# *Data Collection Management & Analysis* *in the Humanities*

Katarzyna Klessa

UNIWERSYTET
IM. ADAMA MICKIEWICZA
W POZNANIU

Workshop Presentation, Poznań, 7th July 2017

# Introduction

Katarzyna Klessa

- Phonetics
- Phonology
- Language databases
- Tools & Resources for the analysis of speech phenomena

More about me:

www.katarzyna.klessa.pl

www.annotationpro.org

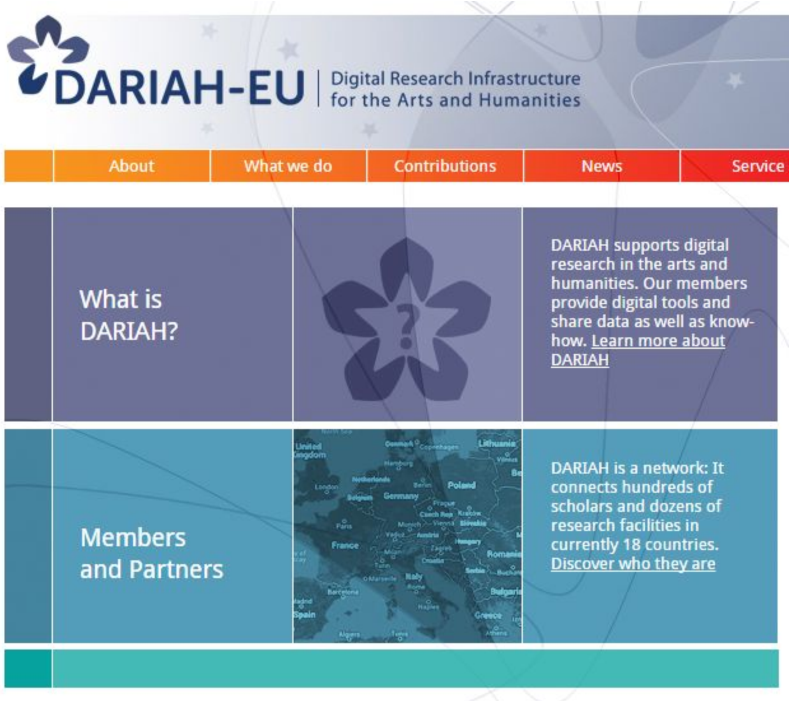
# Data.... in the **humanities**

*Humanities, those branches of knowledge that concern themselves with human beings and their culture or with analytic and critical methods of inquiry derived from an appreciation of human values and of the unique ability of the human spirit to express itself.* (Britannica)

*Humanities are academic disciplines that study aspects of human culture (...) Today, the humanities are more frequently contrasted with natural, and sometimes social, sciences as well as professional training.* (Wikipedia)

*The humanities can be described as the study of how people process and document the human experience.* (Stanford Humanities Centre)

# Digital humanities


DARIAH-EU | Digital Research Infrastructure for the Arts and Humanities
About
What we do
Contributions
News
Service

What is DARIAH?

DARIAH supports digital research in the arts and humanities. Our members provide digital tools and share data as well as know-how. Learn more about DARIAH

Members and Partners

DARIAH is a network: It connects hundreds of scholars and dozens of research facilities in currently 18 countries. Discover who they are

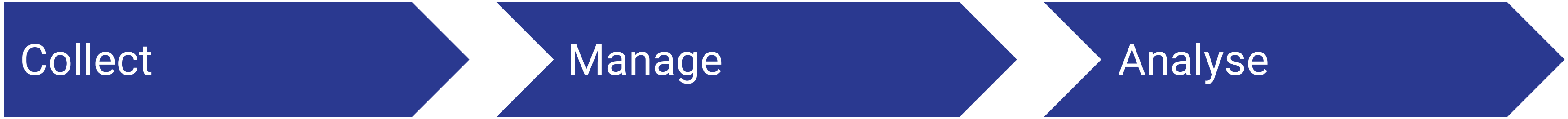
DH - Academic field dealing with the application of computational tools and methods to traditional humanities disciplines.

DARIAH-PL

*Digital Research Infrastructure for the Arts and Humanities*

- Supports digital research in the arts and humanities.
- Members provide digital tools and share data as well as know-how
- Collaboration support.
- Conferences, meetings, workshops.

# Topics for today: Data-related tasks



**Data**

Texts, audio/video recordings, “dark data”

**Metadata**

“Data about data”, additional information about the materials, speakers, environment...

**Technology**

- Hardware and software tools, data management solutions (file collections, relational databases, SQL)
- Backup copies
- Sharing options

**Description**

annotation, transcription => interpretation

**Processing**

Data & metadata processing

**Exploration**

Information extraction, data mining

# Who can be interested in data collection & more?

Linguists Phoneticians Psychologists

Teachers Archivists Psycholinguists

(Speech) Therapists Sociologists Students

Historians (Language) Documenters Lawyers

Speech Technology Specialists

?

………………… ………………… …………………

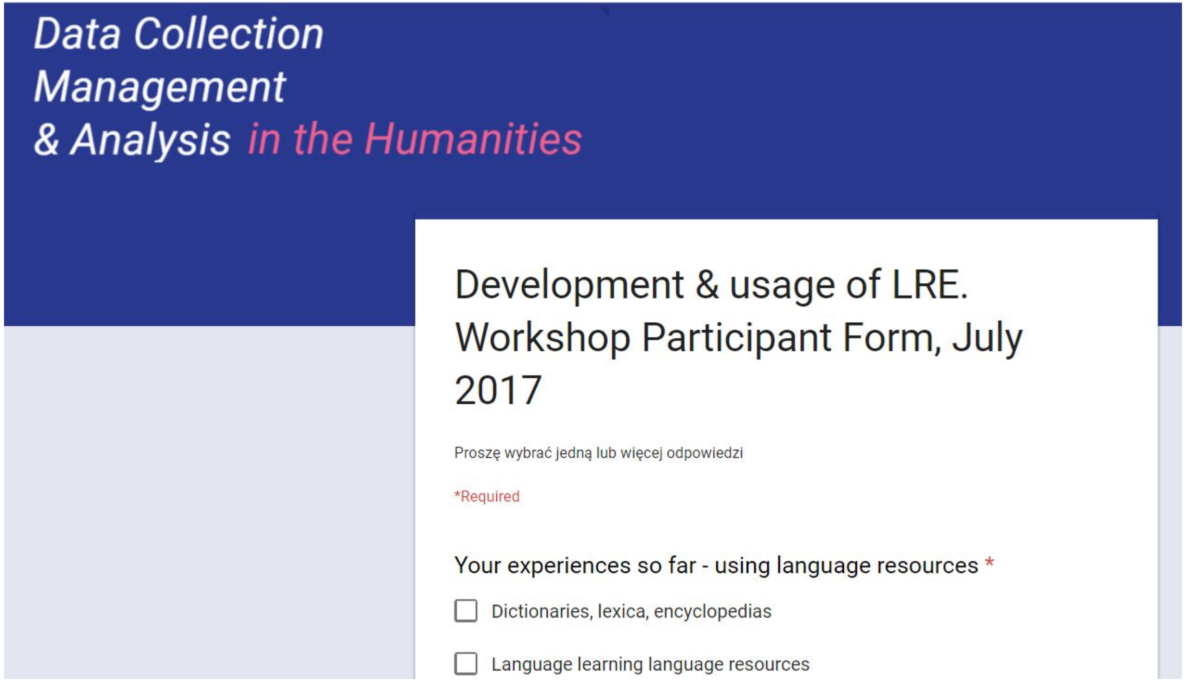
# Diversity: challenge & chance

*It is precisely in the challenge of this divergence between disciplines that one might expect to arrive at a more significant reward*

*– that the contributions of each might ultimately be greater than the sum of their parts.*

Cox, Ch. (2011). Corpus linguistics and language documentation: challenges for collaboration. Language and Computers-Studies in Practical Linguistics, 73(1), 239.

# Workshop survey



*Data Collection*
*Management*
*& Analysis in the Humanities*

Development & usage of LRE. Workshop Participant Form, July 2017

Proszę wybrać jedną lub więcej odpowiedzi

*Required

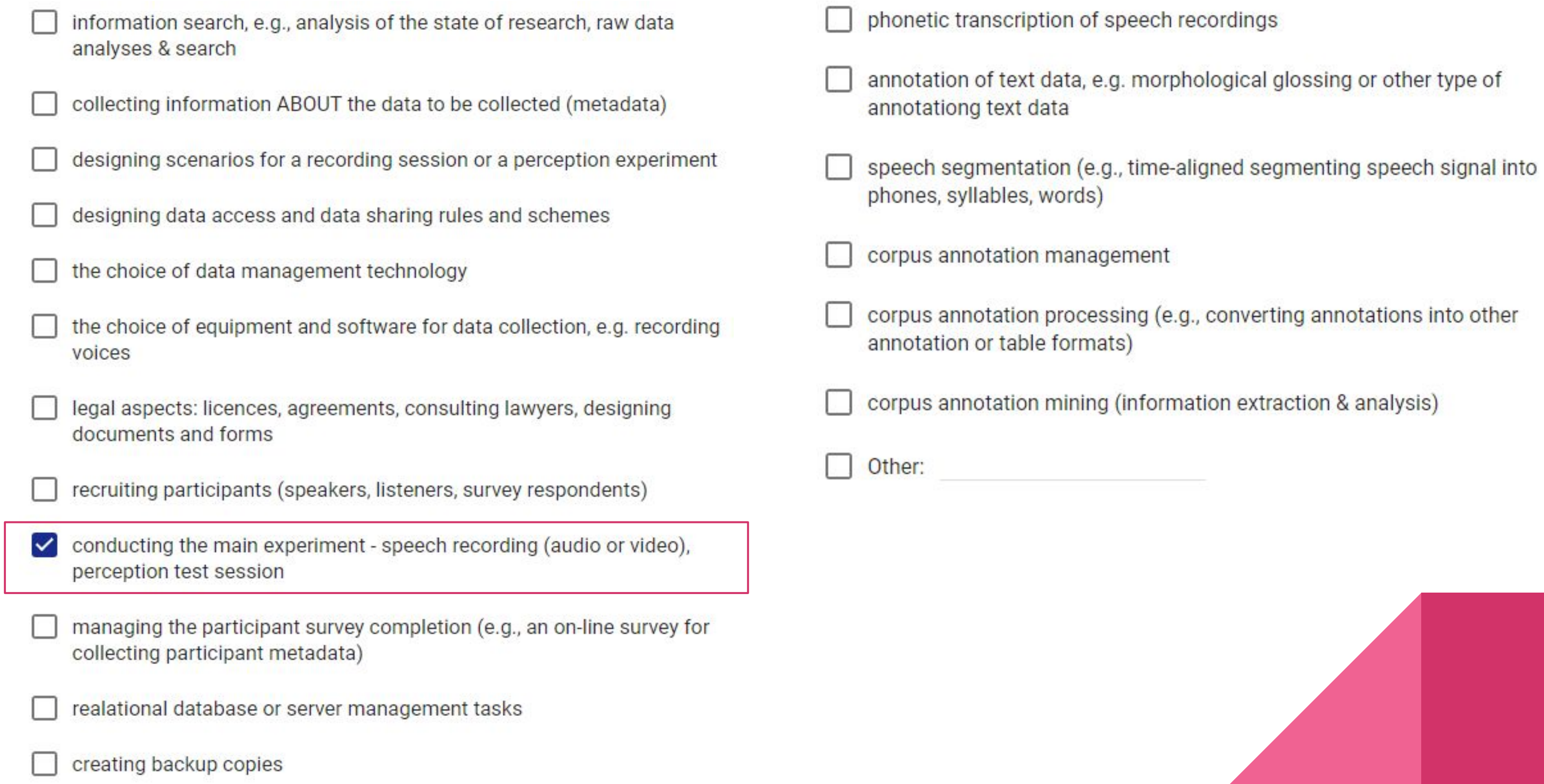
Your experiences so far - using language resources *

□ Dictionaries, lexica, encyclopedias

□ Language learning language resources

Survey link:
https://goo.gl/forms/X2gh1ZvgSErr43VV2

# Workshop survey: not only the target data



□ information search, e.g., analysis of the state of research, raw data analyses & search

□ collecting information ABOUT the data to be collected (metadata)

□ designing scenarios for a recording session or a perception experiment

□ designing data access and data sharing rules and schemes

□ the choice of data management technology

□ the choice of equipment and software for data collection, e.g. recording voices

□ legal aspects: licences, agreements, consulting lawyers, designing documents and forms

□ recruiting participants (speakers, listeners, survey respondents)

☑ conducting the main experiment - speech recording (audio or video), perception test session

□ managing the participant survey completion (e.g., an on-line survey for collecting participant metadata)

□ realational database or server management tasks

□ creating backup copies

□ phonetic transcription of speech recordings

□ annotation of text data, e.g. morphological glossing or other type of annotationg text data

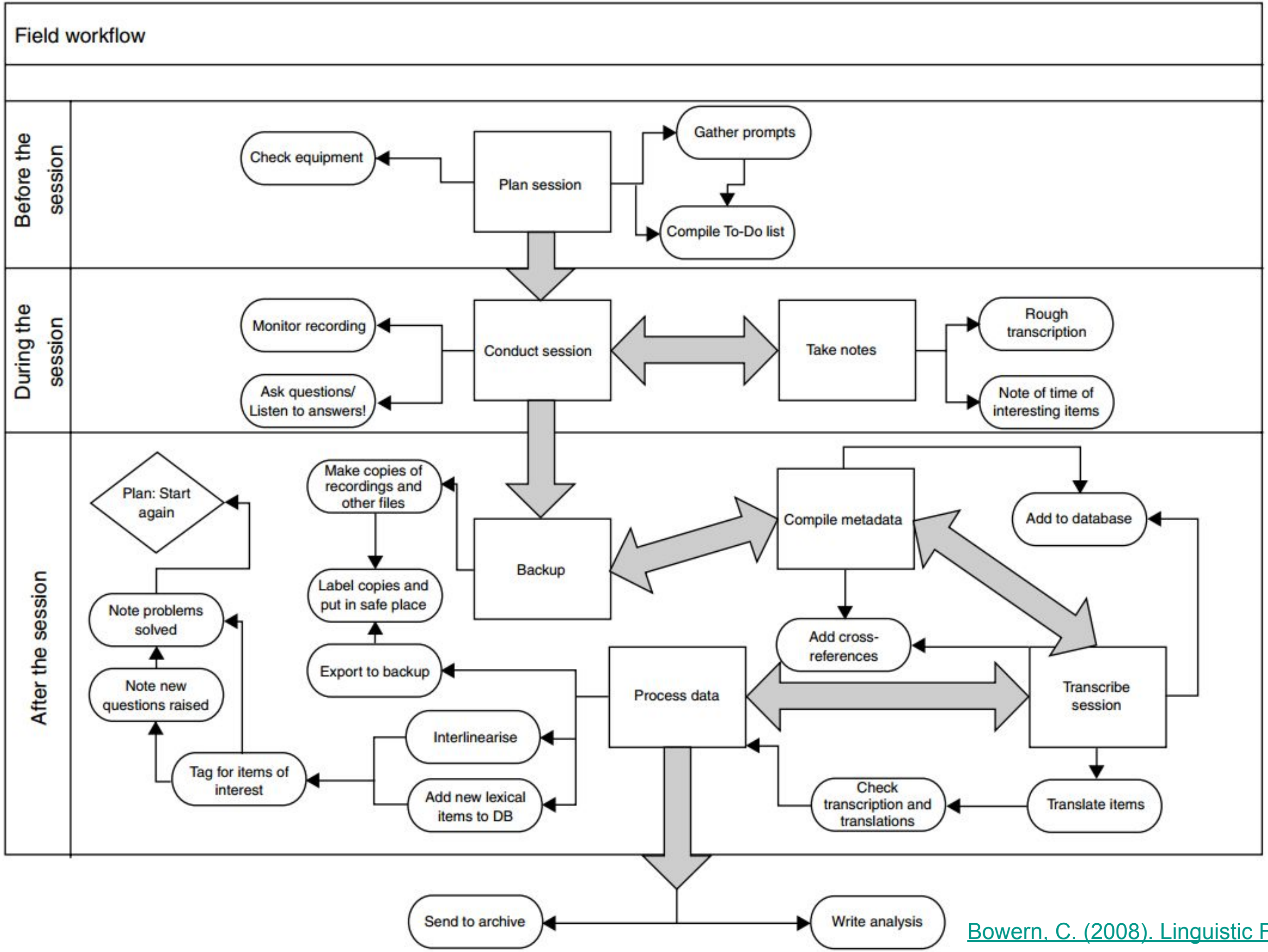
□ speech segmentation (e.g., time-aligned segmenting speech signal into phones, syllables, words)

□ corpus annotation management

□ corpus annotation processing (e.g., converting annotations into other annotation or table formats)

□ corpus annotation mining (information extraction & analysis)

□ Other: ____________



Field workflow

Before the session

Check equipment

Plan session

Gather prompts

Compile To-Do list

During the session

Monitor recording

Ask questions/ Listen to answers!

Conduct session

Take notes

Rough transcription

Note of time of interesting items

After the session

Plan: Start again

Note problems solved

Note new questions raised

Tag for items of interest

Make copies of recordings and other files

Label copies and put in safe place

Export to backup

Interlinearise

Add new lexical items to DB

Backup

Process data

Compile metadata

Add cross-references

Add to database

Transcribe session

Check transcription and translations

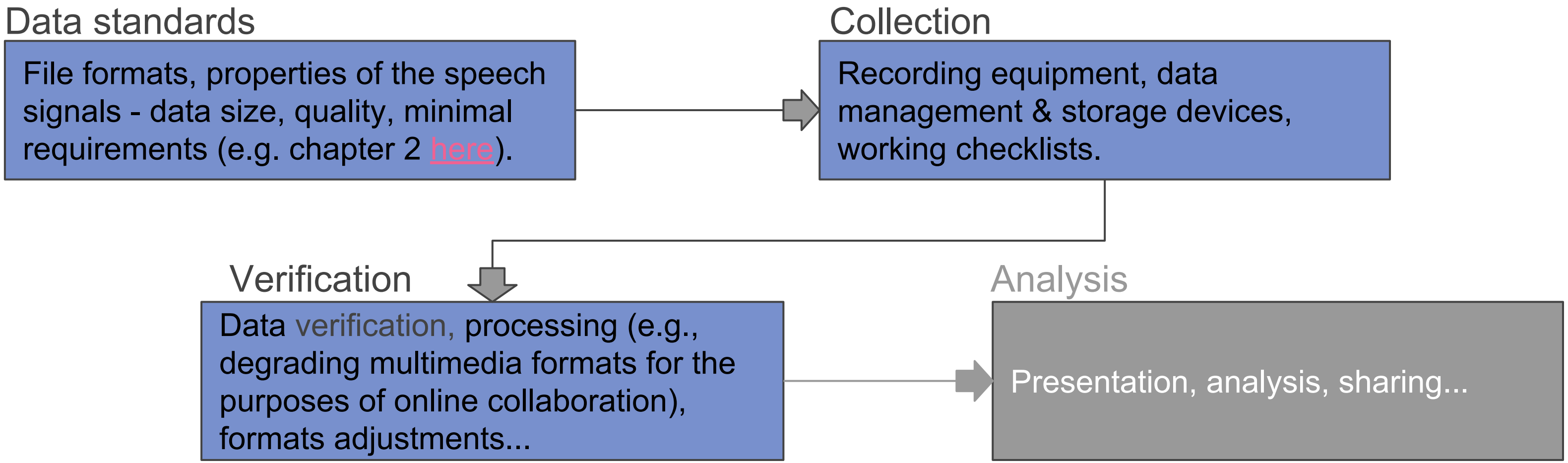
Translate items

Send to archive

Write analysis

Bowern, C. (2008). Linguistic Fieldwork. A Practical Guide, page 48.

# Steps in collecting data



Data standards

File formats, properties of the speech signals - data size, quality, minimal requirements (e.g. chapter 2 here).

Collection

Recording equipment, data management & storage devices, working checklists.

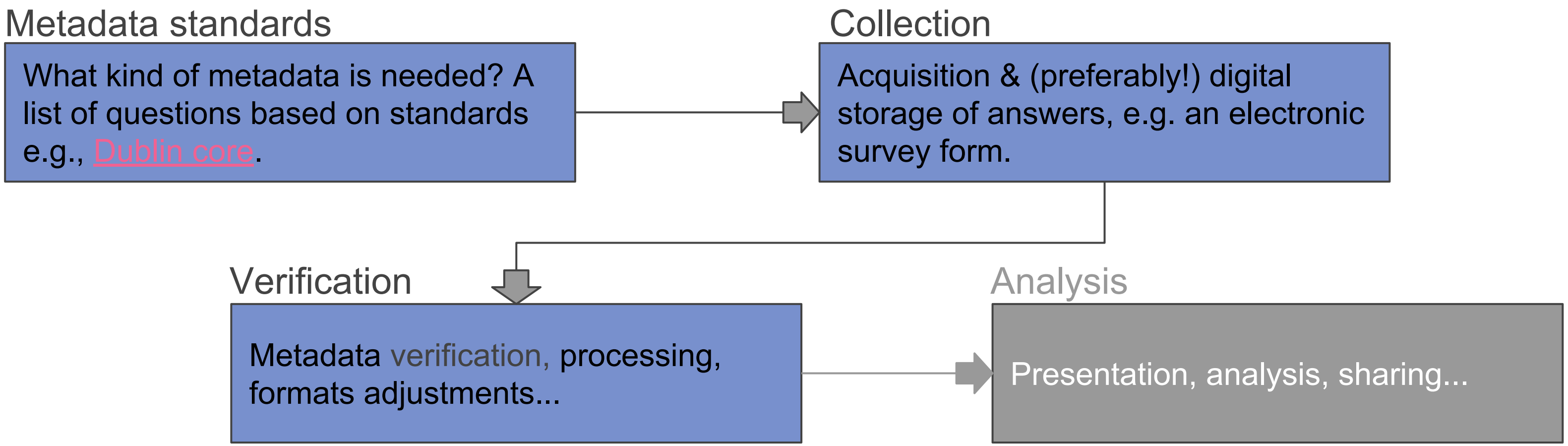
Verification

Data verification, processing (e.g., degrading multimedia formats for the purposes of online collaboration), formats adjustments...

Analysis

Presentation, analysis, sharing...

# Steps in collecting metadata



Metadata standards

What kind of metadata is needed? A list of questions based on standards e.g., Dublin core.

Collection

Acquisition & (preferably!) digital storage of answers, e.g. an electronic survey form.

Verification

Metadata verification, processing, formats adjustments...

Analysis

Presentation, analysis, sharing...

# Metadata - Exercise

Let's suppose that we wish to collect recordings of old songs.

What kind of information would we wish to preserve - besides the recordings themselves?

One of the earliest music recordings here: https://upload.wikimedia.org/wikipedia/commons/8/86/Kham_Hom_-_Sweet_Words.ogg (read more)

# Metadata - Exercise

One of the earliest music recordings.

Here:

https://upload.wikimedia.org/wikipedia/commons/8/86/Kham_Hom_-_Sweet_Words.ogg

What would we like to know about the recording? Where is the metadata?

Read more

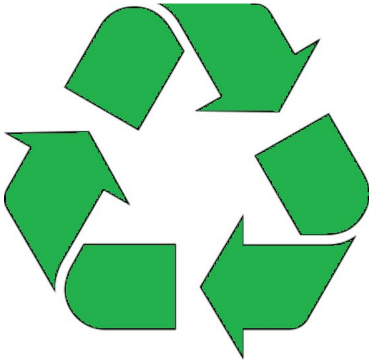
# Metadata - Exercise

- title of the song
- performer/singer (name and maybe more data - in a separate subset)
- date of recording
- length/duration (seconds)
- when was it recorded?
- where was it recorded?
- category (e.g., religious, wedding)
- original key
- language/dialect/whatever
- ... ???

Note: Details for some of these metadata fields may be unavailable for each of the items included in the collection...

# Metadata & **re-usability**

- An efficient metadata system may bring about more re-usability, also for the users from new other areas of interest, e.g. recordings made for phonetic studies may serve as useful resource for psycholinguists, sociologists, speech therapists, speech technology…;
- Proper description & documentation -> better usefulness;
- Sharing data vs. commercial uses -> awarness of legal issues is needed.



# Multi-tasking, mutual dependencies & **re-usability**

# Data acquisition

Scenarios, issues
& example solutions

- What are the possible approaches to collecting data?
- What are the crucial notions, problems/issues?
- Is it always straightforward to collect the data strictly corresponding to our needs?

# From **raw data** to a **corpus**

**Raw data**

Data collected randomly or without any strictly defined planning, not ordered.

The same set of raw data can potentially serve as a source collection for various types of analyses conducted by researchers specializing in a wide range of fields.

**Corpus**

A (usually structured) set of data which have been collected **on purpose** of studying certain phenomena occurring in the domain of interest. A linguistic corpus may be dedicated to enable studies of  language in general, dialect or other 'sub-language'.

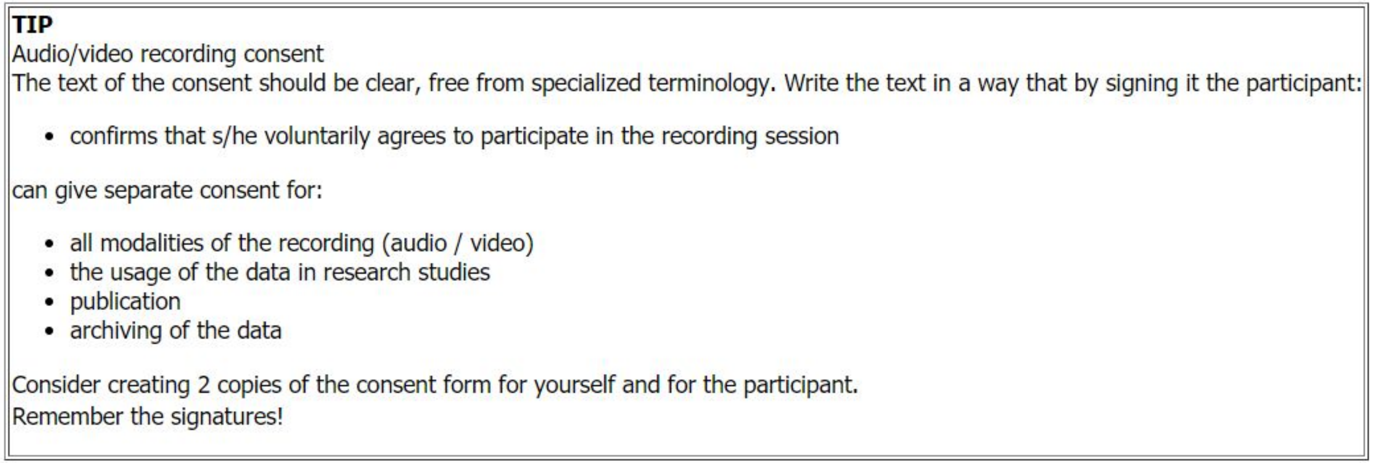
# What do we actually learn based on corpus data?

?

- How general can our conclusions be when derived from a certain type of data?
- We should always be aware that the choice of our data, the way we approach them can significantly influence the results obtained.

# Legal issues in data acquisition

In most countries it is illegal to collect data without the consent of the copyright owner(s); e.g. speech recordings, even with respect to own own conversations with third parties.



**TIP**
Audio/video recording consent
The text of the consent should be clear, free from specialized terminology. Write the text in a way that by signing it the participant:

- confirms that s/he voluntarily agrees to participate in the recording session

can give separate consent for:

- all modalities of the recording (audio / video)
- the usage of the data in research studies
- publication
- archiving of the data

Consider creating 2 copies of the consent form for yourself and for the participant.
Remember the signatures!

# Goals of the humanities

We would like to analyse true human experience, behaviour, real conversations, authentic situations. How can this be achieved? Is it possible at all?

# The Observer’s Paradox

*The aim of linguistic research in the community must be to find out how people talk when they are not being systematically observed; yet we can only obtain this data by systematic observation.*

W. Labov

https://farlane.files.wordpress.com/2007/09/observer-and-theobserved.jpg

## (Semi)spontaneous

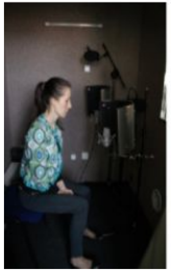
- Conversational speech, dialogues
- Narratives, story-telling
- Can be obtained in real-life situations, during fieldwork or pre-arranged conditions, also in a recording studio

## Controlled

- Read speech
  - isolated utterances
  - word-lists
  - continuous texts
- Elicited dictation
  - e.g. speech (re)produced based on stimuli presented to speakers immediately before recording
- Often collected in studio

# Quality

- **Studio** recordings vs. **fieldwork**
- High quality vs. spontaneous, “natural” character



# Scenarios example: emotion portrayals

How to investigate emotions in speech?

- Polish *Paralingua* corpus -> Emotion portrayals, por. GEMEP Geneva corpus (Scherer et al.)
- Lexically neutral utterances produced with varying emotional load
- Basic emotions characterized by varying valence & activation

Intended anger, Intended joy

Teraz wszystko rozumiem.

Dzisiaj jest poniedziałek.

Od rana pada deszcz.

Powiedział, że nic się nie stało.

Jedziemy na wycieczkę do Grecji.

# Scenarios example: authentic motherese

What are the specific features of speech directed to infants in comparison with adult directed speech?

- Mothers talking to their babies (below 1 year-old)
- Home recording conditions, during everyday activities, feeding the infants, dressing up, playing with them, etc…
- Subsequently: studio recordings, but….!

# Scenarios example: task-oriented dialogues

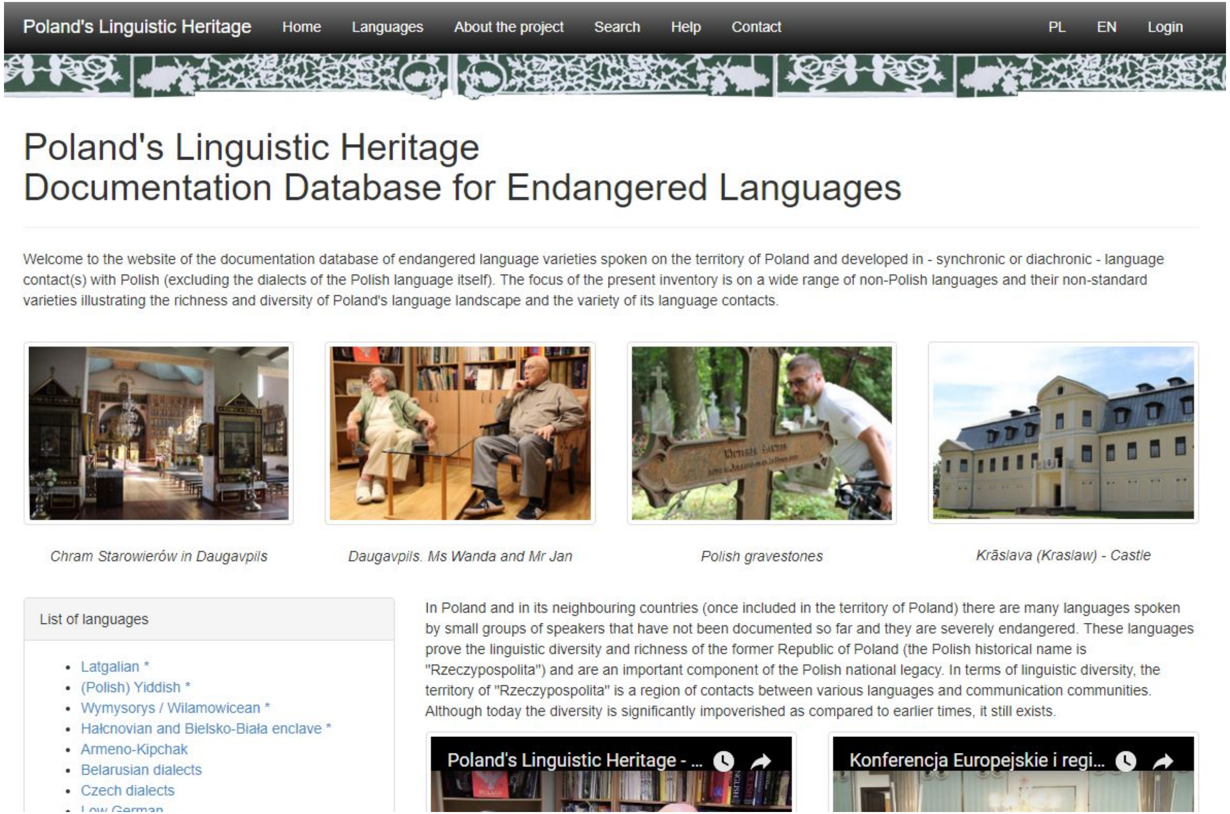
*Borderland (http://borderland.amu.edu.pl/ ) a* corpus of Polish & German speech:

- cooperative / competitive dialogues: let’s built an imaginative tower!
- let’s decide about a gift for a person (based on varying input info)!
- The focus of the speaker is **on the task rather than being recorded,** at least that’s what is intended -> let the participants forget about the artificial conditions...

# Another perspective - language archives

- A specific type of data collections.
- Unlike language corpora it is often characterized by **highly diversified** and **not always well-balanced** contents. The reason for this is that the contents frequently depends on availability.
- For example, in case of a repository for endangered languages, the size and complexity of the collections will be conditioned by the number of speakers, source texts available, as well as the cultural or social limitations -> some communities might be reluctant to share their heritage.

# Language archives - www.inne-jezyki.amu.edu.pl



## Poland's Linguistic Heritage
## Documentation Database for Endangered Languages

Welcome to the website of the documentation database of endangered language varieties spoken on the territory of Poland and developed in - synchronic or diachronic - language contact(s) with Polish (excluding the dialects of the Polish language itself). The focus of the present inventory is on a wide range of non-Polish languages and their non-standard varieties illustrating the richness and diversity of Poland's language landscape and the variety of its language contacts.

*Chram Starowierów in Daugavpils*

*Daugavpils. Ms Wanda and Mr Jan*

*Polish gravestones*

*Krāslava (Kraslaw) - Castle*

List of languages

- Latgalian *
- (Polish) Yiddish *
- Wymysorys / Wilamowicean *
- Hałcnovian and Bielsko-Biała enclave *
- Armeno-Kipchak
- Belarusian dialects
- Czech dialects
- Low German

In Poland and in its neighbouring countries (once included in the territory of Poland) there are many languages spoken by small groups of speakers that have not been documented so far and they are severely endangered. These languages prove the linguistic diversity and richness of the former Republic of Poland (the Polish historical name is "Rzeczypospolita") and are an important component of the Polish national legacy. In terms of linguistic diversity, the territory of "Rzeczypospolita" is a region of contacts between various languages and communication communities. Although today the diversity is significantly impoverished as compared to earlier times, it still exists.

Poland's Linguistic Heritage - ...
Konferencja Europejskie i regi...

# Data management

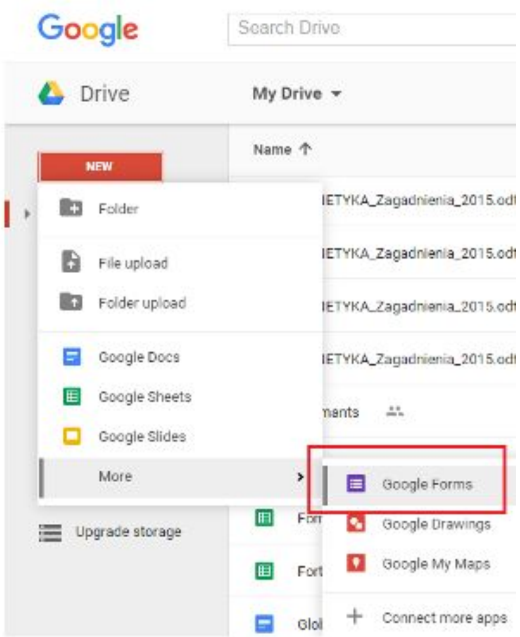
Selected tools & approaches

Crucial role of management for teamwork & collaboration.

Careful planning of data management. When is it a useful addition or a must?

The answer depends on data type, data sizes, and the number of users.

# From file collections to relational databases



Data can be organized in various ways

- Files
- Folders
- File / folder collections, calculation spreadsheets
- Relational databases

# Is it just a useful addition or a **MUST**?

- 1 person, 1 text
- 1 person, more data
- Many people, moderate amount of data
- Many people, lots of data
- Several / many people, remote access to data

# Is it just a useful addition or a **MUST**?

- 1 person, 1 text - not necessary
- 1 person, more data - useful
- Many people, moderate amount of data - useful
- Many people, lots of data - **MUST**
- Several / many people, remote access to data - useful or **MUST**

# From file collections to relational databases

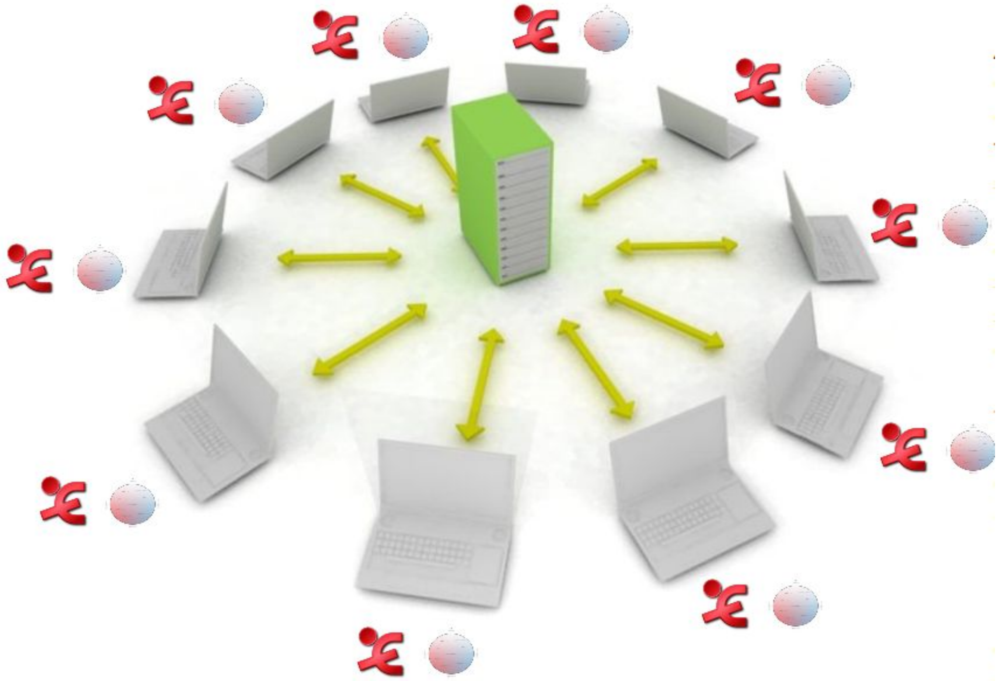
Some general purpose tools available for data management purposes (e.g., git-based version control systems).

Some management options included in annotation & analysis tools:

- file collection management,
- workspaces etc.

However: for **large corpora & simultaneous usage** by many people **more robust tools are needed** such as dedicated data & workflow management software + relational database solutions.

# Client-server architectures: Collaboration support



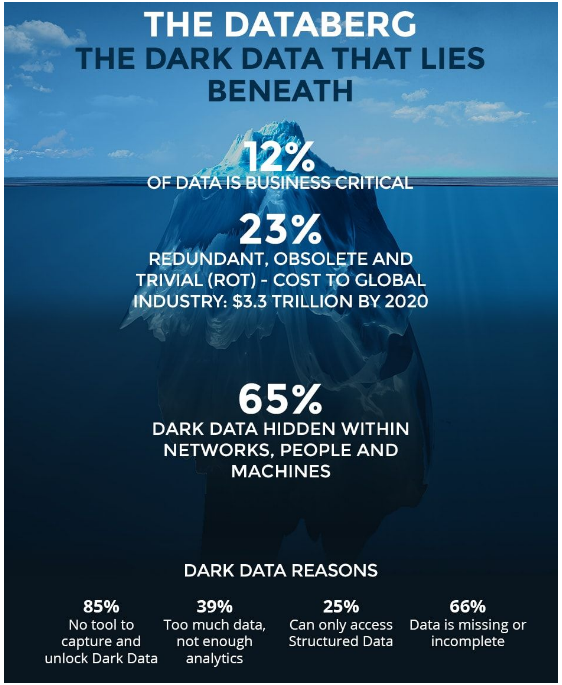
## SQL

/ˈɛs kjuː ˈɛl/ or/ˈsiːkwəl/,
Structured Query Language

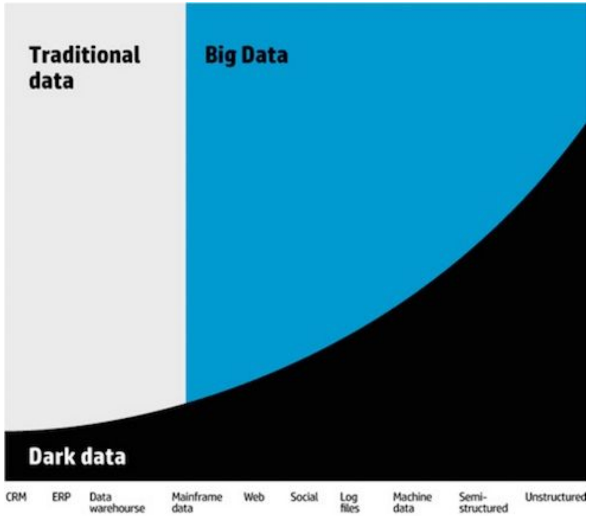
# Data analysis

- Can we be sure what’s in the data?
- Should we be afraid of “dark data”?
- Interoperability of data formats and tools
- Selected aspects involved in speech data analysis
- Tools & approaches

# Data mining: what can we find there?

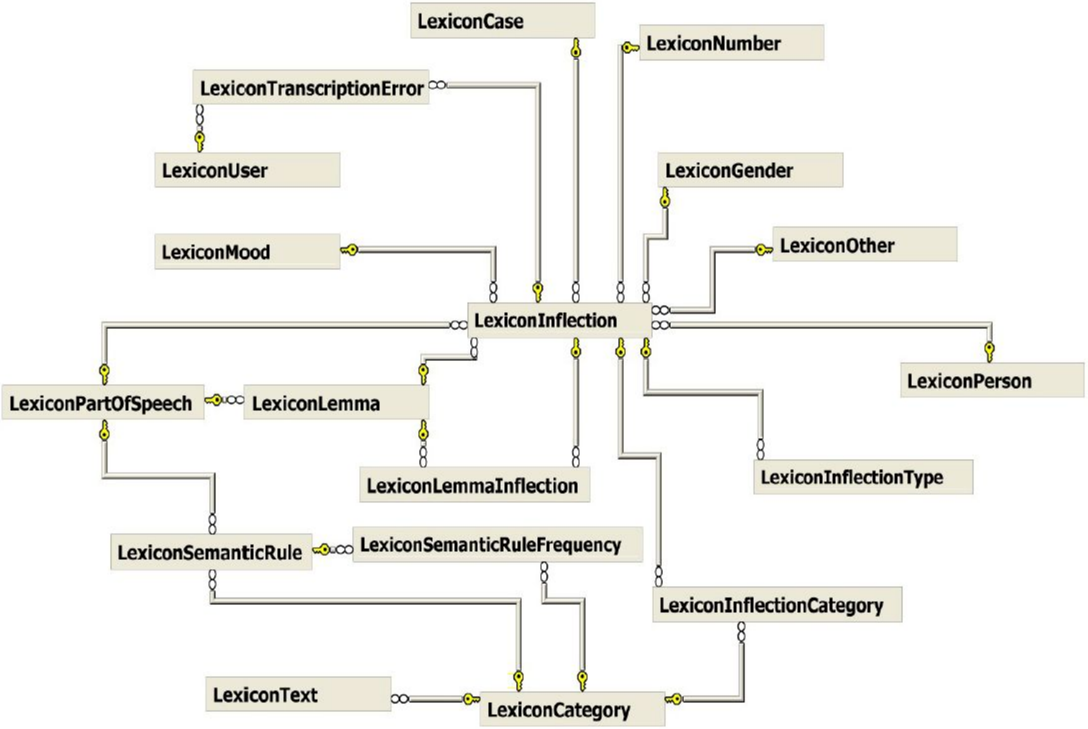

THE DATABERG
THE DARK DATA THAT LIES BENEATH
12%
OF DATA IS BUSINESS CRITICAL
23%
REDUNDANT, OBSOLETE AND TRIVIAL (ROT) - COST TO GLOBAL INDUSTRY: $3.3 TRILLION BY 2020
65%
DARK DATA HIDDEN WITHIN NETWORKS, PEOPLE AND MACHINES
DARK DATA REASONS
85%
No tool to capture and unlock Dark Data
39%
Too much data, not enough analytics
25%
Can only access Structured Data
66%
Data is missing or incomplete

Picture: https://datumize.com/evolution-dark-data/



Picture:
http://www.kdnuggets.com/2015/11/importance-dark-data-big-data-world.html

# Data mining: what can we find there?

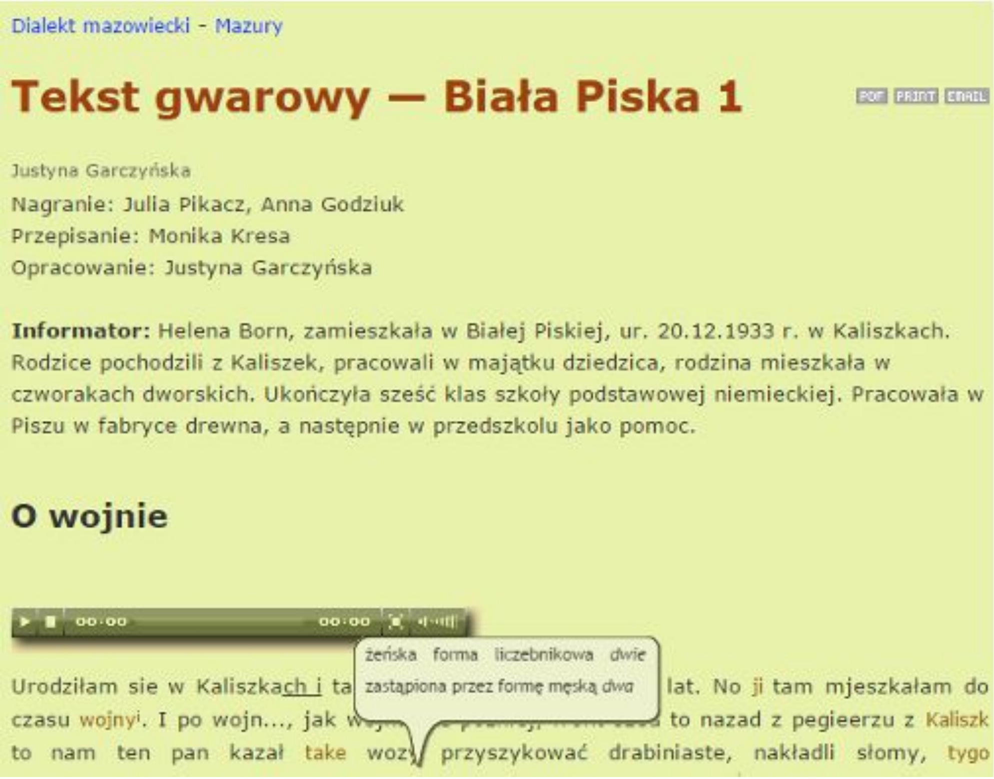


- Discover relationships, dependencies, patterns, rules
- In order to do that we often need to combine multiple types of information within one workspace

# Data formats & tools **interoperability**

- International standards for metadata formats: IMDI, Dublin Core, OLAC
- Best practices for file formats, e.g., XML-based file formats are very popular in text and speech annotation tools
- It is always worthwhile to check to import/export options available in the tools we plan to utilize: the more the better -> towards better interoperability

# Annotation, transcription, time-alignment...



Dialekt mazowiecki - Mazury

## Tekst gwarowy — Biała Piska 1

Justyna Garczyńska

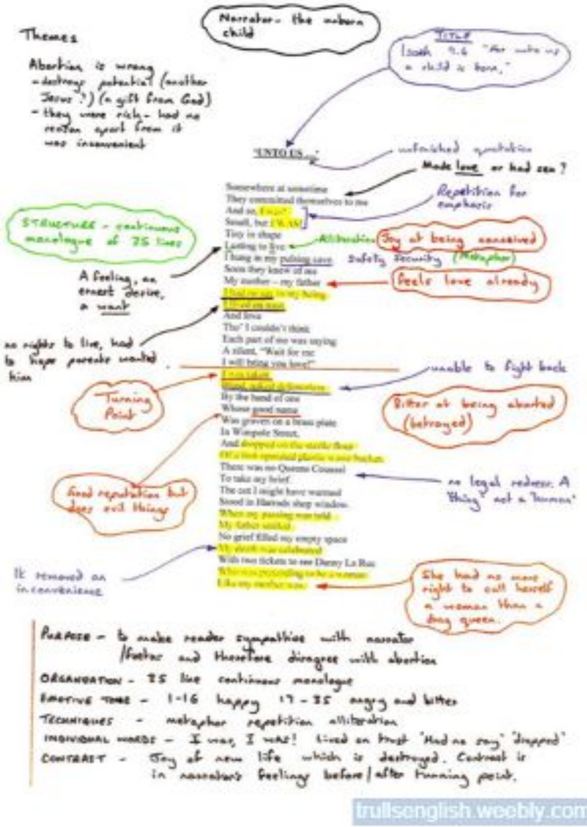
Nagranie: Julia Pikacz, Anna Godziuk
Przepisanie: Monika Kresa
Opracowanie: Justyna Garczyńska

**Informator:** Helena Born, zamieszkała w Białej Piskiej, ur. 20.12.1933 r. w Kaliszkach. Rodzice pochodzili z Kaliszek, pracowali w majątku dziedzica, rodzina mieszkała w czworakach dworskich. Ukończyła sześć klas szkoły podstawowej niemieckiej. Pracowała w Piszu w fabryce drewna, a następnie w przedszkolu jako pomoc.

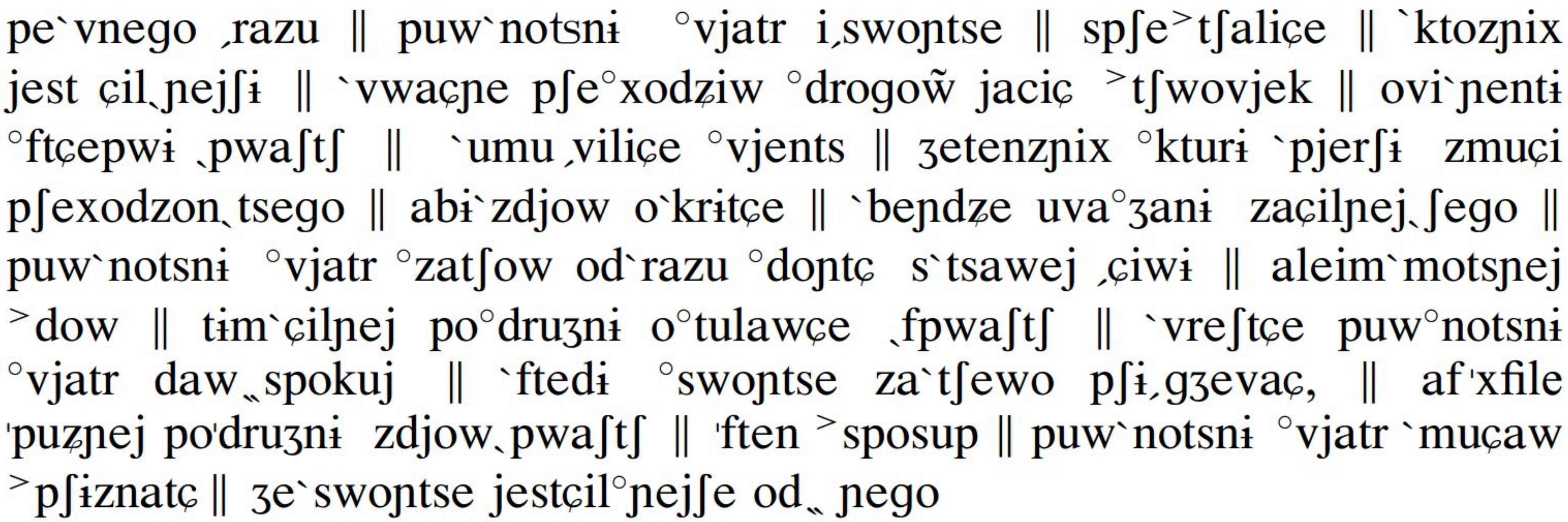
### O wojnie

żeńska forma liczebnikowa dwie zastąpiona przez formę męską dwa

Urodziłam sie w Kaliszkach i ta lat. No ji tam mjeszkałam do czasu wojny[i]. I po wojn..., jak w to nazad z pegieerzu z Kaliszk to nam ten pan kazał take wozy przyszykować drabiniaste, nakładli słomy, tygo

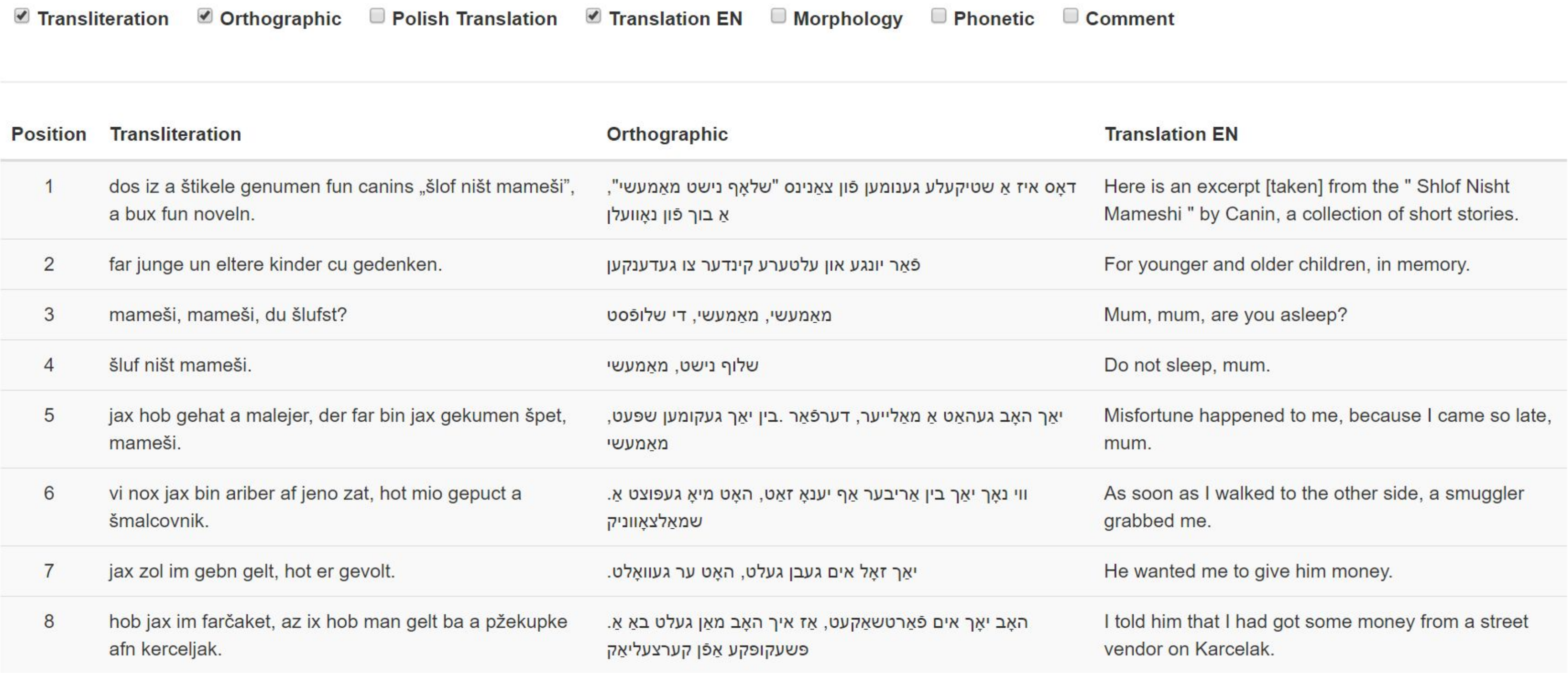

Themes
Abortion is wrong
- destroys potential (another Jesus?) (a gift from God)
- they were rich - had no reason apart from it was inconvenient
Narrator - the unborn child
TITLE
Isaiah 9.6 "For unto us a child is born,"
"UNTO US..."
unfinished quotation
Made love or had sex?
Repetition for emphasis
STRUCTURE - continuous monologue of 35 lines
Alliteration
Joy at being conceived
Safety Security (Metaphor)
Feels love already
A feeling, an ernest desire, a want
no rights to live, had to hope parents wanted him
unable to fight back
Turning Point
Bitter at being aborted (betrayed)
Good reputation but does evil things
no legal redress. A 'thing' not a 'human'
It removed an inconvenience
She had no more right to call herself a woman than a drag queen.
Somewhere at sometime
They committed themselves to me
And so, I was!
Small, but I WAS!
Tiny in shape
Lasting to live
I hung in my pulsing cave
Soon they knew of me
My mother – my father
And love
Tho' I couldn't think
Each part of me was saying
A silent, "Wait for me
I will bring you love!"
By the hand of one
Whose good name
Was graven on a brass plate
In Wimpole Street,
There was no Queens Counsel
To take my brief
The cot I might have warmed
Stood in Harrods shop window.
No grief filled my empty space
With two tickets to see Danny La Rue
PURPOSE - to make reader sympathise with narrator /foetus and therefore disagree with abortion
ORGANISATION - 35 line continuous monologue
EMOTIVE TONE - 1-16 happy 17-35 angry and bitter
TECHNIQUES - metaphor repetition alliteration
INDIVIDUAL WORDS - I was, I WAS! Lived on trust 'Had no say' 'dropped'
CONTRAST - Joy of new life which is destroyed. Contrast is in narrator's feeling before/after turning point.
trullsenglish.weebly.com

# Annotation, transcription, time-alignment



pe`vnego ˏrazu ‖ puw`not͡snɨ °vjatr iˏswoɲtse ‖ spʃe˃tʃaliɕe ‖ `ktozɲix
jest ɕilˎɲejʃɨ ‖ `vwaɕɲe pʃe°xodʑiw °drogow̃ jaciɕ ˃tʃwovjek ‖ ovi`ɲentɨ
°ftɕepwɨ ˎpwaʃtʃ ‖ `umuˏviliɕe °vjents ‖ ʒetenzɲix °kturɨ `pjerʃɨ zmuɕi
pʃexodzonˎtsego ‖ abɨ`zdjow o`krɨtɕe ‖ `beɲdʑe uva°ʒanɨ zaɕilɲejˎʃego ‖
puw`notsnɨ °vjatr °zatʃow od`razu °doɲtɕ s`tsawej ˏɕiwɨ ‖ aleim`motsɲej
˃dow ‖ tɨm`ɕilɲej po°druʒnɨ o°tulawɕe ˎfpwaʃtʃ ‖ `vreʃtɕe puw°notsnɨ
°vjatr daw˵spokuj ‖ `ftedɨ °swoɲtse za`tʃewo pʃɨˏgʒevaɕ, ‖ afˈxfile
ˈpuʑɲej poˈdruʒnɨ zdjowˎpwaʃtʃ ‖ ˈften ˃sposup ‖ puw`notsnɨ °vjatr `muɕaw
˃pʃɨznatɕ ‖ ʒe`swoɲtse jestɕil°ɲejʃe od˵ɲego

# Annotation, transcription, time-alignment

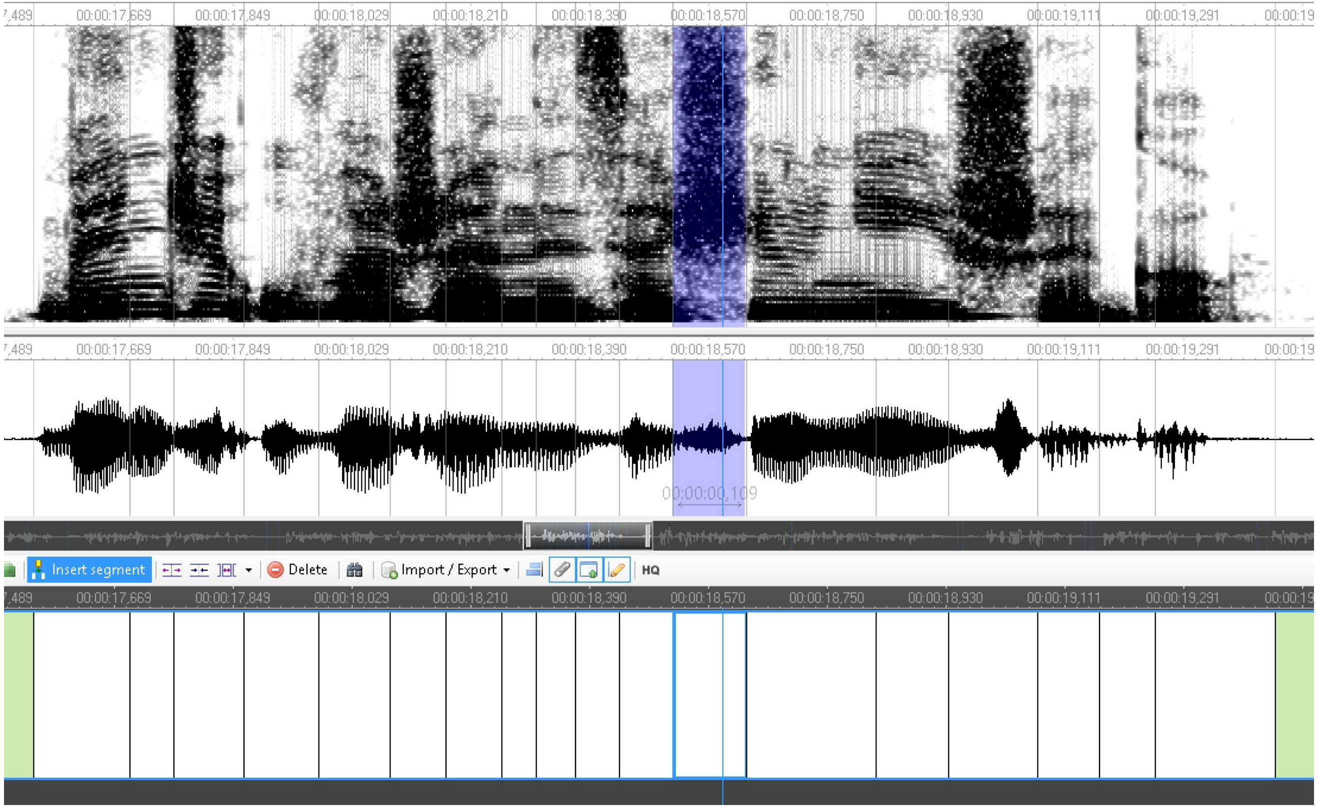


☑ Transliteration ☑ Orthographic □ Polish Translation ☑ Translation EN □ Morphology □ Phonetic □ Comment

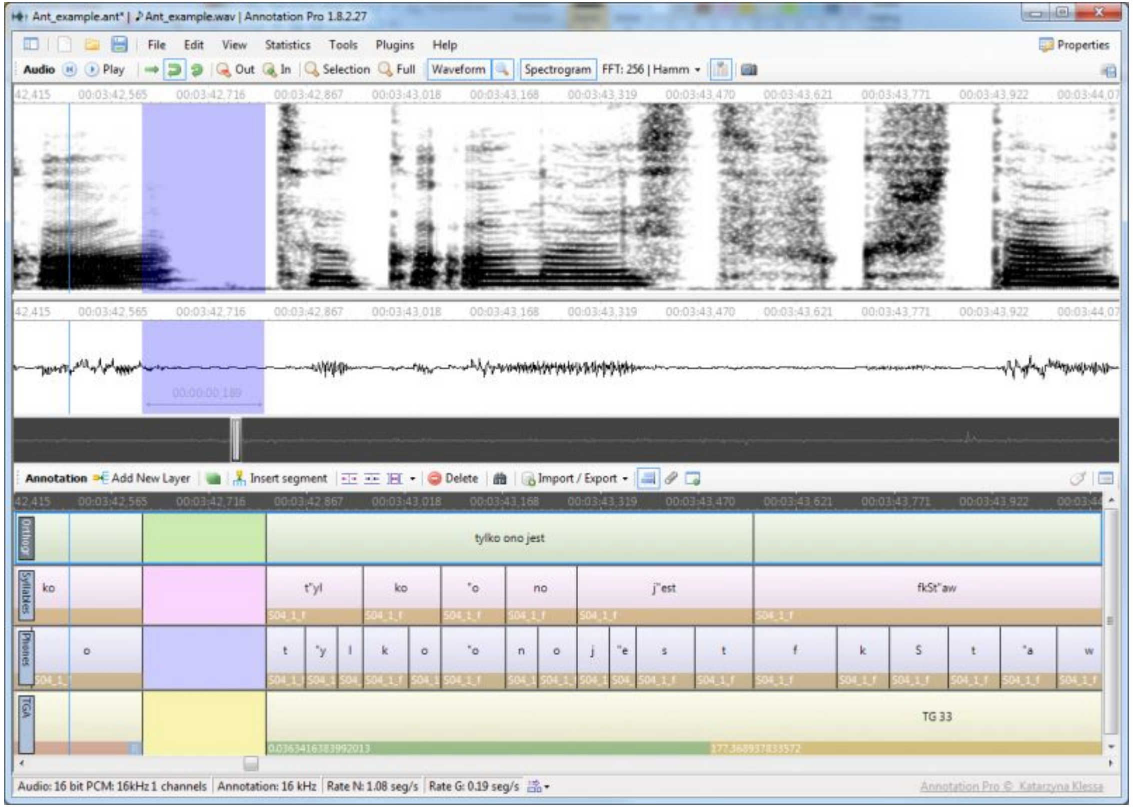
| Position | Transliteration | Orthographic | Translation EN |
|---|---|---|---|
| 1 | dos iz a štikele genumen fun canins „šlof ništ mameši”, a bux fun noveln. | דאָס איז אַ שטיקעלע גענומען פֿון צאַנינס "שלאָף נישט מאַמעשי", אַ בוך פֿון נאָוועלן | Here is an excerpt [taken] from the " Shlof Nisht Mameshi " by Canin, a collection of short stories. |
| 2 | far junge un eltere kinder cu gedenken. | פֿאַר יונגע און עלטערע קינדער צו געדענקען | For younger and older children, in memory. |
| 3 | mameši, mameši, du šlufst? | מאַמעשי, מאַמעשי, די שלופֿסט | Mum, mum, are you asleep? |
| 4 | šluf ništ mameši. | שלוף נישט, מאַמעשי | Do not sleep, mum. |
| 5 | jax hob gehat a malejer, der far bin jax gekumen špet, mameši. | יאַך האָב געהאַט אַ מאַלייער, דערפֿאַר .בין יאַך געקומען שפעט, מאַמעשי | Misfortune happened to me, because I came so late, mum. |
| 6 | vi nox jax bin ariber af jeno zat, hot mio gepuct a šmalcovnik. | ווי נאָך יאַך בין אַריבער אַף יענאָ זאַט, האָט מיאָ געפּוצט אַ. שמאַלצאָווניק | As soon as I walked to the other side, a smuggler grabbed me. |
| 7 | jax zol im gebn gelt, hot er gevolt. | יאַך זאָל אים געבן געלט, האָט ער געוואָלט. | He wanted me to give him money. |
| 8 | hob jax im farčaket, az ix hob man gelt ba a pžekupke afn kerceljak. | האָב יאָך אים פֿאַרטשאַקעט, אַז איך האָב מאַן געלט באַ אַ. פשעקופקע אַפֿן קערצעליאַק | I told him that I had got some money from a street vendor on Karcelak. |

http://inne-jezyki.amu.edu.pl/Frontend/Text/Details/115

# Annotation, transcription, time-alignment



# Multilayer annotation of speech, gesture & text

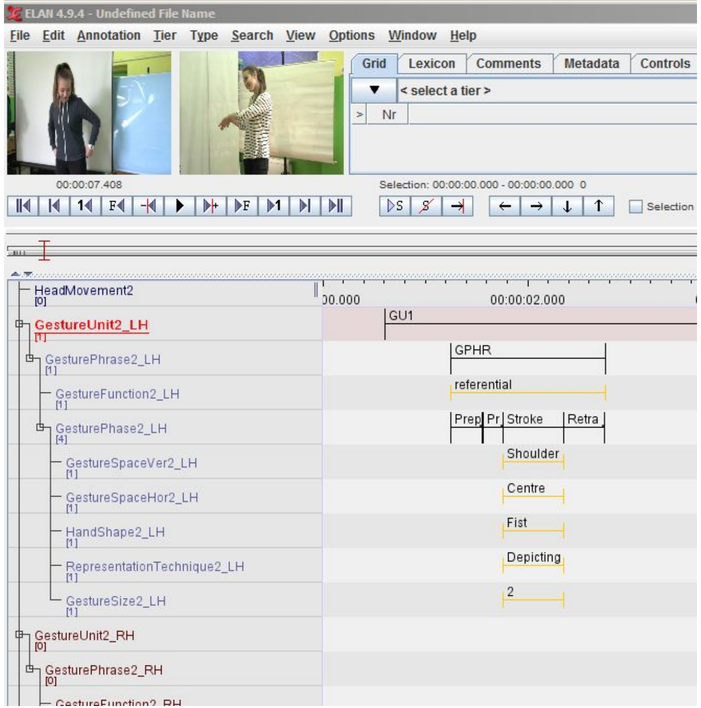


**Annotation Pro:**

- Speech annotation
- Annotation mining
- Perception tests

www.annotationpro.org

# Multilayer annotation of speech, gesture & text

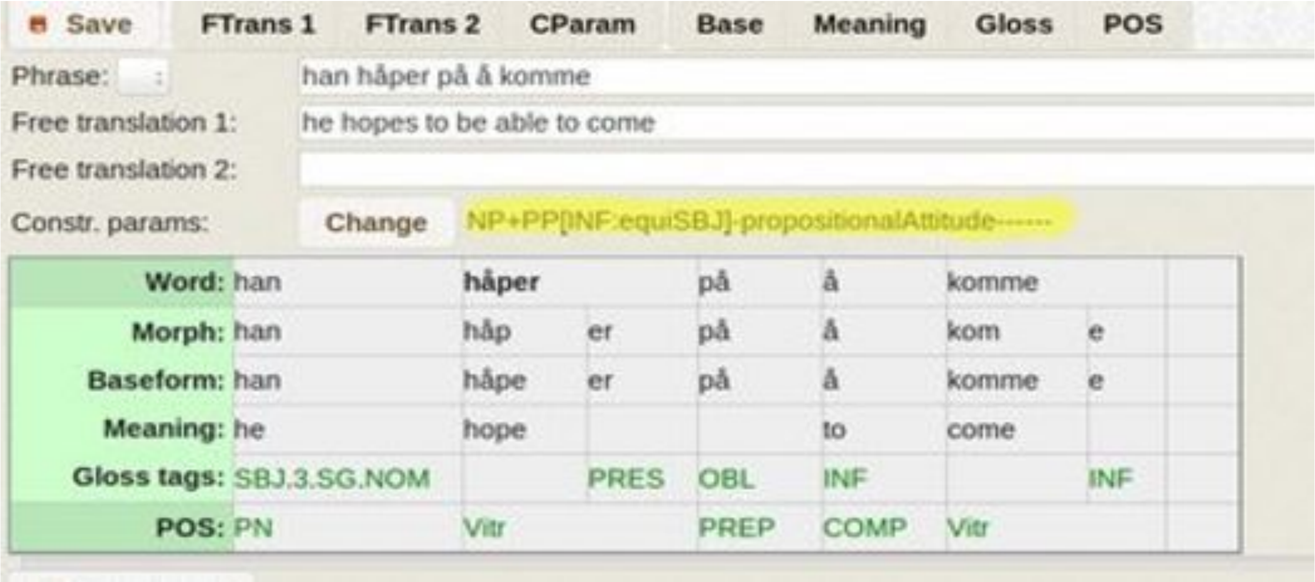

ELAN 4.9.4 - Undefined File Name
File Edit Annotation Tier Type Search View Options Window Help
Grid Lexicon Comments Metadata Controls
< select a tier >
Nr
00:00:07.408
Selection: 00:00:00.000 - 00:00:00.000 0
Selection
00.000
00:00:02.000

| Tier | Annotation |
|---|---|
| HeadMovement2 [0] | |
| GestureUnit2_LH [1] | GU1 |
| GesturePhrase2_LH [1] | GPHR |
| GestureFunction2_LH [1] | referential |
| GesturePhase2_LH [4] | Prep \| Pr \| Stroke \| Retra |
| GestureSpaceVer2_LH [1] | Shoulder |
| GestureSpaceHor2_LH [1] | Centre |
| HandShape2_LH [1] | Fist |
| RepresentationTechnique2_LH [1] | Depicting |
| GestureSize2_LH [1] | 2 |
| GestureUnit2_RH [0] | |
| GesturePhrase2_RH [0] | |
| GestureFunction2_RH | |

ELAN https://tla.mpi.nl/tools/tla-tools/elan/ image: *Borderland* project

Image: *Borderland* project, M. Karpiński

# Multilayer annotation of speech, gesture & text



Phrase: han håper på å komme

Free translation 1: he hopes to be able to come

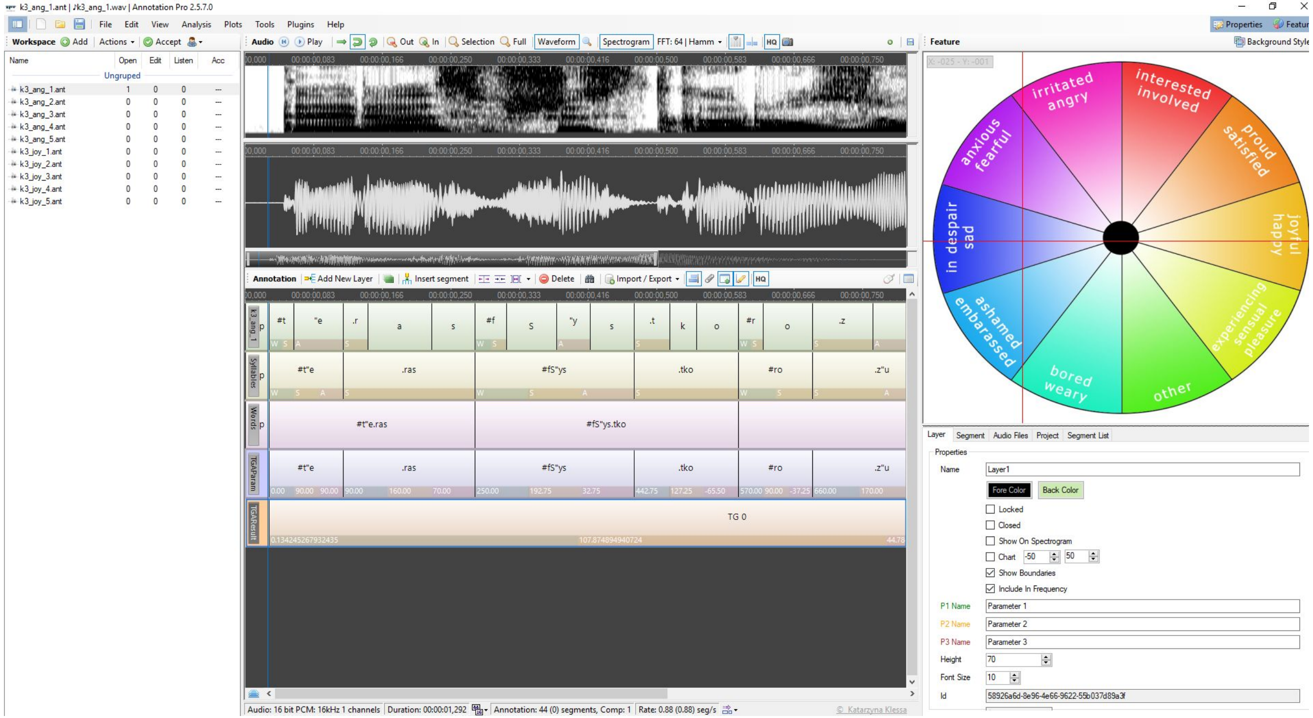
Free translation 2:

Constr. params: NP+PP[INF:equiSBJ]-propositionalAttitude-------

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Word: | han | håper | | på | å | komme | |
| Morph: | han | håp | er | på | å | kom | e |
| Baseform: | han | håpe | er | på | å | komme | e |
| Meaning: | he | hope | | | to | come | |
| Gloss tags: | SBJ.3.SG.NOM | | PRES | OBL | INF | | INF |
| POS: | PN | Vitr | | PREP | COMP | Vitr | |

TypeCraft annotation & text glossing tool https://typecraft.org/

# Annotation data export, import & processing



k3_ang_1.ant | ♪k3_ang_1.wav | Annotation Pro 2.5.7.0

| Name | Open | Edit | Listen | Acc |
|---|---|---|---|---|
| Ungruped | | | | |
| k3_ang_1.ant | 1 | 0 | 0 | --- |
| k3_ang_2.ant | 0 | 0 | 0 | --- |
| k3_ang_3.ant | 0 | 0 | 0 | --- |
| k3_ang_4.ant | 0 | 0 | 0 | --- |
| k3_ang_5.ant | 0 | 0 | 0 | --- |
| k3_joy_1.ant | 0 | 0 | 0 | --- |
| k3_joy_2.ant | 0 | 0 | 0 | --- |
| k3_joy_3.ant | 0 | 0 | 0 | --- |
| k3_joy_4.ant | 0 | 0 | 0 | --- |
| k3_joy_5.ant | 0 | 0 | 0 | --- |

Properties

Name: Layer1

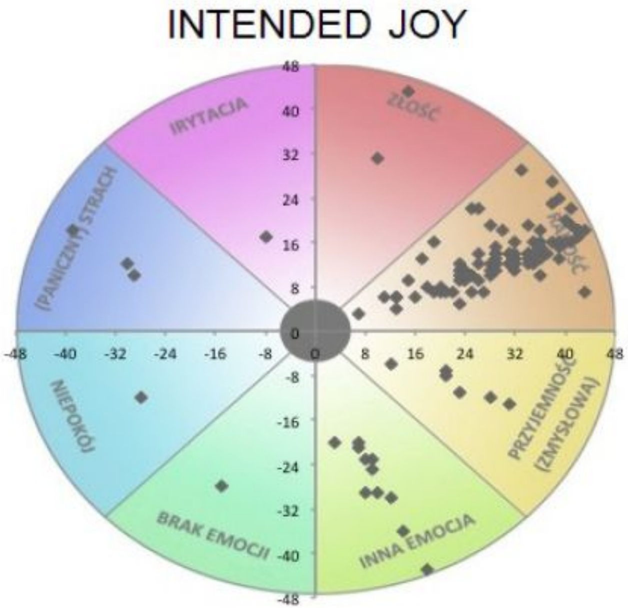
□ Locked
□ Closed
□ Show On Spectrogram
□ Chart -50 50
☑ Show Boundaries
☑ Include In Frequency

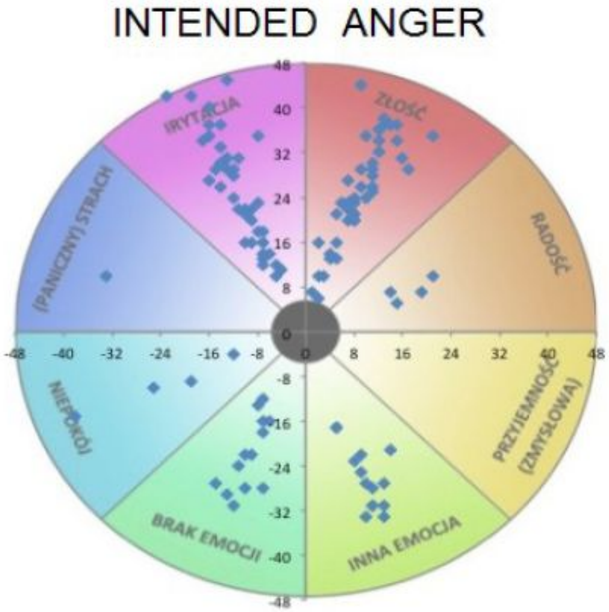
P1 Name: Parameter 1
P2 Name: Parameter 2
P3 Name: Parameter 3
Height: 70
Font Size: 10
Id: 58926a6d-8e96-4e66-9622-55b037d89a3f

Audio: 16 bit PCM: 16kHz 1 channels | Duration: 00:00:01,292 | Annotation: 44 (0) segments, Comp: 1 | Rate: 0.88 (0.88) seg/s

© Katarzyna Klessa

# Approaching emotions & affective states



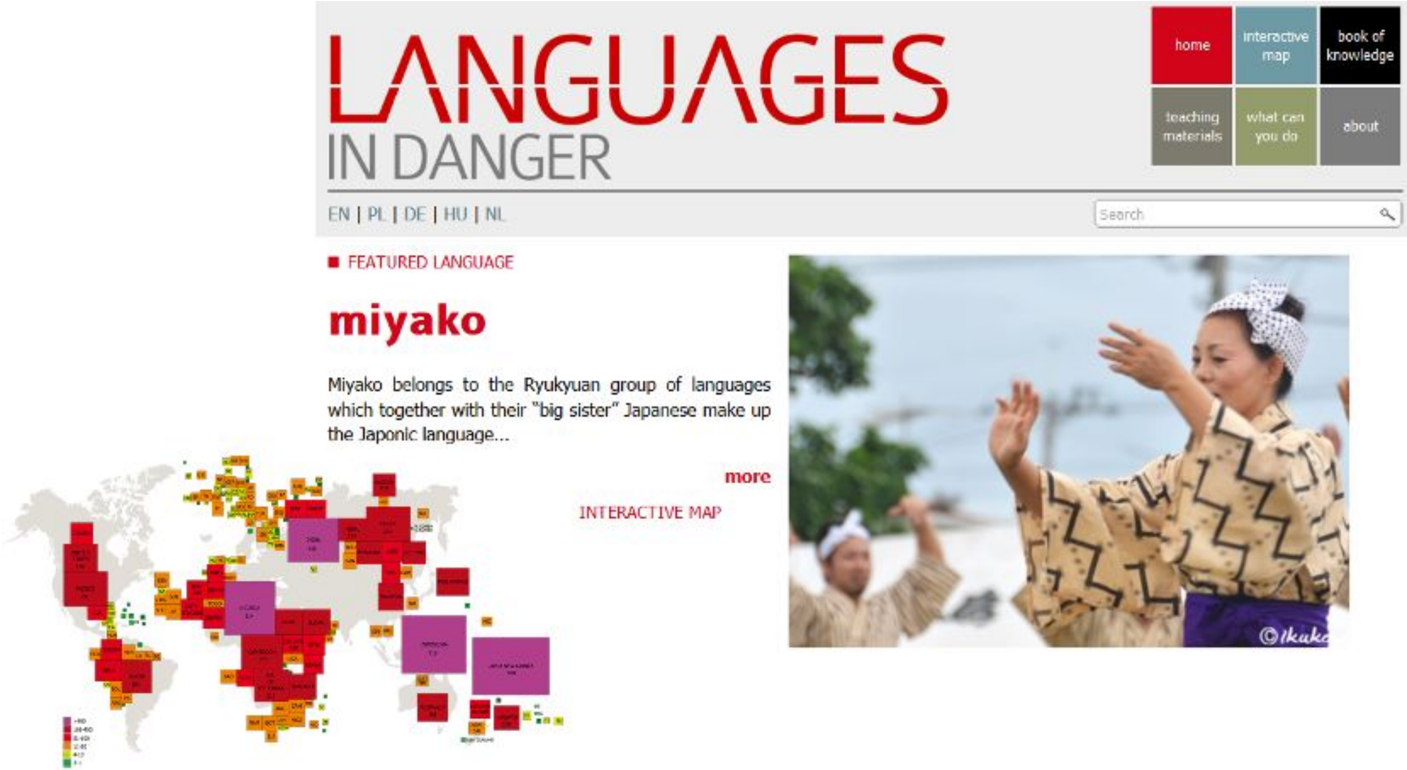


# Summary

- **Data & metadata** collection, management & analysis are subsequent steps in the process of dealing with resources but only to a certain extent.
- The three steps are mutually related, and sometimes they **overlap** and need to be re-defined depending on a particular application.
- According to the contemporary best practices for digital humanities, data collections need to be **re-usable**, and should be stored with the use of data formats enabling **interoperability**.

# MA Studies @ AMU Poznań: www.elldo.amu.edu.pl

# www.languagesindanger.eu



LANGUAGES

IN DANGER

FEATURED LANGUAGE

## miyako

Miyako belongs to the Ryukyuan group of languages which together with their "big sister" Japanese make up the Japonic language...

more

INTERACTIVE MAP

# Thank you!

Katarzyna Klessa

E-mail: klessa@amu.edu.pl

More about me:

www.katarzyna.klessa.pl

# Acknowledgements

The presentation refers to the work conducted within the following projects and funded by the respective grants:

- *NeuroPerKog: Development of the phonematic hearing and working memory in infants and children* funded by National Science Centre, Poland for years 2013-2018.
- *Language of Boundaries and Boundaries of Language* funded by the Polish Ministry of Science and Higher Education within "The National Programme for the Development of Humanities" in the years 2014-2016.
- *Poland's Linguistic Heritage* funded by the Polish Ministry of Science and Higher Education within "The National Programme for the Development of Humanities" (grant No. 11H 11 001480) in the years 2012-2014.
- *INNET (Innovative networking in infrastructure for endangered languages)*, European FP7 grant agreement no: 284415, in the years 2011-2014.