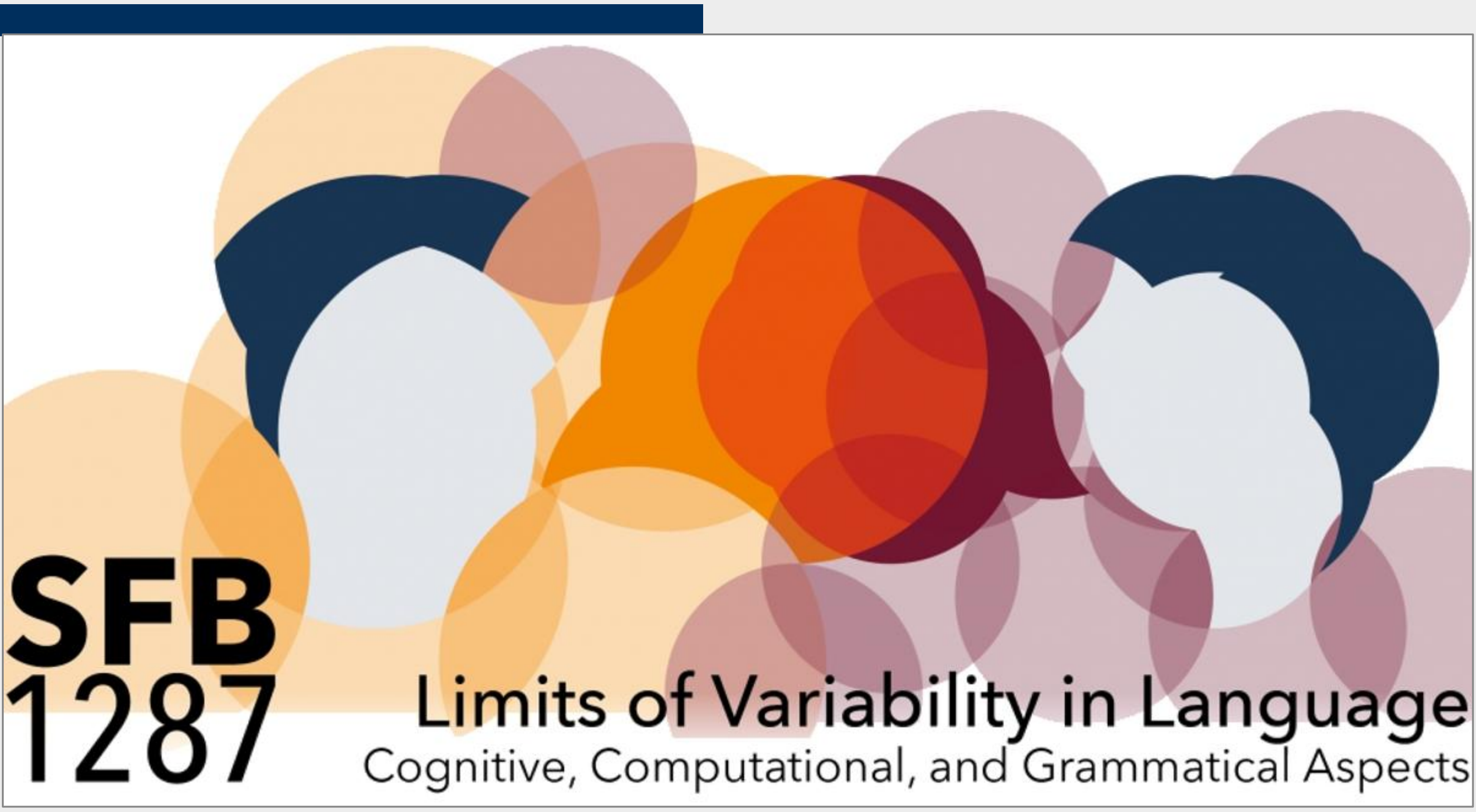
SFB
1287
Limits of Variability in Language
Cognitive, Computational, and Grammatical Aspects

DATA MANAGEMENT PLAN

CRC 1287 | 2025 | Phase 2

Project C06

TABLE OF CONTENTS

GENERAL INFORMATION

Overview

Project number

C06

Name of Experiment / Acronym / Number

Diachronic investigation of dependency lengths

Involved persons

-

PI or responsible person (head of the study)

Ulrike Demske

Subject area

German Studies

Method / Type of data

Data: newspaper texts (1600 - 1900) Annotation: semi-automatic (CoNLLU format (Buchholz & Marsi 2006), Stanford Parser (Manning et al. 2014), Arborator (Gerdes 2013)) Calculation of dependency lengths: using own Python scripts (will be provided) Buchholz, Sabine & Erwin Marsi. 2006. CoNLL-X shared task on multilingual dependency parsing. In Proceedings of the tenth conference on computational natural language learning (CoNLL-x), 149{164. New York City: Association for Computational Linguistics. https://www.aclweb.org/anthology/W06-2920 Gerdes, Kim. 2013. Collaborative dependency annotation. In Proceedings of the second international conference on dependency linguistics (depling 2013), 88{97. Charles University in Prague Matfyzpress Prague, Czech Republic. https://www.aclweb.org/anthology/W13-3711.pdf. Manning, Christopher D. et al. 2014. The Stanford CoreNLP natural language processing toolkit. In Association for computational linguistics (acl) system demonstrations, 55{ 60. http://www.aclweb.org/anthology/P/P14/P14-5010.

Participants (of the study)

-

Short description (of the study)

The original texts are converted into the 10-column CoNLLU format using the Stanford parser. Compliance with sentence boundaries is then manually checked, the lemma level is annotated, and dependencies are annotated using the Arborator tool. A text is processed by two annotators, and the annotations are compared. In case of discrepancies in the annotation, they agree on the most plausible solution.

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

- Data Management in Psychological Science

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?

reused

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

Text corpus (Iskra Fodor) is published at https://weblicht.sfs.uni-tuebingen.de/Tundra.

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?

- 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.

- Stage 1: Internal documentation: Data from researcher's laptop and lab computer will be stored on PRIM server (backed-up) and CRC-File-Server

- Stage 2: External documentation: Data will be stored on OSF platform.

Which file formats are used?

.conllu.

Which measures of quality assurance are taken for this dataset?

Four-eyes principle, good documentation.

Data Analysis

When does data analysis start? (approximately / tentatively)

01.12.2023

When does data analysis end? (approximately / tentatively)

29.02.2024

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?

Anyone who works with linguistic corpora, is interested in historically annotated data for German, or is interested in dependency annotations in general.

....

LEGAL AND ETHICS

Personal Data

Does this dataset contain personal data?

no

If yes, 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?

CC BY (Attribution)

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