The development of lexical biases in grammatical variation. Exemplar-driven and index-driven lectal contamination.

A Data Management Plan created using DMPonline.be

Creator: Dirk Pijpops

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Grant number / **URL:** 1212623N

ID: 198432

Start date: 15-09-2024

Project abstract:

Mechanisms of language variation and change often take as starting point lexical biases in grammatical variation, i.e. the finding that particular words engender speakers to prefer one construction over another while forming utterances. For example, a frequent word that is biased towards a construction may attract similar words through analogy, creating a new conjugation class. What is unclear, however, is how such lexical biases develop in the first place.

To understand this, the project introduces two mechanisms that can create such lexical biases, viz. exemplar-driven and index-driven lectal contamination. Both mechanisms start from language contact between two varieties of the same language, but differ in how such contact leads to lexical biases within the varieties. Exemplar-driven contamination relies on the cognitive storage of exemplars, while index-driven contamination assumes that the words and constructions act as social indices.

A pilot study focusing on nominal morphological variation has already been completed, with positive results. The project will conduct four more corpus-based case studies that test the effect of both mechanisms among other types of variation. Next, I will build an agent-based simulation of each mechanism. This will allow me to validate both mechanisms in-silico and derive exact theoretical predictions for each mechanism. Finally, these predictions will be put to the test through corpus research and a forced-choice and receptive experiment.

Last modified: 06-04-2023

The development of lexical biases in grammatical variation. Exemplar-driven and index-driven lectal contamination.
FWO DMP (Flemish Standard DMP)

1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

Dataset case study 1: observational. The data will be newly collected. These are digital data that will be stored in a tab-delimited file with a txt-extension. This file should not be bigger than 1 GB.
Dataset case study 2: observational. The data will be newly collected. These are digital data that will be stored in a tab-delimited file with a txt-extension. This file should not be bigger than 1 GB.
Dataset case study 3: observational. The data will be newly collected. These are digital data that will be stored in a tab-delimited file with a txt-extension. This file should not be bigger than 1 GB.
Dataset case study 4: observational. The data will be newly collected. These are digital data that will be stored in a tab-delimited file with a txt-extension. This file should not be bigger than 1 GB.
Dataset index-driven simulation: simulation data. The data will be newly generated. They are digital data that will be stored in tab-delimited files with a txt-extension. In total, these files should not be bigger than 10 GB.
Dataset exemplar-driven simulation: simulation data. The data will be newly generated. They are digital data that will be stored in tab-delimited files with a txt-extension. In total, these files should not be bigger than 10 GB.
Dataset questionnaire: questionnaire data. The data will be newly collected. These are digital data that will be stored in a tab-delimited file with a txt-extension. This file should not be bigger than 1 GB.

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

NA

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- No

There is a questionnaire which will be distributed online and will be filled in by humans.

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes

The participants filling out the questionnaire will be asked to provide general data about themselves, such as where they grew up and where they currently live. They will not be asked to give their names or provide any sensitive information. They will also be able to refuse to give this information

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- No

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

2. Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

The data and scripts will be documented through manuals, README.txt files and documentation within the scripts themselves.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be

used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- No

3. Data storage & back-up during the research project

Where will the data be stored?

On the storage facilities of the KU Leuven.

How will the data be backed up?

Through the back-up system of the KU Leuven.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.
If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

- Yes

The project has not started yet, since it has been postponed from November 1, 2022 to September 15, 2024, but there will be sufficient storage and backup capacity when the project does start.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

By personal computer is password-protected and the storage facilities of the KU Leuven are also well protected.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

About €500. They will be covered by my bench fee.

4. Data preservation after the end of the research project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

All data.

Where will these data be archived (stored and curated for the long-term)?

On the file storage of the KU Leuven.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

€500. This will be covered by my bench fee.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Yes, in an Open Access repository
- No (closed access)

Part of the data or the scripts may be made available through my personal GitHub.

If access is restricted, please specify who will be able to access the data and under what conditions.

If the access is restricted, I and my supervisor will have access under all conditions.

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

- No

Where will the data be made available? If already known, please provide a repository per dataset or data type.

This is not yet known, as the project has not started yet. It has been postponed from November 1, 2022 to September 15, 2024

When will the data be made available?

This is not yet known, as the project has not started yet. It has been postponed from November 1, 2022 to September 15, 2024

Which data usage licenses are you going to provide? If none, please explain why.

This is not yet known, as the project has not started yet. It has been postponed from November 1, 2022 to September 15, 2024

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- Yes

What are the expected costs for data sharing? How will these costs be covered?

None

6. Responsibilities

Who will manage data documentation and metadata during the research project?

Dirk Pijpops

Who will manage data storage and backup during the research project?

Dirk Pijpops

Who will manage data preservation and sharing?

Dirk Pijpops

Who will update and implement this DMP?

Dirk Pijpops