ParlaCLARIN Workshop: Creating and Using Parliamentary Corpora

Miyazaki, 7 May 2018

The Danish Parliament Corpus
- a gender study

Dorte Haltrup Hansen,
Costanza Navarretta &
Lene Offersgaard

CLARIN

ParlaCLARIN, LREC, 7th MAY 2018

UNIVERSITY OF COPENHAGEN

We have made a small study of gender differences in the new Danish Parliament Corpus. A study that was never possible before.

Question:

Is it possible to determine gender differences in the Danish parliamentary speeches by analysing:

- **Amount of speeches**
- **Duration of speeches**
- **Age of speaker**
- **Party of speaker**
- **Role in the party**

A study without linguistic or NLP analysis.

Outline

- Other relevant studies

- The Danish Parliament Corpus

- Our analysis

- Conclusions

- Future plans

Gender Studies in Political Speeches

- % female MP is generally low (2005 figures): 38% in Scandinavia, 8% in Middle East (Paxton et al. 2007).

- Female MPs tend to use more implicit meaning and more contrasting constructions unless they want to appear strong and therefore communicate in ways usually described as masculine (Sivrić and Jurčić 2014).

- Swedish female MPs talk less than the male MPs, but they talk as much as men in cases of "softer" policy areas for which they are also often the responsible in the party (Bäck et al. 2014).

Gender Studies in Political Speeches

- Features analysed by LIWC (among 74 linguistic categories) are indicating gender difference in: long words, swear words, articles, pronouns, and social words (Pennebaker et al., 2001).

- Automatic identification of the gender in the Swedish Parliament was higher for right-wing politicians than for left-wing ones and was in the range 72.8-80.1%.

Description of the Corpus

- The Danish Parliament Corpus 2009 – 2017 consists of Hansards (transcripts of parliamentary speeches) from the sittings in the Danish Parliament.

- A speech is here defined as a single intervention by a MP; It can be a question as well as a longer debate contribution.

- The corpus is marked for meetings, speeches, name of speakers, party of speakers and timing of the speeches.

- The date of birth and the gender of the speakers for this study have been extracted from additional sources.

- The Hansards are available online in pdf format but only because of personal relations it has been possible for us to achieve the data in xml format.
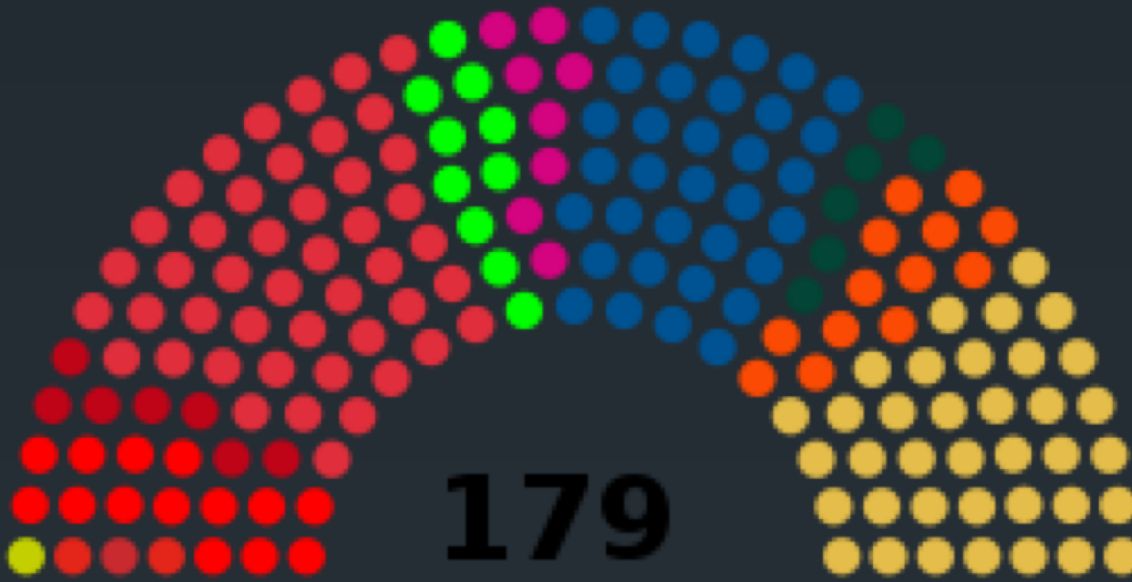
Description of the data - License

- The corpus follows the license for Open Data:

The Danish Parliament grants a world-wide, free, non-exclusive and otherwise unrestricted right of use of the data in the Danish Parliament's open data catalogue. The data can be freely:

- *copied, distributed and published,*
- *adapted and combined with other material,*
- *exploited commercially and non-commercially.*

Following the copyright act the speeches can be distributed without the consent of the speaker but only in a way where the author/speaker of each text/speech is clearly stated. Furthermore, the Danish Parliament must be acknowledged as the source.

- This correlates to CLARIN PUB BY or CC-BY
- We suggest that other CLARIN Parliament corpora could follow the same license as far as possible.

Description of the data – language characteristics

- The Hansards are exact transcripts of the speeches, but slightly edited following the guidelines:

  - ✓ The spoken language is adapted into a colloquial and syntactically coherent written language with a liberal approach to what is deemed correct language.

  - ✓ The editing is done carefully to ensure that the intentions of the speaker are clear.

  - ✓ Factual errors and slips of the tongue are corrected.

  - ✓ The appropriate formal requirement rules are observed.

- Punctuation marks are added, pauses and hesitations etc. are omitted. Therefore stylistic analysis and investigations, which include factors such as "sentence" length must take into account that spoken language has been artificially converted to written language.

Number of words in the corpus

- We have excluded comments from the chairman since they are not relevant for this study

Year	Chairman	Male	Female	Total
2009-17	1,168,618	25,104,094	13,394,453	39,667,165

- The female MPs speak app. 1/3 of the time (1/3 of the words), but the amount varies from year to year

Year	Range female	Total female
2009-17	31 – 40%	34.8 %

Seats and speeches

- Female MPs are responsible for 34.8 % of the words said, do they also represent 34.8 of the MPs?

Election	Male	Female	Total	% Female
Election 2007	113	66	179	36.9
Election 2011	111	68	179	38.0
Election 2015	113	66	179	36.9
Avg.	112	67	179	37.3

- This indicates that female speak less than male MPs

- (For the records) The amount of words corresponds to time of speaking and to the amount of speeches

Speeches	Male	Female	Total	% Female
Speeches	119,441	62,751	182,192	34.4
Speaking time, hours	2484.59	1324.78	3809.37	34.8
Words	25,104,094	13,394,453	38,498,547	34.8

Age, seats and speeches

Age	% Female MPs	% Female speeches
20-29	56.1	52.2
30-39	47.9	47.0
40-49	42.7	35.4
50-59	29.8	22.0
60-69	24.7	15.0
70-79	33.3	30.8
Total	38.7	34.4

- Young female are strongly represented while women over 50 are not.
- But from the age of 40 they do not speak as much as their representations allow them to.

Party, seats, speeches and ministers

Party	% Female MPs	% Female speeches	% Female ministers
EL	36.7	**27.2**	-
SF	53.3	52.0	33.3
ALT	40.0	39.9	-
S	36.5	35.4	36.8
Left-wing	**40.1**	**36.6**	**35.1**
RV	62.5	54.5	40,0
KF	33.3	27.2	54.6
V	29.9	52.0	41.7
DF	37.8	39.9	-
LA	29.6	35.4	50.0
Right-wing	**32.6**	**33.8**	**36.4**

- The right-wing parties have lowest pct. of female MPs.
- The female MPs of the right-wing parties speak relatively much compared to their representation while the left-wing are almost in balance except for the far left party EL.
- The right-wing parties also have more representation of ministers compared to their seats.
- So the right-wing females seem to manage their seats quit well

Male vs. female prime ministers

2 election periods (2011-15, 2015-17)	% Female under female prime minister	% Female under male prime minister
MPs elected	38.0	36.9
Ministers	37.8	38.5
Ministers, hours	44.7	30.1
Other MPs, hours	30.5	33.7

- Even though a bit more female ministers under the male prime minister, the ones under a female minister spoke much more.

- This could be due to personalities, but it could also be that female leadership encourage others to speak more.

Ministers and spokespersons

- Not only ministers play a central role in the parliament, so do spokesmen of the standing committees.

- 62.7% of the male MPs and 67.1% of the female MPs are spokesmen in one or more committees in the current election period.

- The Standing Orders of the Danish Parliament state the time allotted is: max. 20 minutes (ministers), max.10 minutes (spokesmen) and max.5 minutes (others).

- Therefore the debates are mainly done by spokespersons of the committees and the ministers.

Speaking time (2015 – 2017)	% Female
Ministers, hours	30.1
Spokesmen, hours	33.1
Other MPs, hours	36.7

Spokespersons and subjects areas

▪ Subject areas the male and female spokespersons represent and thus speak about.

Policy areas	% Female
Foreign affairs	13.8
Domestic affairs	20.0
Labour and industry	20.0
Economy	23.5
Infrastructure	24.2
Defence	25.0
Food and agriculture	26.3
Church and culture	37.5
Environment	41.2
Law	41.7
Immigration	53.3
Social	54.9
Education	60.0

Conclusion

- More Female MPs < 30y than male MPs <30y.

- Women speak less frequently than male MPs in proportion to their seats in Parliament (statistically significant difference).

- Female MPs from the left-wing speak less frequently than female MPs from the right-left party compared to their seats in Parliament.

- Ministers and spokesmen speak more frequently than simple MPs.

- Female ministers under a male prime minister give fewer speeches than female ministers under a female prime.

- Relatively more male than female spokesmen.

- Female MPs often speak about "softer" political areas for which they were responsible. Male MPs speak about "harder" subjects.

Conclusion

YES!

It is possible to determine gender differences in Danish parliamentary speeches.

Women in Danish politics tend to:
- **speak less**
- **work with other subject areas (and therefore)**
- **speak of other topics than their male colleagues**.

Future work

- Experimenting with manual and automatic classification of the speeches.

- Could we have a common classification scheme in ParlaCLARIN?

- Should we strive for a common type of license for parliamentary corpora in CLARIN?

Thank you!