When to lose one’s senses? Semantic loss vs polysemy sustenance in the lexicon of English

A Data Management Plan created using DMPonline.be

Creator: Hilke Ceuppens

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

ID: 206238

Last modified: 15-04-2024

When to lose one's senses? Semantic loss vs polysemy sustenance in the lexicon of English
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, .csv,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Corpora	Early English Books Online, British National Corpus, Corpus of Late Modern English Texts, Corpus of Historical American English	Reuse existing data	Digital	Compiled data	.txt .xml	<100GB	
Dictionaries (online)	Online dictionaries: Oxford English Dictionary, Merriam-Webster	Reuse existing data	Digital	Compiled data	NA (online access)	NA (online access)	
Dictionaries (print)	Print dictionaries: Macmillan English Dictionary, Etymological Dictionary of Proto-Germantic	Reuse existing data	Physical	NA	NA	NA	2 print books

Samples	Random samples of adjectives and verbs (WP1)	Generate new data	Digital	Compiled data	.txt	<100MB	
Concordances	Concordance lines extracted from corpora	Generate new data	Digital	Compiled data	.txt	<100MB	
Annotated spreadsheets	Spreadsheets with annotated concordances	Generate new data	Digital	Aggregated data	.xlsx	<100MB	
Data coding protocols	Protocols accompanying annotated spreadsheets	Generate new data	Digital	Compiled data	.docx, .pdf	<100MB	
Perl scripts	The Perl scripts created by my supervisor used to search the corpora.	Reuse existing data	Digital	Software	.pl	<100MB	
R scripts	The R scripts used to analyze the data statistically	Generate new data	Digital	Software	.R	<100MB	
Personal notes	Personal notes about literature	Generate new data	Digital	Compiled data	in Zotero database	<100MB	
Summaries	Summaries of literature per topic + all collected and created information per topic (e.g. tables, interpretations of results etc.)	Generate new data	Digital	Compiled data	.one (parts of which are to be converted to .pdf)	<100GB	
Visualizations	Semantic networks, graphs	Generate new data	Digital	Aggregated data	.png, .jpg	<100MB	

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:

Corpora (links lead to the online interface, while my project uses the .txt version)
EEBO: https://www.proquest.com/eebo
BNC: http://www.natcorp.ox.ac.uk/
CLMET: https://perswww.kuleuven.be/~u0044428/clmet3_0.htm
COHA: https://www.english-corpora.org/coha/

Dictionaries
Oxford English Dictionary: https://www.oed.com
Merriam-Webster Dictionary: https://www.merriam-webster.com/
Macmillan English Dictionary: ISBN : 0-333-96667-8
Etymological Dictionary of Proto-Germanic: ISBN : 978-90-04-18340-7

Perl scripts

Scripts provided by supervisor

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

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.

- 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/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).

Each dataset will be accompanied by a README.txt file to explain how the data are structured. Spreadsheets contain comments on abbreviations (if applicable) and scripts contain explanatory comments.
The data coding protocols contain information about the way the data were annotated. They are to be used in combination with the spreadsheets.

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

I will be using KU Leuven RDR; hence DataCite will be used as a metadata standard.

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

Where will the data be stored?

I make use of the Onedrive for Business cloud service provided by KU Leuven.

How will the data be backed up?

OneDrive makes automatic backups.

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

I received 2 TB in OneDrive storage, which is more than the estimated volume specified in the data table.

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

OneDrive's security level (with multifactor authentication activated) is appropriate for the confidentiality level of my data. My laptop has a Bitlocker pre-installed as an additional safety measure (on top of my password).

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. OneDrive for Business is free for staff of KU Leuven.

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 preserved for at least 10 years (according to KU Leuven RDM policy).

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

I will use KU Leuven RDR to archive my data. (if applicable, data that is not suitable for preservation in a repository will be stored using a KU Leuven storage solution, specifically large volume storage)

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

The use of KU Leuven RDR is free. (If need be, I can purchase large volume storage offered by KU Leuven for € 104,42 / TB / year. The bench fee would cover this cost.)

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

I will make my data openly available for reuse by depositing it in KU Leuven RDR (without conditions as to who can access the data).

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

NA

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.

I will use KU Leuven RDR.

When will the data be made available?

Upon publication of research results.

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

CC BY-NC-SA

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

A DOI is added to my data when deposited in KU Leuven RDR.

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

There are no expected costs for data sharing: deposition of smaller datasets in data repositories is covered by the repository.

6. Responsibilities

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

Hilke Ceuppens

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

Hilke Ceuppens

Who will manage data preservation and sharing?

Hendrik De Smet

Who will update and implement this DMP?

Hilke Ceuppens