parlScot: a dataset of 1.8 million spoken contributions from the Scottish Parliament 1999-2021

Daniel Braby

27/02/2021

Description

parlScot contains full-text vectors for all 1.8 million spoken contributions for the Scottish Parliament (up to 2021/02/03), including both speeches in parliament and in committees. Meta-data included in the main datasets include party, gender, constituency and region. Accompanying this is three additional files providing daily coverage of positions held by MSPs, membership of committees and cross-party groups. A variable 'wikidataid' is also provided for easy integration with the Comparative Legislators Database. Additionally, "Corp_Holyrood_V2.rds", provides a version compatible with Rauh & Schwalbach "ParlSpeech V2" dataset. (2020-02-27)

Schema Info

x = row number

daily_order_no = Order number by session date

order_no = Session-level order number

is_speech = Binary (0 = speech/ 1 = non_verbal)

committee = Name of committee or procedure

date = date

item = Focus of Debate

type = Type of proceeding

office held = Generic position category

display_as = Original name in transcript

name = Formatted name

speech = Full text of spoken contribution

parl_id = unique MSP identifier

party = Party name

gender = Gender

constituency = Constituency

region = Region

msp_type = Constituecy or Regional

wikidataid = wikidata identifier/crosswalk for CLD

party.facts.id = Crosswalk to PartyFacts

Supplementary

- Included in rds

gov_post = Government Position(s) held

party_post = Party Position(s) held

committee = Name of members' committee(s)

cross_party = Name of Cross Party Group(s)

cross_party_role = Role in Cross Party Group(s)

date = Date

ParlSpeech V2 Compatible Version

date = Session Date

agenda = Focus of debate

speechnumber = Order number by session date

speaker = Name of speaker

party = Name of party

party.facts.id = Crosswalk to PartyFacts

chair = TRUE/FALSE: Session Chair (Convenor/Officer/Clerk)

text = Spoken contribution

parliament = "Sco-Holyrood"

iso3country = "GBR"

Notes

Government positions, Party Positions, Cross-party Group and Committee memberships are provided as separate JSON files. All are included in the rds files, but not the csv. This is due to data structure which is incompatible with flat-file CSVs, as some MSPs hold multiple positions or affiliations on any given date.

Here is an example for joining in R

```
require(tidyverse)

parlScot <- left_join(parlScot, positions, by = c("parl_id", "date"))
```

Some positions, such as The Convenor and Deputy Presiding Officer are held by multiple people simultaneously. This requires manual coding by date and session which will be expanded on in future versions.

Non-verbal contributions have been retained (n = ~45k) for archival research of specific sessions. These can be removed to provide only speeches by subsetting is_speech == 1

```
parlScot <- filter(parlScot, is_speech == 1)
```

An additional version compatible with Rauh & Schwalbach's "ParlSpeech V2" data is available as "Corp_Holyrood_V2.rds". This has not been fully reviewed and caution is advised if analysis depends on the ordering of speeches for each day.