Codebook for the ParlText Czech Legislative Speech database

This dataset contains information on the speeches held in the Czech parliament's lower house for the time period from 1 January 1993 to 11 July 2024, with 573,784 observations. The source of the data is the official website of the Chamber of Deputies of the Parliament of the Czech Republic.
The dataset contains the following variables:

speech_id: The unique identifier of the speech. It is composed of the two-digit ISO country code, 'S' for speeches, the date of the speech's presentation in YYYYMMDD format and the running order of the speech within the given parliamentary session. For example, CZ_S_20020327_00160 for the Czech parliamentary speech held on 27 March 2002 as the 160^{th} speech of the given parliamentary session.

link: Link to the given speech on the official website of the Chamber of Deputies of the Parliament of the Czech Republic.

agenda: Agenda item title under which the speech was given. 9999 if there is none, like in the case of pre/post-agenda speeches.

electoral_cycle: The electoral cycle during which the speech was given.

speechnumber: The rank order of the speech within the given session.

speaker: The name of the speaker.

chair: Dummy variable indicating whether the speaker is the parliamentary chair. 1 if the speaker is the chair, 0 if the speaker is not the chair.

date: The date when the speech was held (YYYY-MM-DD).

speech_text: The text of the speech.

bill_ID: The id of the bill(s) discussed. It is composed of the two-digit ISO country code, 'B' for bills, the starting year of the electoral cycle in which the bill was introduced, and the official record number under which the bill was introduced. For example, CZ_B_1998_01097 for the Czech bill introduced in the electoral cycle started in 1998 under the official record number 1097. If multiple bills are related to the speech, their IDs are separated by a comma; if the speech is not connected to any bills, it is marked with 9999.

major_topic: The policy content of the bills coded based on the methodology of the Comparative Agendas Project (https://comparativeagendas.net/pages/master-codebook). The codes are mutually exclusive, all-encompassing, and all laws are classified into one category. The codes mean the following policy content of the law:

1 – Macroeconomics
2 – Civil rights
3 – Health
4 – Agriculture
5 – Labour
6 – Education
7 – Environment
8 – Energy
9 – Immigration
10 – Transportation
12 – Law and crime
13 – Social welfare
14 – Housing
15 – Domestic commerce
16 – Defence
17 – Technology
18 – Foreign trade
19 – Foreign affairs
20 – Government operations
21 – Public lands
23 – Culture
999 – No policy content