Replication materials for

"Politicians Unleashed? Political Communication on Twitter and in Parliament in Western Europe"

forthcoming in *Political Science Research and Methods*

by Bruno Castanho Silva and Sven-Oliver Proksch

This README provides an overview of the replication materials for the article. The **Data** section describes the main dataset required to reproduce the figures and tables in the paper and in the Online Appendix, and the auxiliary datasets containing the document frequency matrices generated with tweets, the parliamentary speech corpora, the autotranslated sentiment dictionaries from Proksch et al. 2019, and the hand-coded dataset for validation. The **Analysis** section summarises the purpose of the R script to reproduce the analyses in the paper and the Online Appendix, and the R script to process the parliamentary and Twitter data.

Note: In compliance with Twitter's Terms of Service (https://developer.twitter.com/en/developer-terms/agreement-and-policy, accessed on February 26, 2021), we cannot publicly share the raw files containing the full text of individual tweets and their associated metadata. We therefore share the document-feature-matrix obtained for each country, where each document is the full text of all tweets by one MP during the whole period of study. We also share a list of tweet IDs, which can be used to retrieve the original tweets through the Twitter API. One function that does that in R is explained in the following link: https://rdrr.io/github/ashoksiri/rtweet/man/lookup_statuses.html. It must be noted that the API will not give access to tweets that have been deleted since the date when we collected in real time, reason why the exact numbers of tweets downloaded in this way will not match those reported in the paper.

Data

The main dataset for reproducing tables and figures in the paper is **data20200228.csv**. It contains 3918 observations, with the following variables:

- **name:** name of the MP;

- **name_country**: name + country of the MP;

- **twitter_cty**: Twitter screen name of the MP + country;

- **tw**: binary whether the MP has a Twitter account (1) or not (0);

- **spoke_all**: binary whether the MP spoke at all in parliament during the period of study (1) or not (0);

- **spoke_on_eu**: binary whether the MP gave at least one speech mentioning an EU-keyword during the period of study (1) or not (0);

- **n_speeches_all:** number of speeches given by the MP during the period of study.

- **n_speeches_eu:** number of speeches given by the MP during the period of study which mention a EU-keyword;

- **country.r:** country name, where Spain is divided into Spain1 (up to June 1st, 2018), and Spain2 (from June 1st 2018 onwards);

- **party_name:** MP's party (abbreviated);

- **sent.parliament.kwic5**: sentiment on parliamentary speeches within a window of +/- 5 words around the occurrence of an EU-keyword. The calculation of this estimate from the parliamentary corpora is described and done in the **Dataprep.Rmd** file.

- **sent.parliament.kwic15**: sentiment on parliamentary speeches within a window of +/- 15 words around the occurrence of an EU-keyword;

- **sent.parliament.kwic30:** sentiment on parliamentary speeches within a window of +/- 30 words around the occurrence of an EU-keyword;

- **sent.parliament.long**: sentiment on parliamentary speeches which mention an EU-keyword, estimated from the entire speech;

- **cabinet:** whether the MP's party is in government (1) or not (0), from the ParlGov dataset;

- **screen_name**: Twitter handle;

- **pid_country:** party ID from the CHES data, pasted with the country name;

- **sentiment:** sentiment on all EU-related tweets by that MP. The calculation of this estimate from the Twitter dfm's is described and done in the **Dataprep.Rmd** file.

- **diff_to_party:** absolute distance between the MP's EU twitter sentiment (**sentiment**) and the average EU sentiment by their party on Twitter (**party_sentiment**). The calculation of this estimate from the Twitter dfm's is described and done in the **Dataprep.Rmd** file.

- **party_sentiment:** the average EU Twitter sentiment of that MP's party.

- **n.tweets.all**: total number of tweets by that MP in the period of study;

- **n.tweets.eu**: number of tweets by that MP in the period of study that mention at least one EU-keyword;

- **log_tweets.all**: natural logarithm of **n.tweets.all**;

- **log_tweets.eu:** natural logarithm of **n.tweets.eu**;

- **scaled.count.eu**: scaled number of speeches mentioning an EU-keyword (**n_speeches_eu**) given by the MP, where 0 is the country average. The calculation of this estimate from the parliamentary corpora is described and done in the **Dataprep.Rmd** file.

- **eu_salience:** EU Salience variable from the Chapel Hill Expert Survey;

- **eu_diss:** EU dissent variable from the Chapel Hill Expert Survey;

- **eu:** EU position variable from the Chapel Hill Expert Survey;

- **diff_to_party_overall:** absolute distance between the MP's EU twitter sentiment (**sent.overall**) and the average overall sentiment by their party on Twitter.

- **terms:** number of terms in parliament the MP had already served at that point;

- **male:** whether the MP is male (1) or female (0);

- **min:** whether the MP ever held a cabinet position (1) or not (0);

- **leader:** whether the MP had at the time a leadership position within the party or parliamentary group (party leader, deputy leader, parliamentary group leader, or whip – 1), or not (0);

- **size:** share of parliamentary seats held by the party, from the ParlGov dataset;

- **sent.overall:** sentiment on all other tweets by the MP, which do not mention an EU-keyword;

- **lr:** left-right scale (lrgen) from the CHES dataset;

- **country_uk:** dummy for whether the MP is from the UK (1) or not (0);

Notes on sources:

- CHES: For all countries which were part of the 2017 Flash CHES, whose surveys were fielded in January and February 2018, we used that version, since they were contemporary to the period of the tweets and speeches in this study. For Denmark, which was not part of the 2017 CHES, and the Swedish party Liberalerna, which was not covered in 2017, we used the most recent 2019 version.

- Individual MP level variables (**male, terms, min**) were scraped from MPs' Wikidata pages, and for those who did not have it the information was compiled by a research assistant based on their Wikipedia page or their page in the national parliament website.

Auxiliary Datasets

In the Data folder, the dataset **coded_data.csv** contains the hand-coding by student assistants on whether each of 3,385 tweets that our dictionary approach flagged as mentioning Europe did in fact refer to Europe. This is necessary for replicating Table A.11 and Figures A.2 and A.3 in the Online Appendix. It contains the following variables:

- **text:** the full text of the tweet;

- **Europe.x**: whether coder 1 decided that the tweet refers to Europe (1) or not (0);

- **Europe.y**: whether coder 2 decided that the tweet refers to Europe (1) or not (0);

- **lang**: language of the tweet;

- **user_id_str**: Twitter user ID of the MP who tweeted;

- **screen_name**: Twitter handle (screen name) of the MP who tweeted;

- **cty**: country

- **pty**: Collates party abbreviation + country abbreviation. For example, the AfD is marked as “afdger”, and the Conservatives as “consuk”.

The file **dictionaries_PSRM_replication.RData** contains the seven language dictionaries used in this study (English, German, French, Italian, Spanish, Swedish, and Danish) from Proksch et al. 2019.

The **corpora** folder contains one .RData file with the parliamentary corpus for the period of study for each country. These were either compiled by the authors (Italy, France, Spain, UK, Germany), or made available to us (Denmark and Sweden) in advance of the official release of the ParlSpeech v2 (Rauh and Schwalbach 2020).

The **dfms** folder contains two **.RData** files per country:

1) The files named “**country_full.RData**” contain each an R object (a list) with four document feature matrices (dfm’s) for each country: the entire dfm in the national language (dfm.cty); the dfm for EU tweets in the national language (dfm.eu); the dfm with all tweets in English from that country’s MPs (dfm.cty.en); the dfm with EU tweets in English from that country’s MPs (dfm.cty.en.eu). The UK file only has the first two (dfm.cty and dfm.eu); In all those dfm’s, each document is one MP, and the features are all words used by that MP in all tweets between February 26 and December 31, 2018 (with the exceptions listed in the paper: Italian MPs only after the March elections; Swedish MP’s only up to the September elections).

Each list also contains a dataframe called “**ids**”. This is the list with **tweet ID’s** for all tweets used in the analysis, and which can be used to download back the original tweets from the Twitter API.

2) The files named “country_EUspeeches_kwic5.RData”, which each contains a list with two objects: one called “**sent”**, which is the sentiment estimates from the speeches; and the **dfm.kwic5**, which is the dfm created from the speech data with 5 words before/after each occurrence of an EU-keyword. The code to reproduce these “**country_EUspeeches_kwic5.RData”** files from the speech data is in the **Dataprep.Rmd** script.

Analysis

The *Scripts* folder contains one Rmd script to reproduce the figures and tables (**Analysis.Rmd**) and one script to estimate the MP-level sentiment from tweets and the speeches (**Dataprep.Rmd)**. The codes in both will run as long as the working directory is set to the “Scripts” folder, and the folder structure is kept as it is uploaded here. The **Analysis.Rmd** script reproduces all Tables and Figures from both the main manuscript and the Online Appendix.

All analyses were conducted with R 4.0.2 running on macOS Mojave 10.14.6.

References

Proksch, S. O., Lowe, W., Wäckerle, J., & Soroka, S. (2019). Multilingual sentiment analysis: A new approach to measuring conflict in legislative speeches. Legislative Studies Quarterly, 44(1), 97-131.

Rauh, Christian, and Jan Schwalbach. "The ParlSpeech V2 data set: Full-text corpora of 6.3 million parliamentary speeches in the key legislative chambers of nine representative democracies." (2020).