

DOCUMENTARY LINGUISTICS I

prof. Nicole Nau, UAM winter 2016/2017

Fourth lecture
07 November 2016

TOPICS OF THE DAY

1. Reflection on the first task
2. Warming-up questions
3. How to collect primary data
 - ❖ Basic considerations
 - ❖ Some aspects of written sources
 - ❖ Some aspects of spoken sources
4. Bonus track

EXAMPLES OF LANGUAGE DOCUMENTATION

Your homework?

Which archives did you explore?

How userfriendly did you find them?

What problems did you encounter?

WARMING-UP QUESTIONS (2 MINUTES FOR REFLECTION / CONSULTATION)

- 1) Why is it impossible to record all communicative events in a given speech community?
- 2) “In some instances a documentation project may contribute to the demise of the very linguistic and cultural practices it proposes to document.” (Himmelman 2006: 17)
Comment, give (invent) an example.

HOW TO GET PRIMARY DATA — BASIC CONSIDERATIONS

- ❖ content: what exactly do you want to collect, and why?
- ❖ respect: ownership, privacy, laws
- ❖ technology: what do you need (i) to gather and (ii) to use the data?

=> method

RESPECT

- ❖ be aware of legal issues
- ❖ treat participants respectfully
- ❖ get participants' consent (formal/less formal)
- ❖ think about the future of the data
- ❖ reflect and discuss the necessary level of anonymization

example of a simple consent form:

<http://languagelandscape.org/resources/>

WHAT KIND OF LINGUISTIC DATA DO WE RECORD, COLLECT, WORK WITH LATER?

- ❖ events / **texts** / non-texts (words, isolated sentences)
- ❖ **elicited** (primarily produced for documentation) / **observed** (produced for other purposes) / in between
- ❖ grade of preparation, **planning** (spontaneous / planned)
- ❖ **mode(s)**: spoken / written / signed / multimodal

TYPES OF **WRITTEN** SOURCES FROM THE POINT OF VIEW OF DATA COLLECTION AND PROCESSING

published in print or static electronic formats (PDF)

computer-mediated communication (dynamic electronic formats): blogs, chats, social media, forum posts

unpublished digital texts (computer-written manuscripts)

unpublished non-digitalized written texts

- type-written, hand-written texts (letters, diaries, school essays)
- other: public signs, inscriptions

WHAT TO DO
IF YOUR SOURCE
LOOKS LIKE THIS?

WHAT DOES IT CONTAIN?
WHAT INFORMATION DO
YOU WANT TO DIGITALIZE?



FROM SOURCE TO CORPUS ENTRY: WHAT INFORMATION IS (NOT) PRESERVED? (WWW.KORPUSS.LV/SENIE/)



A2A. lpp.

- 1: Die fünff Heupt-
- 2: stück des heiligen Cate-
- 3: chismi: Neben
- 4: Der Beicht / Frage vnd Antwort /
- 5: die da sonderlich den einfeltigen Bawren vor
- 6: vnd nach der Predigt sollen fein deutlich vnd
- 7: langsam vorgelesen vnd gebetet werden.
- 8: I.
- 9: Die heiligen zehen Gebot.
- 10: Te Desmette Dewe Bowfle. Exodi
- 11: tan zo. wetan.
- 12: Tas pirmais Boußlis.
- 13: Toew nhe buus czittes Dewes turreth
- 14: prexkan man.
- 15: Tas Oteers
- 16: Toew nhe buus Dewe touwe kunge
- 17: waerde nhepatthese walkoot.
- 18: Tas Treßches
- 19: Thoew buus to sweete Dene sweetyt.
- 20: Tas cettortz
- 21: Tow buus touw Thewe vnde touwe

WHAT TO DO IF YOUR SOURCE LOOKS LIKE THIS?

(for example, you want to
investigate the use of English
in public signs in Poznan)



COLLECTING **SPOKEN** LANGUAGE: TECHNICAL CONSIDERATIONS

- ❖ purpose: for phonetic analysis (**speech** analysis) / for other purposes
- ❖ collected in natural situations / laboratory etc.
- ❖ sound file formats
- ❖ recording devices
- ❖ microphones
- ❖ surrounding

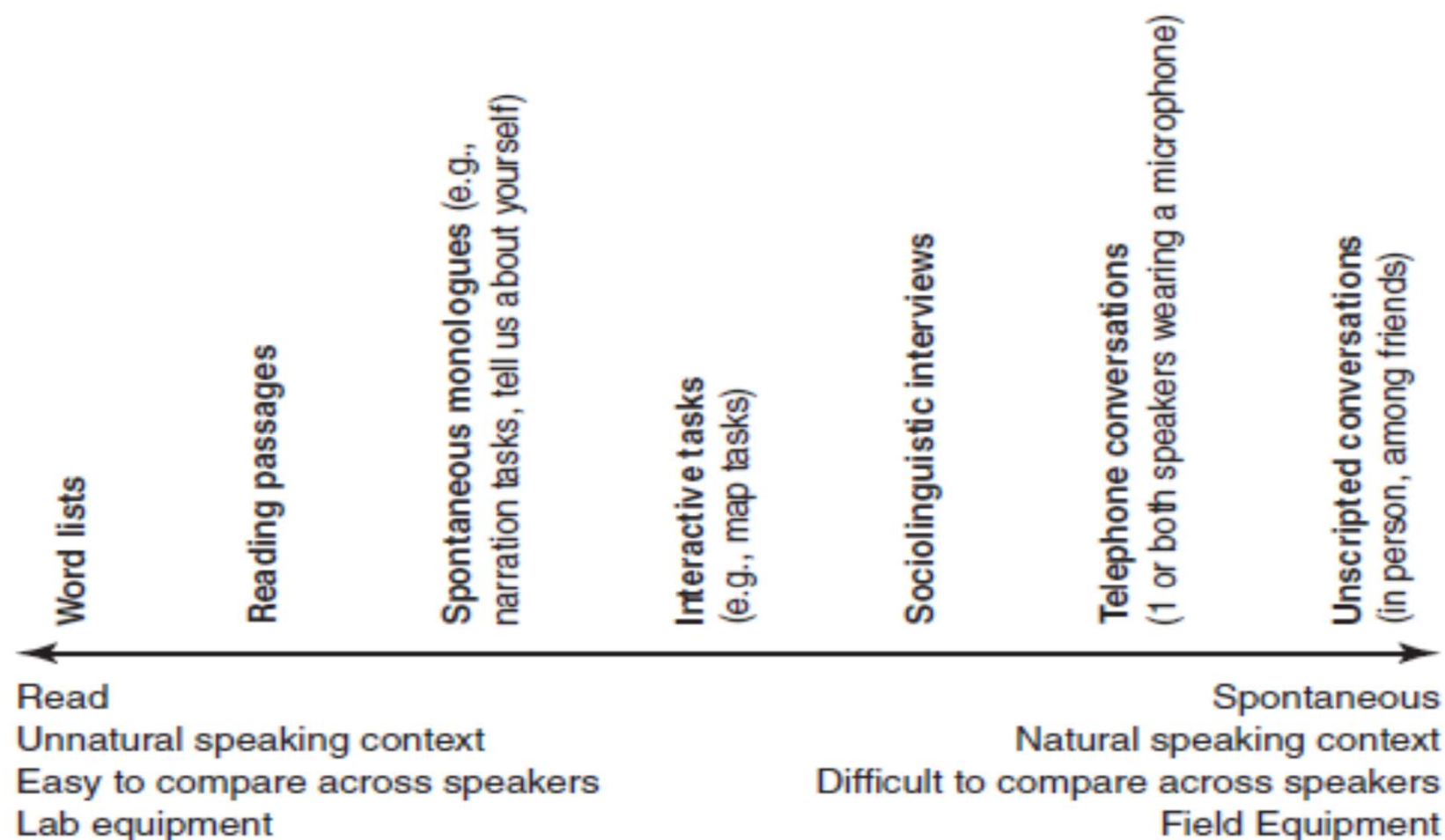
Figure 9.4. *Range of data collection scenarios*



Figure 9.1. *Common microphone mounts: stand-mounted (left), head-mounted (middle), and lavalier (right)*

RECORDING THE
TEOP LANGUAGE
IN BOUGAIN-
VILLE (PNG)

Photo by Ulrike
Mosel



TYPES OF SPOKEN LANGUAGE — MONOLOGUE, DIALOGUE, AND IN BETWEEN

- ❖ strictly monologic: directed to non-human audience (examples: prayer; talking to animals)
- ❖ strictly dialogic: all participants are physically present and may interact (conversation, interview)
- ❖ tele-dialogic: main audience is not present, but communication is bi-directional (telephone, skype)
- ❖ transmitted, one-sided communication: main audience is not present and cannot interact (radio broadcast, video posted on YouTube)
- ❖ mixed forms: radio interview, TV talk show, staged recorded communication, elicitation

TASK

Make recordings for a small corpus of spoken Polish.

Speakers: students.

Length of each recording: 2-5 min.

What do you want to record and how will you do it?

BONUS TRACK

Watch a film about a new method of gathering data (spoken texts) from endangered languages: Aikuma

How does this method work?

What are the advantages of the method?

<http://www.aikuma.org/media.html>