Codebook for the replication data of "Compromise or Differentiate? The Role of Partner Popularity in Coalition Parties' Strategic Parliamentary Speech"

This file provides an overview of the CSV files included in the paper's replication materials and explains the variables they contain.

C1_main_data.csv

date
Month of the data in the format yyyy-mm-dd. The day (dd) is set to "01" for all rows as a placeholder.

K_K_3s–C_C_4s
Variables follow the pattern A_B_*x*s, indicating the number of speeches by legislators from party A classified as those of party B, using the last *x* sentences of each speech.

Party abbreviations: K = Komeito, L = Liberal Democratic Party (LDP), D = Democratic Party of Japan (DPJ), C = Japanese Communist Party (JCP).

For example, K_K_3s represents the number of Komeito speeches (correctly) classified as Komeito speeches using the last three sentences, while L_D_2s represents the number of LDP speeches (mis)classified as DPJ speeches using the last two sentences.

LDP
LDP support rate (%).

cabinet
Cabinet approval rate (%).

pmid
Prime minister identifier: 1 = Yoshiro Mori, 2 = Jun'ichiro Koizumi, ..., 9 = Shinzo Abe.

DPJ.gov

Dummy variable indicating DPJ government (1 = yes, 0 = no).

HOR
Months elapsed since the previous House of Representatives election.

HOC
Months elapsed since the previous House of Councillors election.

C2_feature_importance.csv

feature
A trigram (three-word sequence).

importance
Feature importance of the trigram for the LightGBM classifier.

LDP
Frequency of the trigram in LDP legislators' speeches.

Komeito
Frequency of the trigram in Komeito legislators' speeches.

DPJ
Frequency of the trigram in DPJ legislators' speeches.

JCP
Frequency of the trigram in JCP legislators' speeches.