



# DOCUMENTARY LINGUISTICS I

prof. Nicole Nau, UAM winter 2017/2018

**Second lecture**  
10 October 2016

# TOPICS OF THE DAY

- ❖ What is [a] language documentation (LD) and what is it good for?
- ❖ Principles of documenting languages today
- ❖ Contents of a LD: primary data and metadata
- ❖ Discussion of some details
- ❖ Examples of language archives (today only the addresses, presentation next week)

# DOCUMENTARY LINGUISTICS AND LANGUAGE DOCUMENTATION

«**Documentary Linguistics** [..] is concerned with the making and keeping of **records** of the world's **languages** and their **patterns of use**.» (Woodbury 2003)

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

# 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)

«documentation» in Polish: (a) dokumentowanie, (b) dokumentacja

Why and by whom is this work done?

Who may need the result?

# 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

...

# WHAT IS **A** LANGUAGE DOCUMENTATION?

«a comprehensive corpus of primary data which leaves nothing to be desired by later generations wanting to explore whatever aspect of the language they are interested in» (Himmelman 2006: 3)

«Adequacy in documentation must include all the basic phonology, morphology, syntactic constructions with context, lexicon, a full range of textual genres, registers, and dialects, and data from diverse situations and speakers, with enough tokens and a large enough database to be useful in both present and future research.»  
(Berge 2010: 57)

Of course, this is an **ideal!**

# PRINCIPLES OF LANGUAGE DOCUMENTATION (BEGINNING OF 21ST CENTURY)

- ❖ done in interdisciplinary teams: linguists + other specialists,
- ❖ outsiders and insiders: speech community is involved not only as „informants”
- ❖ 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

# LANGUAGE DOCUMENTATION EARLIER (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)
  - ❖ grammar: contains isolated sentences, 20th c often elicited/constructed
- => what was wrong with that?



# 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?

# CONTENTS OF A LANGUAGE DOCUMENTATION

## BASIC DISTINCTIONS

- ❖ Primary data
- ❖ Metadata
- ❖ Annotation

- 
- ❖ Description (results of analysis: grammar, dictionary)

# DATA, ANNOTATION AND METADATA IN AN EARLY RECORD (THE LORD'S PRAYER IN LAVIAN IN MÜNSTER'S *COSMOGRAPHIA* 1588)

## Von dem Teutschen Landt.

## Reue

Ein andere Statt in Lyfflandt mit nammen Derpst/hat ire Rauffmanschafft mit den Die Statt Derpst. Reuel. Reussen. Reuel aber ist weiter gegen Mitnacht zu dem Ostsee gelegen/gebraucht sich nicht weniger der Rauffmanschafft als Riga/ist auch schier derselb brauch da. Sie haben Lübeckische Recht/welcher sich die andern Stettlein auch gebrauchen. Es sind viel Sprach in Lyfflandt/deren sich keine mit der andern vergleicht. Zu Reuel vnd zu Derpst reden sie ein Sprach/heißt Estensch/vñ vmb Riga braucht man Lnuisch/da ist die rechte Lyffländische Sprach/vnd betten daselbst die Bawren/so Gottes Wort berichtet haben/auff folgende weiß ihr Vatter vnser:

Vatter vnser der du bist in Himmeln/ geheiliget werd dein Name/ zutomm vns dein  
Läbes mus kas tu es eckshan debbessis/ Schwetis tows waark/ enack mums tows  
Reich dein will geschehe wie in Himmeln also auff Erden. Vnser täglich  
walsstibe/ tows praats bus ka eckshan Debbes/ta wursan semmes. Musse denische  
Brot gib vns heut vergib vns vnser schuld als wir vergeben vnsern  
Mäyße düth mums schodeen/ pammate mums musse gräke kames pammate musse  
Schuldigen/ nicht eynfähr vns bösen versuchung sonder behüt vns vor allem bösen.  
parradueken/ Ne wedde mums loune badeckle/ pett passarga mums ni wüsse loune.

Dis alles oder zum größern theil/hab ich von dem erfarnen Johann Hasentödter/ so manich jar in Lyfflandt an den Herren Höfen vñ Cansleyen gewesen/vñ viel erfahre hat.

# THE LORD'S PRAYER IN A SORBIAN DIALECT, MODERN RECORDING WITH ANNOTATION (IN ELAN)

The screenshot displays the ELAN software interface for audio annotation. At the top, there is a playback control bar with various icons for navigation and volume. Below this is a timeline showing the audio waveform for two channels. The main area of the interface is divided into several horizontal tracks, each representing a different layer of annotation. The tracks are labeled on the left side with their respective languages and identifiers:

- topics. [1]**: A single line of text: "Lord's Prayer in the dialect of Schleife; German translation: current ecumenical version, Arbeitsg".
- spch.d [13]**: A line of text in the Sorbian dialect: "Wóšce naš, kiž sy swěćone bydź pćidź k nam Twój Twója wóla se stanjo kaž na nj".
- ft.deu [13]**: A line of text in German: "Vater unser im Him geheiligt werde Dein Reich komm Dein Wille geschehe, wie im Hi".
- ft.eng [13]**: A line of text in English: "Our Father in heav hallowed be yo your kingdom co your will be done, on earth as i".
- nt.deu [3]**: A line of text in German: "00:00:03# der du bi".

Each line of text is segmented by vertical bars, indicating the time intervals for each word or phrase. The time axis at the top of the annotation area ranges from 00:00:00.000 to 00:00:16.000.

# PRIMARY DATA: WHAT IS / BELONGS TO [A] LANGUAGE?

Language is...

- an abstract system of signs
- a cognitive capacity
- a means of communication
- ...

Language **is** or language **happens**?

Language exists/happens in the **individual** or in the **community**?

# 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 (in all of the different senses of “language”). 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)

# OBSERVABLE LINGUISTIC BEHAVIOR

What is a communicative activity? Give examples!

What constitutes a communicative activity / a speech event?

What part of an activity / event is **linguistic** behavior?

What part(s) should be contained in a documentation?

+ one event = one language? No!

# PRIMARY DATA: 2. METALINGUISTIC KNOWLEDGE

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

Give examples that show your metalinguistic knowledge!

How can metalinguistic knowledge be accessed and documented?



# EXAMPLES FOR METALINGUISTIC KNOWLEDGE

- ❖ I know that a Christian prayer ends with the word «Amen».
- ❖ I know that «pyra» is used in the Poznan dialect of Polish.
- ❖ I know where the stress is on Polish words.
- ❖ I know that «I know that ...» is a possible and normal way to start a sentence in English, but «I know if ...» is not.
- ❖ I know that Polish words ending in «(i)arnia» designate places such as shops and cafes.

# METALINGUISTIC KNOWLEDGE VS. ENCYCLOPEDIC AND CULTURAL KNOWLEDGE

**Piwiarnia** – lokal gastronomiczny przeznaczony głównie do spożywania piwa.

Oprócz funkcji gastronomicznej piwiarnie pełnią funkcję społeczną, integrując bywalców. Lokale te służą swobodnym pogawędkom i zabawie. Często w piwiarniach istnieje możliwość gry w karty, kości, darty, czy piłkarzyki. W piwiarniach często można usłyszeć lekką muzykę (piosenka biesiadna).

Do wielu gatunków piw, nalewanych z beczek, jak również podawanych w butelkach, serwuje się zwykle ciepłe i zimne przekąski.

(<https://pl.wikipedia.org/wiki/Piwiarnia>)

# METADATA AND ADDITIONAL INFORMATION

**Annotation** of primary data may contain

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

Tools for annotating speech data: [ELAN](#), AnnotationPro, Toolbox, FLex...

# METADATA FOR SESSIONS AND RECORDINGS

**Metadata** for spoken data usually contain information on:

- participants (speakers and bystanders, their roles)
- time and location
- recorded by
- recorded with
- ...

What are metadata for?

Standards for metadata: OLAC, IMDI (more next week)

Primary data	Apparatus	
recordings/records of observable linguistic behavior and metalinguistic knowledge (possible basic formats: session and lexical database)	Per session	For documentation as a whole
	<p><i>Metadata</i></p> <ul style="list-style-type: none"> <li>– time and location of recording</li> <li>– participants</li> <li>– recording team</li> <li>– recording equipment</li> <li>– content descriptors</li> <li>...</li> </ul> <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> <ul style="list-style-type: none"> <li>– location of documented community</li> <li>– project team(s) contributing to documentation</li> <li>– participants in documentation</li> <li>– acknowledgements</li> <li>...</li> </ul> <p><i>General access resources</i></p> <ul style="list-style-type: none"> <li>– introduction</li> <li>– orthographical conventions</li> <li>– ethnographic sketch</li> <li>– sketch grammar</li> <li>– glossing conventions</li> <li>– indices</li> <li>– links to other resources</li> <li>...</li> </ul>

# HIMMELMANN'S «APPARATUS»

(HIMMELMANN 2006)

EXAMPLE: LOWER  
SORBIAN IN THE  
DOBES ARCHIVE

[HTTP://DOBES.MPI.NL/](http://DOBES.MPI.NL/)

Table 2. Extended format for a language documentation

Himmelman (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>

# 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.

# EXAMPLES OF PROFESSIONAL LANGUAGE DOCUMENTATIONS IN ARCHIVES/DEPOSITES

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


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

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

<http://catalog.paradisec.org.au/> (**PARADISEC** = Pacific And Regional Archive for Digital Sources in Endangered Cultures)

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



- 
- Homework for next 2 weeks:
    - [Johnson](#), Heidi. 2004. Language documentation and archiving
    - [Mosel](#), Ulrike. 2006. Fieldwork and community language work (in *Essentials of Language Documentation*)
    - + get familiar with language archives
  - Your first task for course credits (due 07.11.2017 – later submission = worse grade) => see [handout](#)