



# **ENDANGERED LANGUAGES**

**PROF. NICOLE NAU, UAM 2017**

Seventh & eighth  
lectures  
(03+10/04/2017)

# LANGUAGE DOCUMENTATION (1)

What is (a) language documentation?

Why are languages documented?

What does a modern language documentation contain?

What problems may arise when collecting language records?



# WHAT IS LANGUAGE DOCUMENTATION?

«linguists generally view language documentation [...] as **work** undertaken by an amateur or professional linguist that **results** in some **record** of a language for the **purpose** of furthering the study and understanding of that language.» (Berge 2010)

«language documentation is the **creation**, **annotation**, **preservation** and **dissemination** of transparent records of a language» (Woodbury 2011)

# WHY DOCUMENT A LANGUAGE? FOR WHOM?

To have a record for the future (and the present) -

- ❖ for **speakers** of the language: to preserve the language as part of the culture, the community's heritage
- ❖ for **researchers**: to have material for their investigations
  - general linguistics, language typology, historical linguistics,
  - psychology, ethnography, ...
- ❖ for **educators**: to have material for teaching the language and about the language

...

# «PREHISTORY» OF LANGUAGE DOCUMENTATION

16th – 18th/19th century, for example

- ❖ European travelers and scholars collect the Lord's Prayer in various languages
- ❖ Spanish missionaries in Latin America write grammars and dictionaries of indigenous languages
- ❖ German and Russian scholars describe languages of Russia / Siberia

Why?

# HISTORY: LANGUAGE DOCUMENTATION IN THE LATER 19TH AND THE 20TH CENTURY

**Franz Boas** (1858-1942)

pioneer of modern  
anthropology

founded a standard for  
documenting languages:

- text collection +
- dictionary +
- grammar



# HISTORY: THE TRADITIONAL LANGUAGE DOCUMENTATION (19TH-20TH CENTURY)

- ❖ done by linguists (often only 1 person)
- ❖ done by outsiders (not part of the community)
- ❖ published on paper (usually in small editions)
- ❖ result inaccessible to the speech community
- ❖ includes a dictionary, a grammar, and a collection of texts
- ❖ text collection: most often narrative texts (traditional stories)

# CURRENT IDEAS ABOUT LANGUAGE DOCUMENTATION

«a language documentation should strive to include as many and as varied records as practically feasible [...] **Ideally**, then, a language documentation would cover **all registers and varieties**» (Himmelman 2006)

«our field of enquiring is limited by our understanding: who would have imagined the importance of documenting mundane conversational speech a century ago?» (Berge 2010)



# QUESTIONS

If your language were endangered, what would YOU want to document for the future?

Given the limits of time, money, and persons: what should be documented first?

# THE CURRENT LANGUAGE DOCUMENTATION (BEGINNING OF 21ST CENTURY)

- ❖ done in interdisciplinary teams: linguists + other specialists,
- ❖ outsiders and insiders: speech community is involved not only as „informants”, but as experts
- ❖ done for various purposes
- ❖ published electronically (and on paper)
- ❖ stored in digital archives
- ❖ contains written, audio and video material
- ❖ documents the use of the language in various situations

# SOME THOUGHTS ABOUT ADEQUACY: WHAT IS A «GOOD» LANGUAGE DOCUMENTATION? (BERGE 2010)

«Adequacy in documentation must address the **needs of non-linguists**, particularly the needs of the users of the language being documented, as well as the **needs of linguists**.»

«Adequacy in documentation by linguists should be **based on current advances in linguistic understanding and technology**, it should include adequate and timely efforts to preserve data and make them available to the greater world, and it should include **creative approaches** to the study of linguistics.»

«One measure of adequacy in documentation and description might be how **learnable** the language is as a result, since acquiring fluency in a language requires enough data with enough descriptions to reproduce the language outside its normal context.»

# PRIMARY DATA: 1. WHAT CAN BE OBSERVED

«a language documentation, conceived of as a lasting, multipurpose record of a language, should contain a large set of primary data which provide **evidence for the language(s) used** at a given time in a given community [...]. Of major importance in this regard are specimens of **observable linguistic behavior**, i.e. examples of how the people actually communicate with each other.» (Himmelman 2006)

# PRIMARY DATA: 2. METALINGUISTIC KNOWLEDGE

*manifest in the native speakers' ability to provide interpretation and systematizations for linguistic units and events*  
(Himmelman 1998)

Examples:

- know how to form the plural of a given noun
- know that you use diminutives when speaking to children
- know that the people in the next village speak another dialect, and how it differs from yours

How can metalinguistic knowledge be accessed and documented?

# **METADATA: DATA ABOUT (PRIMARY) DATA**

**Metadata** for spoken data usually contain

- participants (who speaks, who listens)
- time and location
- recorded by
- recorded with (tool, technology)
- ...

**What are metadata for?**

# METADATA STANDARDS

## **OLAC Metadata Standard** – explanations:

<http://www.language-archives.org/NOTE/usage.html>

(Standard, more formal: <http://www.language-archives.org/OLAC/metadata.html>)

Example: <http://www.language-archives.org/language/ain>

(or simply google: OLAC Ainu)

### **see also:**

«A gentle introduction to metadata» by Jeff Good (2002):

<http://linguistics.berkeley.edu/~jcgood/bifocal/GentleMetadata.html>

# ANNOTATION: ANOTHER KIND OF METADATA

**Annotation** of primary data may contain

- transcription / transliteration
- translation
- grammatical annotation («glossing», «tagging»)
- comments
- ...

Tools for annotating speech data: ELAN, Toolbox, FLex...

Table 2. Extended format for a language documentation

Himmelmann (2006)

Primary data	Apparatus	
recordings/records of observable linguistic behavior and metalinguistic knowledge	Per session	For documentation as a whole
	<p><i>Metadata</i></p> <p><i>Annotations</i></p> <ul style="list-style-type: none"> <li>– transcription</li> <li>– translation</li> <li>– further linguistic and ethnographic glossing and commentary</li> </ul>	<p><i>Metadata</i></p> <p><i>General access resources</i></p> <ul style="list-style-type: none"> <li>– introduction</li> <li>– orthographical conventions</li> <li>– glossing conventions</li> <li>– indices</li> <li>– links to other resources</li> <li>...</li> </ul>
		<p><i>Descriptive analysis</i></p> <ul style="list-style-type: none"> <li>– ethnography</li> <li>– descriptive grammar</li> <li>– dictionary</li> </ul>

# ARCHIVES CONTAINING DOCUMENTATIONS OF ENDANGERED LANGUAGES

<http://dobes.mpi.nl/> (DOBES = **D**okumentation **b**edrohter **S**prachen)

<https://elar.soas.ac.uk/> (ELAR = **E**ndangered **L**anguages **A**rchive)

<http://www.ailla.utexas.org/site/welcome.html> (**A**ILLA is a digital archive of recordings and texts in and about the indigenous languages of Latin America)

<http://catalog.paradisec.org.au/> (**P**ARADISEC = **P**acific and **R**egional **A**rchive for **D**igital **S**ources in **E**ndangered **C**ultures)

[http://lacito.vjf.cnrs.fr/pangloss/index\\_en.htm](http://lacito.vjf.cnrs.fr/pangloss/index_en.htm) (**P**ANGLOSS collection)

# MAKING RECORDINGS FOR LANGUAGE COMPARISON: SOME TECHNIQUES

Gathering basic vocabulary with a Swadesh list.

[http://en.wiktionary.org/wiki/Appendix:Swadesh\\_lists](http://en.wiktionary.org/wiki/Appendix:Swadesh_lists)

Recording a recitation (or reading) of the Lord's Prayer.

Recording a reading of The North Wind and the

Sun. [http://clas.mq.edu.au/oz\\_transcription/NorthWind.html](http://clas.mq.edu.au/oz_transcription/NorthWind.html)

Recording a retelling of the film „The Pear Story”.

<http://www.pearstories.org/docu/ThePearStories.htm>

# REFERENCES

Berge, Anna. 2010. Adequacy in documentation. In: Furbee-Losee, Louanna & Grenoble, Lenore A. eds. 2010. *Language documentation: practice and values*, 51-66. Amsterdam: Benjamins.

Himmelman, Nikolaus. 2006. Language documentation: What is it and what is it good for? In: Gippert, Jost & Himmelman, Nikolaus & Mosel, Ulrike, eds. 2006. *Essentials of language documentation*, 1-30. Berlin, New York: Mouton de Gruyter.

Woodbury, Anthony C. 2011. Language documentation. In: Austin, Peter K. & Sallabank, Julia, eds. 2011. *The Cambridge Handbook of Endangered Languages*, 157-186. Cambridge: Cambridge University Press.