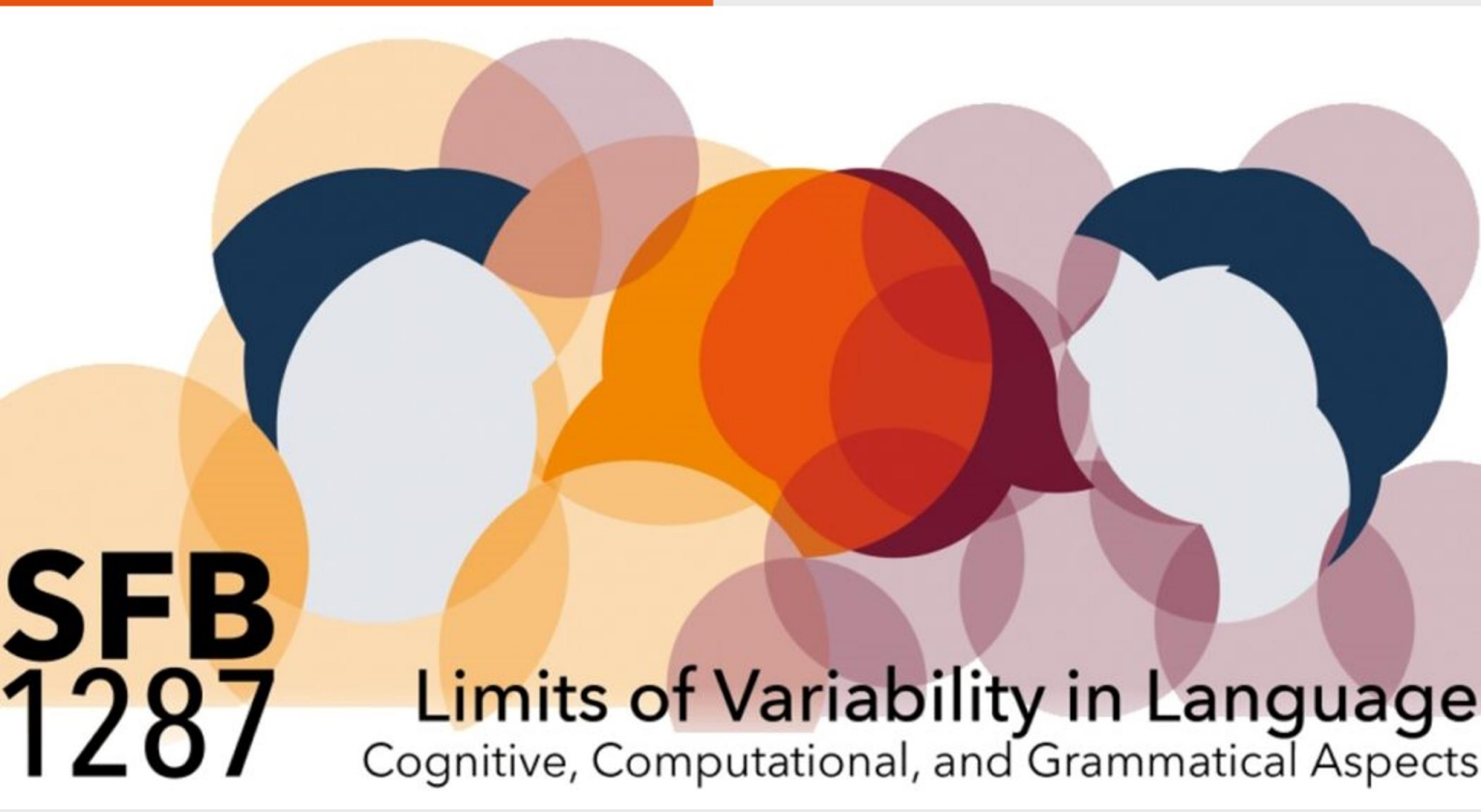

DATA MANAGEMENT PLAN
SFB 1287 / 2025 / PHASE 2

PROJECT B04

Table of contents

General Information

Overview

Project number

B04

Name of Experiment / Acronym / Number

Work Packages 1&2: Individual profiles of morphological processing in L1 German and L2 English

Involved persons

Prof. Dr. Harald Clahsen, PD Dr. Claudia Felser, Sabia Costantini

PI or responsible person (head of the study)

Prof. Dr. Harald Clahsen, PD Dr. Claudia Felser

Subject area

Psycholinguistics

Method / Type of data

- *EEG data*
- *Language History and English Language Tests:*
    - *Language History Questionnaire. The questionnaire is divided into 5 sections, namely "personal information", "language acquisition", "language use", "handedness", and "health" (i.e.: language disorders and vision). The questionnaire gathers biographical and language-related data.*
    - *Adapted Oxford Quick Placement Test. The Adapted Oxford Quick Placement Test used in this study is an online-based (i.e.: Google Forms) version of the publicly available 2001 Oxford Quick Placement Test (see here). It is used to establish the English language level of the study's participants.*
    - *Self devised Past Participle Knowledge Test https://forms.gle/3Ekc7bDfZyrjFku86 This self-devised, online-based (i.e.:*

Google Forms) test targets participants' knowledge of English past participles in sentence contexts.

- *Verbal Fluency (Animal, Food, 60 seconds) Participants are asked to name as many animals/food items in English as possible in 60 seconds. Data is quantified in terms of correct responses, first and subsequent response time.*
- *Working Memory Task (Digit Span Backward) https://eprimego.com/download/7R45N5X6*

Participants (of the study)

- *100 participants were recruited with the following requirements:*
- *be aged between 18-35*
- *have German as their only mother tongue*
- *be right-handed*
- *have normal or corrected eyesight*
- *have no medical history of a neurological nature (migraine, epilepsy) or current mental/physical illness*
- *have an upper intermediate English proficiency level (i.e.: B1-C2 CEFR, see https://www.coe.int/en/web/common-european-framework-reference-languages/table-1-cefr-3.3-common-reference-levels-global-scale)*
- *have no previous knowledge of Turkish*

Short description (of the study)

Individuals show considerable levels of variability in the way they process language (Kidd, Donnelly, & Christiansen, 2017). Event-related brain potentials (ERPs) have been rather helpful in understanding how variable language processing is. For instance, Tanner (2019) noticed that, when confronted with the same linguistic stimuli, monolingual language speakers reported contrasting performances, with some individuals displaying a negativity-dominant ERP response resembling the one of an N400, and others eliciting a positivity-dominant ERP response mirroring a P600. Similar findings were reported by Caffarra et al. (2019) who observed a biphasic LAN/P600 response pattern in some individuals, P600 effects and N400 effects in others during morphosyntactic processing. Although most studies in the field have been conducted with monolingual speakers (Kim et al., 2018; Tanner & Van Hell, 2014; Nakano et al., 2010; Osterhout, 1997), variability in language processing abilities is particularly appreciable in bilingual individuals, who appear to have more varied ERP response patterns compared to the ones of their monolingual counterparts (Pélissier 2020; Fricke et al., 2019). However, research including bilingual speakers has

mainly compared different groups of bilingual speakers, with only a handful of studies examining variability across the bilinguals' spoken languages (Grey, 2022; Bice & Kroll, 2021; Wampler et al., 2014). To contribute to a better understanding of variability in bilingual language processing, the current study explores variation in language processing abilities within the same group of German-English bilinguals, focussing on ERP responses across their spoken languages during morphological violations processing. References Bice, K. and Kroll, J.F., 2021. Grammatical processing in two languages: How individual differences in language experience and cognitive abilities shape comprehension in heritage bilinguals. Journal of neurolinguistics, 58, 100963. Caffarra, S., Mendoza, M. and Davidson, D., 2019. Is the LAN effect in morphosyntactic processing an ERP artifact?. Brain and Language, 191, pp.9-16. Fricke, M., Zirnstein, M., Navarro-Torres, C. and Kroll, J.F., 2019. Bilingualism reveals fundamental variation in language processing. Bilingualism: Language and Cognition, 22(1), 200-207. Grey, S., 2022. Variability in native and nonnative language: An ERP study of semantic and grammar processing. Studies in Second Language Acquisition, 1-30. Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. Trends in Cognitive Sciences, 22(2), 154-169. Kim, A.E., Oines, L. and Miyake, A., 2018. Individual differences in verbal working memory underlie a tradeoff between semantic and structural processing difficulty during language comprehension: An ERP investigation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 44(3), 406. Nakano, H., Saron, C. and Swaab, T.Y., 2010. Speech and span: Working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. Journal of Cognitive Neuroscience, 22(12), 2886-2898. Osterhout, L., 1997. On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. Brain and language, 59(3), 494-522. Pélissier, M. (2020). Comparing ERPs between native speakers and second language learners: Dealing with individual variability. In A. Edmonds, P. Leclerq, & A. Gudmestead (Eds.), Interpreting language-learning data (pp. 39–69). Language Science Press. Tanner, D. and Van Hell, J.G., 2014. ERPs reveal individual differences in morphosyntactic processing. Neuropsychologia, 56,.289-301. Tanner, D., 2019. Robust neurocognitive individual differences in grammatical agreement processing: A latent variable approach. Cortex, 111, 210-237. Wampler, E. K., McLaughlin, J., & Osterhout, L. (2014). How Gender, Handedness, and L1 Processing Strategy Influence L2 Grammatical Processing. Annual Meeting of the Society for the Neurobiology of Language.

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.*
- *Internal Data Management Requirements (PRIM) As part of the internal data management requirements at PRIM, the following need to be provided and stored on the PRIM database:*
    - *Study Description - Material (items, experimental lists)*
    - *Anonymised participant information*
    - *Procedure & Experiment Instructions*
    - *Experiment presentation scripts (NBS presentation)*
    - *EEG preprocessing scripts*
    - *Data files (in a readable/usable format)*
    - *Data documentation*
    - *Statistical analysis scripts*
    - *Statistical results*

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)

02.05.2022

When does data collection end? (approximately / tentatively)

31.03.2023

Data Handling

Where is the dataset stored during the project?

The dataset is being stored locally on the laboratory's computer.

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

Each EEG recording produces three files, respectively in .eeg, .vhdr, .vmrk formats. For each participant, these three files are backed up on the researcher's work USB stick, alongside the University of Potsdam BOX.UP server.

Which file formats are used?

EEG Data is saved in three different file formats: .eeg, .vhdr, .vmrk. English language related data and working memory scores are saved as .csv files. Scripts are coded with: Presentation NBS (.sce and .exp files) for stimuli presentation Python (.py) for EEG data pre-processing (regression-based baseline correction) Matlab (.m) for EEG data pre-processing Documentation is saved as plain text (.txt)

Which measures of quality assurance are taken for this dataset?

Automatisation, additional impedance checks, and external revision of analysis scripts are implemented to ensure quality of the data: Screening Tests Automatising the scoring of language-related data with a Python script and, where possible, with Excel formula to avoid hand-coded errors. EEG Signal Documenting EEG impedances at the start and end of the experiment. Given the lengthy duration of the experiment (3h), impedance levels are additionally checked and fixed during breaks. Scripts To increase reproducibility of the experiment and its analysis, R scripts will be externally reviewed.

Data Analysis

When does data analysis start? (approximately / tentatively)

06.06.2022

When does data analysis end? (approximately / tentatively)

31.08.2023

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?

Researchers in the field of psycholinguistics and language acquisition might be interested in re-using the data. Moreover, current/future PRIM researchers might be interested in reusing the data. The reuse of the data has implications for both the study's researchers and interested researchers: Study's Researchers:

- *Store the data on a platform which will allow it to reach a wide scientific audience*
- *Provide detailed explanation of the pre-processing codes in Python and Matlab*
- *Provide detailed explanation of the data analysis codes in R Statistics*
- *Provide results and their interpretation Interested Researchers:*
- *Have a license for Matlab EEGLAB to use the pre-processing script*
- *Have access to Python Community*
- *Have access to R Statistics*

Legal and Ethics

Personal Data

Does this dataset contain personal data?

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?

yes

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

The project uses data/resources that are protected by intellectual rights:

- *Oxford Quick Placement Test http://www.rigsieradz.pl/uploads/news/id22/test%20-%20QPT%20Paper%20and%20pen%20V1.pdf*
- *NBS Presentation for stimulus presentation https://www.neurobs.com/*
- *Matlab EEGLAB toolbox for data-pre processing https://sccn.ucsd.edu/eeglab/index.php*
- *E-PRIME Go Software for remote data collection (Working Memory Task) https://pstnet.com/eprime-go/*
- *SoSci Survey for Professional Online Questionnaires (Language History Questionnaire) https://www.soscisurvey.de/*

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?

-

If yes, when will the data be published?

Recommended procedure: Upload data and obtain digital identifier (e.g., DOI, OSF link) when submitting the first paper; thus, you can cite the data in the paper. If necessary, restrict public access (embargo) until last paper published (max. 2 years).

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

coordinated by:

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