2

4

1

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

2nd Trondheim Workshop on Language Description and Documentation







DOBES – Dokumentation Bedrohter Sprachen



http://www.mpi.nl/DOBES/volkswagen foundation/



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.

http://www.hrelp.org/

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

e Ec	dit <u>A</u> nno	tation ∏ier	Type <u>S</u> ea	rch ⊻	iew Option	s <u>W</u> indow	Help						
rid	Text	Subtitles	Lexicon	Audio	Recognizer	Metadata	Control	s					
•	t@Asu												
Nr						Annota	tion				Begin Time	End Time	Dura
16												00:05:09.711	
		iava toku vae										00:05:11.878	
					a sue ki bona ate bene Kur							00:05:15.715 00:05:19.419	
		amate vai or		unvan	ale perie Kul	IKakamu a.						00:05:21.362	
				000 000	ogono ri ma:	a hana amisi	no					00:05:24.424	
					ravaakoto va							00:05:28.730	
169 vavaakoto ri bono vahara belko meori paa van tagune warwan							00:05:33.637						
170 eve kotokoto vakamis ri bari meori paa van tagune,							00:05:36.421						
17	171 eori he tea sue bata, "Hehl Eara koma paa goroho vuru,							00:05:36.421	00:05:39.325	00:00:0			
172 a goroho vuru mea tabae a taemeam, eam he paa -""E-													
5:28.	.730	F∢ - (• } +) F		N N	Selection: 00	₩ ←		07 1 Selectio		00:05:39:325 Mode 4 0	00:05:43.972	00:00:0
5:28.	.730	F∢ - (• } +) F	M	N N	Selection: 00	* *	- → ↓	1 Selectio		Mode 40		33.000
6:28.	14 14	.000 .000 	• • • • • • • • • • • • • • • • • • •			Selection: 00	K→ ←	- → ↓	↑ Selectio 00:05:30.000 • • • • • • 00:05:30.000		Mode 40	00 00:05:	- E-
5:28.	.730	.000 .000 	00:05:26.0		00:05:27.000	Selection: 00	K→ ←	- → ↓ 00:05:29.000 00:05:29.000 Asu_01R:16	1 Selectio	00:05:31.00	Mode 40	00 00:05	33.000
6:28.	ref@	E4	00:05:26.00		00:05:27.000	Selection: 00	8.000	- → ↓ 00:05:29.000 00:05:29.000 Asu_01R:16	↑ Selectio 00:05:30.000 • • • • • • 00:05:30.000	00:05:31.00	Mode 40	00 00:05	33.000
6:28.	ref@ \\ph@	F4	00:05:26.00	▶1 00 00	N N 00:05:27.000	Selection: 00 S S S 00:05:2 00:05:2 no 'xhara 'bc	★ ← ← (8.000 (0) (0) (0) (0) (0)	-) →) ↓ 0):05:29.000 0):05:29.000 Asu_01R:16 g/gaxh,khotho	1 Selectio	00:05:31.00 00:05:31.00 00:05:31.00	Mode 40	00 00:05: 00 00:05: 00 00:05: an'bata	33.000
6:28.	ref@ \\ph@	F4 -14 000		DO DO DO Hatajiki ia ma v		Selection: 00 SS S SS S 00:05:2 00:05:2 no 'xhara 'bc: na ri bono va	► ← ₩8.000 ₩8.000 ₩8.000 ₩8.000 ₩8.000 ₩8.000 ₩8.000	- → ↓ 0::05:29.000 0::05:29.000 Asu_01R:16 g/@axh,Motho vavaakoto ri	COUDE:30.000 OUDE:30.000 OUDE:30.	00:05:31.00 00:05:31.00 00:05:31.00 ск+о: 'ттул рэа @	Mode 40 0 00:05:32.0 0 00:05:32.0 an *ta guni 'gang tagune wanwan	000 00:05 000 00:05 an 'bat a	33.000
6:28.	ref@ \\ph@ \\ph@	F4 -14 000		DO DO DO Hatajiki ia ma v	00:05:27.000 00:05:27.000 00:05:27.000 00:05:27.000	Selection: 00 SS S SS S 00:05:2 00:05:2 no 'xhara 'bc: na ri bono va	★ ← ★ 8.000 ★ 8.00	- → ↓ 0):05:29.000 Asu_01R.16 g/gaxh,⊮orho vavaakoto ri made them I	Setectio O0:05:30.000 O0:05:30.000 O0:05:30.000 O0:05:30.000 I' ibano 'xhara 'bell boono vahara bell boote all the children	00:05:31.00 00:05:31.00 00:05:31.00 ckto: 'myri 'p'a @ ko meori paa var	Mode 40 0 00:05:32.0 0 00:05:32.0 an 't*a guni 'gang at tagune wanwan oke up one atter	000 00.05; 000 00:05; an 'ba†a the other, sayin	33.000
5:28.	ref@ \\ph@ \\ph@ \\m@ \\m@	F - H 0000 1000 1000 Asu Asu ae ami 1751 ae ami lack an 1759 1000 1751 1000 1751 1000 100 1000 1		00 00 1 +ałąrkt na ma v ame (ar		Selection: 00 SSECTION: 00 SSS 00:05:2 00:05	► ← ₩8.000 ₩8.000 ₩8.000 ₩8.000 ₩8.000 ₩8.000 ₩8.000	- → ↓ 0):05:29.000 Asu_01R.16 g/gaxh,⊮orho vavaakoto ri made them I	Setectio O0:05:30.000 O0:05:30.000 O0:05:30.000 O0:05:30.000 I' ibano 'xhara 'bell boono vahara bell boote all the children	00:05:31.00 00:05:31.00 00:05:31.00 ckto: 'myri 'p'a @ ko meori paa var	Mode 40 0 00:05:32.0 0 00:05:32.0 an *ta guni 'ĝanĝ tagune wanwan	000 00.05; 000 00:05; an 'ba†a the other, sayin	33.000

a sa how how hangeon o puse heard, eur na haven ion a Kubisi rio abar Juane nimaa a Kubisi Kuring nove nie". Schata pen vance nibon Kubisi Kurince me pra bona overe . Evbrana pa sue pete " a vasurei boboha cun a overe (bobooboha) hee i ena ne boleta nivan a Kaho vio ala, vine in a kano man sie bobohn a overe!" i subvara par vara hel band ban Kaho nae nepa boun over e vepa gaga. na voi bona oriere mepaa rake pete tea on bona e ne subvara pa ne bona bonafi kena melaa ahiki! ena na Izana kana ninom a nihi vio!

A raw transcription of a Teop legend by a native speaker

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

Components of a language documentation



typological profile ♦

- definition
- ♦ collocations, idioms
- ♦ examples with references

APPARATUS ♦ index of gram. terms

♦ abbreviations

♦ keywords

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 >> new genres
orthography	based on linguistic principles (phonological)	Orthography similar to that of the dominant language
lexicon	electronic lexical database	encyclopedic dictionary on paper

Cooperative fieldwork

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

with respect to

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

LEXICAL DATABASE ♦ head word ♦ part of speech

Teop Legends

ELAN versions PDF versions

community edition on paper without translation

recording of reading + ELAN version + PDF version



	10,7,5,10	and the second s					
	Lub Arrist Ubabil	Naporf Nof					
	Learning and meaningly. Bench lines: Second in Color Hand with Kinese Kapana, as meaning by Color March.						
E Las Teninimter	photos.	Der vorm dit version					
147 shares		har only one."					
Intel 7		Annual, the Lorensel 1 Sectors on the Lorensel related on					
tai mai 7		manify been marined #					
And Broundship		And so, the other space."					
144 million an ap		and the people had all left to adduce for - and					
La1 marines to	an an an bears had	the people had left - last to address if					
		while the still women had percent affective in a work.*					
Lat Tiper ad		Dec 242					
Lat repairing		and standard how and particle interface as the days					
114 Summers)		Put (c) into the sec.7					
iid repainent		and (char) alternate sets.)					
124 hebeneitie		she stand the (for-and) 2					
parata.	and particular product	And high same, and the most indice, shap along a					
	an in she what with her	hu) he might set of millions app					
the second		there wait (
if a second of the		that the compression					
memory at 7 %		"Do and not as any could be dealed at range, bits represent to the back	-				
114 American St.		Then the mar were, 2					
the parenties.		brought the line word of the old - array."	-				
114 million has		ship hould came canalog?					
11.4 mentioners	bene served.	and they for all all					
114 maplified		and marked threads as 2					
And Plant Pages	The state only a set						

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



 4. proofreading
 5. illustrations 6. printing





Workflow

1. recording of data

2. transcription 3. translation







7. archiving

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

Metadata:





- of printed materialis as PDFs

Building a corpus of texts



- translations
 - glossing ♦
- comments on form & content (keywods)

Archiving Metadatadescription

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

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



(b) Translations

Which target language? OR languages?

 a major European language for international academic purposes
 the national language

3) the regional lingua franca

what kind of translation?

1) free translation

2) literal translation

3) interlinear morphemic translation

23

21

[TEOP] Mutanae komanae hata komana vai. [free transl.] 'Her husband became suspicious.' [TEOP] muta- na- e koma-nae husband-POSS-38G inside-POSS-38G hata komana vai POSS = possessive marker EMPH = emphatic particle bad EMPH now [lit. transl.] 'husband of her, inside of his very bad now.' No gender? > sketch grammar What else goes into the sketch grammar?

Grammatical information in glossings

24



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

27

- 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 propesctive 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 decsribe 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