

The Praxis of Language Documentation in the Paunaka Documentation Project with Special Emphasis on Tools for Linguists

Lena Terhart

2nd Trondheim Workshop in Language Description and
Documentation

12th September 2012

Aspects of language documentation

- Fieldwork
- Data collection and analysis
- Archiving



Aspects of language documentation

- Fieldwork
- Data collection and analysis
- Archiving

- Funding
- Tools
- Products for the community
- ...



Commitments

- Commitments towards the funding agency (collecting, transcribing, translating, interlinear glossing, metadata, archiving)
- Commitments towards the speakers (compensation for their work, production of material in the language)
- Commitments towards yourself (your academic career)



Before going to the field

- You need to choose a language to work on
- You need funding
- You need to borrow or buy equipment
- You should spend an hour on thinking about appropriately labeling your data



Before going to the field

- You need to choose a language to work on
- You need funding
- You need to borrow or buy equipment
- You should spend an hour on thinking about appropriately labeling your data



Funding

- ELDP: <http://www.hrelp.org/grants/apply/>
- National research foundations
- Minor grants: Foundation for Endangered Languages (FEL), The Endangered Language Fund (ELF), Gesellschaft für bedrohte Sprachen (GbS)



Applying for funding

- Information needed for application:
 - Permissions of the speakers and any relevant institutions
 - List of costs: journeys, accommodation, payment of speakers, equipment etc.
 - Importance of documentation of that specific language
 - Institutional support
- Small project before larger project
- Try to get a copy of at least one successful application!
- Austin 2010 gives some useful tips on application writing



Before going to the field

- You need to choose a language to work on
- You need funding
- You need to borrow or buy equipment
- You should spend an hour on thinking about appropriately labeling your data



Equipment

- Minimum: high quality audio recording device and photo camera
- video camera with connection for external microphone + microphone + tripod
- Other equipment: Rechargeable batteries + recharger, cases, flashlight, silica gel, waterproof bags, storage media, etc.



Before going to the field

- You need to choose a language to work on
- You need funding
- You need to borrow or buy equipment
- You should spend an hour on thinking about appropriately labeling your data



Labeling

- Choose names that are unambiguous!
- In the PDP, we use codes that contain: abbreviation for the speaker(s), genre, date and abbreviation of the fieldworker in the session, and possibly another number, if there is more than one recording of the same kind, e.g.
- jxx-n120516l-1 → narrative 1 with speaker j (x is a placeholder) and fieldworker l on the 16th of May 2012



Fieldwork

- Start with word lists
 - Base your analysis on a mixture of elicitation and more natural speech events
 - Double-check your data
- not always possible



Fieldwork with the Paunaka



- Paunaka is a Southern Arawakan language spoken in Concepción, the communities Santa Rita y San Miguelito de la Cruz y Santa Cruz, Bolivia



Fieldwork with the Paunaka

The speakers



How do you choose the speakers to work with?

- We try to work with everyone
 - the ones with very limited language skills
 - the ones who are a bit reluctant
 - the ones who are very interested in the project, but are certainly not the best speakers
 - and, of course, the best speakers!
- Speakers have different skills: story-telling, elicitation, correction, large vocabulary, spontaneous speech



How much to record? What to record?

- We try to record everything!
- You never know when an interesting construction shows up and speakers may not be capable of repeating it when asked for!
- We have to do a lot of editing and cutting of recordings!



Ethics: Compensation

- One of the most delicate questions in fieldwork
- In the PDP, we pay people by hour (10 BOB \approx 8 NOK)
- Speakers can earn some money (subsistence agriculture)
- But:
 - Difficult for the anthropologist who does a lot of participant observation
 - Local students may want to work with the people, too, and cannot afford paying them
 - People may get dependent on your payment



Ethics: Data and informed consent

“Informed consent is a negotiation between researcher and data producer/consultant of all future uses of the material: who will access the data, where will the data be housed, in what form will it be stored, and who will make future decisions over it.” (Dwyer 2006:43)



Ethics: Data and informed consent

- We conducted a workshop with the speakers to present the archive and talk about their rights about their data
- After each recording, we ask the speaker if (s)he liked the recording and if it was okay, if we passed it on to other people to listen to it.



Other kinds of work in the project

- Try to get help from specialists! Look for assistants or interns in other departments!
 - Website
 - Booklets for the speech community
 - Educational material



Tools in Language Documentation

- Cut your recordings
- Transcribe your recordings
- Segment and gloss your recordings



Tools: Cutting tools

- Audacity (freeware)
 - Cutting audios
 - Merging audios
 - Rising the volume
- MPEG Streamclip (freeware)
 - Converting videos, incl. export of audio files
 - Cutting videos
- iSkysoft Video Converter (35 US\$)
 - Merging videos
 - Converting videos (supports more formats than MPEG Streamclip)



Tools: Transcription tools

- Transcriber/ TranscriberAG
 - Very easy to learn
 - No support of Unicode
 - Only one tier per person
- Elan
 - More complex
 - Supports Unicode
 - Unlimited number of tiers
- Praat
 - Phonetic/phonological analysis



Tools: Glossing tools

- Toolbox
- FLEx (FieldWorks Language Explorer)
- TypeCraft



Tools: Glossing tools: TypeCraft

- Advantages
 - Very easy
 - Excellent search function
 - Easy to share data with a group
 - Possibility of publishing your language material and link it with additional information (grammar sketch, project pages, etc.)
- Disadvantages
 - No dictionary
 - Low compatibility with Elan (very difficult export process, loss of information)
 - Given number of tiers (not flexible)
 - Only usable online
 - Slower compared to a well-working Toolbox project



Tools: Glossing tools: FLEx

- Advantages
 - Seems to have a very good parser (balance of automatic and manual segmentation)
- Disadvantages
 - Low compatibility with Elan
 - Given number of tiers (not flexible)
 - Stored data is not easily accessible
 - Several people can share a common project only via a network → high-speed internet connection necessary



Tools: Glossing tools: Toolbox

- Advantages
 - Very flexible
 - Fast
 - Good import/export with Elan
- Disadvantages
 - Very complex configurations needed before it works properly
 - Erroneous automatic parsing not easily replaceable by manual parsing
 - Easy to mess up



Literature

Austin, Peter K. (2010): Applying for a research grant. In: *Language Documentation and Description*. Vol. 7, London: Hans Rausing Endangered Language Project. 285-299

Dwyer, Arienne M. (2006): Ethics and practicalities of cooperative fieldwork and analysis. In :Gippert, Jost, Nikolaus P. Himmelmann & Ulrike Mosel (eds.): *Essentials of Language Documentation*. Berlin: Mouton de Gruyter. 31-66



A wide, reddish-brown dirt road stretches into the distance, flanked by tall green grass. In the background, there are several simple houses with tiled roofs, palm trees, and utility poles under a bright blue sky with scattered white clouds. The scene is peaceful and rural.

¡Chapiebiyu!
Thank you very much!