

Privacy notice for the use of the Lahjoita puhetta dataset at the MaiNLP lab (LMU Munich)

Document creation date and last modification date

Created and last modified on: 03 July 2025

Name and contact details of the data controller

Ludwig-Maximilians-Universität München (LMU Munich)
Geschwister-Scholl-Platz 1
80539 Munich
Germany

Name of the principal investigator or contact person responsible for the project

- Principal investigator: Barbara Plank
- Contact person: Verena Blaschke
- {b.plank, verena.blaschke} AT lmu DOT de

Purpose of processing personal data

Our purpose for using the dataset is to analyze how good current language/speech processing systems are at processing written vs. spoken and standard-like vs. dialectal language in the context of classification tasks. To this end, we want to use the metadata about the utterance theme/topic (to use topic classification as the language/speech processing task), the metadata of the speakers' dialects (to analyze the effect of the speaker dialect). We may also analyze gender and age in case the speech processing models' performances are sensitive to them.

Legal basis for processing personal data

The legal basis is scientific research at a public university carried out in the public interest involving the re-use of scientific data (Art. 6(1)(e), Art. 6(2, 3) General Data Protection Regulation (GDPR) read in conjunction with Art. 4(1) Bavarian Data Protection Act (BayDSG), Art. 2 Bavarian University Innovation Act (BayHIG), Art. 6(2)(3)(c) BayDSG, Art. 89 GDPR read in conjunction with Art. 25 BayDGS, Art. 5(1) GDPR).

Description of the personal data processed

Resource: Donate Speech Corpus, version 1.0,
<https://clarino.uib.no/comedi/editor/lb-2020090321>, obtained from the Language Bank of Finland (Kielipankki).

Data to be processed:

- Audio recordings
- Transcriptions
- Metadata/annotations: theme/topic, dialect, gender, age

Recipients / Disclosures of data

The data will not be processed outside LMU Munich.

Transferring data outside the EEA

The data will not be transferred outside the EEA.

Protection of personal data

We will store the data on a university server with access restricted to project members.

Retention period of personal data

03 July 2025 (or as soon thereafter as we receive the data) to 31.01.2026.

How data subjects can exercise their rights in relation to personal data

According to the General Data Protection Regulation (GDPR), a data subject has the right to

- obtain information on the processing of their personal data
- access their data
- rectification of inaccurate personal data
- request that the processing of personal data be restricted
- erasure of personal data
- object the processing of personal data.

See also the [Finnish Data Protection Ombudsman's website](#).

If data subjects have questions about their own personal data relating to resources made available via the Language Bank, their enquiries should be addressed to the Language Bank. If necessary, the Language Bank can then forward the questions to the data controller of the original resource.