguistic research. There is a possible unintended drawback to the requirement for research teams to be interdisciplinary. It may reproduce an anachronistic division of labour, allowing the representatives of the two disciplines to sit back and leave things as they are (or used to be), namely by considering ethnography to be something that concerns anthropologists only. Therefore, one of the first questions to deal with is as to why linguists should bother about ethnography given that there are so many other things to worry about. Secondly, what exactly do we mean by ethnographic material? And thirdly, where and how do we place ethnographic data in the archival record? The present article deals with these questions. It is based on a presentation given at the most recent DoBeS training workshop and is published here with the hope that reports on new ideas and experiences concerning ethnography in the course of joint work carried out in language documentation projects will become a regular feature of this newsletter.

Why bother about ethnography?

If language documentation is to serve as a basis for long-term archiving it needs to be “meta-theoretical” (the term used by the VW-DoBeS programme). “Meta-theoretical” is not to be confused with “non-theoretical” because strictly speaking every documentation has a theoretical frame, however implicit it may be. Rather, “meta-theoretical” translates into “amenable and useful to a broad spectrum of researchers and theoretical interests”. This goal cannot be achieved if data collection follows solely the current, and to some extent arbitrary, interests of the researchers involved. Moreover, the problem is not only one of *what* kind of data should be collected but also one of *how* it should be organized for the archive, in other words how it should be “chopped up” into sessions as searchable units of such an archive. The problem may be approached by orienting the research less according to the positions of researchers in their contemporary intellectual landscape and more according to the positions of speakers in *their* social landscape. The research strategy that promises to open up this social landscape to us is ethnography. In other words, ethnography helps to reduce the arbitrariness in data collections by drawing on the cultural context of speakers and it helps to make the language documentation materials meta-theoretical enough to be suitable for a long-term archive.

In researcher-directed data collection, as for instance in structured interviews or stimuli-generated contexts, data sessions are demarcated by decisions of the individual researcher – however arbitrary these may seem to others – as to what and who is to be included in the record. When moving away from this established, narrow mode of data collection the question of what constitutes a session has to be resolved through ethnography. Ethnography suggests an orientation

Ethnography in Language Documentation

Thomas Widlok

MPI, Nijmegen, DOBES project #Akhoe Hai//om

Ethnography for linguists

It is one of the innovative aspects of the DoBeS programme that the Volkswagen Foundation requires project teams to involve anthropologists as well as

towards decisions made by speakers “on their turf” in their specific situations, their positions and their locations. The result is that data sessions do not form a fixed mosaic of non-overlapping units but may be cross-cutting the pre-conceived categories of analysis of individual researchers.

Take the Hai//om healing dance as an example. On the basis of ethnographic work one may define wide sessions that are larger than a single video-tape and in fact larger than the dancing behaviour itself because it includes preparations and subsequent talk. Within this broad session ethnography also allows to define narrow sessions. These can be much shorter than the performance of a dance when they focus on the micro-interaction between healer and patients. The healer “rubbing the chest” of participants, or the healer applying scented powder to participants may be examples of such narrow sessions. Both types of sessions, broad and narrow, are grounded on the “turf” of the speakers instead of on pre-conceived ideas as to what constitutes “the healing dance” or the elements within it.

When defining data sessions, researchers can follow the typical course of ethnographic fieldwork. In ethnography we look out for how speakers distinguish elements on their own turf and we begin with a wide definition of units that are then funnelled down as our understanding increases. In other words: *Cast a wide ethnographic net and follow a funnel approach by defining wide sessions that can be narrowed down or broken down further in the course of research.*

What counts as ethnographic material?

It will have become clear from the discussion of the first question that ethnography is not a separate domain in the record (such as botany or technology) but it is the stuff that connects materials in different domains. It helps to gear language documentation towards a holistic perspective. One of the advantages of this holistic perspective is that relevant, and at times unexpected, connections appear. For instance the researcher may come to realize that kin-talk may involve plants or place names or the dead, cutting across pre-conceived domains. However, if ethnography is not simply to be added on top of a traditional linguistic description but in a sense saturates the language documentation as a whole, how do we manage this comprehensive whole and the many potential interconnections within it? In other words, if the body of materials gains a certain comprehensiveness (as is the aim in the DoBeS documentation and similar research programmes) the question of how the sessions are connected becomes important. If everything is *potentially* connected it becomes even more important to be able to trace such connections. Since ethnographic material relates to the process of involvement with the community of speakers in field research as well as to the social relationships that saturate cultural acts and objects it can provide a lead for establishing these connections.

The links between sessions that are highlighted by ethnography are based on the fact that it allows us to see specific speakers, places and institutions that are related to one another beyond the single speech event. In other words, any language documentation that makes use of ethnography will have in-built connections between sessions because the speakers (and addressees as well as bystanding listeners), the places and settings (spatial ones and non-spatial ones) and the social institutions are part of an ongoing process of mutual engagement which we tap when working ethnographically.

Take the Hai//om kinship system as an example. It is possible to systematically elicit a Hai//om kinship categorization by collecting genealogical data and by then asking informants as to what kinship term they would use to refer to the various individuals in this genealogical chart. Using ethnography as a research strategy it is possible to trace the relations of this informant (and others) not only on an abstract chart but as it is constituted across time and space. It will emerge that not only is there a difference between the kinship system of terms of reference and that of kinship terms of address used in practice. Differences also occur in the ways in which relatedness is being managed through diversified kin-talk. In small groups, in which everyone is related somehow genealogically, the idiom of kinship is much more than a label for a certain position in a genealogical chart. In a sense the kinship system is spread across a host of situations (economic, political, ritual) in which people refer to one another not necessarily in kinship terms but in what may be called “para-kin terms”, i.e. terms that define rights, demands and obligations of kin as an idiom of relatedness.

When looking for the “glue” that connects sessions, ethnography counts because it is more than just another domain. Ethnography helps to re-connect what has been archived as separate sessions. The connections are already there because unlike in anonymous survey-data or in large regional comparisons a collection of ethnographic data keeps the continuity of speakers (persons), places and settings across individual speech events on record. This fact can be exploited for the purpose of connecting sessions. In other words: *Do not elicit an abstraction of “cultural systems” (be it the kin categorization or any other system for that matter) in order to then subsequently search for its application but begin by collecting the pragmatics of (kin-)talk that allows to re-connect chunks of data and which still allows systematizations later on.*

How to locate ethnography in the record?

Defining sessions and being able connect these records to a holistic view are problems that researchers face when compiling a language documentation archive. The problem as to how the data is to be accessed after it has been archived may seem less pressing at first but nevertheless has to be resolved from the start: If the

data sets are to be accessible by other researchers and across individual projects some standardized tools for data documentation need to be introduced (such as the IMDI editor and browser in the DoBeS programme). Metadata is an essential requirement from an archivist point of view but it is also an obvious place where ethnographic sources can be connected to the database. This applies to both, pre-existing ethnographic sources as well as materials collected in the project. In some cases the ethnographic material can be directly tied to the metadata categories if some flexibility is allowed for.

Metadata categories should – from the archivists' perspective – be standardized to some extent across the projects. However, this does not necessarily exclude the productive and – to some extent – also creative use of these categories by the individual researcher (or the individual research team). Surely, some free-text search function is – from the researchers' perspective – essential, allowing not only for the idiosyncrasies of each case but also for unforeseen changes in matters to do with the metadata. But even within the confines of the metadata templates used across cases there are opportunities for filling the categories with ethnographic life.

Take Hai//om spatial categories as an example. Spatial categories, just like categories of group or person, emerge at a number of levels and should therefore receive more than one entry in the metadata descriptions. In this case they would comprise (at least) toponyms, landscape terms and deictic language. Toponyms are manifold because there are several ethnic or language groups in one area and there is a dynamic history of settlement which is reflected in the naming of places. To note down the places of a recording (or of places referred to in conversation) is therefore a glimpse into a complex history of connected settlements that ethnography will bring to the surface. In Hai//om landscape terms are ubiquitous because they also feature in spatial language as absolute space markers. They connect people, landscape features, vegetation, social relations and directions and constitute one of the features that connect sessions that otherwise may be grouped separately as pointed out above. The use of deictic language and of gestures similarly involve both social and spatial elements and shorthand descriptions may occur in the metadata category of place as much as in the category of persons/groups involved.

When administering a growing body of data ethnographic input can balance the need for standardization. Metadata categories may seem to be a necessary evil but the fact that anyone using the archive in the future will encounter the metadata first should be incentive enough to “frontload” the metadata descriptions to the extent that outstanding ethnographic features should not be tucked away in remote folders but should be presented through a productive use of metadata descriptions. In other words:

Do not be judgmental about the diversity of space/time/person/group categories but make use of this diversity to facilitate access to the database.

There is no doubt that every experienced ethnographer will have more insights to convey which help to provide a better ethnographic record. The points raised in this short contribution are limited specifically to show how ethnography is a necessary and potentially rich feature in the work of interdisciplinary research teams working in the field of language documentation. There are other important entailments with regard to the relationship between researchers and the community of speakers which I have not touched upon but which deserve equal attention.

Technical Reports

Sound Recording: Microphones

David Nathan
SOAS, London

For language documentation and archiving, it is important to record the best possible sound quality for the following reasons:

- your recordings may form the only documentation of the language or specific performances in it
- people will want to use your recordings to study or learn the language in detail, including pronunciation and intonation
- it is likely that products will be made from your recordings that make unexpected juxtapositions of sounds (e.g. a speaking dictionary or concordance)
- unlike film, there is no contextualisation, such as a view out of a window establishing that there is a busy street outside. Such contextualisation orients the film's viewer to the environment, so that environmental sounds do not detract from the audio. This is not usually possible when recording sound only.

The microphone is the most important item of your recording equipment. For most purposes, the main microphone should cost at least 30% of the cost of the recorder, and at least US\$100.

Different microphones have different capabilities, strengths and weaknesses, and a fieldworker should have a (small) variety of microphones for handling different situations and choose appropriately.

In any case, remember that the usefulness of your recordings will depend overwhelmingly, and equally,

on two factors: *quality* and *consistency*. Consistency is important because psychological factors are so great in the perception of sound, and because the eventual usages of your sound resources may result in arbitrary segments played in juxtaposition.

Factors in choosing microphones

Choose microphones in regard to factors including:

1. Directionality

Choose a directional (also known as unidirectional or cardioid) microphone, or an omnidirectional microphone, depending on the circumstances.

Directional microphones are typically better for linguistic work, because you can reduce sound pickup from sources other than the speaker. For stronger directionality, use a shotgun or hypercardioid microphone. All of these will give good results but require more work, as they must be well-positioned and aimed. Microphones with greater directionality do not have a pickup zone that extends further, they simply have a more narrowly restricted zone of pickup. Directional microphones are more sensitive to wind noise and popped aspirations, so you may need a foam windshield or equivalent.

Omnidirectional microphones pick up sound from all directions with more or less the same sensitivity. In many cases this is not what you want, although lapel microphones are often omnidirectional.

2. Condenser or dynamic

Condenser microphones are also called *capacitor microphones*; some types of condenser microphone are called *electret*.

Condenser microphones are more accurate and provide more output but require a power source.

Dynamic microphones are less accurate and provide less output, but they do not need power, and are more robust in adverse physical and climatic conditions such as high humidity.

3. Impedance

Low impedance microphones are more accurate and pick up less interference, but tend to be more expensive and more fragile.

4. Power requirements

You need to consider the availability of mains and battery power for your microphones, recorders, and rechargers.

Condenser microphones require batteries, which may need to be replaced often, or use "phantom power" from the recorder itself, which adds a power burden to the recorder if it is running on batteries.

Dynamic microphones don't require power but are not as accurate. Nevertheless a good dynamic microphone may be appropriate and can provide a good second or backup microphone, if kept close to the speaker's head when recording.

5. Size/form

Microphones come in several shapes and sizes. Shotgun microphones are often long and thin because they are normally attached to a boom. Lapel microphones, also known as lavalieres, may be useful but risk the dangers of handling or movement noise as the speaker rustles his/her clothes or moves about, or moves the cable, which can destroy the recording. Alternatives are to attach the microphone clip to a speaker's hat or glasses. Use of lapel microphones may require "wiring up" the speaker and/or having a radio transmitter; they are perhaps better used in controlled studio situations where the speakers can be instructed and are sitting still.

6. Stereo/mono

Mono recording is usually OK for linguistic work (and may provide the advantage of halving the amount of data you have to store). However, make sure that you are aware of the equipment you are connecting; for example, some computer microphone input jacks and some microphone plugs are mono miniplug. In some combinations, this can result in recording only one channel of a stereo signal, which will result in a poor recording. (On the other hand, a mono microphone with a stereo plug will split its signal and there should be no problem). For higher quality equipment, XLR (or "Canon") connectors are often used; you should make sure that you have and test the right combinations and extensions etc.

7. Use of stands/booms

It is necessary for any high-quality recording to get the microphone as close as possible to the speaker's head.

The ideal is a hand-held boom or extension-piece enabling someone to continuously locate the microphone within 30cm of the speaker's mouth and pointing at the speaker's forehead. However, this is not always possible, and requires more people.

Simple handheld mode is a good way to manage a microphone if you have limited equipment and crew, but can be tiring for the person holding the microphone after about ten minutes. Using a handheld microphone gives the most control over its position and direction. The person holding the microphone should take special care in handling it. Some microphones (and their cables) are *very* susceptible to handling noise, such as caused by the hand tensing or moving around the microphone. Never, of course, cover the meshed end or any holes around it. If you ask a speaker to hold the microphone, show them where and how to hold it (it may help to record some good and bad examples to listen to).

Table stands (or tripods) for microphones are effective if they allow the microphone to be located close enough to the speaker's mouth. If the table is too far from the speaker, background noise levels will be higher. Take care not to disturb the table: vibrations from tapping on the table, kicking the table leg, or

walking on the floor, will be transmitted through the stand to the microphone. A thick insulating mat between the table and stand will reduce this type of noise transmission. The presence of a table may encourage noise from shuffling papers, which will damage the recording.

8. *Balanced or unbalanced*

Some professional equipment uses balanced connections, where the voltages in the recorder, cables and microphone are arranged so that cables do not pick up extraneous interference from the electrical/magnetic environment. For many situations, especially where connecting cables are not long, this is not necessary, but beware of incompatibilities and that, at the worst, unsuitable combinations can damage components.

9. *Automatic level control (ALC)*

Although ALC is a function built into some recorders and cameras, not microphones, it affects the way you use your microphone. ALC will try to raise the input level when the input volume is low; however, ideally, you want to position the microphone close to the speaker's head so there will naturally be low input volume when the speaker speaks very quietly or is silent between utterances. Allowing ALC may cause annoying changes in background sound at the beginning and end of utterances, and unexpected levels of background sounds to be recorded. It is best avoided.

10. *Using a video camera*

If using video, don't forget about sound recording. Substituting video recording for sound recording without attention to sound quality may leave you without any good record of the language sound. If at all possible, do not use the video recorder's internal microphone. However, a video recorder with a good, compatible, well-located, external microphone can make very good recordings, and can serve as a second or backup sound-recording system. Take care with the microphone cable and avoid any mechanical noise made by the camera itself.

Microphones should be selected to be compatible with recording equipment. You will need to take into account the recorder's price, quality, connections, power availability, manufacturer's recommendations, and compatibility advice from vendors, professional users, and others.

Preparing and practising

The fieldworker should spend at least half a day thoroughly practising and testing equipment *and recording methods* by recording in a variety of situations. For example, check how the equipment performs when recording at various distances from a speaker, how it handles background noise, its response to physical handling (e.g. moving hands on the microphone or its stand, boom etc).

The field worker should have an idea of the sound properties (such as echo) of various environments and surfaces, and understand that sound perception is strongly psychological and different to the "reality" that will actually be recorded. Although your mind can focus on a particular sound source to the exclusion of others, or ignore a sound coming from a direction you are not interested in, a microphone cannot: "what you hear is not what you get" (Rose 1999: 94). For example, microphones pick up more echoes and reverberation than you would normally consciously hear.

Adapting to the fieldwork situation

Prepare strategies to deal with situations such as playing children, refrigerators, traffic noise, wind, fans, and electrical interference from cabling, computers etc. When at the recording site, survey the expected recording locations in regard to their acoustics – not only within the building or immediate area but also the direction of noise sources such as roads. Some simple steps can help: if indoors, move the speaker away from walls, and, if hard reflective walls cannot be avoided (for example by working near curtains, bookshelves, furniture or other textured surfaces), then try to face the speaker at an angle (i.e. not square-on) to walls in order to lessen sound reflections. The sound environment changes throughout the day, so if possible, acquaint yourself with the daily sound rhythms of the location (e.g. of traffic, household activities, homecomings, sounds of birds and animals) and plan accordingly. If using a radio microphone, it is especially important to test it at the actual location and time of recording because this technology is the most vulnerable to interference, loss of signal, flat batteries. Although a radio-transmitted lapel microphone might seem like a good choice for field situations, Rose (1999) recommends that radio-transmission is only reliable in controlled studio situations. (It may also not be legal in some countries.)

Turn off all mobile phones (cellphones); not only can they ring, but they regularly emit signals to page their base station which may seriously interfere with your recording.

For outdoor use (and for indoor use of shotgun microphones that are sensitive to pops and aspiration), you may need to protect the microphone against wind noise. Again, familiarity with the equipment and testing and monitoring are important. Typically, foam windshields are used, although in an emergency you can wrap some open weave or other low-density fabric around the head of the microphone.

You should also take into account the people you expect to record; whether there will be one or more people, whether they are likely to perform simultaneously (e.g. singing), their mobility, acceptance of guidance, comfort with technology, possible sensitivity of content, and the privacy of the location. Alert people you work with that touching the microphone, its stand, cables, or the table on which it

is sitting, can damage the recording. Take special care with paper – shuffling papers can easily make parts of your recording next to useless.

If recording in multiple sessions, keep a record of the microphone used, location, orientation etc, so that you can achieve consistency where necessary, such as when recording for a talking dictionary, where arbitrary sounds can get sequenced when the final product is in usage.

Listening and monitoring

You will need some ways to listen to the sounds you record at various stages. Have at least one pair of headphones. Use closed headphones to check a sample of your initial recordings to check sound level and quality, paying special attention to background sounds, echoes and reverberation. You might use open headphones during recording to monitor the input from time to time. In addition, it is worth having a small pair of speakers to share and review the recordings with participants.

Brands

Quality brands include (in alphabetical order) AKG, Audio Technica, Beyerdynamic, Sennheiser, Shure, and Sony.

References and useful information

Rose, Jay. 1999. *Producing Great Sound for Digital Video*. San Francisco: Miller Freeman

Michael Hartkopf's Microphone website at <http://members.aol.com/mihartkopf/>

Bartek Plichta's website at <http://bartus.org/contents/>

Bartek Plichta's presentation at <http://saussure.linguistlist.org/cfdocs/emeld/workshop/2002/presentations/bartek/bartek.ppt>

Ladefoged, P. 2003. *Phonetic Data Analysis*. Oxford: Blackwell.

Nathan, David – updated version of this article: <http://www.hrelp.org/archive/advice/microphones.html>

Acknowledgements

Thanks to the following people for helpful input: Rob Kennedy (Language Centre, SOAS), Jerry Glasgow (AV, SOAS), and Gary Hammond (AV, SOAS).

Safeguarding the Documentary Heritage of Cultural and Linguistic Diversity

Dietrich Schüller

Phonogramm Archive Vienna, Chair of the Memory of the World Sub-Committee on Technology, UNESCO

Ever since their development audiovisual sources have played a predominant role in cultural and social sciences, specifically in linguistics and dialectology, ethnography, ethnomusicology and, more recently, oral history. In fact, phonetics, dialectology and ethnomusicology have developed interdependently with emerging sound recording techniques from the late 19th century onward. The availability of portable tape recorders since the mid-1950s supported systematic fieldwork in every part of the world. In the late 1970s videography came within practical reach of the anthropological disciplines. Consequently, substantial amounts of audio and video recordings have been produced for research purposes, which are the sources of today's knowledge of the linguistic, musical and cultural diversity of mankind. Because of the general progress of acculturation and globalisation, a lot of the stored documents are of historic character, representing languages, musics, and cultural expressions which, since their recording, have been subjected to substantial change or have become extinct.

Contrary to traditional paper documents, audiovisual records are extremely vulnerable and chemically unstable. Life expectancies are at best limited to a few decades. The fact that they are machine-readable documents adds another dimension to their retrievability, namely the availability of dedicated replay equipment. With advanced sophistication of modern, specifically digital audiovisual formats this situation becomes increasingly aggravated. The old analogue materials, specifically the enormous amounts of magnetic tape, are approaching the end of their lives. Contents can only be kept alive if they are transferred to the digital domain and subsequently migrated from one format to the next, before obsolescence of hard- and software strikes. Thus, long-term preservation of audiovisual records is an indispensable and demanding job for audiovisual archives, calling for substantial skills, associated with considerable infrastructural and financial resources. Radio, Television and National Archives are aware of this task and are getting prepared to adequately respond to the enormous challenge. It can be anticipated that at least the most valuable documents collected so far by this group of archives will be safeguarded in the long term.

However, an estimated 80% of the world-wide holdings representing the cultural and linguistic diversity of mankind are not held by audiovisual archives proper. Originally brought about in order to advance research and knowledge, they are now part

of research institutes, museums, and often private researchers not necessarily aware and/or equipped to adequately handle and preserve these vulnerable sources. The media, mainly tapes, were shelved and used like books, and as long as relatively sturdy materials were concerned this system worked with sufficient reliability. More recently, however, problems are increasingly beginning to occur: The oldest tapes become brittle, more recent tapes, audio and video, suffer from pigment binder breakdown, rendering tapes difficult to replay, sometimes even leading to a complete loss of the magnetic layer which carries the information. The instability of the documents is complemented by the threat of equipment obsolescence, as most manufacturers of analogue magnetic tape players have ceased production, and spare part supply is generally only maintained for another 10 years.

The numerous small research collections, which reflect by far the greatest part of documents related to the linguistic and cultural diversity of mankind, are notoriously underfunded and sometimes not even aware of the threats facing their unique and irreplaceable holdings. In the course of the last years several projects have been undertaken to improve the situation of endangered documents—most of them under the buzzword of *Digitisation*—but there is little to none coordination between such projects, often regrettably little professionalism, and, generally, no strategic and financial provisions to keep digital carriers alive in the long term. Consequently, many of these newly produced digital copies will be lost very soon again.

The challenge facing audiovisual collections in general, and the holdings of small research institutions in particular, which reflect the cultural and linguistic diversity of mankind as recorded over the past 100 years, can only be met by an international action plan which coordinates governmental, scientific, and private interests and activities in this field. Unless systematic efforts are made on an international scale to establish adequate preservation schemes for those collections, most of these documents will be lost in the next 10–30 years, either by deterioration of the materials or by obsolescence of replay equipment.

This pessimistic outlook reveals that any activities to strengthen cultural and linguistic diversity must likewise be associated with effective measures to safeguard the respective documents that have been accumulated so far.

Natural allies in such strategies are researchers and cultural promotion agencies as organised in a variety of NGOs as well as Intergovernmental Organisations, most prominently UNESCO, which maintains several programmes in this context. The European Institutions are promoting several initiatives in this area. The private sector, including various foundations, has also developed specific activities.

It is therefore proposed that, under the leadership of UNESCO, a worldwide Action Plan for Safeguarding the Documentary Heritage of Linguistic and Cultural Diversity be established. This Action Plan should, in their relevant activities, coordinate:

- UNESCO Programmes such as:
 - Memory of the World
 - Oral & Intangible Heritage
 - Living Treasures
 - Endangered Languages
- International and Regional Intergovernmental Organisations (EU, etc.)
- Non-Governmental Organisations like
 - International Council of Traditional Music
 - IUAES and other relevant linguistic and anthropological NGOs
 - International Association of Sound and Audiovisual Archives (IASA), et al.
- Private Sector and Foundations:
 - Ford Foundation
 - Rockefeller Foundation
 - Soros Foundation
 - Paul Getty Trust
 - Volkswagen Stiftung
 - WennerGren Foundation
 - Asia-Europe Foundation, et al.
 - Recording, Media, and Computer Industry

Endangered Languages Events

Endangered Languages Conference: Endangered Languages Research in the Netherlands

Marc Linsse

NWO Council for the Humanities, Den Haag

The Netherlands Organization for Scientific Research (NWO) has created an Endangered Languages Program, which is actually sponsoring three endangered languages projects: two in Africa, and one in South-America. On August 26, 2004, a one-day workshop will be organized in the Cultural Center of the Vrije Universiteit (Amsterdam, The Netherlands). During this workshop these three projects will be presented and the future of the program will be discussed. The workshop will also offer an opportunity for an exchange of ideas about the objectives and the importance of endangered languages documentation,

and about more practical problems that researchers encounter while doing this type of research. A number of national and international specialists will present their ideas about a number of issues relevant to these themes.

News in Brief from HRELP and ELAR

David Nathan
SOAS, London

HRELP's Endangered Languages Documentation programme has recently announced its 2004 grants awards to support language documentation projects. A total of 20 projects across 5 continents will be supported, for a total of more than 1 million pounds. For further information, see <http://www.hrelp.org/grants/>

HRELP has a new website! Led by ELAR, we fully designed our website to make a more streamlined site that is fast to load, easy to navigate, and will provide better paths for expansion. We intend the site to become a major information provider right across the field of endangered languages. You can access it at <http://www.hrelp.org>, and feedback and suggestions are more than welcome.

DOBES: Next Application Deadline 1 November 2004

Vera Szoelloesi-Brenig
Volkswagenstiftung, Hannover

Applicants who wish to submit a proposal in the funding initiative "Documentation of Endangered Languages" are asked to please note that the next application deadline is 1 November 2004.

Applications will be accepted for documentation projects for a period of up to three years. The data are to be recorded and processed in accordance with the technical, linguistic and juridical framework of the program; the outgoing language documentation is to be transferred to the databank at the Max-Planck-Institute in Nijmegen. For more information please see <http://www.volkswagenstiftung.de> or contact Dr. Vera Szöllösi-Brenig (szoelloesi@volkswagenstiftung.de).

DOBES Presentation in the Nobel Museum, Stockholm

Vera Szoelloesi-Brenig
Volkswagenstiftung, Hannover

The exhibition "Science + Fiction — Between Nano-World and Global Culture" is hosted at the Nobel Museum Stockholm beginning June 18th, 2004.

The exhibition is organized by the Volkswagen Foundation. In the past, it has been held in Hannover, Karlsruhe, Bonn and Dresden. The DOBES-demo, a computer based presentation, has been part of this exhibition since 2002. Many aspects of the DOBES work (such as data examples, computer tools and linguistic analyses) are presented in a computer simulation. The newest version of the Demo is multilingual: German, English and partly in Portuguese. It shows a new feature "the audio globus" that allows the user to browse through the world map, to zoom in and to listen to the languages from different DOBES projects. Future sites for the exhibition are Eindhoven (The Netherlands) and Tokyo.

Endangered Languages Public Event

Marloes Telle
NWO, Council for the Humanities, Den Haag

Endangered Languages are languages that are threatened with extinction. On August 28, a public event on endangered languages will take place in the Museon, Den Haag. This public event is aimed at a (non-specialised) public in or curious about the preservation and recording of endangered languages and cultures. Interactive workshops are being organised on the click language Sandawe, spoken in central Tanzania; the West papuan language Mpur, the Russian language Nenets and the Surinam language Trio. UNESCO's decision to declare the oral and graphic expressions of the Wajâpi Indians as a 'Masterpiece of the Oral and Intangible Heritage of Humanity' will also be spotlighted. A member of the Wajâpi group and an external expert on the Wajâpi culture have been invited to attend the event. Together they will explain the cultural, mythological and social significance of the Wajâpi painting tradition, which includes body painting.

For the programme and to register, see <http://www.nwo.nl/bedteigdetalen/publieksevenement>.

International Expert Meeting on Access Management for Distributed Language Archives

Peter Wittenburg
MPI, Nijmegen

In the realm of the DELAMAN (Digital Endangered Language and Music Archives Network) initiative, the MPI team will organize an expert meeting to discuss strategies for the long-term preservation of material in endangered languages and music archives and for the integration of access management. These topics are of great importance for many archives and several initiatives such as the Internet2 or Global GRID communities are aware of the challenges and are brainstorming and working on solutions. The expert meeting will bring together linguists to discuss their requirements also taking into consideration the needs of the speech communities with which they work, archivists to present the current archive solutions and their expectations and technologists who will present the state and perspectives of existing and emerging software packages. At the end of the workshop we aim to draw conclusions as to how the integration of the different archives collaborating in DELAMAN can best be achieved. We are thinking of federated collections where each archive remains a self-standing entity with all rights to define access rules. The workshop will be held in Nijmegen in November and will be funded by the VolkswagenFoundation.

News in Brief

ELAN Native Media Handling

Albert Russel & Peter Wittenburg
MPI, Nijmegen

Similar to other projects the ELAN team based its media handling software components on Java Media Framework and Quicktime for Java (QTJ). For some time it has been obvious that the idea of creating a platform independent media framework has failed. There have been severe problems with JMF and QTJ and the utterly complex issue of handling video in a smooth fashion has not been solved. Those who have used ELAN will have recognized the low degree of precision when playing media and may have attributed this to ELAN. We have invested much time in improving the media handling in two steps: (1) We cleaned up the media handling code to get the best performance that can be achieved with JMF. (2)

We have now implemented an alternative method of controlling video by using native platform dependent software. Under Windows the MS DirectX library is used, which is also applied by companies generating computer games. Tests have shown that video handling by ELAN on a Windows machine is now much smoother and yields the required precision. This is solved for MPEG1, and we are now looking for good codec for MPEG2. Currently, we are busy using native methods on the MAC as well. Results will be presented in the next LAN issue. It should be noted here that the precision for MPEG encoded data is limited to the frame duration. For a higher precision one has to refer to the audio file as the primary media file in ELAN.

Report from DoBeS training week, 2004

Paul Trilsbeek
MPI, Nijmegen

From May 10 to 14 a training course was held at the MPI Nijmegen for new DOBES teams and other interested people. The course was also announced on the Gesellschaft für bedrohte Sprachen mailing list, which resulted in a large number of participants. In total there were 21 participants, 6 of which were new DOBES team members.

After some general linguistic topics on the first day, the course covered the whole technical trajectory from collecting data in the field to making the data in the archive accessible. Topics that were presented include Audio and Video recording and editing, Power Management in the field, IMDI metadata & tools, annotation using ELAN and Transcriber, Shoebox, PRAAT, workflow principles and access management. In general, feedback from the participants about the course was very positive, especially the more practical sessions were very much appreciated.

Some documents and handouts that were used during the course can be found on the DOBES website: <http://www.mpi.nl/DOBES/training/>

LREC conference 2004

Peter Wittenburg
MPI, Nijmegen

The Language Resource and Evaluation Conference (Lisbon, May 2004) is certainly one of the most relevant conferences for all matters regarding linguistics and language engineering resources. A number of issues were discussed that were of direct relevance for the work of the MPI team and also for the technological

aspects in the DOBES programme. Of course, we can only report a few selected topics:

- The ISO TC37/SC4 initiative is about to come up with first standards. The nature of the Data Category Registry that is seen as a repository for all relevant linguistic concepts to increase semantic interoperability is widely agreed. The IMDI metadata concepts for example have already been inserted. The Lexicon Markup Framework is currently being worked out as a flexible lexicon format. Its design is influenced by the requirements of the DOBES programme with all its differing lexicons. The MPI team is busy building one of the first lexicon tools based on LMF.
- In the area of metadata description, IMDI and OLAC are well-known and widely recognized by the community. Increasingly more people realize the importance of describing their resources with the help of metadata to make them visible to others.
- It seems that the number of video annotation tools is decreasing. Some major developments in the last few years such as the ATLAS project from NIST and LDC have been stopped. ELAN seems to be one of the few that will be continued. Yet it is not clear why several projects have been stopped but video handling is still not simple and requires many resources.
- A very interesting panel organized by Brian McWhinney (Childes CMU) showed that there are two major ideas concerning how to achieve flexible collaborative environments: (a) peer-to-peer frameworks that allow collaboration via instant messaging, i.e. collaborative commentary or annotation is done in real-time; and (b) collaborative commentary on web-based presentations that are not done in real-time. The MPI team was one of the first demonstrating a peer-to-peer collaborative annotation framework based on an ELAN extension.

The MPI team gave 13 presentations at the conference and conference workshops on topics such as metadata, multimodal annotation, the relevance of XML for archiving and distributed archiving as intended in DELAMAN. See <http://www.lrec-conf.org/lrec2004/>

LinguaPax Forum on Language Diversity, Sustainability, and Peace

Peter Wittenburg
MPI, Nijmegen

The 10th LinguaPax Congress was devoted to the topics language diversity and language sustainability and their relevance to peace. B. Comrie, N. Hornberger, S. Romaine, A. Bastardas, F. de Varennes and M. Siguan were invited to present their ideas on the three major conference topics. Five workshops were organized in parallel, to give a spectrum of the ongoing work and to work out perspectives: (1) Positive Models of language policy and planning, (2) Case studies of language revitalization and standardization, (3) Evaluation of the current sociolinguistic research, (4) Language law and language rights, (5) Agents in favor of language diversity. The papers presented and the results can be viewed at the LinguaPax web-site. The DOBES programme was invited to present the DOBES documentation and archiving model and its relevance, for example, to language revitalization. U. Mosel and P. Wittenburg shared the presentation, which covered both linguistic and technological aspects. The DOBES contribution was very well received, we got many positive responses. <http://www.linguapax.org/congres04/indexang.html>

ELAN 2.2 Now Available

Hennie Brugman
MPI, Nijmegen

ELAN version 2.2, for native Windows Media, JMF and Quicktime, is now available on the MPI tools website (<http://www.mpi.nl/tools>). In addition to the capability to play video using native software (discussed separately in this issue), it adds the following features:

- The search user interface and functionality are substantially improved.
- The Grid Viewer is extended. It can now optionally show annotations from a number of dependent tiers in a tabular way, where different columns contain annotation values from different tiers. This can be done for all tiers where the annotations are one-to-one associated with annotations on dependent tiers. Where no child annotations exist they can be created by just filling in an empty cell.
- The Timeline Viewer and Interlinear Viewer now support several new ways to hide or display tiers and to sort tiers.

- Controlled Vocabularies can be defined and associated with Linguistic Types. When annotation values of such a Type are created or modified a list of alternative values is shown for selection. CVs are stored inside ELAN's EAF files. They can be reused for other documents as part of a template document, or by importing CVs from existing template documents using ELAN's built-in CV Editor.
- Export of ELAN's current document view to a JPEG image is now supported.
- Export of an annotation document, or a part of it, to a SMIL 'clip' is now possible. This clip can for example be embedded and played within a PowerPoint presentation.
- Import of CHILDES CHAT files, both pre-UTF-8 and UTF-8, is now supported. Existing time alignment following a couple of notation conventions is maintained. Where time alignment applies to a number of consecutive CHAT utterances the available time interval is equally divided over the utterances. When no time alignment is present by default one second time intervals are assigned to consecutive utterances.
- Import of Shoebox/Toolbox is improved: when Unicode is used it is maintained during import. As alternative to using a .typ file the user can now manually specify which tier markers are used in the imported Shoebox file.
- The RTR input method for IPA is revised a bit: more characters are made available under lower case keys.