

IN THIS NUMBER:

Editorial

Endangered Languages

- | | |
|--|---|
| New in DoBeS: Documentation of Jaminjungan and Eastern Ngumpin
<i>Eva Schultze-Berndt</i> | 2 |
| New in DoBeS: Documentation of Enets and Forest Nenets
<i>Kaur Mägi, Florian Siegl</i> | 2 |

Technical Section

- | | |
|--|---|
| Review: Roland Ediol R-1
<i>Anthony Jukes, David Nathan</i> | 3 |
| Review: Mayah Flashman
<i>Gerd Klaas, Nick Wood</i> | 5 |
| Review: Maycom Handheld II
<i>Rob Kennedy</i> | 6 |

News in Brief

- | | |
|--|---|
| Second HRELP Grantee Training Workshop, London, June 2005
<i>Peter K. Austin</i> | 7 |
| EMELD 2005, Harvard University, July 2005
<i>David Nathan</i> | 7 |
| LSA: Language Documentation, Harvard University, July 2005
<i>Peter K. Austin</i> | 8 |
| IFLA: The Multicultural Library, Stockholm, August 2005
<i>Peter K. Austin</i> | 8 |
| ELAN – Latest Changes
<i>Han Sloetjes, Hennie Brugman, Alex Klassmann</i> | 9 |
| MPI Training Courses for Field Linguists
<i>Romuald Skiba, Paul Trilsbeek</i> | 9 |

Announcements

- | | |
|--|----|
| DoBeS: Application Deadline November 1, 2005 | 10 |
| Workshop on Endangered Languages and Literacy, SOAS, London, December 2005 | 10 |
| ELAR at SOAS Begins Archiving | 10 |

Editorial

Dear Readers,

LAN 6 puts the spotlight on audio recording, with a bumper issue containing reviews of three different solid-state recorders. We believe this is timely – recently several mailing lists have been busy with questions from linguists about the new recording technologies.

Solid-state is rapidly replacing older technologies. We estimate that although in 2004 a negligible number of language documenters were using solid-state recorders, by mid 2005 perhaps 20% or more are doing so. If you are amongst them, and have useful information to share, LAN wants to hear from you! In addition, news, information, and articles of any type related to the technological, archiving or related aspects of language documentation are warmly welcomed.

Best wishes,

David Nathan, Romuald Skiba, Marcus Uneson

text collection include significant sites, plant use, and oral history, which are likely to be of interest to the speakers and their descendants as well as to linguists, anthropologists, biologists, ecologists, and historians.

The project director is Eva Schultze-Berndt (University of Graz), who has worked on Jaminjungan languages and Ngarinyman for several years. Principal investigators are Patrick McConvell (AIATSIS, Canberra), a linguist and anthropologist with long experience with the languages and culture of both groups, and Felicity Meakins (University of Melbourne), a linguist with expertise in Eastern Ngumpin languages. PhD students Kristina Henschke and Candide Simard will focus on the description of prosody and code-switching. The project team will be supported by ethnobotanist Glenn Wightman and ethnomusicologists Alan Marett and Linda Barwick, as well as by Nikolaus Himmelmann (University of Bochum). The project will involve intensive collaboration with the Diwurruwurru-Jaru Aboriginal Corporation, an Aboriginal Language Centre based in Katherine (NT), and will also include community members as trainees and co-investigators.

Endangered Languages

New in DoBeS: Documentation of Jaminjungan and Eastern Ngumpin

Eva Schultze-Berndt
University of Graz

Jaminjungan and Ngumpin are language families traditionally associated with neighbouring regions in the Victoria River District in Northern Australia. The three-year project “Jaminjungan and Eastern Ngumpin: A documentation of the linguistic and cultural knowledge of speakers in a multilingual setting in the Victoria River District, Northern Australia” seeks to document the linguistic and cultural knowledge of the remaining few hundred speakers of several varieties: Jaminjung and Ngaliwurru (Jaminjungan), and Gurindji, Ngarinyman, Bilinearra, and Mudburra (Eastern Ngumpin). These varieties (together with English and Kriol) form part of a network of multilingual communicative practice in the region, since the speakers have been in close contact for centuries and now share the same settlements throughout the region. One aim of the project therefore is to carefully document dialectal and ideolectal variation and code-switching in actual language use in such a multilingual setting. Topics for

New in DoBeS: Documentation of Enets and Forest Nenets

Kaur Mägi, Florian Siegl
University of Tartu

Enets and Forest Nenets are two indigenous languages of Western Siberia and belong to the Samoyedic branch of the Uralic language family. According to the most recent data, the Enets live on the Lower Jenisej, on the Taimyr Peninsula and in the tundras south of the town Dudinka in the villages Vorontsovo, Potapovo and Tukhard (Taimyr Autonomous Region). The Forest Nenets live along the upper courses of the Pur, the Kazym, and the Agan Rivers in the Yamalo-Nenets and Khanty-Mansiisk Autonomous Region (Tjumenskaja oblast'). Although the Forest Nenets live mainly in the forests and forest tundra, they cluster around three villages – Varjogan, Numto, and Khalesovoj.

According to recent estimations, about 50–75 older people still use Enets, but it is no longer spoken by children. Forest Nenets is spoken by approximately 1000–1500 persons and in some areas it is still learned by children, but it is clearly losing ground. Whereas in earlier days, trilingualism prevailed in both areas, today bilingualism (Enets/Forest Nenets and Russian) or even monolingualism (Russian only) prevails. Basic documentation of both languages was made during the Soviet Period, but there is no comprehensive modern description of either language.

The Tartu–Göttingen project “Documentation of Enets and Forest Nenets” aims to cover the following:

- documentation and archiving of contemporary Enets and Forest Nenets according to DoBeS standards;
- investigation of the linguistic competence of the remaining speakers;
- sociolinguistic aspects of language usage;
- situational language use;
- contemporary multilingualism;
- socio-cultural documentation of the communities in post-communist Russia (especially land use and kinship patterns).

The project is hosted by Prof. Ago Künnap (Chair of Uralic languages, University of Tartu), and co-hosted by Prof. Eberhard Winkler (Dept of Finno-Ugric studies, University of Göttingen). Fieldwork will be conducted by PhD students Kaur Mägi and Florian Siegl. Their positions within the project are funded by the DoBeS project.



Roland Edirol R-1. Front view

Technical Section

Review: Roland Edirol R-1

Anthony Jukes, David Nathan

HRELP, SOAS, London

We have been waiting for a replacement for our trusty minidisc recorders, and the new breed of solid state (flashcard) digital recorders has looked promising. Initial impressions of the Marantz PMD670 were mixed (quality sound, but rather big, too expensive, and poor battery life), and while the newer PMD660 is smaller, at around €600 it is still rather expensive. Roland's Edirol R-1 is cheaper (about €500) and smaller than the PMD660, and has been eagerly awaited by many linguists. Here we evaluate the R-1's suitability for language documentation in real field situations, with particular attention to its usability and its performance using various microphones and recording formats.

Initial impressions

The R-1 seems generally flimsy. The case does not appear well-assembled and the door to the CF slot operates clumsily and is likely to break off. The large "VALUE" knob can be knocked off easily. There is no carry strap bracket. The unit's buttons are simple enough to use but do not have a satisfying feel. The mix of controls is odd: it has button-operated menus, analogue-deck-style control buttons, mechanical slide switches, and potentiometer wheels. Monitoring of input levels is limited by the lack of marks on the input level control and the basic LCD display. Microphone input is only via stereo miniplug, so microphones with XLR connectors will require an adaptor. The R-1 has no rubber feet and slides around uncontrollably on smooth surfaces. Given its less than rock-solid construction, we recommend sticking some feet on it, or, better, keeping handy a sheet of grippy rubber "anti-slip" mesh that is sold in many hardware shops.

Ease of use

The R-1 is easy and convenient to use if simple recording with the inbuilt microphone is all that is required. Some of the controls are annoying; for example display mode settings are not retained and have to be toggled after each recording.

Care needs to be taken – for example, turning the R-1 off before stopping recording (easy to do!) loses the entire recording, or even the memory card format. This in turn can cause the R-1 some confusion; the manual warns that the unit could even be permanently damaged. The manual itself is very clearly written, and is compulsory reading since it describes such situations where data loss might occur (however, we found



Roland Edirol R-1. Top and left view

inaccurate information, e.g. the memory card does not have to be formatted in the Edirol).

Digital storage and connectivity

The R-1 uses widely available Compact Flash memory cards. It handles cards up to 2GB (larger cards, including microdrives, will not provide any increased capacity). Having experienced card incompatibility problems with the Marantz PMD 670 (see LAN 4), we tried the R-1 with a number of cards. All worked perfectly, even though one of them was a Buffalo RCF-X 256MB, which looks suspiciously like one of the cards listed by Roland website as incompatible.

The R-1 has a USB output and quickly transfers files to a computer using USB2. Unfortunately the unit must be running on mains power to use the USB port, which could be a disadvantage in the field (e.g. if you need to move recordings off a full card to continue recording). The R-1 has no optical input. Provision of optical input would have made the R-1 useful for transferring sounds from other devices in the field.

Evaluation of test recordings

We made test recordings of the R-1 with various microphones and settings, and also made a brief comparison with a Marantz PMD660. Each of the microphones – the R-1's inbuilt stereo microphone, a Sony ECM-MS 957 stereo microphone, and an AT803b clip-on lavalier – was used to record speech at several settings: 24-bit/44.1kHz WAV (the R-1's highest) and MP3 at 320, 256, 192, 128 and 64 kbps. Recordings were evaluated by listening via the R-1 and Mac and IBM notebooks using Grado SR80 and Victor HP-DX1 headphones.

The R-1 is capable of making good quality recordings. Recordings at high bit rates MP3s (320 and 256 kbps) were perceptually indistinguishable from the 24-bit/44.1kHz WAV recordings. Predictably, lower bit rates yielded progressively less quality. At 192 kbps, artefacts became noticeable (as a fluttering of the higher frequencies in the background noise). By 64kbps, the sound was telephone quality. If conserving file space is the highest priority, there may be no real harm in using high bit rate MP3 settings (we can already hear some people throwing up their hands in horror!). Much more important factors influencing the quality of recordings were the choice of microphone, its placement and handling, and attention to the recording environment to reduce unwanted noise.

Microphones

We were disappointed with voice recordings made using the inbuilt microphone. They were rather noisy, with an unacceptable amount of hiss, significant handling noise, and oversensitivity to ambient sound. The R-1's microphone characteristics may reflect its musical origins and better suit musicians than linguists. The Edirol website (links: Edirol) has some fine recordings of musical instruments recorded with its inbuilt microphone. However, field recording of spoken voice has different requirements.

Using the R-1 with a Sony ECM-MS 957 microphone resulted in good, natural sounding recordings, with less noise. This combination would prove quite suitable for recording a single speaker, or a number of speakers in conversation. Best results for recording a single speaker were obtained using an AT lavalier microphone. It allows closer placement to the speaker's mouth which greatly attenuates background noise sources such as dogs, poultry, children, and other noisy creatures found in typical village settings. A great combination for many field situations would consist of using the R-1 with two microphones recorded in separate channels: a lavalier for the primary speaker and an omnidirectional microphone for audience responses. However, an adaptor would be required to connect the two mono microphones to the R-1's single stereo miniplug, and the R-1 does not allow the channel recording levels to be set independently.

The R-1 has a switch for selecting dynamic or condenser microphones, but the "dynamic" setting was required for best results using the Sony ECM-MS957 Electret condenser unit. This seems to be something to do with supply of "plug-in power" for MD (or so-called "digital") type microphones. Perhaps the condenser setting applies to those types only, and all others, including the R-1's internal, use the dynamic setting. Although we queried Edirol UK about this, we were not correctly advised, and we concluded that the switch changes microphone powering as well as sensitivity and frequency response, although not in the expected ways.

We also compared recordings made using the Edirol and a Marantz PMD660, both using a Sony ECM-MS 957 microphone. The Marantz showed a higher sensitivity (or better match to the microphone); other than this, there was almost no difference in the clarity or noisiness of the two recordings.

Finally, we found the R-1 quite susceptible to electrical interference when using its internal microphone. If

it was within 3 metres of a laptop computer, or even within 5 metres of a normal refrigerator (in another room!), a buzz was recorded. This effect was much less noticeable when using the external microphones.

Battery life

A pair of fresh alkaline batteries lasted for only 45 minutes of recording, although others have reported up to two hours. We were able to record for over two hours using fully charged NiCads. Some have reported that much better results are obtained with lithium batteries, and there are reports of over four hours of recording using rechargeable NiMH batteries. The R-1 gives only a few minutes warning when the battery level is low. It then saves the current recording and shuts down.

Conclusion

The Edirol R-1 offers the ability to make acceptable digital recordings by simply placing the unit on the table and pressing record. Using external microphones, high quality recordings can be made. However, many of the R-1's strengths lie in its musical heritage; for linguists, its apparent lack of robustness and limited microphone connectivity make it a second choice to the (admittedly more expensive) Marantz PMD660. Those who predicted that the Edirol's arrival would provide the perfect solution to linguistic field recording may be disappointed.

The MPI technical group adds:

The R-1 records and plays back in nine modes ranging from 64 kbps MP3 to 24-bit uncompressed WAV. It can record via its two inbuilt omni-directional electret microphones, or through external microphone and line inputs. Its headphone jack doubles as an S/PDIF out, so you can stream audio to digital equipment such as DAT and MiniDisc recorders. As per the main review, we also found that the microphone inputs had low sensitivity.

Initial testing found that there was a linear behaviour of the recording meters, which made them unreliable for setting an accurate recording level. Edirol have resolved this problem with an upgrade available from their website (which offers good support and downloadable upgrades). After upgrading, the meters are logarithmic and provide better visual indication of the incoming signal.

Links

Edirol: <http://www.edirol.com/products/info/r1.html>

LAN back issues available at:
<http://www.mpi.nl/LAN>

Review: Mayah Flashman

Gerd Klaas, Nick Wood
 MPI, Nijmegen

Solid-state (flash) memory recorders are a leap forward from the MiniDisc and DAT recording devices that have been the industry standard for the past ten years. Like MiniDisc and DAT recorders, they are small in size and record in digital formats; however, unlike those devices, solid-state recorders have no moving parts, which eliminates mechanical noise and wear and tear. They also have a potentially larger capacity, and store recordings as files that can be directly transferred to a computer, eliminating the need for an additional capture process.

The Mayah Flashman was originally designed as a recorder for in-the-field radio reporters. Released about 3 years ago, it was the industry standard for solid-state recording and retailed at approximately €1500. Thanks to its age and competition from other producers, it now retails at approximately €700–800.

The Flashman has no internal microphone – it is designed to be held in the left hand leaving the right hand free to operate an external microphone. As it is meant to be held in the left hand, most of the controls can be operated comfortably with the left thumb. Recording is via 2 microphone inputs (XLR and 3.5mm minijack) or line input. Recording modes range from 56kbps through to 256kbps (MP3) and 16-bit linear WAV. It has an S/PDIF digital output and an RS232 I/O port for connecting to a computer (to transfer settings, not data). There is no USB connection, so the memory card needs to be removed and read with a card reader. Although memory cards are currently up to 8 GB capacity (with 16 GB cards on the way), the Flashman can only handle flash cards up to 2 GB. Future software updates may remove this limitation.

The 4 AA batteries allow approximately 3 hours of recording. An AC adaptor is also provided, although rechargeable batteries will need an external charger, as the Flashman has no built in charging.

The Flashman's expense is not necessarily matched by its capabilities. Although marketed as a professional unit, recordings are not always of a higher quality than other less expensive recorders. However, the Flashman's microphone preamp does have more sensitivity than some less expensive models, so it does have the potential for making better quality recordings. It has high power requirements, and needs a good supply of batteries if not being used on mains power.

Solid-state recorders are still an evolving technology. At the time of writing, a new device, the M-Audio Micro Track 24/96, is due for release (links: m-audio), and aimed at the semi-professional market. Once tested by the MPI Technical Group a review will be posted.

Links

Mayah website: <http://www.mayah.com>

m-audio: <http://www.m-audio.com>

Review: Maycom Handheld II

Rob Kennedy
SOAS, London

The Maycom Handheld II is a self-contained handheld device offering high quality audio recording in WAV and MPEG 2 Layer II formats. About the size and weight of two mobile phones, it contains a mono microphone, operating controls, LCD display, compact flash card, USB connector, batteries, tiny loudspeaker, and a headphone socket. Additional connectivity is via the recorder's cradle unit, which provides a pair of XLR sockets for stereo microphone or line inputs, an extra headphone socket, stereo minijack (3.5mm) line output, speaker, USB "B" socket, and charging for the recorder's NiMH battery pack. Docking the recorder into the cradle requires care; the cradle connector looks easy to damage.

Controls

Most operating controls are situated on the handheld unit. However, the supplied *HHConfig* software (Windows only) must be used to make settings such as file format, sampling frequency and bit rate. These recording parameters are defined within *profiles* which are set within the software and transferred to the handheld unit via its (non-standard) USB 1.1 connector. For example, you might define a best quality profile for field recordings such as BWF with 48 kHz sampling rate, AGC off, no limiter, and LCD backlight off to conserve battery power. A second profile for audio note-taking might use MPEG compression, with both AGC and backlight on. Once the profiles are defined and named, they can be uploaded to the handheld. In the field, the appropriate profile can be chosen using the handheld's LCD menu. This system means that if you do need to change a profile's parameters, perhaps just before starting an important recording session, you will also need a computer handy.

On the handheld unit, switches and buttons are minimally labelled. Play, pause and wind/rewind (borrowing the analogue terminology) are marked with the normal symbols, and the pre-record button is red. However, record gain, playback volume and record mark are not labelled. I would advise sticking on labels to help identify these for the first weeks of use. The small LCD menu screen has a backlight which can be turned on or off via profiles or via the menu. But a separate backlight on-off switch is needed since in dark conditions you cannot see the menu screen to switch on



Maycom Handheld II

the backlight if it's already off! The manual (supplied as PDF in several languages on the software CDROM) notes that the backlight increases battery consumption by 5%.

Microphone

The Maycom's microphone is good at handling plives down to about 20cm. Although omni-directional according to its specifications, we found that higher frequencies are noticeably reduced for off-beam recordings, so be sure to aim the microphone directly at the subject when recording. It is highly susceptible to wind noise, even indoors, for example from the air moving across it when moving the recorder quickly

Suggestions and contributions welcomed at:

LAN@mpi.nl

Next deadline for copy:

December 1, 2005

between two subjects. Wind noise is reduced but not eliminated by using the supplied windshield. Adjusting gain is easily done via the up-down controls beside the display, though during adjustment the level readout changes disconcertingly from a bar graph to a numerical format, and then takes a full second to revert.

Handling noise was a problem; any movement of fingers against the plastic casing was clearly perceptible in the recording; this was not reduced by wrapping a soft cloth around the case. As with all microphones, the only way to avoid handling noise is to not move your hand on the unit during recording. If the situation allows it, the Maycom can be used with the supplied mini tripod, or placed in its cradle on the plastic mounting plate. Of course the use of such tripods or stands should not remove the need for close mic-ing.

Audio quality

Audio quality of the WAV recordings was excellent as long as care had been taken to avoid handling and wind noise. There was very slight hiss, probably introduced by the microphone preamp. However, MPEG 2 Layer II sample files suffered an unacceptable low-level whine when transferred to either of two PCs via USB or a CF card reader. Strangely, this whine was not present on the Maycom's analogue outputs. The internal speaker is convenient for making quick playback checks but does not provide much volume, so in noisy environments it needs to be pressed to the ear, or else substituted by closed headphones.

A 256MB Compact Flash card held just under 30 minutes of mono WAV data at 48 kHz, and three hours for MPEG2 Layer II data at 192kbps. The manual and profile editor make no explicit mention of bit depth.

Battery life

The supplied NiMH battery pack lasted just over four hours during moderately intensive work. A handy button shows the remaining charge via four LEDs. If an additional battery pack is purchased, one pack can be charged in the cradle while the other is in the handheld. Alternatively the Maycom runs on four alkaline AAA batteries, though we did not test the life of these.

Conclusion

Overall we found the Maycom Handheld to be stylish and generally easy-to-use, giving excellent quality WAV recordings. We have doubts about the ruggedness of the cradle docking connector; we would like to see explicit information on bit depth, and are mystified by the digital noise in our MPEG recordings. The all-in-one configuration is extremely convenient – as long as the recorder is held carefully so as to minimise handling noise. The Maycom should well serve a regular user who knows its strengths and weaknesses.

Links

Maycom website: <http://www.maycom.nl/handheld.html>

News in Brief

Second HRELP Grantee Training Workshop London, June 2005

Peter K. Austin
ELAP, SOAS, London

HRELP held its second training workshop for ELDP grantees in June at SOAS in London. The workshop was open to all new grantees and those from previous years who had not attended the first training workshop in September 2004. A wide range of expertise and experience was represented, from PhD students about to begin their projects to experienced researchers nearing the end of their funded research. While this presented the trainers with some challenges it also gave unique opportunities for the grantees to share their diverse experiences. Participants came from a variety of countries including Russia, China, Australia, US, India, Indonesia, and Germany. Session topics were both practical and theoretical, including making quality sound recordings, lexicography, software tools and techniques, data formats, fieldwork methodologies, archiving, ethics and intellectual property rights, and documentation outcomes. Presenters included ELAP and ELAR staff from SOAS, as well as outside specialists in anthropology and film making, and some sessions were led by grantees. An evening round-table discussion on "Current directions in language documentation in Latin America" complemented the training sessions, and social events were included this year to enable interaction outside the classroom context. Participants gave positive evaluations of the week.

The next ELDP training workshop will be held in mid-June 2006. For more details please contact Jacqueline Arrol-Barker (ja30@soas.ac.uk). See also <http://www.hrelp.org/events/workshops/eldp2005/>

EMELD 2005 Harvard University, July 2005

David Nathan
ELAR, SOAS, London

Over 70 international participants attended EMELD 2005 "Linguistic Ontologies and Data Categories for Language Resources", held at Harvard University in Cambridge, MA. Faced with growing interest in some quarters for machine-processable linguistic data, ontologies have surfaced as a focal methodology, in an attempt to align linguistic data handling with projects

like the semantic web and automated translation systems. A short but varied and entertaining conference, it consisted predominantly of two amicable but opposed groups – those of a computational bent who would like linguists to produce their data in a more computational-friendly form, and “ordinary working linguists” (OWLs) who would like the computational experts to produce more linguistic-friendly software for them.

The conference consisted of about 30 short papers on related topics (link below), interspersed with Working Group sessions. Most of the papers fell into either of the camps above, i.e. either promoted the value of ontologies, in particular the GOLD ontology produced by EMELD in collaboration with Terry Langendoen and others, or else raised complex or fuzzy linguistic phenomena that linguists struggle hard enough to nail down in the relatively unconstrained space of a particular language’s grammar, let alone using a predefined set of categories.

This dialogue was amplified in the Working Groups, which brought the two constituencies head to head in an attempt to refine GOLD’s categories and mechanisms. In one group, the OWLs perhaps felt unqualified, intimidated or frightened of the consequences, to take on the full task of pouring their morphosyntactic knowledge into GOLD’s pan. Ultimately this was the sticking point, because while the computational agenda spiralled in on establishing the superset of morphosyntactic labels that might appear in, say, a conventional (Shoebbox-generated style) interlinearisation (under the title IGT “interlinear glossed text”), the linguists sought both to free themselves from such simplistic, item-and-arrangement, single level of description, and were rather hoping that the computationalists would liberate them by inventing new software that could handle linguistic flourishes such as discontinuity, gradation, phrase and clause structure, and even semantics and pragmatics. The epitome of this disconnect was the circularity of the computationalists’ promotion of the value of linguists using GOLD for glossing web-based interlinear data on the basis that such data could then be computationally harvested to induce grammars from such datasets. This left us wondering whether the GOLD ontology is primarily aimed at providing interoperability to typologists wishing to amass and compare large datasets, optimising opportunity for computational linguistic induction, or providing a termset for glossing in Shoebbox.

However, despite such tensions and contradictions, such foundational interfaces and discussions between linguists and computationalists are welcome and indeed needed. More than that, the whole conference was delightfully organised, suffused with good humour, and thoroughly thought provoking.

Links

EMELD <http://emeld.org/workshop/2005/program.html>

LSA: Language Documentation Harvard University, July 2005

Peter K. Austin
ELAP, SOAS, London

This conference on the theory, practice, and values of language documentation was organised by the Linguistic Society of America Archivist Professor N. Louanna Furbee and members of LSA Conversation on Endangered Languages Archiving. There were about 80 attendees. Six themes were treated via conventional paper presentations, poster sessions, and panel presentations with open discussion:

- the requirements of field linguistic and language documentation training;
- the concerns and involvement of heritage language communities;
- the question of what is adequate documentation;
- the uses of documentation in speaker communities;
- training and careers in field linguistics; and
- ethics and archiving best practices.

A wide range of endangered language situations was covered, including Native America (US, Mexico), Australia, Siberia, the Pacific, and East and South-East Asia. A particularly pleasing aspect of this conference was the involvement of speakers of endangered languages, including Karuk, Tohono O’odham and Nez Perce (North America) and Kurna (Australia). A selection of papers arising from the conference is to be edited for publication by N. Louanna Furbee and Lenore Grenoble. For more information, see <http://www.lsadc.org/languagedocumentation/>

IFLA: The Multicultural Library Stockholm, August 2005

Peter K. Austin
ELAP, SOAS, London

In association with the International Federation of Library Associations (IFLA) annual conference in Oslo, Norway, the IFLA section on library services for multicultural populations organised a satellite conference on “Staff competence for success in the multicultural library” in Stockholm, Sweden. Several sessions touched on language documentation, archiving and endangered languages issues, especially under the general theme of “Language Rights”. Peter Austin (SOAS) gave a presentation on “Endangered Languages

and Languages Documentation”, followed by Leena Huss (Centre for Multiethnic Research, University of Uppsala, Sweden) on “Nordic language policies in transition”. Edgardo Civalero (National University of Córdoba, Argentina) discussed “The Sound Library” project set up to collect archival and contemporary recordings of Argentinian indigenous languages and to return them as cassette libraries to local communities. The audience of librarians familiar with larger minority languages was struck by this passionate presentation illustrated with sound recordings of a number of extinct and threatened languages of Argentina. Other sessions dealt with issues including Unicode, multilingual cataloguing, metadata, and delivery of multilingual materials to dispersed and often disadvantaged clients. For more details and the full programme, see <http://www.ifla-stockholm2005.se/>

ELAN – Latest Changes

Han Sloetjes, Hennie Brugman, Alex Klassmann
MPI, Nijmegen

ELAN has recently seen two minor releases (2.4.1 and 2.4.2), bringing new features and bug fixes. The main new features and improvements are:

- Export as Interlinear Text. This option generates a UTF-8 text file of interlinearised (“Shoobox style”) annotations on a selection of tiers. The output can be limited to a selected time interval, and can be extensively customised, including the sort order of the tiers. A preview shows how the export will look in a text editor. The printing feature now implements line wrapping. Overall performance in this area has been improved.
- The maximum number of constraints in the search dialog has been increased to 10. The results of a search on multiple files are now presented in browsable pages of 50 hits each.
- On Mac OS X the default appearance of the user interface now more closely resembles the native Mac look and feel. It is still possible to choose the previous “cross platform” look and feel.
- Shoobox/Toolbox export now allows the user to specify a subset of tiers, their order, and the number of characters for block wrapping. On import a custom duration per block or record can be defined.

The main bug fixes are:

- Blockwise wrapped lines in Shoobox files are now correctly converted to annotations on import to ELAN.

- When searching over multiple files, file-read errors are now handled more gracefully. Unaligned annotations in the search result table no longer show invalid (negative) time values.
- The behaviour of several interface elements has been improved, e.g. windows that did not fit on the screen, disappearing elements etc.

In August, we released a new and improved source distribution of ELAN. This distribution includes source, libraries and an Ant build file for easy compilation. ELAN executables and source files are available at <http://www.mpi.nl/tools>

MPI Training Courses for Field Linguists

Romuald Skiba, Paul Trilsbeek
MPI, Nijmegen

For 5 years, the MPI for Psycholinguistics has regularly run training courses for field linguists. These courses started with the arrival of the DOBES project, when we started to develop new procedures and tools for archiving, together with courses for training the DOBES documentation teams on these procedures and various tools for language documentation. The courses were initially based on experience that MPI developers had gained since the 1980s from working with locally-created corpora on bilingualism. However, a lot of knowledge was also gained from specialists within the DOBES projects during its pilot phase. The transformation of all this knowledge into theoretical and practical course materials has resulted from close collaboration between linguists, software developers and corpus managers.

The lecturers include both technically and linguistically trained people. Participation in the courses has been open not only to DOBES teams, but also to other interested people. Participants have come from different fields within empirical linguistics, including endangered languages, language acquisition and sign language typology.

The course program typically runs over 5 days and follows the standard order of fieldwork events: audio/video recording, data segmentation, metadata creation, annotation, and exploration, as follows:

Day 1: Gives a general overview of the archive. Corpus examples are shown, the roles of the various groups involved in archiving (depositor, archivist, other users) are explained; plus an in-depth look at all technical aspects of the archive.

Day 2: Deals with recording techniques for both audio and video. Various recording devices are shown; power management for field situations is discussed.

Day 3: Covers digitisation (capturing data to the computer), the creation of metadata descriptions and the connection between data segmentation and the metadata. The IMDI editor is presented as a tool for making metadata descriptions. Digitising and editing of data are covered both practically and theoretically.

Day 4: Covers data annotation and exploration using a number of widely used tools (e.g. ELAN, Shoebox), as well as conversion between different annotation formats and associated problems.

Day 5: Offers a recapitulation of all previous steps plus practical exercises.

All topics have written course materials, some of which are available on the web at

<http://www.mpi.nl/corpus/a4guides/>

Workshop on Endangered Languages and Literacy SOAS, London, December 2005

Peter K. Austin

ELAP, SOAS, London

The Endangered Languages Academic Programme at SOAS will be organizing a one-day workshop on the theme of literacy and endangered languages on December 3, 2005. There will be a number of presentations by invited speakers on topics such as orthography development, literacy and orality, literacy in education, and literacy and multimedia. Further details including a programme and registration form, will be available in October at

<http://www.hrelp.org/events/workshops/>

ELAR at SOAS Begins Archiving

David Nathan

ELAR, SOAS, London

ELAR's start-up for archiving services was set back due to delays at SOAS; however, progress was made over summer, so that we will now be able to take archive deposits at the end of October 2005. We recently moved into our newly-built work area, deposit forms and information system are under construction, and our mass data storage systems (including a four terabyte disk array mirrored at another campus, a robotic tape backup unit, and an agreement with the Oxford Text Archive to supply further backup), are currently being commissioned. ELAR welcomes digital materials on all endangered languages; for further information, please see

<http://www.hrelp.org/archive/>

Announcements

DoBeS:

Application Deadline November 1, 2005

Applications within the DoBeS project (Documentation of Endangered Languages) will be accepted for documentation projects for a period of up to three years. For more information please contact Dr. Vera Szöllösi-Brenig (szoelloesi@volkswagenstiftung.de) or see <http://www.volkswagenstiftung.de>

LANGUAGE

ARCHIVES

NEWSLETTER

www: <http://www.mpi.nl/LAN/>

Mail: LAN@mpi.nl

ISSN: 1573-4315

Editors: David Nathan (ELAR, SOAS, London)

Romuald Skiba (MPI, Nijmegen)

Marcus Uneson (Lund University)

Layout: Marcus Uneson

Contributions welcomed at:

LAN@mpi.nl

Last submission date for the next issue:

December 1, 2005