Dataset- and variable descriptions
Version: v1

Dataset

The deposited file `dataset_for_main_analyses.Rdata` contains the dataframe `df.agg.sub` which is the dataset used for the analyses reported in the main manuscript. The dataset was created by merging two other datasets:

1. Transcripts of all parliamentary speeches made in the Danish parliament (*Folketinget*) in the calendar year 2018. These data, and some associated meta-data, are obtained directly from the `ParlSpeechV2` (Rauh and Schwalbach 2020).

2. A corpus of press releases (broadly understood) collected by the authors of the paper.[1]

The resulting dataset has been augmented with additional meta-data. Please refer to the paper for more details on the dataset, its creation, its structure etc.

Loading the dataset

It may be necessary to install and load the {`arrow`} package (Richardson et al. 2023) prior to loading `df.agg.sub`.

Variables

This section contains short descriptions of each of the variables contained in `df.agg.sub`. Please refer to the paper for more details on the measurement of the variables.

- `speechindex`: Parliamentary speech unique ideitifier.
- `speech_date`: Date of the parliamentary speech.
- `speech_agenda_type`: Type of agenda item under which a parliamentary speech was delivered (1: deliberation over a bill; 0: otherwise).
- `speech_party`: Party affiliation of the politician who held the parliamentary speech.

1. To ensure compliance with the General Data Protection Regulation (GDPR) of the European Union, the raw data is not made publicly available.

- `speech_length_words`: Length (in words) of the parliamentary speech.
- `in_government`: Whether the politician is a member of a government party.
- `in_main_opposition:` Whether the politician is a member of a main opposition party.
- `in_main_opposition_alternative`: Alternative operationalization of `in_main_opposition`.
- `is_minister`: Whether the politician was a minister at the time of the delivery of the parliamentary speech.
- `is_party_leader`: Whether the politician was a party leader at the time of the delivery of the parliamentary speech.
- `member_origin`: Whether the politician was elected in Denmark or Greenland/Faroe Islands (i.e., on a Nordic mandate).
- `pressreleaseindex`: Press release unique identifier.
- `pressrelease_date`: Date of the publication of the press release.
- `pressrelease_organization`: Organization that published the press release.
- `pressrelease_length_words`: Length (in words) of the press release.
- `organization_type`: Whether the organization that published the press release is an interest group or a think tank.
- `organization_ideological_leaning`: The ideological leaning of the organization that published the press release.
- `information_flow_continuous`: A measure of information selection in the parliamentary speech from the press release (i.e., the dependent variable).
- `dyad_id`: Unique identifier for politician × organization dyads.

References

Rauh, Christian, and Jan Schwalbach. 2020. The ParlSpeech V2 Data Set: Full-text Corpora of 6.3 Million Parliamentary Speeches in the Key Legislative Chambers of Nine Representative Democracies. *Harvard Dataverse,* 1–14. https://doi.org/10.7910/DVN/L4OAKN.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, and Apache Arrow. 2023. *Arrow: integration to 'apache' 'arrow'.* R package version 13.0.0.1. https://CRAN.R-project.org/package=arrow.