DOCUMENTARY LINGUISTICS I
prof. Nicole Nau, UAM winter 2018/2019

Nineth lecture
27 November 2018

TODAY

- Some remarks about video recordings
- First steps after making a recording
- Annotation: overview; transcription and translation
- Types of transcription
- The problem of segmentation

SOME REMARKS ON VIDEO RECORDING

«Video has become increasingly relevant to language documentation (Austin 2007b:27; Seifart 2012:1) because of its utility in capturing aspects such as turn-taking, gesture, facial expressions, eye gaze and lip movement (Nathan 2011:269; Margetts & Margetts 2012:32). **However, its ease of creation also raises questions of quantity versus quality, and to its value if not utilised effectively**.» (Brickell 2018)

Making useful video recordings has to be learnt!

SOME RECOMMENDATIONS FOR VIDEO RECORDING SPEECH

The following slides are taken from

a presentation by

Marta Ostajewska

The Basic Principles of Creating

Audiovisual Material

Marta Ostajewska, Artes Liberales, UW

Video and language documentation

* You have to discuss sharing and permissions with the community, and the individuals you are recording with and have the permission to record

* Setting up any recording situation is obtrusive

* Being comfortable with your equipment and making people feel comfortable with your presence mitigates many of those problems. Practice setting up as many times as you can before you begin the project. Record your friends and family. Know your equipment well enough now to continue chatting throughout the setup.

Rotation from horizontal level to vertical level

Do not turn the camera, shoot horizontally, it makes your montage easier

Bad lighting

- Do not film a caller under the sun and against the window, in a place with variable light. Shot a scene preferably on a neutral background

- Record under daylight in a well lit room

- Do not mix artificial light with natural light

Moving to much

- Give a scene time to resound (do not change the position of the camera constantly). Try to be calm during filming

- **Do not make jumpy zoom in – zoom out**

- Get closer to the subject with the camera instead of zooming

- If you zoom in with a camera on the tripod, set a good frame and give the scene a time
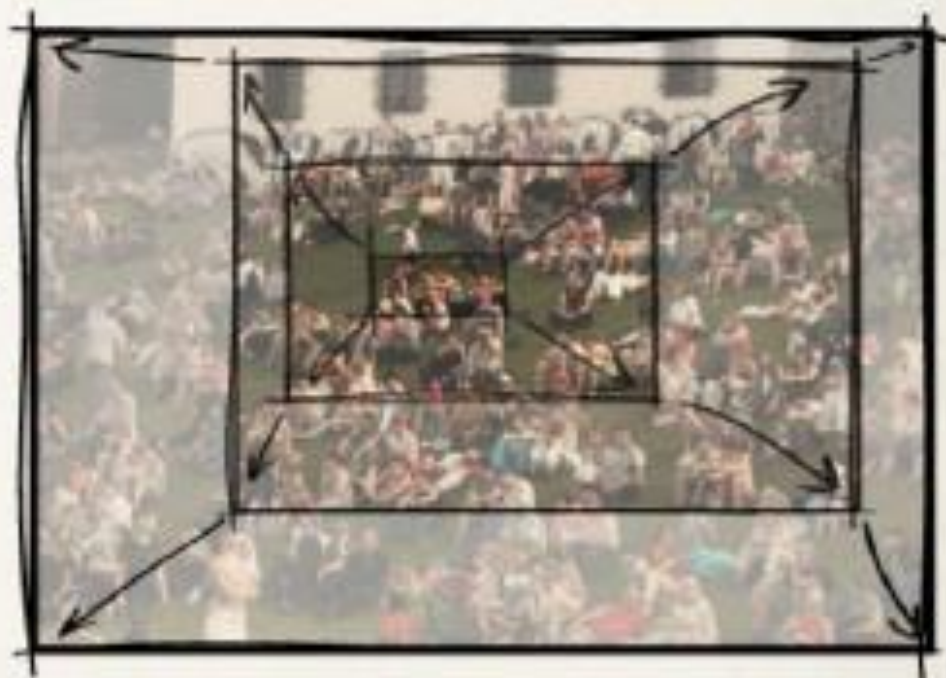
Sound problems

* Do not record sound only on the camera, have an external microphone

* Try not to record in the strong wind

* Dress up your microphone outdoor

* Avoid overdrive

In video documentation of the language it is important that the gestures fit in the frame, do not go beyond the frame, are not cut

Reasons you should use video in language documentation

* Gesture is an important part of communication

* Gesture is an important part of cognition

* Gesture is an important part of culture

* Having video as well as audio makes transcription easier, and ensures that the elements of discourse that are not in the spoken channel are still collected

VIDEOS WITH AND WITHOUT GESTURES

Example 1. Language: Yiddish (Speaker lives in England)

Interview with Sonia Pinkusowicz-Drata (from «Poland's Linguistic Heritage»), http://inne-jezyki.amu.edu.pl/Frontend/Text/Details/1114

Example 2. Language: Bakola (Cameroon)

Story of personal experience told by Mba (from DOBES Bakola Corpus)

https://hdl.handle.net/1839/00-0000-0000-0017-DC3E-E

REMEMBER

All **recordings are reductions** – they can never render the total of the speech event, all aspects that may be important for the recorded language use.

Audio recordings reduce still more than video recordings.

All **transcriptions** of a recording **are further reductions**.

You are forced to make a **selection** – make it consciously!

WHAT ARE THE FIRST THINGS TO DO WITH A RECORDING?

1. **Back up**: store original at a save place. Use an appropriate, unique file name.

2. **Metadata**: describe the recording and keep the description with the original (and remember where you have the **consent form**)

Only then you should go on and work with a copy of the recording. This may involve:

3. **Cut** the recording into manageable and coherent pieces.

How to cut? Recommended program for audio: «Audacity»

4) ANNOTATE CHOSEN (PIECES OF) RECORDINGS. ANNOTATION MAY INCLUDE:

- ❖ transcription(s)
- ❖ translation(s)
- ❖ grammatical analyses
- ❖ comments on the language/speech/code (use of slang, whispers, code-switching…)
- ❖ comments on the content
- ❖ notes on situational factors (a telephone rings, the speaker pauses to drink coffee, talks to their dog…)
- ❖ video: description of gestures, facial expression etc.
- ❖ …

WHY ANNOTATE?

«It is self-evident that the task of documenting a communicative event does not stop at simply recording it (by producing, e.g., an audio- or videorecording). Especially in the case of languages only spoken by a small group of people, such a recording would not be interpretable by the majority of people with a potential interest in the language – e.g. linguists, anthropologists, historians, or the general public. In the case of endangered languages, the recording would possibly not even be interpretable to the descendants of the speakers themselves. Therefore, a recording has to be accompanied by further information, in a format that is accessible to a wider, possibly non-specialist, audience.» (Schulze-Berndt 2006)

ANNOTATION SHOULD BE

❖ multi-tiered

❖ time-aligned

❖ expandable

«Another important point to remember is that "language documentation is an inherently ongoing process" and that annotations may be produced or corrected "multiple times by one or multiple authors" (Holton 2003: 6; cf. also Edwards 2001: 322). It is thus quite possible that, for example, an annotation consisting of a transcription in a practical orthography and a translation will be supplemented, many years later, with a prosodic annotation by a research project on prosody, and with grammatical annotation by someone working on a reference grammar of the language.» (Schulze-Berndt 2006)

EXAMPLE OF A MULTI-TIERED NON TIME-ALIGNED HARDLY EXPANDABLE ANNOTATION (OTHERWISE NICE :)

ix hob es ništ gefilt vi epes vos iz hexer oder nideriker.
/ɪx hɔb ɛs nɪʃt gɛfɪlt vɪ ɛpɛs vɔs ɪz | hɛxɛʀ ɔdɛʀ nɪdɛʀɪkɛʀ ‖
Nie czułam, aby było to czymś, co jest lepsze albo gorsze.
I don't think that it is a matter of being better or worse.
ix   **hob**   **es ništ** **ge-fil-t**   **vi epes**   **vos** 1SG.NOM have.PRS.1SG it NEG PTCT-feel-PTCT as something that **iz**   **hex-er**   **oder niderik-er** bePRS.3SG   high-COMP or   low-COMP

Free translation:
how free can it be?

un es hot mix gemaxt filn efšer štolc afile, az me hot gehert a pojliš jidiš un dos iz geven a litviš jidiš, iz majn štolc geven cu redn a pojlišn jidiš, jo.
ən ɛs hɔt mɪx gɛmaxt fɪln ɛfʃɛʀ ʃtɔlts afɪlɛ az mɛ hɔt gɛhɛʀt a pojlɪʃ jɪdɪʃ ən dɔs ɪz gɛven a lɪtvɪʃ jɪdɪʃ ɪz majn ʃtɔlts gɛven tsʊ ʀedn a pojlɪʃn jɪdɪʃ | jɔ ‖
Czułam się chyba raczej dumna z mojego polskiego jidysz.
In fact I feel rather proud of my Polish Yiddish.
un **es hot**   **mix**   **ge-max-t**   **filn**   **efšer** **štolc** **afile,** and it have.PRS.3SG I.ACC PTCT-make-PTCT feel.INF maybe proud even **az**   **me**   **hot**   **ge-her-t**   **a**   **pojliš jidiš** when 3SG.INPERS have.PRS.3SG PTCT-hear-PTCT a.SG Polish Yiddish **un dos iz**   **ge-ven**   **a**   **litviš** **jidiš,** and it be.PRS.3S PTCT-be.PTCT a.SG Litvak Yiddish **iz**   **majn štolc** **ge-ven**   **cu redn**   **a**   **pojliš-n**   **jidiš, jo.** be.PRS.3SG my   proud PTCT-be.PTCT to speak.INF a.SG Polish-DAT Yiddish, yes.

TWO PROBLEMS WITH FREE TRANSLATIONS

«The first is that a free translation, especially when the translation is that of a whole paragraph, tends to assume the stylistic features of written as opposed to spoken language. This is not a major issue if the translator is aware of it and if the translation is regarded as an aid for the interpretation of the original utterance by later users, not as a faithful rendition of the original. In special cases however, e.g. when translating ritual speech events or verbal art, the translator may well strive to represent aspects of the original discourse structure»

«The second issue is that of adding information not present in the original»

(Schulze-Berndt 2006)

TRANSCRIPTION: DIFFERENT SYSTEMS FOR DIFFERENT PURPOSES

❖ orthographic transcription(s)
- ➢ using an established writing system
- ➢ using a newly developed / developping writing system
- ➢ using an adhoc-writing «system»

«If an orthography for the language under investigation is already established and acepted by the speech community, it is virtually an obligation for a documentary linguist to provide an orthographic transcription asa part of the annotation, since this greatly adds to the accessibility of the documentation for the members of the speech community themselves.» (Schulze-Berndt 2006)

Do we agree?

❖ phonetic transcription(s)

➢ broad / narrow phonetic transcription

➢ IPA or a regional/local traditional phonetic script (for example, slavists's system for transcribing Slavic languages)

➢ SAMPA

(3)   Phonetic transcription using IPA and SAMPA symbols (German)

\phonet_ipa	ˈpyŋkt.lıç
\phonet_sampa	ˈpYNkt.lIC
\orth	pünktlich
\ft	punctual

\phonet_ipa	ˈʃø.nə
\phonet_sampa	ˈS2:.n@
\orth	schöne
\ft	beautiful (F)

(Schulze-Berndt 2006)

❖ conversation-analytic transcription

(1) NSPCC DG Daughter and Fleas. 1.59

```
1    CPO:        It's a bit acrimonious [is it,]
2    Caller:                            [Ye:s  ] iddis.
3    CPO:        Ri:ght. Okay, .hh ER:m (0.3) tk (.) Ri:ght.
4                (0.5)
5    CPO:        .hh I MEAN THERE MAY BE (.) VARIous rea:sons,
```

WHY TRANSCRIBE? DIFFERENT PURPOSES

❖ phonetic and phonemic analysis, (remember: transcription **is** analysis)

❖ conversation analysis,

❖ other linguistic purposes,

❖ making content easily accessible (faster reading than listening),

❖ finding words and constructions,

❖ other form of "storage", of having a "record" (but remember a transcript is not primary data),

❖ preparing further steps in annotation, e.g. translation

SOME QUESTIONS

❖ How many different tiers? How much / which information on one tier? (phonetic, prosodic)

❖ How much time does it take to transcribe?

❖ How accurate, detailed should a transcription be?

❖ How to use punctuation marks?

❖ Who should do the transcription?

❖ How can we ensure future readers will understand the transcription?

«Transcriber effects are unavoidable. Anyone who has transcribed recorded data has experienced surprise at discovering that chunks of audio material have been entirely overlooked in the transcript, frequently due to the natural facility of attending to the salient constituents of a message and tuning out material perceived to be irrelevant to the message.» (Nagy & Sharma 2014)

WHY TRANSCRIPTION IS GOOD FOR LINGUISTS

«linguists are never as close to their object of study as when they are transcribing. The very act of transcription helps the researcher find and understand patterns in the data, seeing elements that may be elusive and fleeting in the original oral form.»

(Nagy & Sharma 2014)

TRANSCRIPTION AS TRANSLATION FROM ORAL TO LITERATE

Sometimes, you see his hand at close up, as he is snatching the pears from the trees, and you hear the sound really strongly.

and… you see his hand sometimes at close up – snatching the pears from the tree. And you hear the sound really strongly.

and (0.4) you see his hand sometimes at close up ;
(1.1) uh snatching the pears from the tree /
(0.8) and you hear the sound really: strongly \

UNITS IN WRITTEN AND SPOKEN LANGUAGE

Units of written language	Units of spoken language
words	words
	intonation units
sentences	?
paragraphs	paragraphs?

CAN YOU HEAR INTONATION UNITS IN AN UNKNOWN LANGUAGE?

ja aš tonin gredahma na sa rekla toni homa po na palatu je reka što maš po či e hočam po vi na mala ferju kaka rivivahma di skola ja nisa bi vidila ka profesor sidaša nonda di parket alor tonin sa vračaša na mala naza vidim ka biša profesor je mu

Orthographic transcription:

Ja aš Tonin gredahma na... Sa rekla: «Toni', homa po na Palatu!»
Je reka: "Što maš po či?"
«È! Hočam po vi na mala ferju!»
Kaka rivivahma di skola,
ja nisa bi vidila ka Profesor sidaša nonda di parket.
Alor Tonin sa vračaša na mala naza, vidim ka biša Profesor.

Breu, Walter. *Le voyage à Palata*, Na-našu corpus. Pangloss Collection, LACITO-CNRS. Programme franco-allemand Euroslav 2010 : ANR(09-FASHS-025) - DFG(BR 1228/4-1), http://lacito.vjf.cnrs.fr/pangloss/corpus/show_text_en.php?id=crdo-SVM_PALATA_SOUND&idref=crdo-SVM_PALATA

NIKOLAUS HIMMELMANN (2006) THE CHALLENGES OF SEGMENTING SPOKEN LANGUAGE: HOW TO DETERMINE INTONATION UNITS?

«In most languages, evidence for intonation unit boundaries is provided by changes in pitch and rhythm.»

1. Pitch

1.a) **boundary tone** at the end of an intonation unit, change in pitch on the last syllable

1.b) new onset at the beginning of the next unit, **jump in pitch**

1.c) reset of the baseline: early pitch peaks in the new unit are higher than the final pitch peaks in the preceding one.

2. Rhythm

2.1 pause

2.2 lengthening of the final segment

2.3 anacrusis (accelerated delivery of unstressed syllables at the beginning)

Obligatory in many languages: 1.a) and 1.b)

Most practical diagnostics: 1.a) and 2.a)

TRANSCRIPTION CONVENTIONS
USED BY HIMMELMANN

/   rising boundary tones

\   falling boundary tones

;   unit ends somewhere in the middle, neither rising nor falling

(.) short pause

(..) longer pause

(0.5) measured pause of 0.5 seconds

:   lengthening

=   latching

EXAMPLES FROM HIMMELMANN

(1)   PEAR-L-36FF

    36. and (0.4) you see his hand sometimes at close up ; (1.1)
    37. uh snatching the pears from the tree / (0.8)
    38. and you hear the sound really: strongly \ (0.8)

(2)   PEAR-L-48FF

    48. he climbs down the ladder / (0.5)
    49. and he puts a couple of the pears– (=)
    50. well: (=)
    51. as he's standing there ; (=)
    52. couple of the pears fall \ (0.4)

A USEFUL PROGRAM…

… for analyzing the prosody (pitch, pauses, rhythm) of small audio files:

Praat

http://www.fon.hum.uva.nl/praat/

REFERENCES AND RECOMMENDED READING

Naomi Nagy & Devyani Sharma: Transcription (in: Podesva & Sharma, eds. 2013. Research Methods in Linguistics)

Eva Schultze-Berndt. 2006. Linguistic annotation (in: *Essentials of Language Documentation*)

For the lectures in December you need a laptop with ELAN installed!