DOCUMENTARY LINGUISTICS I
prof. Nicole Nau, UAM winter 2016/2017

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

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

…

WHAT IS A LANGUAGE DOCUMENTATION? IDEAL

«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» (Himmelmann 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)

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 (additional information)

❖ Description (results of analysis: grammar, dictionary)

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 **indiviual** 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.» (Himmelmann 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 linguistc units and events (Himmelmann 1998)

Give examples that show your metalinguistic knowledge!

How can metalinguistic knowledge be accessed and documented?

«[…] corpus observation is best done in conjunction with meta-corporeal and metalinguistic discussion; for example, if you are making a thesaurus, you don't want to just find the names of different grasses in your corpus, you also want the resultant list to be discussed and gone over by speakers who are authorities on grasses to make sure you have the field properly covered and to generate good definitions. Rather, what a documentation-oriented view says is that the discussions of grass names should themselves be videotaped or tape recorded and should themselves become a part of the whole corpus; as should any and all grammatical elicitation of the traditional kind. Moreover, years from now, it will be the grass name attestations and grasses discussion tapes, and not the dictionary, that you will consider as the final document on grass names.» (Woodbury 2003)

METADATA AND ADDITIONAL INFORMATION

Annotation of primary data may contain

- transcription / transliteration

- translation

- grammatical annotation («glossing», «tagging»)

- comments

- …

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

METADATA FOR SESSIONS AND RECORDINGS

Metadata for spoken data usually contain

- participants (speakers and bystanders)

- time and location

- recorded by

- recorded with

- …

What are metadata for?

Standards for metadata: OLAC, IMDI

Primary data	Apparatus	
	Per session	For documentation as a whole
recordings/records of observable linguistic behavior and metalinguistic knowledge (possible basic formats: session and lexical database)	*Metadata* – time and location of recording – participants – recording team – recording equipment – content descriptors … *Annotations* – transcription – translation – further linguistic and ethnographic glossing and commentary	*Metadata* – location of documented community – project team(s) contributing to documentation – participants in documentation – acknowledgements … *General access resources* – introduction – orthographical conventions – ethnographic sketch – sketch grammar – glossing conventions – indices – links to other resources …

HIMMELMANN'S «APPARATUS»

(HIMMELMANN 2006)

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.

ON THE BENEFIT OF STANDARDIZING METADATA AND ADDITIONAL INFORMATION

You want to find records of the language Aleut.

Compare the following web-pages:

Non-standardized ressource list:
http://www.native-languages.org/aleut.htm

OLAC ressource catalogue:
http://www.language-archives.org/language/ale

EXAMPLES OF LANGUAGE DOCUMENTATIONS

Professional documentation (deposits of larger projects)

http://dobes.mpi.nl/ (DOBES = **Do**kumentation **be**drohter **S**prachen)

https://elar.soas.ac.uk/ (ELAR = **E**ndangered **L**anguages **Ar**chive)

Non-professional documentation (small samples of languages)

http://languagelandscape.org/