



**SFB
1287**

Limits of Variability in Language
Cognitive, Grammatical, and Social Aspects

DATA MANAGEMENT PLAN

CRC 1287 | 2021 | Phase 1

Project A03

.....

TABLE OF CONTENTS

General Information	3
Overview	3
Data Management Requirements	4
Financial Support	4
Dataset Information	6
Data Origin	6
Data Collection	6
Data Handling.....	6
Data Analysis	7
Data Reuse	8
Legal and Ethics	9
Personal Data.....	9
Property Rights.....	9
Publication	10
Publishing or Sharing Data	10
Storage and Long-Term Preservation	11
Archive	11

.....

GENERAL INFORMATION

Overview

Project number

A03

Name of Experiment / Acronym / Number

Anaphoric Distance: Story Continuation Experiment

Involved persons

Berfin Aktas

PI or responsible person (head of the study)

Manfred Stede

Subject area

discourse, coreference

Method / Type of data

Narrative continuations in text and audio format

Participants (of the study)

native speakers of English (recruited via crowd sourcing)

Short description (of the study)

The impact of anaphoric distance on referential choice is tested across oral and written media. Native speakers of English are recruited via a crowdsourcing company. Participants continued to short stories either by providing oral or written responses.

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
- 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 317633512

.....

.....

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)

October 2020

When does data collection end? (approximately / tentatively)

November 2020

Data Handling

Where is the dataset stored during the project?

- CRC file server
- Box.UP university cloud
- researcher's computer
- computer in the laboratory

.....

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

- Sharing data through box UP
- Regular backups are taken to an external hard drive.
- The university cloud and CRC File Server is backed up regularly.

Which file formats are used?

- Participants' text responses: .xlsx
- Participants' audio responses: .webm, .ogg, .wav
- Audio Stimuli: .mp3

Which measures of quality assurance are taken for this dataset?

The participants qualified via a qualification experiment where their responses are evaluated manually. Engagement of the participants are checked via asking the result of mathematical expressions and/or asking basic questions about a text. All the audio

Data Analysis

When does data analysis start? (approximately / tentatively)

May 2021

When does data analysis end? (approximately / tentatively)

July 2021

.....

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?

The data collected via the crowdsourcing experiment that might be of interest to researchers at or outside of University of Potsdam, for example to:

- Perform further experiments on the impact of anaphoric distance on referential choice across different production media
- Analyze the differences in oral vs written narratives

.....

.....

LEGAL AND ETHICS

Personal Data

Does this dataset contain personal data?

yes

If yes, are these data anonymised?

yes

Property Rights

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

no

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?

yes

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

yes

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

CC0 (Public Domain)

If yes, when will the data be published?

when the paper is 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?

- CRC 1287 File-Server
- OSF
- Research Data Server from Project IN-FDM-BB (a.t.m. not available)

.....

coordinated by:



Universität Potsdam

<https://www.uni-potsdam.de>

funded by:

DFG Deutsche
Forschungsgemeinschaft

<https://www.dfg.de/>

in cooperation with:

**RUHR
UNIVERSITÄT
BOCHUM**

RUB

<https://www.ruhr-uni-bochum.de>