

## Documentary and descriptive Linguistics

Ulrike Mosel, Kiel University  
10 September 2012

2<sup>nd</sup> Trondheim Workshop on  
Language Description and  
Documentation



## DOBES – Dokumentation Bedrohter Sprachen

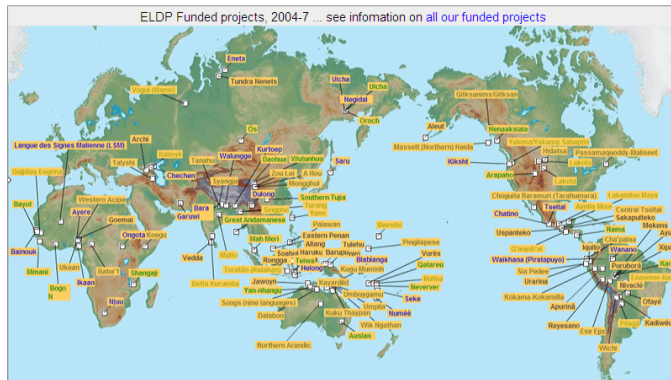


[http://www.mpi.nl/DOBES/volkswagen\\_foundation/](http://www.mpi.nl/DOBES/volkswagen_foundation/)

2

## Endangered Languages Project

*Because every last word is another lost world*



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

3

## The aims of documentary linguistics

Produce and archive documentations of endangered languages

- that provide primary data not only for linguistics, but also for other disciplines of the humanities and social sciences;
- that can be understood without prior knowledge of the documented language;
- that is accepted by the speech community and can be used for language maintenance and revitalisation.

Develop and test new methods of researching, processing and archiving linguistic and cultural data.

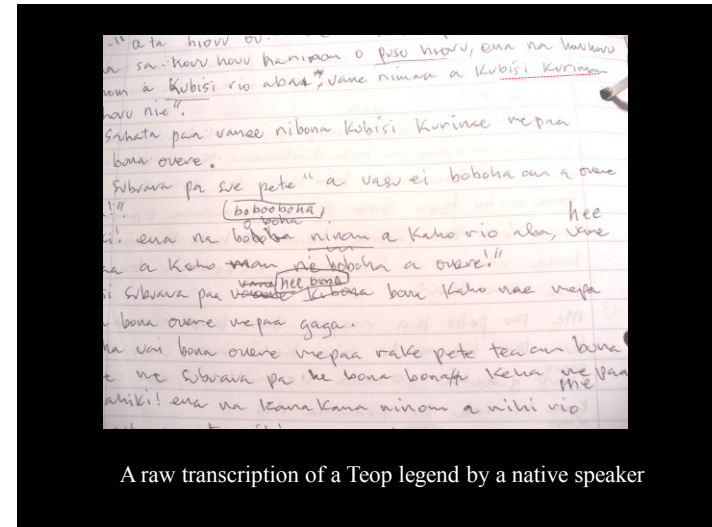
4

## First aim of documentary linguistics: Provide primary data

- audio / video recordings,
- fieldnotes,
- elicited wordlists, paradigms, clauses
- photographs, ....

everything linguists  
like to keep in drawers  
and never publish

5



A raw transcription of a Teop legend by a native speaker

| Nr  | Text  | Subtitles | Lexicon | Audio Recognizer | Metadata | Controls |
|-----|---|-----------|---------|------------------|----------|----------|
| 162 |   |           |         |                  |          |          |
| 163 | E subuava toku vaevuru ??   |           |         |                  |          |          |
| 164 | E Taagoraa paa rosin vooou nao mepaa sue ki bona maamihu                |           |         |                  |          |          |
| 165 | Meori paa nomaa mepaa ma asun vamate bane Kurikaakana.                  |           |         |                  |          |          |
| 166 | Asun vamate vai ori.  |           |         |                  |          |          |
| 167 | meori paa nao vaha nao meori paa gonogono ni maa bono amisina.          |           |         |                  |          |          |
| 168 | amaa karee ae amisina meori paa ma vavaakoto vaha ni bono vahara beiko. |           |         |                  |          |          |
| 169 | vavaakoto ni bono vahara beiko meori paa van tagune wanwan              |           |         |                  |          |          |
| 170 | eve kotokoto vakamis ni bani meori paa van tagune.                      |           |         |                  |          |          |
| 171 | leoni he tea sue bafa. "Hehi Eara koma paa goroho vuru.                 |           |         |                  |          |          |
| 172 | a goroho vuru mea tabae a taemseam, eam he paa -"E-                     |           |         |                  |          |          |

## Second aim of documentary linguistics:

useability without prior knowledge of the documented language

### 1. Introductory information

#### 1.1. What kind of language?

- genetic affiliation,
- socio-linguistic context,
- prominent, typological features.

#### 1.2. What kind of language documentation?

- circumstances of research and documentation
- aims
- size and content

8

### Third aim of documentary linguistics:

produce and archive a language documentation that is accepted by the speech community and can be used for language maintenance and revitalisation.



Teop children reading a preliminary draft of the legends book

### The linguists' and the speech community's documentation

|                  | Linguists   | Speech community                                     |
|------------------|---|--|
| kind of language | spontaneous language; great variety of genres and registers   | planned language; selected important genres          |
| content          | greatest possible variety                                     | suitable for educational purposes                    |
| format media     | audio and video recordings with transcription and translation | written texts<br>>> new genres                       |
| orthography      | based on linguistic principles (phonological)                 | Orthography similar to that of the dominant language |
| lexicon          | electronic lexical database                                   | encyclopedic dictionary on paper                     |

### Components of a language documentation

#### ANNOTATED CORPUS OF RECORDINGS

audio/video recordings ♦  
**transcriptions** ♦  
**translations** ♦  
**glossing** ♦  
 comments on form & content ♦  
 (keywords)

#### LEXICAL DATABASE

♦ head word  
 ♦ **part of speech**  
 ♦ **definition**  
 ♦ collocations, idioms  
 ♦ examples with references

#### SKETCH GRAMMAR

**phonology, orthography** ♦  
**parts of speech** ♦  
**grammatical categories** ♦  
 examples with references ♦  
 typological profile ♦

#### APPARATUS

♦ **index of gram. terms**  
 ♦ abbreviations  
 ♦ keywords

### Cooperative fieldwork

- > active participation in the language documentation project
- training, capacity building, empowerment
- respect the community's expectations

with respect to

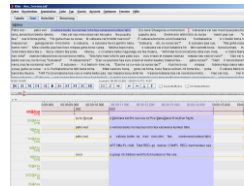
- the format the selection of genres
- the content
- the format and media

**Teop Legends**

ELAN versions  
 PDF versions

community edition  
 on paper without  
 translation

recording of reading  
 + ELAN version  
 + PDF version



Format and amount of grammatical information should not impede the user's pleasure of reading.



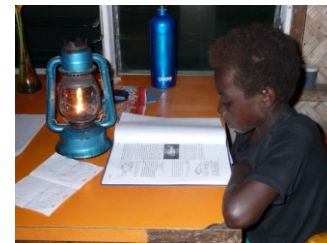
**Workflow**

1. recording of data
2. transcription
3. translation



15

4. proofreading
5. illustrations
6. printing



16

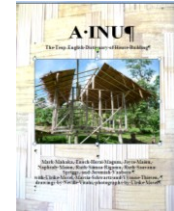


## 7. archiving

- of annotated sound/video files and written texts in ELAN with metadata

### Metadata:

- names of language(s), dialect(s)
- date of recording
- technical details of recording
- names of speakers
  - & gender, age, place(s) of residence
  - & education
- topics
- name of recorders, translators, illustrators ...
- ...



- of printed materials as PDFs

18

## Building a corpus of texts

### ANNOTATED CORPUS OF RECORDINGS

- audio/video recordings ◆
- transcriptions ◆
- translations ◆
- glossing ◆
- comments on form & content ◆
- (keywords)

**Archiving**  
Metadatasdescription

19

## Metadata descriptions of recording sessions

1. Provide information on the language or languages and the context of the recording

- Who did the recording?
- Who was/were the speaker/speakers?
- Who else was present?
- Where?
- When?
- Discourse type - Monologue, interview, casual conversation?
- Topic

20

**Annotations**

audio or video recordings plus

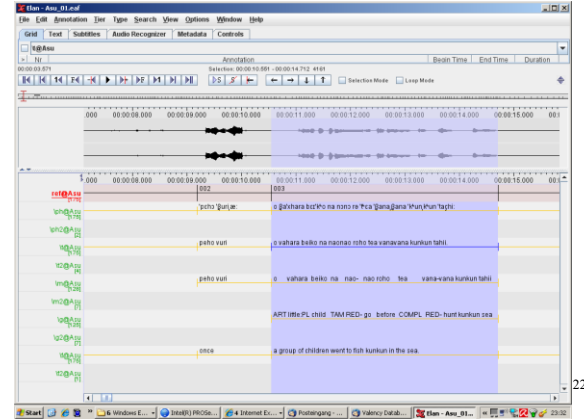
- (a) transcriptions,
- (b) translations,
- (c) comments on content and linguistic form

**(a) transcriptions:**

- a small subcorpus in IPA
- most of the corpus in a practical orthography accepted by the speakers

21

- 2. Facilitate the accessibility of the corpus in an electronic archive from all over the world via the internet.



22

**(b) Translations**

Which target language? OR languages?

- 1) a major European language for international academic purposes
- 2) the national language
- 3) the regional lingua franca

what kind of translation?

- 1) free translation
- 2) literal translation
- 3) interlinear morphemic translation

23

**Grammatical information in glossings**

[TEOP] **Mutanae komanae hata komana vai.**  
 [free transl.] 'Her husband became suspicious.'

[TEOP] **muta- na- e koma-na- e**  
 husband-POSS-3SG inside-POSS-3SG

**hata komana vai**  
 bad EMPH now

POSS = possessive marker  
 EMPH = emphatic particle

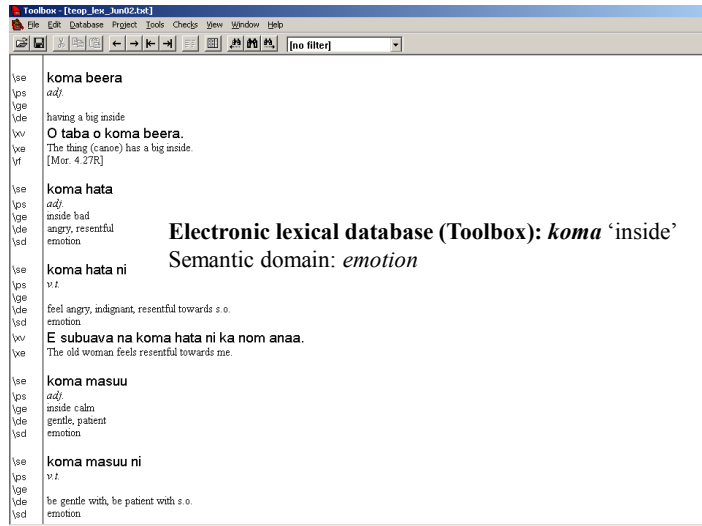
[lit. transl.] 'husband of **her**, inside of **his** very bad now.'

No gender?

> sketch grammar

What else goes into the sketch grammar?

24



### Keep norms of form and content of language documentations at a minimum:

- diverse languages and cultures
- diverse linguistic and philological traditions
- sustainable theories and methods in the new field of documentary linguistics can only evolve in a non-restrictive atmosphere.
- diverse resources in terms of time, money, man power
- diverse expectations of the speech communities and capacities

27

### Norms and standards in language documentation

- Rules of conduct that respect the intellectual property rights and the cultural norms of indigenous people
- Long-term preservation of the documentations in digital archives requires continuous updating to new storage formats
- Generally accepted linguistic conventions and shared practices contribute to userfriendliness

BUT

there is no "one-size-fits-all" language documentation format,

### Choose a form of presentation

that guarantees sustainability and is suitable for the various kinds of prospective users

have all data in electronic form  
store them in an archive  
that guarantees transfer to future formats

But remember:  
Not all prospective users will be linguists.

Have most data in a userfriendly format

Not all prospective users have electricity and computers.  
Have print outs!

28

## Impact of documentary linguistics on general linguistics

Documentary linguistics will revolutionise linguistics  
because language documentations

(Whalen 2004)

- provide natural data in their context
  - (linguistic theories heavily rely on grammars that describe  
invented isolated clauses)
- allow others to scrutinize the researcher's analysis
  - (in most grammars and other linguistic publications  
the sources of examples are not identified and made accessible to others)
- provide data for future research on questions  
that have not been asked at the time of documenting the language

## The global perspective

International cooperation with respect the development of  
technical standards  
rules of conduct  
legal issues (intellectual property rights)

International cooperation in language documentation projects  
(share experiences, hire specialists on short term contracts)

Close cooperation with the speech communities  
Capacity building

Summerschools, training courses,  
workshops, conferences

30