DATA MANAGEMENT PLAN
SFB 1287 / 2025 / PHASE 2

PROJECT T01

Table of contents

General Information

Overview

Project number

T01

Name of Experiment / Acronym / Number

RSTmulti: Corpus of multiple rhetorical text structure annotations

Involved persons

Manfred Stede, Freya Hewett

PI or responsible person (head of the study)

Manfred Stede

Subject area

Linguistics

Method / Type of data

Corpus data

Participants (of the study)

-

Short description (of the study)

The corpus contains texts which have two separate Rhetorical Structure Theory annotations, for the purpose of studying disagreement in the context of discourse structure.

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)

31.12.2024

When does data collection end? (approximately / tentatively)

10.04.2025

Data Handling

Where is the dataset stored during the project?

Box.Up -Cloud Server

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

-.

Which file formats are used?

XML (more specifically, .rs3)

Which measures of quality assurance are taken for this dataset?

Multiple annotators and thorough reviews of the annotations

Data Analysis

When does data analysis start? (approximately / tentatively)

02.03.2025

When does data analysis end? (approximately / tentatively)

10.04.2025

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 corpus is a collection of Rhetorical Structure Theory (RST) annotations with multiple annotations for German and English, and is highly relevant for the topic of "perspectivism", which has been gaining traction over the last few years. RST annotations are known to be subjective and also contain multiple layers of annotations, making the study of disagreement in this context particularly interesting. The corpus is interesting for both researchers working specifically on discourse structure, but also for those working on perspectivism and disagreement in general, as well as psycholinguists or cognitive scientists who are interested in the comprehension of discourse.

Legal and Ethics

Personal Data

Does this dataset contain personal data?

no

Are these data anonymised?

-

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?

-

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

Creative Commons Attribution-ShareAlike (CC-BY-SA)

If yes, when will the data be published?

27.05.2025

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?

- *SFB 1287 File-Server*

- *Research Data Server from Project IN-FDM-BB (a.t.m. not available)*

- *GitHub: https://github.com/fhewett/RSTmulti/*