

---

# 'It's like y'know my mother tongue'. The impact of L1 and individual variation on the use of pragmatic markers by learners of English in peer-to-peer interaction

*A Data Management Plan created using DMPOnline.be*

**Creators:** Lieven Buysse, n.n. n.n.

**Affiliation:** KU Leuven (KUL)

**Template:** KU Leuven BOF-IOF

**Principal Investigator:** n.n. n.n.

**Data Manager:** n.n. n.n.

**Project Administrator:** Lieven Buysse

**Grant number / URL:** C14/23/053

**ID:** 205843

**Start date:** 01-10-2023

**End date:** 30-09-2027

## **Project abstract:**

Pragmatic markers are words or phrases that are ubiquitous in spoken English and that do not contribute to the truth conditions of an utterance. Yet, they contribute to discourse coherence, express speakers' attitudes, index various types of relations, and help interactants to build rapport. Examples in English are *so, well, you know, I mean* and *like*. Research on pragmatic markers has surged over the past thirty years, although much of the field has been rather anglocentric. In recent years the research scope has been broadened to also include non-native discourse. Because pragmatic markers are rarely taught explicitly in foreign language classes, language learners have been shown to use them to widely varying degrees. It has proved hard to explain in any reliable way why specific markers are used extensively by some learners and others are hardly used at all. The most obvious explanation for learner behaviour is always how learners use similar items in their mother tongue. Researchers are, however, faced with the obstacle that in most cases no similar data are available for these languages, making such explanations highly tentative.

This project will, therefore, monitor how native speakers of Dutch who are learners of English use pragmatic markers in English and how this compares (i) to native English speech, and (ii) to these learners' use of pragmatic markers in their mother tongue (Dutch). For this we will study recordings of 45 dyad conversations between familiars. These conversations will be equally distributed across learners of English in English, these same learners but in their mother tongue (Dutch), and native speakers of English. The native English conversations are already available from a previous C1 project, whereas the others will be collected within the proposed project. The data will be analysed both quantitatively (e.g. comparing frequencies between speakers/speaker groups) and qualitatively (e.g. functional analyses of pragmatic marker use).

Apart from shedding light on the influence of transfer from learners' mother tongue in how they use pragmatic markers in their target language, the project will make three further major contributions to the research field. First, the approach adopted in the study will allow for a uniquely combined perspective yielding insight into speakers' individual pragmatic preferences as well as into how these contribute to tendencies at group level, which diverts from the typical tendency to consider learners as a homogeneous group. Second, contrary to most other studies of learner language, this investigation will be able to rely on data from peer-to-peer interaction, since pragmatics is best observed in settings that are not artificially constrained. Third, our knowledge of pragmatic markers in Dutch will be greatly expanded through an unprecedented corpus-based analysis of peer-to-peer interaction, which will also act as a catalyst for further studies in this field on Dutch.

In short, the project will empirically show the nature of mother tongue impact on learner pragmatics, lay bare contrasts between English and Dutch in pragmatic marker usage, and will enable educators to rethink pedagogical strategies to the acquisition of pragmatics.

**Last modified:** 26-03-2024

## 'It's like y'know my mother tongue'. The impact of L1 and individual variation on the use of pragmatic markers by learners of English in peer-to-peer interaction

### 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
		<i>Indicate: N(ew data) or E(xisting data)</i>	<i>Indicate: D(igital) or P(hysical)</i>	<i>Indicate: Audiovisual Images Sound Numerical Textual Model SOftware Other (specify)</i>		<i>Indicate: &lt;1GB &lt;100GB &lt;1TB &lt;5TB &gt;5TB NA</i>	
L2 English videos	Corpus of videos of native speakers of Dutch in English	N	D	A	mp4	<100GB	
L2 English transcriptions	Corpus of video transcriptions of native speakers of Dutch in English	N	D	A	eaf	<100GB	
L2 English annotations	Corpus of annotations of pragmatic marker usage by native speakers of Dutch in English	N	D	T	xlsx	<100GB	
L1 Dutch videos	Corpus of videos of native speakers of Dutch in Dutch	N	D	A	mp4	<100GB	
L1 Dutch transcriptions	Corpus of video transcriptions of native speakers of Dutch in Dutch	N	D	A	eaf	<100GB	
L1 Dutch annotations	Corpus of annotations of pragmatic marker usage by native speakers of Dutch in Dutch	N	D	T	xlsx	<100GB	
L1 English videos	Corpus of videos of native speakers of English in English	E	D	A	mp4	<100GB	
L1 English transcriptions	Corpus of video transcriptions of native speakers of English in English	E	D	A	eaf	<100GB	
L1 English annotations	Corpus of annotations of pragmatic marker usage by native speakers of English in English	E + N	D	T	xlsx	<100GB	

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:

A native speaker English corpus gathered for a previous BOF project (project code 3H170338) carried out with the same supervisor will be reused.

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)

E-2023-4104

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).

- Yes (Provide PRET G-number or EC S-number below)

G-2023-7010

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).

The approach is described extensively in the project outline that accompanies this project as well as in Blanchard (2021), whose methodology in terms of corpus compilation is being replicated in the current project. The corpus transcription and annotation guidelines will be available in a separate document.

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.

- Yes

DataCite

#### Data Storage & Back-up during the Research Project

Where will the data be stored?

- OneDrive (KU Leuven)

**How will the data be backed up?**

- Standard back-up provided by KU Leuven ICTS for my storage solution

**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?**

Access is restricted to the PI and the PhD researcher. The regular security measures are in place for access to Onedrive under KU Leuven policy, including multifactor authentication.

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

As it stands, no extra data storage or backup would appear to be necessary. Should this need arise, the project budget can provide for any expenses related to this.

**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

The collected corpora and related metadata will be preserved.

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

- KU Leuven RDR

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

None.

**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 embargoed data (temporary restriction)

The data will be made available through KU Leuven RDR. Exported transcriptions and annotations will be made available immediately after the end of the project, but access to the video recordings will remain restricted for 5 years following the completion of the project, because the corpora will first and foremost be further explored within the research group.

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

During the period of embargo data will be made accessible (i) to researchers with whom the research group collaborates, and (ii) to journal editors that request access in the light of a publication based on the data.

**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.**

- No

For privacy reasons, neither their names nor any personal information about the participants will be shared. All participants did, however, give permission for their videos to be reused for future research purposes.

**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)

Since there are no disciplinary repositories readily available, KU Leuven RDR will be used.

**When will the data be made available?**

- Specific date (specify below)

At the end of the project (30 September 2027)

**Which data usage licenses are you going to provide?**

**If none, please explain why.**

- Other (specify below)

CC-BY-NC-ND-4.0

**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?**

None

**Responsibilities**

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

Line Winkelmans (PhD researcher)

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

Line Winkelmans (PhD researcher)

**Who will manage data preservation and sharing?**

Lieven Buysse (supervisor)

**Who will update and implement this DMP?**

Lieven Buysse (supervisor)