DATA MANAGEMENT PLAN
SFB 1287 / 2025 / PHASE 2

PROJECT C10

Table of contents

General Information

Overview

Project number

C10

Name of Experiment / Acronym / Number

Longitudinalkorpus Eltern-Kind-Interaktion

Involved persons

Martin Pfeiffer, Aaron Schmidt-Riese, Anton Köring, Mars Mentzel, Jan Ragwitz, Niklas Grüninger

PI or responsible person (head of the study)

Prof. Dr. Martin Pfeiffer

Subject area

German Linguistics, Developmental Sociolinguistics, Language Acquisition, Corpus Linguistics

Method / Type of data

Corpus of natural interactions between parents and their children (aged approx. 1;6-4;6)

Participants (of the study)

3 parents with their children, living in the High Alemannic region in the very Southwest of Germany (Federal State of Baden-Württemberg)

Short description (of the study)

The project explores the limits of phonological dialect-standard variability in young children who acquire two varieties of German, a High Alemannic dialect and standard German. Drawing on LEKI, a longitudinal audiovisual corpus of spontaneous family interaction involving three children regularly recorded within the age span of approx. 1;6 to 4;6, we aim to uncover the language-internal and language-external factors that constrain this variability and the way it develops over time. To this end, we combine methods of interactional linguistics and corpus linguistics.

Comments (optional)

-

Data Management Requirements

Are there requirements regarding the data management from your scholarly / scientific community?

yes

If yes, what are the requirements?

- *DFG Guidelines on the Handling of Research Data*

- *Data Management in Psychological Science*

- *„Handlungsempfehlungen zum Umgang mit Forschungsdaten" University of Potsdam*

- *„Technische und organisatorische Maßnahmen" (TOM) gemäß Art. 32 Abs. 1 DSGVO.*

Financial Support

Who is funding the project?

DFG - Deutsche Forschungsgemeinschaft e.V. (German Research Foundation) - https://www.dfg.de/en/

In which funding line and / or which funding program is the project funded?

Collaborative Research Centre 1287 - Project number 317633480

Dataset Information

Data Origin

Is the dataset being created or re-used?

created

If re-used, who created the dataset and under which address, PID or URL is the data set available?

-

Data Collection

When does data collection start? (approximately / tentatively)

-

When does data collection end? (approximately / tentatively)

-

Data Handling

Where is the dataset stored during the project?

We developed our own data management server and implemented the moca3 (multimodal oral corpus administration 3) at the University of Potsdam.

If data is stored on lab or personal computers, please describe the backup strategy.

All video recordings, audio recordings, and transcripts of the LEKI corpus are stored on two different external storage systems on different locations.

Which file formats are used?

.mp3, .mp4, .wav, .eaf (ELAN, a linguistic transcription computer program), .docx, .xlsx, .csv, .R

Which measures of quality assurance are taken for this dataset?

-

Data Analysis

When does data analysis start? (approximately / tentatively)

-

When does data analysis end? (approximately / tentatively)

-

Data Reuse

Which individuals, groups or institutions could be interested in re-using this dataset? What consequences does the reuse potential have for the provision of the data later?

-

Legal and Ethics

Personal Data

Does this dataset contain personal data?

-

Are these data anonymised?

-

Property Rights

Does the project use and/or produce data that is protected by intellectual or industrial property rights?

-

If yes, please explain which data protected by intellectual or industrial property rights?

-

Publication

Publishing or Sharing Data

Will this dataset be published or shared?

Internal only

If yes, the principal investigator of the study ensured that the consent form / subject information sheets support publishing of the data?

-

If yes, under which terms of use or license will the dataset be published or shared?

-

If yes, when will the data be published?

-

If no, please explain why not. Please differentiate between legal and contractual reasons and voluntary restrictions.

-

Storage and Long-Term Preservation

Archive

Does this dataset have to be preserved for long-term?

yes

How long does the data need to be stored?

The DFG expects primary data that is the basis of a publication to be stored in the researcher's own institution or an appropriate nationwide infrastructure long-term (for at least 10 years).

What are the reasons this dataset must be preserved for the long-term?

- *Use in a publication / Evidence of good scientific practice*

- *Reuse (if anonymizable data) in subsequent projects or by others*

- *Legal obligations*

- *Documentation because it is socially relevant*

- *Self-commitment*

- *Evidence of good scientific practice*

- *DFG requirements*

Where will the data (including metadata, documentation, and relevant code) be stored or archived after the end of the project?

- *http://moca3.corpora.uni-potsdam.de/moca3/*

coordinated by:

funded by:

DFG Deutsche
Forschungsgemeinschaft

in cooperation with:

RUHR
UNIVERSITÄT
BOCHUM

RUB