BJHP [Version?] Data Statement

Author(s): Yvette Oortwijn, Thijs Ossenkoppele, Arianna Betti, Annotator 1 (A1), Annotator 2 (A2)

Version: 1.1

Date of creation: March 26th 2021

Date of last modification: March 30th 2021

Document origin:

https://docs.google.com/document/d/11esJonIrzaz_I0npkpduD4HiyjAm2E2Oq4MFyc9DEAo/edit

Aim: To release a data statement cf. Bender & Friedman (2018) for the BJHP [VERSION?] dataset.

Released together with: The HumEval2021 paper 'Interrater disagreement resolution; A systematic procedure to reach consensus in annotation tasks' (2021), written by Oortwijn et al.

On: Forthcoming

A. **Curation Rationale.** The BJHP dataset consists of:
   i. A collection of 54 annotated ID's that correspond to journal articles from the *British Journal for the History of Philosophy* (BJHP) from the period between 2017 and 2019. The annotations were done by two annotators, A1 and A2 (A1's annotations can be found here, A2's here). Of these articles, only the abstract, the introduction, and the sections that contain info on methodology were read by the annotators. After reading, they scored every article a 1 or a 0 on the following questions:
      1. Does the article use a reproducible methodology with explicit inclusion and exclusion criteria for identifying and finding primary literature?
      2. Does the article use a reproducible methodology with explicit inclusion and exclusion criteria for identifying and finding secondary literature?
      3. Does the article explicitly attempt to identify all the available primary literature relative to the research question?
      4. Does the article explicitly attempt to identify all the available secondary literature relative to the research question?
      5a. Does the article argue for wide-scope historical claims, i.e., claims spanning multiple decades or periods or intellectual movements?

   In cases where the annotators answered 5a positively, they also fulfilled the following tasks:
      5b. If 5a is answered positively, then transcribe the passage.
      5c. If 5a is answered positively, does the article qualify the wide-scope claims? (1 or 0)
      5d. If 5c is answered positively, then transcribe the passage.
   ii. An overview of the disagreements between A1 and A2, including an indication of the type of disagreement (A-E) per case. These indications correspond to the types of interrater disagreements presented in Oortwijn et al. A more detailed description of all the disagreement cases can be found here.
   iii. A calculation of the interrater agreement between A1 and A2.
B. **Language Variety.** The exact data on the various languages used in the texts is not available, but the standard language of the texts is British English (en-GB). Incidentally, short fragments of US English (en-US,) Latin (la-Latn), Greek (grc-Grek), German (de-DE), and French (fr-FR) occur.

C. **Speaker Demographic.** An exact demography of all 178 historians of philosophy that are included as speakers in the dataset is not readily available. However, since the percentage of female authors in the BJHP decreased from ±33% in 2004 to ±21% in 2014/2015 (Wilhelm et al. 2018), we have good reasons to expect a significant gender imbalance. There is no other gender-related data available, nor can we say anything about the representation of ethnicity or native languages in the dataset. Further, we do not have information about the academic positions of all authors. Given the large differences in income among academics, ranging from undergraduate students to professors, we are therefore uncertain about the socioeconomic statuses that are represented in the BJHP dataset.

D. **Annotator Demographic.** From Bender2018: "What are the demographic characteristics of the annotators and annotation guideline developers? Their own "social address" influences their experience with language and thus their perception of what they are annotating. Specifications include: Age, Gender, Race/ethnicity, Native language, Socioeconomic status, Training in linguistics/other relevant discipline." A1 is 40 years old, male, Dutch, white, a Dutch native speaker, and has a training in philosophy.

E. **Speech Situation.** The articles were all published between 2017 and 2019 in the BJHP, which is a scientific journal, and thus all texts are intended for an academic audience.

F. **Text Characteristics.** The texts are all within the domain of the history of philosophy, and contain discussions of texts by historical philosophical actors, as well as reviews of secondary literature about these actors. The domain of philosophy that is covered by the BJHP is rather broad, and among included topics are historical philosophical texts and authors in epistemology, metaphysics, ethics, aesthetics, and the philosophy of science. Initially, the corpus contained irrelevant items, e.g. comments and editorials. These items were removed from the dataset.

G. **Recording Quality.** N/A

H. **Other.** N/A

I. **Provenance Appendix.** N/A

References:

Bender, Emily M., and Batya Friedman. 2018. 'Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science'. *Transactions of the Association for Computational Linguistics* 6: 587–604. https://doi.org/10.1162/tacl_a_00041.

Wilhelm, I., Conklin, S.L. & Hassoun, N. 2018. 'New data on the representation of women in philosophy journals: 2004–2015'. *Philosophical Studies* 175: 1441–1464.