Data information

Variables

Columns for stylistic fronting

- `year` : The speech's year
- `date` : The speech's date (YYYY-MM-DD)
- `speech_type` : This column marks what type of speech is being given. The top 6 (out of 13) values are:
  - *ræða*: This seems to be used as a default value and is therefore not particularly useful. Makes up 69% of the data.
  - *flutningsræða*: This type is usually reffering to a prepared speech. Makes up 16% of the data.
  - *andsvar/svar*: These types most likely refer to the same speech type. This would typically be an unprepared reply. Makes up 10% of the data.
  - *um fundarstjórn*: A value associated with comments about the nature of the meeting, which suggests unpreparedness. Makes up roughly 1.5% of the data.
  - *óundirbúinn fyrirspurnatími*: A session where minority MPs ask majority ministers questions. This suggests that the MPs are prepared while the ministers are not. Makes up around 1% of the data.
- `person` : A unique ID for each MP, genreated from first name and surname. The birthyear is appended to the ID if more than one MP shares the same first- and surname.
- `sex` : The MP's sex.
- `year_born` : The MP's birthyear
- `role` : The person's role within the parliament
  - *minister*: The speaker is a minister.
  - *member*: The speaker is a member of parliament (MP).
  - *replacement*: The speaker is a replacement MP.
  - *president*: The speaker is the president of Iceland (not to be confused with the chair of parliament (forseti alþingis)).
- `speaker_type`
  - *regular*: A member of parliament that is not the chairman.
  - *chair*: The speaker is the chairman of parliament.
  - *guest*: The speaker is a guest, and not a member of parliament.
- `party_id` : A short ID, of the form `party.XX`, where `XX` is the party's ID.
- `party_name` : The full name of the party, resolved from the `party_id.

- `party_status` : The status of the speakers party in government: *Minority* or *majority*.
- `gov` : Denotes which electoral period the speech is from.
  - **LV**: The period is part of the republic of Iceland (1944-present)
  - **FV**: The period is part of the sovereignity. (1918-1944)
  - **HS**: The period is part of the home rule (1874-1918)
- `is_stylized` : 1 if stylization occurs, 0 otherwise.
- `finite_verb` : Four types of finite verbs can occur: *hafa*, (have), *vera* (be), *munu* and *skulu* (both marked as 'mod').
- `non-finite_verb` : The non-finite verb of the sentence.
- `nfv_freq` : The frequency of the non-finite verb in the Icelandic Gigaword Corpus (IGC).
- `full_text` : The full sentence where the stylistic environment occurs.
- `speech_source` : The URL link to the speech on the parliaments website https://www.althingi.is
- `speech_id` : The speech's ID from the IGC-Parla corpus.

Extra columns for corrected data

- `correction` : The correct tag for the occurence. 1 for stylization, and 0 for none. Then a -1 if the occurence does not have an environment for stylistic fronting, a 'xxx' if the speech sound/video file is not found.
- `comment` : Any comments made by the annotator. Usually regarding corrections of the annotations from the automatically tagged data.
- `here` : A column, used when a variaion of the word "hér" (here) comes after the complimentizer. Marked as 'H', but is rarely tagged.

Columns for hardspeech

Instead of the `is_stylized`, `relevant_text`, `finite_verb`, `non-finite_verb` and `nfv_freq` columns from the SF data, the data for hardspeech includes the following columns:

- `word` : The wordform as it appears in the speech, with the possibility of hardspeech application.
- `lemma` : The word's lemma.
- `pos` : The word's part-of-speech. The tags can be resolved with the MIM-GOLD tagset. For nouns the first two categories are used (noun-gender). For others only the pos category is used.
- `word_freq` : The word's frequency in the IGC.

Remarks

- If a minister has no associated party, they are most likely categorized as an extraparliamentary minister (utanþingsráðherra).
- There are occurences where a speech is attributed to another MP with the same name. Two known cases have been found:

- Speeches by EinarGudfinnsson are attributed to EinarGudfinnsson1955. The speech is dated when the latter is 11 years old.

- Speeches by HalldorAsgrimsson1947 are attributed to HalldorAsgrimsson. The speech is dated one year after the latter's death.