Documentary and descriptive Linguistics
Ulrike Mosel, Kiel University
10 September 2012

2nd Trondheim Workshop on Language Description and Documentation

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.
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

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
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

Components of a language documentation

The linguists' and the speech community's documentation

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

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

4. proofreading
5. illustrations
6. printing
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 material as PDFs

Building a corpus of texts

ANNOTATED CORPUS OF RECORDINGS
- audio/video recordings
- transcriptions
- translations
- glossing
- comments on form & content
(keywords)

Archiving
Metadata description

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
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

(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

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

[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

No gender?

What else goes into the sketch grammar?

> sketch grammar

POSS = possessive marker
EMPH = emphatic particle
Electronic lexical database (Toolbox): koma ‘inside’

Semantic domain: emotion

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.

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

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!
Impact of documentary linguistics on general linguistics

Documentary linguistics will revolutionise linguistics because language documentations

- 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

(Whalen 2004)

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