Case description

The Danish Parliament Corpus (2009-2017): <https://repository.clarin.dk/repository/xmlui/handle/20.500.12115/8> contains transcripts of parliamentary speeches. The case consists of doing an exploratory or confirmatory analysis (of your choosing) using the data source, or parts of the data source. You are allowed to use other data sources as well.

Ideas for cases (problem statements) could be:

* Is there an increase in the discussion surrounding e.g. data privacy, technology or artificial intelligence, and to what extend is this linked to other patterns in the data?
* Are the data well described by a set of clusters? And if so, what are they?
* Are there patterns related to the amount of words with e.g. weekday, month, holidays or weather?
* …

The corpus consists of xml files, one for each parliamentary year, running from October to the following June. The files are marked for meetings, item title and number, speeches, name and party of speakers, date, time etc. The Danish Parliament must be acknowledged as the source.

Method choices:

It is recommended that you use your pre-requisite knowledge to pre-process the data, and help each other out as much as possible on DTUinside with data handling and pre-processing issues. The methods you use for analysis, being unsupervised, supervised or a combination of these, should relate to this course (02582). Please make solid arguments for your choice of methods.