Using NLP tools for empirical studies of morality in text

This checklist is designed to help researchers clarify the objectives of their study, to reflect on their research questions and better understand the preconditions on the data that is needed to answer these questions. This checklist does not claim completeness. It is meant as a living document that can be updated as required.

Overview

1. Study title:

2. What is the main purpose of the study?

   A investigate moral framing / moral rhetoric in text ☐

   B investigate the moral values or stances of an individual or group ☐

   C develop, evaluate or improve a theory on morality ☐

   D other ☐

   If you checked D, please specify the main purpose of the study :

3. What are the main research questions?

Data selection

1. What is the population about which the scientific claim is made?

   E.g.: a specific person (the use of moral rhetoric in speeches of US president Biden),
   a social group,
   a specific culture, etc.

2. Please add a detailed description of the population (what are the main characteristics):

E.g.: British working class teenagers between 12 and 16

3. What are the characteristics of the sample needed to study the population defined above?

4. Describe the data you are planning to use in your study and explain why it is representative of the population about which the scientific claim is being made.

5. Are there any reasons why this sample might not be representative of the target population?

Are all members of the population equally well represented in the data? ☐ yes ☐ no
Are there any confounding variables? ☐ yes ☐ no
Are there any variables that could not be controlled? ☐ yes ☐ no

Please add a detailed description of any potential issues regarding the points above:

6. Most research questions include a comparison of two or more populations (e.g., comparing the moral values of conservatives and liberals). While your samples might be roughly representative of each of the target population, are the samples also comparable to each other?

For example:
Are the samples comparable with regard to domain, text type, topic, time of creation, etc.? (e.g., comparing two political actors, based on data for actor A taken from campaign speeches and data for actor B taken from parliamentary debates; or speeches by A on the topic of immigration with speeches by B on the topic of economy)

The samples are from the same domain ☐ yes ☐ no

The samples include the same text types ☐ yes ☐ no

The samples include the same topic(s) ☐ yes ☐ no

The samples are comparable wrt creation time ☐ yes ☐ no

Are the different samples comparable in size? ☐ yes ☐ no

Please describe any differences between the samples that might impact results:

[redacted]

Please note: if those differences are part of the RQ (e.g., comparing presidential debates over time), then the data also needs to encode these different variables. However, when comparing presidential candidates that compete with each other in the same election, then the data for both also needs to be comparable w.r.t. text type, domain, time of creation etc.

Reflecting on the properties of your data/samples, reconsider your RQ and the suitability of your data to answer this question. If necessary, refine your RQ or look for a more suitable dataset.

Sampling procedure

1. Explain how the data will be sampled (sampling method: simple random, stratified, cluster sampling, etc.).

[redacted]

2. Is there any impact on the data / any possible bias arising from the sampling procedure? ☐ yes ☐ no

   If yes, please describe the bias in detail.

[redacted]

Preprocessing

1. Describe how the data will be preprocessed.

[redacted]

2. Do you intend to filter the data / remove outliers? ☐ yes ☐ no

If yes, please make sure to include this process in the description of the preprocessing.

How will the filtering or removal of outliers impact results? Are there any risks of introducing bias?
Please describe any risks below.

[redacted]

Theory and variables of interest

1. What is the theoretical framework that will be used for modelling morality?

[redacted]

2. What is the motivation for chosing this specific theory?

[redacted]

3. What are the constraints that might result for using this particular theory (if any)?

[redacted]

4. Describe the variable(s) to be measured in the study (*e.g., moral values of Twitter users from the US who express COVID-19 vaccine hesitancy*).

[redacted]

5. How are the variables of interest defined in the theory?

   E.g., *MFT defines moral values as moral foundations or intutions that are "causes for judgments" (Haidt & Bjorklund, 2008) while Sinnot-Armstrong et al. (2010) describes moral intuitions as " strong, stable, immediate moral beliefs". The main difference between those views concerns the question whether moral judgments are inferentially justified.*)

6. Describe how the variable(s) will be operationalised.

   E.g., *measuring moral values based on moral foundations (MFT))*.

   Please add a detailed description of the classification scheme (number of variables and their definition) below.

7. What methodology / tools will be used for measuring moral language in text?

   - ☐ A dictionary-based approach
   - ☐ A supervised Machine Learning-based approach
   - ☐ Other approaches

   If you checked “Other approaches”, please describe them below:

Hypotheses

1. Please specify the hypotheses that will be tested in the study.

2. How do you plan to test the hypotheses? Describe the outcome that will allow you to reject the H0.

3. Which tests Name the test(s) and add a short description. Also include any constraints that follow from the distribution of your data.

Dictionaries

Releveant only for dictionary-based approaches for text analysis

1. Name of the dictionary to be used in the analysis.

2. Add a detailed description of the dictionary. Include information on the number of classes/ categories distinguished in the dictionary, the number of entries per class, the entry type (word form/lemma/stem/regex), etc.

3. Has the dictionary been validated? Describe the validation procedure.

4. Can the results from the validation be transferred to your data and setting? Give a short justification. Also include any risks for applying the dictionary.

Generalisability, reproducibility, robustness

1. What are the conclusions that you can draw from your study? What are possible limitations for the generalisability of your results?

2. Describe any contexts in which you do not expect your results to hold.

3. Do you intend to test for robustness? Describe your approach.

[redacted]

4. Try to think of ways to test for external validity to strengthen your results.

[redacted]

References & Links

Taherdoost, Hamed, Sampling Methods in Research Methodology; How to Choose a Sampling Technique for Research (April 10, 2016). Available at SSRN: https://ssrn.com/abstract=3205035 or http://dx.doi.org/10.2139/ssrn.3205035

Developing Linguistic Corpora:
a Guide to Good Practice https://bond-lab.github.io/Corpus-Linguistics/dlc/chapter1.htm

Data and Code Guidance by Data Editors:

https://social-science-data-editors.github.io/guidance/

Egami, N & E. Hartman (2023): Elements of External Validity: Framework, Design, and Analysis. *American Political Science Review*. 2023;117(3):1070-1088. doi:10.1017/S0003055422000880

Sample sizes

Mumtaz Ali Memon, Hiram Ting, Jun-Hwa Cheah, Ramayah Thurasamy, Francis Chuah, Tat Huei Cham (2020): Sample Size for Survey Research: Review and Recommendations. *Journal of Applied Structural Equation Modeling*.

https://doaj.org/article/963f24beb0f54f488f7a2619faa9b99a

Felderer, Barbara, Sand, Matthias, & Bruch, Christian (2022). Sample Size Calculation For Complex Sampling Designs. Mannheim, GESIS - Leibniz Institute for the Social Sciences (GESIS-Survey Guidelines).
DOI: 10.15465/gesis-sg_en_042

Based on: https://reforms.cs.princeton.edu/

https://social-science-data-editors.github.io/guidance/