LANGUAGE ARCHIVE NEWSLETTER

DEC. 2004
VOL. 1, NO. 4

IN THIS NUMBER:

Language Archive Newsletter. www: http://www.mpi.nl/LAN/. Mail: LAN@mpi.nl. ISSN: 1573-4315.
Editors: David Nathan, Romuald Skiba, Marcus Uneson. Layout: Marcus Uneson.

Editorial

Reviews Invited

LAN would like to invite its readers to send in reviews of equipment that they have used for fieldwork, data processing, or other tasks relevant to digital language archiving.

We live in an era which where our work often needs to be done urgently under unstable circumstances. Contributing factors include the rapid rate of language loss, the nascent but unevolved field of documentary linguistics that frames much of our work, and changing and converging technologies. Documentary linguistics typically involves technologies such as audio and video recording, digitisation, specialised software and data encoding, and, of course, digital archiving. A structured approach to these technologies creates an integrated stream of data from elicitation to archiving, an approach to data that could be called "born archival". Other approaches use technologies in less formal ways, or for complementary purposes such as for mobilising data for direct use in pedagogical, community-centred multimedia products. However, almost everyone doing language documentation today will be using equipment and technologies that were unknown only a decade – or even a year! – ago.

Therefore, we invite all types of reviews, giving readers the chance to pick out what they think is interesting or important. We will be pleased to hear about your adventures in this new territory.

On a similar note, we also invite comments and tips on relevant readings – books, articles, useful links, etc – that you believe your colleagues may find useful (or that you wish to warn them about!). Comments could be of any type, ranging from a full-fledged review to just a tip (see for instance the section "Readings and Links" in this issue). Sharing knowledge will help keep us all up-to-date.

Yours sincerely,

David Nathan, Romuald Skiba, Marcus Uneson

Endangered Languages

Documenting Sri Lanka Malay: Linguistic and Cultural Creolization

Umberto Ansaldo
University of Amsterdam

The project aims to provide a full documentation of the restructured Malay varieties spoken in the Malay diaspora of Sri Lanka, with particular emphasis on the variation found across the different speech communities. Now endangered, as they are no longer acquired as a first language with the exception of perhaps just one community, Sri Lanka Malay varieties are valuable for studies of language contact, evolution and creolization as they have evolved in an environment in which no standard European acrolectal variety is involved; the main influences on the evolution of Sri Lanka Malay can be found in Tamil and Sinhala. These varieties offer valuable theoretical perspectives on typological convergence, restructuring under contact and creolization theory.

Headed by Umberto Ansaldo (Universiteit van Amsterdam), whose research focuses on linguistic typology, creolization and languages of East and Southeast Asia, the research team consists of phonetician Lisa Lim (UvA) who oversees the acoustic and visual aspects of the documentation process, Walter Bisang (Universität Mainz) whose specialties include typology and universals, language contact and grammaticalization as well as extensive experience in Asian languages, and Thiru Kandiah (University of Peradeniya, Sri Lanka) who provides the team with expertise in issues of language, culture and ideology. The project also funds doctoral student Sebastian Nordhoff who will start at the Amsterdam Centre for Language and Communication (UvA) in January 2005 and engage extensively in fieldwork and documentation.

Contributions (or proposals for contributions) are welcomed at LAN@mpi.nl
Deadline for copy for the next issue is **February 15, 2005**
Back issues are available on the LAN website at http://www.mpi.nl/LAN

Archiving

Technology and Tools for Language Documentation

Peter Wittenburg, Romuald Skiba, Paul Trilsbeek
MPI, Nijmegen

Language documentation and preservation

When speaking about technology and tools for language documentation, there are two important aspects to keep in mind. First, technology and tools are continuously changing. Something that is state of the art today can be outdated tomorrow. Second, people involved in language documentation do not all have the same needs and preferences, and therefore may have different criteria for making choices. Documentation creators, for example, are typically concerned with familiarity and ease of use, whereas for archivists, technical quality and long term preservation are more important.

The language documentation process consists of several steps (some of these take place in sequence, others in parallel):

- creation of recordings
- transferring and manipulating recordings using computers
- transcribing, annotating and pre-analysing recordings
- integrating materials into an archive infrastructure
- exploring and re-analysing materials for various purposes
- making materials accessible to different user groups
- protecting materials against misuse
- preserving materials for use by future generations

For each of these activities we can identify relevant methods, standards and frameworks, the sum of which we may call technology. Technology can be applied through the use of tools – e.g. DV technology is used in digital video cameras and XML technology is used in the annotation tools Transcriber and ELAN. Language archives need technology and tools to support the ongoing management of data and to provide for various usages of data by different user groups.

How can we satisfy different user groups?

We envisage that the following groups of users may wish to access endangered languages archives, each group bringing their own specific needs:

- language communities and linguists who would like to access language material for educational or related purposes may require an educational style of presentation.
- local centres may want to have complete copies of digital archive materials, including metadata and structural information, in order to provide local, flexible usage of the data.
- linguists and other researchers may want to access a cross-linguistic selection of data for a typological study.
- teachers of linguistics may want to demonstrate to students how languages can differ by comparing annotated recordings of different languages.
- journalists may want to create a story about language diversity for the interested public in collaboration with linguists or members of a speech community.

While archives might aim to support a large variety of usages, it is almost impossible to create customised access tools for every conceivable group of users. What archives can do is provide access to data in such a way that users can easily find and retrieve the data they need. Therefore, an archive can offer:

- well defined and well documented data formats – preferably open
- extensible archive and data structures
- detailed metadata descriptions
- powerful search and exploration tools for content and metadata
- easy, yet secure, access to the data

Users can use search and exploration tools to create their own resources from the archive. Or they can use information about the archive and its data structures to help create their own access and presentation tools for their own local purposes.

Media, codecs, formats—can we predict the future?

Today's language archives almost exclusively archive material in digital form. This has advantages such as perfect reproducibility; however, there are also great issues to deal with such as the limited lifetime of storage media and the continual changes in data formats. To ensure that materials can still be read after some hundreds of years, an archive has to make the migration

of data to the latest storage media, file formats, and codecs as easy as possible.

In traditional archives, materials can be accessed using the human senses (typically, the eyes). Although it may be necessary to understand a special Sumerian encoding system to understand the content of a clay tablet, the eyes alone can identify signs and their patterns. In digital archives, material is stored as magnetic sequences of ones and zeros, so we always depend on computer hardware and software to provide access to the content.

Even after the magnetically encoded data has been accessed, it may not be directly usable. We have to recognise various layers of encoding, such as are required for audio and video:

- codecs that determine how audio and video streams are represented by bit streams (a parallel for text would be how particular glyphs are encoded by particular sequences of bits)
- file formats that determine how these bit streams are packaged into units that are the objects handled by operating systems and application software
- tools that process bit streams or present them on screen or in print

Examples of video codecs are MPEG1, MPEG2, MPEG4 and DV, while video file formats include AVI and MPG. Examples of audio codecs are Linear PCM, MP3 and MD-Atrac, while common audio file formats are WAV and AIFF. This layered system means that an AVI file can contain video streams with different codecs, so the specification "AVI file" does not identify the encoding or quality of its video content. Various software tools support different codecs and formats, so for each tool one has to check what is supported.

For archiving purposes, open and well-documented standards such as MPEG, UNICODE and XML should always be used. It can be assumed, for example, that MPEG codecs will continue to be used for many years and that there will therefore be tools available that can decode MPEG bit streams. For encoding characters, UNICODE is recommended despite current limitations in the range of characters represented. For structuring documents, XML is recommended; in addition to being a widely adopted standard, XML files are human readable, and can be viewed and edited using even the simplest text editor.

Software tools that encapsulate information content in a proprietary format – such as MS Access, FileMaker Pro, MS Word and Excel – do not provide files that are appropriate for archiving. There is a contradiction between the short-term needs of linguists and the long-term needs of archivists. While linguists prefer tools that efficiently provide data entry and presentation, archivists are more concerned with data representation, i.e. encoding and format standards.

For continuously time-varying signals, such as audio and video, archives prefer to store the most simple, direct and high quality digital representations of signals. For audio, a procedure called Linear PCM defines a temporal resolution (such as 44.1 or 48 kHz), measures the value of the signal's amplitude in a particular resolution (e.g. 16 bits) at equidistant points in time, and then stores those values in sequence. The amount of data produced in this way is not so great as to cause problems for storage; compression such as MP3 or MD-Atrac is not necessary. However, many recording devices do apply these compression techniques, which remove spectral and temporal components from the original signal (components which are claimed to be filtered out by the human ear anyway). Due to this unrecoverable reduction in information, and the additional complexity in decoding compressed files, it is generally recommended not to use compressed formats for material to be archived.

Turning to images, all digital still cameras produce files in the JPEG file format which has become the de facto standard even though it applies lossy compression which deletes high-frequency image components. However, many cameras are now offering RAW file format, which is not only uncompressed but allows adjustments of settings (white balance, ISO sensitivity, etc.) after shooting. RAW files are of course larger in size, but with flash memory prices dropping rapidly, this is no longer a problem.

For video, however, current technologies are not able to handle uncompressed video streams. Various compression formats such as MPEG1/2/4 and DV have been developed. Each of these has its own disadvantages. DV is used by almost all camcorders, but its data rate is too high for current storage media. MPEG2 can currently be seen as a good compromise between archiving needs and tractable data rates, while MPEG1 and MPEG4 can be seen as derived formats for special purposes.

The choice of tools and technologies influences the quality, stability and durability of materials. While not all the language documentation steps listed at the beginning of this article are directly related to archiving, the tools and technologies used in all of those steps ultimately affect the quality and stability of the archive that holds the materials. Therefore, to characterise or evaluate an archive, documentation of the technologies and tools used in all phases of material creation is needed.

Technical Section

Video Recording in the Field

Jochen Cholin
(formerly) MPI, Nijmegen

Fieldworkers have to deal with a wide variety of tasks under difficult circumstances. It is extremely important to have the appropriate equipment, suited to your individual needs, and to know how to handle it correctly. The better acquainted you are with the equipment, the more you can concentrate on the actual research. Therefore, choose appropriate equipment, and get to know it well before you go to the field. Set up the equipment fully, and try several potential set-ups until you feel comfortable with it. Make test recordings, including sound. Before you leave for the field, always personally check that all your equipment is present and properly packed.

Equipment list

The following can be used as a "shopping list" for preparing for video recording in the field.

Video

1. MiniDV camera. MiniDV is currently the best format because the cameras and cassettes are well engineered, and, since they are very popular, available almost all over the world.
2. Sufficient MiniDV tapes.
3. UV filter for lens protection.
4. Tripod with correct plate.
5. A wide-angle lens that is compatible with the camera. Choose one that does not cause vignetting (loss of image at the edges of the frame).
6. At least two rechargeable batteries.
7. Stand-alone battery chargers that can also supply power to the camera; a mains powered one, and a cigarette lighter connector for powering from a car. Recharge batteries in the stand-alone charger, not in the camera.
8. Electrical adapters for all countries you will visit, covering both plug types and supply voltages.
9. Remote control.
10. Other batteries, e.g. for the remote control.
11. White paper and a marker to label the clapboard (white paper can also be used to set the white balance).
12. Clean, well packed cleaning tissue and a cleaning tape.
13. Waterproof bags and silica gel.
14. Manuals.

Audio

1. Several microphones with different characteristics, including a lapel microphone.
2. Good quality headphones.
3. Microphone stand.
4. Spare batteries for each microphone (if the microphones are not powered from the camera).
5. Several microphone cables of different lengths, with 3.5 mm stereo jack (miniplug). Good quality cables have each channel individually shielded.
6. Mono-to-stereo-adapter (if you use a mono microphone).

Using the camera

Camera basics

1. Always try to set up the camera and sound equipment at least ten minutes early, so as to avoid stress. This will result in better recordings.
2. Use a tripod whenever possible, even if only a small lightweight one. If a tripod is not available you can use a table, chair, or similar solid object.
3. If it is impossible to use a tripod or substitute, hold the camera with both hands, as straight and still as possible. If you move it, move it slowly and gently unless there is a particular reason to have faster movement.
4. Generally you should switch on the steadyshot (optical picture stabiliser).
5. Never use the digital zoom. It results in extreme loss of picture quality. Turn it off via the camera menu.
6. Never use the LP-mode (long-play). It can result in problems with playback. Turn it off via the camera menu.
7. Fast movement, such as gestures, can be recorded with greater accuracy using a higher shutter speed if there is enough light to do so. For example, a shutter speed of 500 rather than 50 will result in less motion blur.
8. To film yourself, you can rotate the LCD-monitor, switch on the mirror function in the camera menu, and then control the camera via the remote control. This can be useful for setting up

the camera by yourself before asking someone to be filmed.

9. The recording indicator light that comes on when you start recording tends to make people nervous; you could try putting electrical tape over it.
10. The interviewer should not see him/herself while recording—it may be necessary to turn off the LCD monitor.
11. Always record sound using the 16 bit setting (set via the camera menu).

Camera position, image perspective and size

1. Wherever possible, work with the sun behind you.
2. Do not position the camera too far from the subject. However, neither should you approach too close to human subjects because they might be distracted or alarmed (particularly if they are shy). If the subject is a single person, position the camera about one to two metres from them.
3. To get the right perspective, place the camera at the same height as the subject's head.
4. Select an appropriate picture size using the zoom. For example, if you are interested in capturing facial expressions you might zoom close around the subject's head. Or if you wish to analyse gesture, you might leave a space of one arm's length around the head.
5. Avoid zooming while recording.

Main image functions

Three functions are crucial for achieving good images:

1. Focus—for getting a sharp picture.
2. Exposure—for adjusting brightness.
3. White balance—for getting the right colour.

These are commented more in detail below.

You might prefer to use the camera's automatic settings, or to adjust it manually. However, it is generally recommended to allow the camera to make automatic adjustments initially and then to switch to manual (to turn off further automatic adjustments).

Focus

1. The most important feature of the image must be in focus.
2. Autofocus can work very well. However, it can be affected under various circumstances, such as when the subject is not at the centre of the image, the subject is moving, someone walks near the camera, or there are objects in the foreground. If the distance between the camera and subject is unlikely to change (for example during an interview) select autofocus to make the initial focus setting and then switch it off again.
3. To focus accurately on the subject, zoom in closely on the subject's eyes, adjust the focus either manually or automatically, and then zoom out again. This will not work, of course, if the distance between the camera and subject changes during filming.
4. In low light conditions the automatic focus may not function correctly.

Exposure

1. Video cameras have a limited exposure tolerance. This means that if there are both very light and very dark areas in the same picture, either the light or the dark areas (or both) will have a poor level of detail.
2. The most important features in the image must be correctly exposed. This may leave other parts of the picture over- or under-exposed.
3. If possible, make sure that the sun is behind you. Avoid facing into a light source; this may be unavoidable in some circumstances, for example if a person is in front of a lighted window. The person will appear too dark unless exposure is compensated using the backlight function (this does not work in every case).
4. When using semi-automatic exposure, zoom in on the most important part of the picture (such as the face), select automatic exposure, allow the camera to set the exposure, then switch automatic exposure off again, and zoom out again to select a suitable picture size.
5. When using manual exposure, switch off the automatic exposure and set the exposure using the adjustable exposure ring. (Important for many situations!)

White balance

1. What appears "white" may vary in different environments. Just as the human eye adapts to the ambient light quality automatically, the camera also has an automatic white balance function.
2. Automatic white balance works well most of the time. However, if colours do not appear realistic, you need to correct the white balance. Hold a sheet of white paper in front of the camera to completely fill the viewfinder, and then press the white balance button. Or manually choose one of two presets: daylight or artificial light (tungsten).

Making recordings

Preparing to record

1. Before you begin using a new tape or start filming a new scene, write the main aspects of the event to be filmed (subject's name, date and location of filming etc.) on a sheet of paper. Then film the page for 20 to 30 seconds while speaking the same information into the microphone. Rewind and check the recording immediately. This step is important not only because the cassette label can get lost, but also because the tape at the beginnings and ends of cassettes can be faulty.
2. Also film a clap of your hands, which will provide synchronisation of image and sound in case you want to work on them separately later.

Recording

1. Each shot should start at least 5 seconds before the action begins and continue at least 5 seconds beyond the end of the action.
2. DV records time code data as well as the image and sound. If the timecode is not continuous throughout the cassette, different scenes will have the same timecode, which makes digitizing and editing difficult. To produce a continuous timecode, make sure every new shot starts at or before the end of the previous one. If you simply stop recording and then start recording again the camera will take care of the timecode. However, if you have removed the cassette or used the playback controls, you should start recording one or two seconds before the end of the last recording. To find the end of the last recording, use the end-search button if the cassette has not been taken out of the camera; otherwise, use the edit-search button to play back until shortly before the end of recording.

When the tape is full

1. Do not run the tape right to the end, because the last 15 seconds may be faulty.
2. After you take the cassette out of the camera, label it immediately, protect it against re-recording by sliding the record-prevention tab, then store it safely.

Sound recording

It can be more difficult to make good recordings of sound than images. It is for that reason that TV and movie crews have one or more members dedicated to recording audio. The difficulty with sound can be understood by contrasting it with images: with the lens you can choose—and see—what you want (and don't want) in your picture; sound, however, comes from all directions and distances (see also the article *Sound recording and microphones* in LAN 1:3, July 2004).

1. Use an external microphone, if possible.
2. Switch off the zoom function of the camera's internal microphone (set via the camera menu).
3. For each recording, use a microphone with characteristics appropriate to the recording situation and environment.
4. Secure the microphone cable to the camera using a cord grip.
5. Most external microphones use batteries. These microphones have to be switched on before use. When the microphone is switched on, a red light glows momentarily to indicate that the battery has charge.
6. Remember that the human ear is good at ignoring background sounds while the microphone cannot. Therefore, keep the acoustic environment as quiet as possible: switch off air conditioning, refrigerators, radios, and ask bystanders to remain silent.
7. Locate the microphone as close as possible to the source of the sound to minimise the effect of background noises.
8. To avoid aspiration, "popping", or the sound of breathing, aim the microphone slightly offset from the mouth and at an angle to it (and use a windscreen).
9. Always do a careful sound check using quality headphones before starting to record.
10. Coping with wind: when recording sound in windy conditions use a windscreen on the microphone. Windscreens are effective only in light winds; in worse conditions you will have to find a sheltered corner or shield the microphone from the wind with an umbrella or similar.
11. If your recordings have hum, noise or other interference, check all connections and cables. Do not place sound cables on or near power cables. Try using batteries instead of mains power. Avoid touching equipment unless necessary.

Caring for equipment

1. All items of equipment should be handled very carefully, especially the camera, which is a sensitive instrument.
2. Never touch the lens with your fingers; clean it before every recording.

3. Unless you are very experienced, take the camera's manual with you (also, read it first!).
4. If you take the camera to areas of extreme conditions, such as high humidity, or airborne dust or sand, use waterproof bags to protect it. Use silica gel bags to avoid condensation (these can be obtained in most outdoor sports stores). Bags that have turned pink through exposure to damp should be dried in the sun, or near a fire, until they turn blue again.
5. If you are working under extreme climatic conditions, leave the cassette cover of the camera open for at least half an hour to allow the camera to acclimatise. This is also necessary if the humid/moisture sign appears in the camera viewfinder.
6. Never expose equipment to direct sunlight for a long period. Take an umbrella to provide shade. In particular, tapes and batteries should be kept cool and dry.
7. Never let sunlight shine directly into the viewfinder, because the viewfinder lens focuses the light and the viewfinder may be damaged.
8. Never leave an LCD screen exposed to direct sunlight for a long period.
9. When recording in low light conditions, the picture quality will be poor.
10. If you do not use the recording function for a long time, take out the cassette to protect the camera heads and to save power.
11. If you do not use the camera for a long time, switch it off to protect the camera heads and to save power. Many cameras switch off automatically after 5 minutes without use.
12. Use rechargeable batteries until they are empty. Each recharging lessens the life of the battery.
13. Empty rechargeable batteries should be recharged as soon as possible to prevent deep discharging.
14. If the recorded picture or sound is poor due to tape effects such as dropouts, insert a cleaning tape and run it for 10 to 15 seconds. Check the result and, if necessary, run the cleaning tape again. If the problem is not solved, the camera needs to be checked. Do not rewind cleaning tapes.
15. At airports, DV cassettes must be sent through the luggage inspection, not carried through the walk-through personal security gate, which uses magnetic fields. Aluminium foil will not fully protect against magnetic fields.

Summary

Ultimately, the challenges of video and audio recording in the field will require researchers to find the best compromises between their scientific goals and what is actually possible.

Here is a summary checklist that you can use in the field to check the most important steps:

1. Insert charged batteries and new tape into camera.
2. Mount camera on tripod (if tripod is used).
3. If a wide-angle lens is to be used, mount it.
4. Clean camera lens.
5. Adjust tripod to the correct height and aim the camera.
6. Select an appropriate microphone, plug in, set up, switch on, check battery status (red light glows momentarily).
7. Plug in headphones.
8. Write all necessary information on clapboard.
9. Switch on camera.
10. Switch off digital zoom and longplay, switch on steadyshot, select 16 bit sound.
11. Monitor and check sound.
12. Start recording, and record clapboard for at least 20 seconds while speaking the relevant information into the microphone. Rewind and play through to check the picture and sound carefully. Stop just before the end of the recording, to allow the next shot to overlap slightly and generate continuous time code
13. Focus the camera on the subject, adjust exposure, and, if necessary, the white balance.
14. Zoom to an appropriate image size.
15. Start recording, clap hands in front of the camera, and then wait for a few seconds before beginning the interview or action.
16. Do not stop recording immediately after the interview or action ends; let the camera continue recording for a few seconds before stopping it.
17 Ensure to produce continuous time code (see above for details).
18. Stop recording before the end of the tape.
19. Take the recorded cassette out of the camera, label (if not already labelled), and lock the record-prevention tab.
20. Switch off camera and microphone.

Review of Marantz PMD670: Solid State of the Art?

David Nathan
ELAR, SOAS, London

There are now so many recognised advantages of digital sound recording and storage that debates about sound processing are no longer about digital vs. analogue but about data formats, compression, hardware, and techniques. Interestingly, there were also debates about all of these issues in the analogue era, and history tells us that solutions to them were evolutionary rather than absolute, and typically emerged in the form of preferences or standards rather than hard science. Nevertheless, the advantages of solid state over alternative digital recording equipment are compelling.

Solid state digital recording

"Solid state" means, paradoxically, that data is recorded on electronic circuitry ("flash memory") rather than on physical media such as disks or tapes. The technology is not new—computers have long stored basic start-up data in on-board chips—but is now feasible for consumer large volume storage because the price to volume ratio of flash memory cards has reduced dramatically. A variety of consumer devices are moving rapidly towards data storage on flash cards due to their compact size, reusability, and the fact that no motors or transport mechanisms are needed, which in turn offers potential advantages of robustness, reliability, size, battery life, quietness, and less electrical "noise". In addition, the standard data format of flash memory files provides a seamless path from recording to computer.

Flash cards are available in several physical formats including Compact Flash (CF), SD, Smart Media, and Sony Memory Stick. Following the growth of the digital camera market, the latter types had been gaining ground on CF (the oldest and largest of the formats); however, recently CF is making a comeback because its larger size offers more potential data capacity, and Japanese manufacturers of video cameras—the most data-hungry devices of all—are eager to see flash cards replace disks and tapes. In mid 2004, 512 MB and 1 GB cards, which offer data capacity similar to a CD, became relatively affordable, and now 2 GB cards are starting to appear. An interesting alternative is the IBM microdrive, a tiny hard disk in the physical CF format that can store up to 4 GB, although it has greater power requirements and is probably less robust than cards.

Solid-state digital recording looks like an ideal solution for language documentation where quality recordings are essential but fieldwork can take place under challenging physical conditions. Solid state recording does, however, have some disadvantages, and these may be quite serious for some researchers. Firstly, it is expensive. Flash cards are by far the most expensive recording medium of all, at about 200 times more expensive per megabyte than CDs or DVDs and 50 times more than minidiscs. As an example, although the Marantz PMD670 under review in this article is itself expensive, just a few CF cards will exceed its value. Secondly, because flash cards are expensive, the data recorded onto a card will typically have to be moved onto a cheaper medium to allow the card to be reused. In practice, this means that a computer needs to be available to write the data to CD or DVD. Fortunately, most fieldworkers these days will be carrying a laptop computer, and CDs or DVDs are a good solution for cheap short-term storage and distribution of recordings. CF card data is easily copied to the computer's hard disk via a USB card reader, but a better solution in the case of laptops, is to use a PCMCIA card, which is not only cheaper than a card reader (a CF to PCMCIA card costs around 10 euros), but allows faster data transfer and does not require any cables or objects projecting from the computer. If required, CF data can be written directly to CD without writing the data to the hard disk.

Marantz PMD670

Inputs

Marantz has a reputation for the quality of its recording equipment, and many researchers have long relied on their sturdy cassette recorders. Such researchers will not be surprised by the appearance of the PMD670. Others may be struck by its relatively large size, and its XLR microphone sockets (also known as Cannon connectors). These professional-level microphone connectors are used instead of the miniplug (also known as 3.5mm jack or minijack) sockets found on typical fieldwork recorders such as Sony Professional Walkman, DAT and minidisk recorders; they offer several advantages including reminding fieldworkers that microphone quality is extremely important. On the other hand, many good quality recorders and microphones do use stereo miniplug, and researchers are likely to have useful microphones of this type, so it seems unfortunate that Marantz has not also provided connectors for them.

Controls and display

The PMD670 offers three arrays of controls. On its front panel are simple recording controls, including a large red slide for starting recording. This allows the machine to be slung in a bag or placed on a table top and operated simply. On the top panel, close to the recording controls, are the playback controls; also on the top are a number of buttons providing advanced functions such as composing edit decision lists. None of the buttons have a particularly pleasing or quality feel. Other controls are for selecting recording formats. The PM670 can record in various formats such as linear PCM (normally the

same as ".wav" files) and MP3 at various sampling and bit rates. Unfortunately, there is no support for formats such as the better quality, open source Ogg Vorbis. Providing MP3 recording while not providing miniplug connectors seems oddly inconsistent.

After setting up recording options (the selection of recording format interacts in a confusing way with "Algorithm/File Parameter" selection), I found the PMD670 generally easy to use. The sound quality, as expected, is extremely good. There was a very small amount of background hiss when recording at 16 bit, 44.1 KHz, but this may have been due to microphone compatibility, and was not regarded as a problem. The PMD670 should be useful and flexible in the field—it has a speaker, USB connections, and digital and analogue line-in/out (so it could be used, for example, for digitising analogue sources).

The PMD670's LCD display panel is disappointing. It offers the expected array of information, but is small in relation to the machine, displays some text too small to read easily, and is in monocolour green; in other words, it is little better than the display of a portable consumer minidisc recorder. Battery life was not as long as would be expected, at between 3 to 5 hours, which, considering that 8 batteries are needed, will worry some fieldworkers.

Flash card compatibility warning

However, the greatest disappointment that I had with the PMD670 was one that you are unlikely to experience (I hope). During a field trip, following an elicitation session, I found that one track was corrupted and unreadable (both in the PMD670 and using a computer and card reader). While it turned out that the recording was of a speaker reading from a prepared list, and could be repeated the following day, the effects of this failure were more profound, in that I no longer felt secure using the machine, and subsequently used it only in tandem with a minidisc recorder or else did not use it at all. I have never once experienced a similar loss in years of using various recorders, including minidisc. I subsequently discussed the problem with the manufacturer's agent, who blamed the CF card, a Japanese-manufactured Buffalo 512 Mb card that had been used before and since in other equipment without problem. The agent referred me to the scanty advice about cards on the Marantz Pro website, which neither ruled out the card I had used nor approved of the unbranded card that was originally supplied with the machine. I therefore strongly recommend that you conduct extensive testing of any cards that you buy (and you will buy cards, because the 128MB card supplied with the machine can record only 15 minutes of sound at high quality settings)—and run a backup recorder when recording irreproducible events.

Conclusion

The Marantz PMD670 has provided a welcome introduction to solid state recording that offers excellent recording quality and is generally easy to use. Its appearance will, hopefully, convince fieldworkers that quality digital recordings cannot be made by recording direct to their computer's hard disk, which has none of the PMD670's advantages of dedicated digitising circuitry, a clean electronic environment, portability, robustness, proper recording controls, and ease of operation. On the other hand, the machine has an unimaginative design and feels like a conversion of previous generation machines to solid state, with the addition of some digital functions such as MP3 and track editing which will not be especially useful for most fieldworkers. While the recording quality is excellent and it offers the compelling lure of solid state, this is not yet a state-of-the-art machine.

News in Brief

A World of Many Voices, Frankfurt, Sep. 4–5, 2004

Peter Austin, Jost Gippert***
*SOAS, London
**Universität Frankfurt am Main

In conjunction with the summer school on Language Documentation, the international conference "A world of many voices" was held in Frankfurt with two main themes:

- the impact of language documentation techniques and technologies on linguistic methodologies and theories
- the influence of the speech communities' participation on the methods and goals of language documentation.

The conference was attended by the participants of the summer school (about 80) as well as some 70 interested people, among them a remarkable number of members of communities speaking endangered languages. The keynote speakers were:

- Leanne Hinton and Daryl Baldwin who spoke about language revitalisation with the example of the Miami language
- Felix Ameka who addressed ethical issues and the place of speaker communities in documentation agendas
- Maria Viallalon and José Aquino Medina who reported about current trends, approaches

and actions undertaken in connection with the phenomenon of language shift in Venezuela.

All five continents were represented among the endangered languages covered by the presentations, with the communities' points of view being addressed much more than is usually the case in a linguists' meeting. Approximately half of the papers arose from research under the VW DOBES programme. Papers falling under the first theme included several presentations which discussed the interaction between information technology and linguistic documentation.

The conference concluded with a round table discussion involving representatives of the VW Foundation; among the issues raised were:

- challenges and opportunities in the documentation of endangered languages
- the place of documentation within frameworks of university subjects with a focus on present-day Germany.

The full program of the conference is available at http://titus.fkidg1.uni-frankfurt.de/curric/dobes/conf6cir.htm

Summer School on Language Documentation, Frankfurt, Sep. 1–11, 2004

Nikolaus Himmelmann
Universität Bochum

The ten-day International Summer School "Language Documentation: Methods and Technology" took place in Frankfurt from September 1 to 11, 2004, and was attended by 50 students of linguistics and adjacent disciplines (ethnology, anthropology, African Studies, Asian Studies, etc.). About half of the students were from Germany, the others from as far away as Canada, Nigeria and Australia.

The Summer School offered a lecture series addressing the major conceptual, methodological and practical issues that have to be dealt with when compiling a language documentation. Daily field work tutorials, each attended by no more than six students, provided the opportunity to gain some first hand experience in interacting with native speakers. Topics of more specific interest such as Grammar writing, Gesture in language documentation, Prosody, and Orthography development were addressed in a special series of seminars held in the late afternoons. Finally, a daily tutorial offered the chance to informally discuss practical and theoretical issues in language documentation, or to get acquainted with software packages for documentation work such as the IMDI metadata editor, SIL's Shoebox, or ELAN, a program for producing time-aligned annotation of audio and video recordings.

The presenters came from a wide variety of internationally renowned institutions and research areas. About half of them are currently actively involved in the DoBeS program. It is planned to publish some of their major contributions as a book to be titled Essentials of Language Documentation, and edited by Jost Gippert, Nikolaus Himmelmann and Ulrike Mosel. Other contributions will be published in LAN and/or made available on the DoBeS website.

Overall, the school was judged to be a great success both by the students and the presenters, with the highlight for the students being the seminar series and the daily fieldwork tutorials.

First HRELP Grantee Training Workshop, London, Sep. 2004

David Nathan
ELAR, SOAS, London

HRELP held its first training programme for ELDP grantees during September at SOAS in London. The workshop provided an introduction to skills and issues in language documentation, and gave a chance for us to find out more about grantees and for the grantees to share experiences. Participants came from a variety of countries including Senegal, Tibet, Australia, and Nicaragua. Session topics included making quality sound recordings, lexicography, software tools and techniques, data formats, fieldwork, archiving, ethics, and documentation outcomes. Participants found the programme useful, but provided the valuable feedback that the scope of practical hands-on activities could be expanded, and we are looking to do so when we run the next workshop in June 2005. For more details about future ELDP training workshops, please contact Jacqueline Arrol-Barker (ja30@soas.ac.uk) or Professor Peter Austin (pa2@soas.ac.uk).

Final Review of INTERA

Peter Wittenburg
MPI, Nijmegen

The INTERA project, which among others was intended to further develop the IMDI metadata infrastructure and to extend the IMDI domain, had its final review. The reviewers were impressed by the results achieved and agree that the IMDI infrastructure is now very stable and mature. A number of leading European data centres created IMDI descriptions and are organising their archives according to IMDI. Some of the content of

these archives can be seen using the IMDI BC browser. Other metadata sites will be added to the domain in the next few weeks. In total there are now more than 50 institutions using the IMDI set.

News from the Archive of the Max Planck Institute for Psycholinguistics

Peter Wittenburg
MPI, Nijmegen

ISO TC37/SC4

Another TC37/SC4 meeting took place in Pisa recently. This ISO subcommittee has the task of creating standards in the area of language resource maintenance. A number of projects are underway:

- An open and ISO compliant registry for linguistic data categories (concepts that occur in metadata, morphology, phonology etc.), has been defined and will soon be available to the general public via the web;
- A standard for a flexible lexicon model called LMF (Lexicon Markup Framework) is being discussed;
- A standard for a flexible annotation framework called LAF (Linguistic Annotation Framework) is also being discussed.

The MPI presented its lexicon tool, LEXUS, during that meeting. It implements the LMF model and allows us to work with lexicon structures in a manner similar to Lego blocks, and already has a connection with the data category registry. It was accepted as a test and implementation platform.

Archive exploration framework

Web-accessible language archives are becoming increasingly important. Therefore, it will be very important to create frameworks that allow us to easily ingest new resources, or new versions of existing resources, and to explore the archive content in a flexible way via the web. Currently, the MPI is working on an exploration framework that will include working on annotated media with ELAN-type access methods, on lexica with a web-based LEXUS version, and on texts. IMDI metadata will be used to select a set of resources. A first meeting with some interested users was organised to talk about general concepts and ideas. Further discussions with other potential users will be held in the future to consolidate ideas. A first version will be presented at the beginning of the new year with the goal of opening up a broad discussion within the DOBES programme and at the MPIs in Nijmegen and Leipzig. Co-funding from the Max-Planck-Society was granted to work on a system for collaborative commentaries that will allow the drawing of relationships between different corpus fragments, exploration of these relationships, and the ability to make comments on corpus fragments. In 2005 we hope to be able to work on these projects as well.

40,000 IMDI Sessions

Daan Broeder
MPI, Nijmegen

The integration of IMDI metadata for language resources into a single browsable and searchable domain has reached a new level with almost 40,000 sessions now available.

The metadata was produced within several projects and by different institutes and research groups. The contributors are EU funded projects such as ECHO and INTERA as well as privately funded projects such as DOBES. The metadata covers resources from the endangered languages domain and dialect corpora to resources for multi-modal research and lexica.

The metadata is organised in a browsable tree. At the top level is a classification in terms of the contributing institute or project. In some cases the resource data itself is also available on-line. Although the metadata is hosted at several sites (those of the contributing institutes), it can all be accessed through a portal at the MPI for Psycholinguistics.

You can browse and search the available metadata by downloading the IMDI Browser from: http://www.mpi.nl/IMDI/tools. A manual for the IMDI Browser is also available at this location.

More information about the IMDI metadata set itself can be found at http://www.mpi.nl/IMDI.

ELAN Audio Playback

Albert Russel, Paul Trilsbeek
MPI, Nijmegen

Previous versions of the ELAN multimedia annotation tool had some issues regarding the timing accuracy of audio playback. Researchers from the Radboud University Nijmegen have evaluated the accuracy of the latest Windows release of ELAN (v. 2.3) in comparison to PRAAT (http://www.praat.org), taking PRAAT as the reference. A detailed description of ELAN media rendering issues will appear on the MPI web site (http://www.mpi.nl/tools/) at the end of December 2004; however the main points regarding audio rendering in the Windows version are as follows:

- If there is only MPEG1 media data available, no more than frame accurate audio playback can be expected.
- If there is also separate WAV media data available, the same accuracy as PRAAT is possible.

When a new annotation is made, the user has to define a list of media files that will be annotated. These media files can be one or two MPEG videos and/or a WAV file. The first media file in the list is the so-called Master Media. In order to achieve the PRAAT accuracy, one has to assign the WAV file as the Master Media.

ELAN 2.3 Available

Hennie Brugman, Han Sloetjes, Albert Russel, Alex Klassmann
MPI, Nijmegen

ELAN 2.3 has been available since November 2004. The main new features, improvements and bug fixes are:

- Search functionality is improved and extended: continued search on previous results, export of results, saving and restoring of queries and search across all tiers are now possible.
- Regular expression search on annotation documents in a "basket" of local folders.
- Further improvements in media playback on Macintosh.
- For experimental use it is now possible to annotate 2D regions of a video signal. Since this is implemented on the basis of Quicktime, the price paid is that it is not possible to use native media playback at the same time.
- Shoebox/Toolbox import is substantially improved: users can define and reuse their own marker or tier setups (as alternative for using a Shoebox "typ" file), Toolbox import is supported, including (interlinearized) Unicode encoded fields, users can define their own participant markers and exclude markers from import.
- PRAAT connection: when a wave file is opened in ELAN, PRAAT can be opened from the Signal Viewer, viewing either the whole media file or just the current selection in ELAN.
- Undo is implemented for all edit actions where an annotation, tier, type or controlled vocabulary is modified. Also 'clear selection' can be undone. ELAN currently remembers the last 10 commands.
- Built-in tokenizer: it is now possible to automatically generate new annotations for every token in the annotation of some source tier. Any tier can be used as a source tier, and time subdivision tiers or symbolic subdivision tiers can be used as a target. This can be used, for example, to decompose utterances into individual words. Delimiters for tokens are user definable.
- A number of small additions and improvements.

The most important bug fixes for this release are:

- A number of bugs concerning the manipulation of complex graphs of alignable, or sometimes unaligned, annotations are fixed.
- The Annotation Density Viewer now also shows annotations for long media files correctly.
- Input methods on the Macintosh are now available.

We intend to make localized versions of ELAN available for English, Dutch, German, Swedish, Spanish, and Catalan at the end of 2004.

Announcements

New Publication

Peter Austin
SOAS, London

Peter K. Austin (Ed.), 2004.
Language documentation and description, Vol 2.
SOAS, University of London

Language documentation and description is a new collection of papers dealing with three topics in language documentation: training and capacity building for endangered languages communities, archiving, and multimedia documentation. Most of the papers arose from workshops held at the School of Oriental and African Studies in November 2003 and February 2004. They represent important contributions to the theory and practice of the new field of language documentation by some of the leading scholars in the field, along with contributions from younger researchers. The volume will be of interest to anyone concerned with documenting and describing languages..

Of particular interest to LAN readers will be "Language documentation and archiving, or how

to build a better corpus", by Heidi Johnson; and "Reconceiving metadata: language documentation through thick and thin", by David Nathan and Peter Austin.

The volume costs £10 (postage included) and can be ordered by downloading the order form from http://www.hrelp.org/publications/papers and mailing or faxing it.

Also to be published as a companion to the volume will be a new CD-ROM "The disappearing sounds of the world's languages" – an interactive presentation of a public lecture by Peter Ladefoged at SOAS in February 2004, including soundtrack, transcription, and graphics. The CD costs £10 (postage included). The volume and CD can be purchased together for £17.50.

Vienna School of Audio Preservation, Vienna, July 2005

The Vienna School of Audio Preservation will be held in July 2005. For details, see http://www.pha.oeaw.ac.at/home_e.htm

Readings and Links

The following readings on audio digitisation and archiving were kindly provided by Dr. Dietrich Schüller.

- IASA Technical Committee, Standards, Recommended Practices and Strategies. IASA-TC 03 The safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy, version 2, September 2001. Online at: http://www.iasa-web.org/iasa0013.htm

- IASA-TC 04 Guidelines on the Production and Preservation of Digital Audio Objects, edited by Kevin Bradley. Order form at: http://www.iasa-web.org/tc04/index.htm

- Selection Criteria of Analogue and Digital Audio Contents for Transfer to Data Formats for Preservation Purposes. Online at: http://www.iasa-web.org/taskforce.pdf

LANGUAGE ARCHIVE 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 for the next issue welcomed at:

LAN@mpi.nl

before

February 15, 2005