Emotion language in context: A longitudinal, multi-modal investigation of meaning-making in everyday life and its association with expertise and well-being

A Data Management Plan created using DMPonline.be

Creator: Katie Hoemann

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: Katie Hoemann

Data Manager: Katie Hoemann

Project Administrator: Katie Hoemann

Grant number / **URL:** 12A3923N

ID: 198915

Start date: 01-10-2022

End date: 30-09-2025

Project abstract:

Some people are more skilled than others in how they understand and experience emotion: they have greater expertise in emotion. These individuals have better mental, physical, and relational well-being because they make emotional meaning that is context-specific and tailored to the situation at hand, helping them to regulate and cope in healthier ways. Language is a key means of gaining insight into emotional meaning-making. However, research often studies emotion language in a decontextualized manner – focusing only on emotion words, ignoring situational information, and sampling for limited timespans. This approach limits our understanding of expertise as context-specific meaning-making and its ties to well-being. This project puts emotion language back in context by using in-the-moment descriptions of experience to examine multiple features of language, their patterns of use over time and relationship to the physical and social environment. To do this, I create a toolbox that uses experience sampling and mobile sensing to collect data on real-world emotional experiences, together with computational linguistic and network analytic techniques to yield psychologically interpretable models of emotion language in context. This toolbox is a crucial innovation that allows me to link linguistic features to situation-specific dynamics, leading to a more profound understanding of how people make emotional meaning in everyday life and which ways are associated with better outcomes.

Last modified: 25-04-2023

Emotion language in context: A longitudinal, multi-modal investigation of meaning-making in everyday life and its association with expertise and well-being
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.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .cvs,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Pre-screening	Pre-screening data from survey	Generate new data	Digital	Observational	.csv	<100MB	
Participant list	Participant list	Generate new data	Digital	Compiled/aggregated data	.csv	<100MB	
Key table	Key table between participant list and pseudonym	Generate new data	Digital	Compiled/aggregated data	.csv	<100MB	
Surveys	Data from surveys pre-, during-, post-experience sampling	Generate new data	Digital	Observational	.csv	<100GB	
Experience sampling recordings	Event descriptions recorded by participants	Generate new data	Digital	Observational	.m4a, .aac	<1TB	
Experience sampling transcriptions	Transcribed event descriptions	Generate new data	Digital	Observational	.txt	<1TB	
Experience sampling prompts	Structured responses provided at prompts	Generate new data	Digital	Observational	.csv	<1TB	
Experience sampling mobile sensing	Mobile sensing data gathered from smartphones	Generate new data	Digital	Observational	.csv	<1TB	
Pre-processing code	Language and mobile sensing feature extraction, questionnaire compilation, README files	Generate new data	Digital	Other	.py, .R, .txt	<1GB	
Analysis code	Network- and regression-based analyses, README files	Generate new data	Digital	Other	.py, .R, .txt	<1GB	
Pre-processed data	Extracted language and mobile sensing features, questionnaire measures per participant	Generate new data	Digital	Compiled/aggregated data	.csv	<100GB	
Analysis output	Results of network- and regression-based analyses, per participant and overall	Generate new data	Digital	Other	.csv, .txt, .pdf	<100GB	
Study materials	Participant-facing documentation and protocols	Generate new data	Digital	NA	.docx, .pptx, .pdf	<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:

Not applicable.

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.

- Yes, human subject data

The observational datasets above will be collected from healthy, adult volunteers. Ethics approval has been granted for this project by the Privacy and Ethics team and Social and Societal Ethics Committee (SMEC) at KU Leuven, reference number G-2023-6379-R2(MIN).

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 observational datasets collected will include identification information (name, addresses, email address, telephone number), demographics (age, gender, and race/ethnicity), psychological details (emotion-related abilities), audio recordings (including raw language data and ambient sound), electronic activity data (IP addresses, GPS locations), and data about mental health (self-

reported symptoms).
Raw language data may contain identifying information and will be redacted before further analysis. All participant data will be collected under a pseudonym. Pseudonymized recordings will be submitted to algorithms that will identify participant speech and convert it to text. The resulting transcriptions will be reviewed by research assistants, who will listen to the original recordings and make any necessary corrections to the transcriptions. These corrections will include redacting names of specific people and places (i.e., by substituting <NAME> and <PLACE> in the text). Research assistants will sign data access agreements and will only have access to pseudonymized data via KU Leuven accounts.

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).

Study protocol will be documented in MS-Word files (.docx). Data processing steps will be documented via pre-processing and analytic code in R (.R) and/or Python (.py), with accompanying README files (.txt) and codebooks (.csv). All files will be shared via Teams/SharePoint Online within KU Leuven and archived in the KU Leuven research data repository (RDR) when the datasets are final.

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.

- Yes

DataCite is used for the KU Leuven RDR.

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

Where will the data be stored?

Data will be stored on secure KU Leuven network drives and Microsoft Teams/SharePoint Online linked to a KU Leuven account (Katherine Hoemann, katie.hoemann@kuleuven.be). These data storage solutions are suitable for the confidentiality level and volume of data collected (confidential, 1-5TB) and access needs of the researchers.
Consent, questionnaire, and task data will initially be collected using Qualtrics under the KU Leuven agreement, and after collection will be stored as above.

How will the data be backed up?

Data will be back up using the standard back-up provided by KU Leuven ICTS.

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

Data storage through KU Leuven Teams/SharePoint Online covers up to 5TB of data.

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

The secure storage solutions mentioned above are appropriate for the confidentiality level of the data (confidential). Multifactor authentication with the KU Leuven Authenticator app is enabled, ensuring that only appropriate users have access to data via web browsers. No physical data will be collected.
All experience sampling data are uploaded to secured servers at KU Leuven using encrypted SSL connections to ensure that their data cannot be intercepted by third parties. Data processing and primary data analysis will be conducted on KU Leuven servers and encrypted drives. The use of cloud computing infrastructure is not planned.
In the informed consent procedure, participants will be asked if they consent to secondary analysis of their data, and whether this applies to the transcripts as well as variables derived from language analysis, questionnaires, and tasks.
To protect participant privacy, transcripts will never be publicly shared. Researchers interested in working with the transcripts and other raw data will be asked to sign an appropriate data use

agreement, created with approval from SMEC/KU Leuven. Data from participants who have not consented to secondary data analysis will never be shared.

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

There are no expected costs for data storage and backup during the research project. The storage tools mentioned above are sufficient to cover the expected volume of data and are free for KU Leuven researchers.

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 will be stored for 10 years according to KU Leuven RDM policy.

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

All pseudonymized data will be stored in the KU Leuven RDR. All sensitive personal data (participant lists, key tables, audio) will be stored in a OneDrive library accessible only to the PI and fellow.

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

Expected use of the RDR will be below the 50GB limit.

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 a restricted access repository (after approval, institutional access only, …)
- Yes, in an Open Access repository

Raw data, including transcripts of participant speech, will be stored in a restricted access repository (see details below).
Compiled/aggregated and deidentified data - such as language features (e.g., number of emotion words per prompt), locations types visited (i.e., home, work, etc.), etc., and questionnaire responses (e.g., mental health symptoms) - will be made available via an open access repository.

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

The PI and fellow will have access to restricted data at all times.
Research team members from KU Leuven will be granted access to restricted data based on operational needs. In case these are students, they will sign a data use agreement.
In the informed consent procedure, participants will be asked if they consent to secondary analysis of their data, and whether this applies to the transcripts as well as variables derived from language analysis, questionnaires, and tasks.
External researchers interested in working with the transcripts and other raw data will be asked to sign an appropriate data use agreement, created with approval from SMEC/KU Leuven. Data from participants who have not consented to secondary data analysis will never be shared.

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.

- Yes, Privacy aspects
- Yes, Ethical aspects

Personal data will be stored in online KU Leuven storage facilities, which have built-in access management through multifactor authentication. Access to personal data is limited and guarded by the PI and fellow.
All collected and compiled/aggregated data will be pseudonymized. The key tables will be stored with other personal data.
To protect participant privacy, transcripts will never be publicly shared. The volume of language data collected (i.e., ~25,000 words per person) may make it possible to identify participants with tools from forensic linguistics (i.e., identify patterns in speech or language use that help isolate particular individuals).

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

Restricted data will be made available through the KU Leuven research data repository (RDR).
Compiled/aggregated and deidentified data will be made available through the Open Science Framework (OSF) repository.

When will the data be made available?

Data will be made available at the end of the project.

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

Data stored in the KU Leuven RDR will be made available through a data transfer agreement to external researchers only after evaluating their request and purpose of reuse.

Data made available via OSF will be shared under the CC-By Attribution 4.0 International license.

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?

The KU Leuven RDR covers data storage up to 50GB. No additional costs for sharing are expected. There are no costs associated with depositing data in or sharing data via the Open Science Framework (OSF) repository.

6. Responsibilities

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

Katherine Hoemann (katie.hoemann@kuleuven.be)

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

Katherine Hoemann (katie.hoemann@kuleuven.be)

Who will manage data preservation and sharing?

Katherine Hoemann (katie.hoemann@kuleuven.be)

Who will update and implement this DMP?

Katherine Hoemann (katie.hoemann@kuleuven.be)