Plan Overview

A Data Management Plan created using DMPonline.be

Title: JC's Plan

Creator: Jean-Christophe Verstraete

Affiliation: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Project abstract:

When cultures come into contact, words have to be found for novel items and concepts that go along with contact, by borrowing (e.g. dyuga for ‘sugar’ in Ngiyambaa), coinage (e.g. mayi nani ‘sand food’ for ‘sugar’ in Umpithamu) or semantic extension (e.g. tyulkuwa ’ashes’ for ‘flour’ in Yukulta). This process of ‘lexical acculturation’ is well-known, but there is surprisingly little systematic research beyond borrowing strategies. This is all the more surprising given the emergence of lexical typology as a subfield, along with databases for systematic cross-linguistic study of lexical semantics. This project aims to conduct a large-scale study of semantic and areal patterning in lexical acculturation, based on the indigenous languages of Australia. This should allow us to develop a more systematic understanding of lexical acculturation, embedded in recent work on lexical and areal typology.

ID: 213659

Start date: 01-10-2024

End date: 30-09-2028

Last modified: 20-03-2025

JC's Plan

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 name / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
		Indicate: ***N****(ew data) or* ***E****(xisting data)*	Indicate: **D**(igital) or **P**(hysical)	Indicate: **A**udiovisual **I**mages **S**ound **N**umerical **T**extual **M**odel **SO**ftware Other (specify)		Indicate: <1GB <100GB <1TB <5TB >5TB NA	
Basic_data	lexical items with meanings, source information, basic formal analysis, basic semantic categorization	E	D	T	.xslx	<1GB	12.000+ rows
Fieldwork_data	Interviews about concepts with partial referential overlap	N	D	A, T	.wav, .eaf	<1GB	-

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:

- Basic_data will be based on a sample of 200 languages, for which all available secondary sources (mainly grammars and dictionaries) will be analysed for terms for introduced concepts. A full set of bibliographic references will only be available towards the end of the project, as some sources require archival work. A starting point is the bibliography provided in Hoogmartens & Verstraete (2020), the pilot study for this project.
- Basic_data will integrate database underlying Hoogmartens & Verstraete (2020), available as open access resource in supplementary materials of the publication. Categorization may be modified, new sources may be added for the sample used in that publication.

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

- Yes, human subject data (Provide SMEC or EC approval number below)

Fieldwork_data; ethical approval yet to be obtained (data will only be gathered later in the project, building on preliminary results of basic_data)

Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).

- No

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 or 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

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).

- Basic_data: (i) Basic form and meaning: Given the diversity of the sources, even the basic forms and meanings of lexical items can be hard to interpret. Forms and meanings are recorded in the database in a such a way that all decisions can be traced back, by systematically distinguishing between the information found in the source (with a reference that is as precise as possible), and the analysts' interpretations (including the evidentiary basis for their interpretation). (ii) Semantic and formal categorization: Semantic and formal categorization is based on a set of criteria that is constantly updated in a readme.txt file, with particular attention to problematic and borderline cases.
- Fieldwork_data: (i) Basic information about speaker, recorder and recording situation are always recorded in audio form, along with the actual interview, so that basic metadata can never be separated from the data. (ii) The entire interview is audio-recorded and subsequently transcribed in ELAN, which makes it fully searchable.

Will a metadata standard be used to make it easier to find and reuse the data?
If so, please specify which metadata standard will be used.

If not, please specify which metadata will be created to make the data easier to find and reuse.

- No

Data Storage & Back-up during the Research Project

Where will the data be stored?

- Shared network drive (J-drive)
- Other (specify below)

- Basic_data: Shared network drive (OneDrive), in a folder to which supervisor and PhD student have access.
- Fieldwork_data: On the supervisor's personal laptop, with a backup on an external hard drive and in the speaker's personal archive (Flashdrive and memory card).

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution
- Personal back-ups I make (specify below)

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- Yes

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

- Basic_data: Only supervisor and PhD student will have access to the relevant folder on the shared network drive. Students whose BA or MA thesis topic relates to the project, may receive temporary access to part of the database relevant to their topic, subject to restrictions imposed by supervisor and PhD student.
- Fieldwork_data: Only the supervisor and the speaker will have access to copies of the audio recording and the transcript.

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

Costs are minimal, and are included in the project budget.

Data Preservation after the end of the Research Project

Which data will be retained for 10 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 will be preserved for 10 years according to KU Leuven RDM policy

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

- KU Leuven RDR
- Other (specify below)

Fieldwork_data will be stored on the supervisor's personal laptop, with backup on an external hard drive and the speaker's personal archive (Flashdrive and memory card). Subject to the speaker's approval, they will also be stored at AIATSIS, the

Australian Institute of Aboriginal and Torres Strait Islander Studies, which is the standard archive for depositing data related to Australian Indigenous languages, cultures and history. Access conditions are always set in consultation with researchers and speakers.

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

Costs are minimal, and will be covered by the supervisor's basic running expenses for IT (eg occasionally updated hard drives).

Data Sharing and Reuse

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

- Yes, as open data
- Other (specify below)

- Basic_data: At the end of the project, the database will be made available as an open access resource, using the CLDF standard (Forkel et al. 2018; or whatever is its successor by that time) to ensure replicability and interoperability.
- Fieldwork_data: Access conditions will be determined in consultation with the speaker.

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

Fieldwork_data: Access conditions will be determined in consultation with the speaker.

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 per dataset or data type where appropriate.

- Yes, other

It is common (and good) practice when working with speakers of indigenous languages to let them (and/or their descendants) determine access conditions for audio recordings, even if these do not contain any personal information (as will be the case in this project, where the interviews will be concept- rather than person-oriented).

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- KU Leuven RDR (Research Data Repository)

When will the data be made available?

- Other (specify below)

Basic_data: Upon completion of the PhD embedded in the project.

Which data usage licenses are you going to provide?

If none, please explain why.

- CC-BY 4.0 (data)

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- Yes, a PID will be added upon deposit in a data repository

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

No expected costs.

Responsibilities

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

- Basic_data: PhD student and supervisor
- Fieldwork_data: supervisor

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

- Basic_data: PhD student and supervisor
- Fieldwork_data: supervisor

Who will manage data preservation and sharing?

- Basic_data: PhD student and supervisor
- Fieldwork_data: supervisor

Who will update and implement this DMP?

The project's supervisor and PhD student.