Nieminen, Jeremias; Simola, Salla; Tukiainen, Janne

Working Paper

Political representation and the evolution of group differences within parties: Evidence from 110 years of parliamentary speech

Discussion paper, No. 161

Provided in Cooperation with:
Aboa Centre for Economics (ACE), Turku

Suggested Citation: Nieminen, Jeremias; Simola, Salla; Tukiainen, Janne (2023) : Political representation and the evolution of group differences within parties: Evidence from 110 years of parliamentary speech, Discussion paper, No. 161, Aboa Centre for Economics (ACE), Turku

This Version is available at:
https://hdl.handle.net/10419/298494

Jeremias Nieminen, Salla Simola, and Janne Tukiainen

Political representation and the evolution of group differences within parties: Evidence from 110 years of parliamentary speech

Aboa Centre for Economics

Discussion paper No. 161

Turku

June 2023

ISSN 1796-3133

Printed in Uniprint
Turku
June 2023

Jeremias Nieminen, Salla Simola, and Janne Tukiainen

Political representation and the evolution of group differences within parties: Evidence from 110 years of parliamentary speech

Aboa Centre for Economics
Discussion paper No. 161
June 2023

ABSTRACT

We study the long-term evolution of party demographics and the associated changes in parliamentary speech patterns of various within-party groups in Finland during 1907-2018. We find significant speech differences by gender and university education status, while other MP characteristics - age, white-collar job, first-term MP status, or urbanicity - do not predict speech patterns. We find that when female seat share began to rise in the late 1950s, there is a concurrent increase in speech differences by gender. As the representation of women increased, there was also a shift in speech topics female MPs specialized in. Additionally, we observe a sharp increase in speech differences by education when the seat share of university-educated increased in the 1960s. These results suggest that descriptive representation of these groups may play a role in changing speech patterns, and thus, in their substantive representation.

JEL Classification: D72, N44, J16, P00

Keywords: intra-party politics, parliamentary speech, descriptive representation, substantive representation

Contact information

Jeremias Nieminen
Department of Economics, Turku School of Economics, University of Turku.
Email: jeremias.nieminen (at) utu.fi

Salla Simola
Storytel
Email: salla.mari.simola (at) gmail.com

Janne Tukiainen
Department of Economics, Turku School of Economics, University of Turku.
Email: janne.tukiainen (at) utu.fi

Acknowledgements

We thank Oda Nedregård, Pamela Campa, Matti Mitrunen, Eetu Kauria, Matthew Gentzkow, Zeynep Somer-Topcu, and various participants at EPSA 2022 Annual Conference, TSE Economics Research Seminar, and Helsinki GSE Public & Labor Economics PhD Student Workshop, for useful comments or other helpful assistance. The computer resources of the Finnish IT Center for Science (CSC) and the FGCI project (Finland) are acknowledged.

1 Introduction

In recent decades, many countries have seen large changes in the demographic composition of their parliaments, such as more women being elected.[1] These changes may affect policy if different groups have different policy priorities.[2] It is therefore important to know whether these large changes in the *descriptive representation* of specific groups are associated with changes in the *substantive representation* of those same demographic groups, and to understand more generally, how changes in different groups' descriptive representation changes the behavior of MPs belonging to that group. In this paper, we aim to shed light on this by studying 1.) whether MPs with different demographic characteristics speak differently, controlling for party affiliation, 2.) if they do, whether they talk about topics that voters with similar background characteristics find important, and 3.) whether the speech patterns and topics discussed by MPs belonging to certain groups have changed as the descriptive representation of those groups has changed. We use data from a long time period of more than 100 years, which enables us to study the dynamic links between descriptive and substantive representation of different groups in the parliament, and also distinguishes our study from previous work.

To measure group differences in political views, we use 110 years of parliamentary speech data and a LASSO-penalized distributed multinomial regression method developed in Gentzkow et al. (2019).The original application of the method was to measure partisanship in political speech, but the method can be used to measure speech differences between any two groups in the parliament, and thus can also be applied in our context. This approach differs considerably from other approaches, such as that of Peterson and Spirling (2018) or Jensen et al. (2012), that have been used in earlier literature to measure speech divergence between groups. We choose to use the Gentzkow et al. (2019) method as it has been proven to be

[1] For example, according to the World Bank (2022), the average share of women in the parliaments around the world has risen from 12 % in 1998 to 26 % in 2021.

[2] There is also causal evidence that the types of politicians within parties indeed impact policy choices (see, e.g., Hyytinen et al. (2018) and Meriläinen (2022))

very robust in distinguishing the real speech differences from random noise (see Gentzkow et al. (2019) for discussion related to this).

Parliamentary speech data allows us to detect possible intra-party divides better than voting records, because speech is less constrained by party discipline than parliamentary voting (Schwarz et al., 2017). The speech difference measure we use has been used in a number of studies before, also in the Finnish context.[3] Similar to our paper, Fiva et al. (2023) study intra-party group differences in Norway during 1981-2020 using the same method. Our significantly longer time period, however, distinguishes us from Fiva et al. (2023): while their observation period starts in the 1980s, our paper uses more than 100 years of data. The long time span is necessary, because we want to study the historical changes in descriptive representation, and women's descriptive representation began to increase already in the 1950s. In fact we are to our knowledge the first to study the links between the descriptive and substantive representation of women over such a long time period that includes the large demographic shifts in representation taking place in the earlier decades of the last century.

Our main contribution is showing the historical evolution of speech patterns of MPs with different background characteristics (gender, age, district, university education status, white-collar/blue-collar profession, and first-term MP status) in Finland, controlling for party affiliation. We find that these intra-party differences in speech are the strongest in gender and education: in both of these dimensions, we find significant differences between the estimated speech differences and the placebo series. The size of the intra-party speech differences we find for gender and education are larger than between-party (left-wing vs. right-wing) speech differences estimated with the same method. This suggests that within-party demographic composition of MPs may even be more policy relevant than between-party differences. In other intra-party dimensions (first term MP status, white-collar job, age, or urbanicity), we do not find statistically significant group differences in speech.

Regarding the trends in the observed intra-party differences, our main finding is that

[3] Simola et al. (2023) study polarization *between* Finnish left-wing and right-wing parties but did not study differences between MPs belonging to different demographic groups inside parties.

intra-party differences with respect to gender have slightly increased during our observation period, simultaneously as the share of women in the parliament has been rising from around 10% to almost 50%. The largest part of the increase in speech differences occured during the period between the late 1940s and 1970s, during which women's descriptive representation also rose from 10 % to 20 %. After the 1970s, the speech differences between men and women have not increased, but on average have stayed similar to the levels observed in the 1970s. Simultaneously with the increase in speech differences between women and men, women began to talk about many more topics compared to earlier years, and began to differentiate their speech from men's speech. It can also be observed that the topics women specialized in 100 years ago were mostly related to children and family, while the issues they nowadays give speeches about are different, and usually related to education and healthcare, consistent with Hargrave and Blumenau (2022) who find that gender stereotypes do not describe reality in political speech nowadays as much as they did in the past. We also show that the topics women began to talk about, such as healthcare, are also topics women voters care about in voter surveys. We complement our main topic analyses with a structural topic model (STM) analyses. The STM approach yields a different set of topics but also suggests that women favor the "healthcare, family & children" topic, and the prevalence of that topic has been increasing since the 1960s.

The speech pattern changes we observe could stem from a behavioral change among female MPs, from changes in selection, or from both of these. In addition, there are likely also other societal trends influencing the behavior of female MPs, and thus even if the change resulted from behavioral change, it could be potentially be attributed also to many other factors besides the increase in the descriptive representation of women. If the changes resulted from a behavioral change among female MPs, one interpretation that our results would be consistent with is that women need a sufficient number – a critical mass – of other female MPs in order

to be able to state their mind.[4] The concept of critical mass is, however, controversial, and some studies have found no support for it (see, e.g., Grey (2002) and Towns (2003), or for a discussion on the critique, see Childs and Krook (2006) and Childs and Krook (2008)). Regarding the selection of MPs, we show that female MPs elected later have more education, come more often from white-collar backgrounds, and are younger. It is possible that these compositional changes would explain some of our findings as females with these characteristics may speak differently from other females. However, these changes in selection do not rule out the possibility of there being also a behavioral change among female MPs. We attempt to disentangle these different mechanisms by running topic regressions for the sample of women who were in the parliament both before and after the largest change in speech differences occured. Doing this, we find some evidence of a behavioral change as those women who were in the parliament before the increase in women's seat share and stay there, seem to start talking more about healthcare, labor issues and women's issues, and less about the elderly.

There are numerous papers demonstrating that women and men speak about different issues.[5] In a recent paper, Wäckerle and Castanho Silva (2023) conduct a cross-country analysis of the link between substantive and descriptive representation of women in Western Europe, reaching similar conclusions to those of our paper. For example, they suggest that women MPs speak differently from men both in content and style. This is similar to our finding that there are gender speech differences both between and within topics. Our paper complements this cross-country analysis by offering a within-country analysis which we can conduct due to having a long time period during which many changes in the descriptive representation of women have occured.

Our speech difference estimates along dimensions other than gender are also informative

[4]Our descriptive evidence is consistent with a critical mass of around 15-20 percent, because as the descriptive representation of women increased above 15 percent in the late 1940s, women began to specialize in many new topics, and there is an increase in speech differences between women and men. Speech differences not increasing further after the descriptive representation reached 20 % in the 1970s suggests that the critical mass is not higher than 20 %.

[5]see e.g. Lippmann (2022), Fiva et al. (2023), Catalano (2009), Osborn and Mendez (2010), Tam (2017), and Tremblay (1998)).

as they can be used to assess whether gender speech differences are larger or smaller than those found between other groups. The estimates for these various other groups are generally small except for university educated MPs. Moreover, our results suggest that when the number of university-educated MPs increased from around 40 % to around 70 % during 1960s and 1970s, there was a concurrent, relatively sharp increase in the speech differences between those who have university education and those who do not. Explanation for the finding could be related to, e.g., university educated MPs gaining a majority in the parliament, or to the composition of university educated MPs changing. Our results on university educated MPs can also inform other literatures. For example, university educated MPs have been found to contribute to the development of democratic institutions (see e.g. Rasmussen and Dahlum (2021)), which us a result supported by some of our descriptive results. For example, we find that university educated MPs are more represented in constitutional committees and they speak more in the Parliament than non-university educated MPs.

The paper proceeds as follows. Section 2 discusses theory and hypotheses. Section 3 describes institutions and data. Section 4 introduces the methods. Section 4 presents our results. Section 5 concludes.

2 Theoretical background and hypotheses

The concepts of descriptive and substantive representation were introduced in Pitkin (1967), who argued that it is important to focus on assessing whether a group's views are advocated for in the parliament, i.e., substantive representation, instead of focusing only on a group's seat share in the parliament, i.e., descriptive representation. Since then, a large literature on women's political representation has tried to establish whether there exists a link between women's descriptive and substantive representation. Many approaches have been used in the literature to measure substantive representation, but according to Cowell-Meyers and Langbein (2009), many studies in this literature approach the question either by comparing

stated priorities of female MPs to those of female voters[6], looking at policy proposals, voting and committee behavior[7], or looking at the effect of women's seat share on actual policy outcomes.[8]

Instead of these approaches, our paper measures substantive representation in parliamentary *speech*. Legislative speech as an indicator of substantive representation has some advantages and disadvantages. First, as opposed to methods where female politicians are asked about their priorities, looking at speeches shows what actually happens on the parliament floor. For example, we can study whether female politicians talk about issues that female voters find important. Although a group's views being represented in parliamentary speech does not necessarily imply an effect on policy, parliamentary speech data can be used to analyze for whom the representatives *act for*, which is the original definition of substantive representation. Although measuring substantive representation in speech is less common compared to other approaches, several studies have used parliamentary speech data to assess differences in speech patterns between men and women.[9] One potential disadvantage of our approach is that parties may have significant power over who are appointed to different committees. Thus, e.g., gender stereotypes may affect committee seat allocation, and this may feed parliamentary speech, as legislators may talk more about issues that are related to the policy area of the committee they are a member of. Although it is hard to distinguish which part of these differences is due to parties making decisions based on stereotypes, and which part is due to the preferences of the groups, we show in the Online Appendix Figures SI12 and SI13 the descriptive representation of women and university educated people in different committees.

Although we study the speech patterns of many different demographic groups in this

[6]see, e.g., Norris (1996) and Thomas and Welch (1991)

[7]see, e.g., Bratton (2005), Dolan (1998), Hogan (2008), Saint-Germain (1989), Swers (1998), and Tolbert and Steuernagel (2001)

[8]see, e.g., Hessami and da Fonseca (2020), Berkman and O'Connor (1993), Bratton and Ray (2002), Caiazza (2004), Chen (2021), Crowley (2004), and Thomas and Welch (1991)

[9]See, e.g., Clayton et al. (2017), Fiva et al. (2023), Lippmann (2022), Catalano (2009), Bäck et al. (2014), Blumenau (2019), Pearson and Dancey (2011), Bäck and Debus (2019), Osborn and Mendez (2010), Tam (2017), and Tremblay (1998)

paper, our main focus is on women as the concept of substantive representation was originally developed to study women's political representation. Therefore, the first hypothesis of our paper is:

- **H1**: Female and male MPs speak differently, also when controlling for party affiliation

The meaningfulness of the gender speech difference estimate can then be assessed by comparing it to within-party speech difference estimates for other groups. As it is an established finding in the literature that, e.g., female MPs speak about different topics than men, the main focus of our paper is somewhat different. The key measure used in this paper to measure substantive representation is the *speech differences* between different groups of MPs. In the earlier literature on intra-party group differences, the method we use has only been used in Fiva et al. (2023), who focus on contemporary speech differences between various intra-party groups. The metric we use also has the advantage that it is possible to separate within-topic differences from between-topic speech differences. To the extent that the speech differences come from specialization, they should be visible in the between-topic partisanship. Within-topic partisanship, on the other hand, can capture also differences within topics, which can be attributed to either different individual preferences inside topics, e.g., different framings, or to differences in style. Separating these components has the advantage of not only understanding the specialization dynamics but also whether the differences in individual preferences are larger than differences in specialization patterns.

In addition to understanding whether there are intra-party differences in how various intra-party groups speak, we want to uncover the mechanisms behind possible changes in observed speech differences. We expect that if there are changes in how different groups speak, it is linked to changes in the descriptive representation of that group. The mechanism, however, can be either behavioral change of the affected group, or change in the composition of the group (selection). Regarding the development of speech differences between genders, we hypothesize the following:

- **H2**: When the descriptive representation of women increases, women will increasingly talk about topics that women voters care about

- **H3**: Because women talk more about issues that are different from what men talk about (H2), measured speech differences between men and women will increase as a result of this specialization

One important, although controversial, idea in the literature on descriptive and substantive representation is that a certain level of descriptive representation is needed for different groups (usually women) to speak their mind in the parliament (Paxton and Hughes, 2007). There is, however, a significant amount of uncertainty about what the size of this threshold is, and if such a threshold even exists. Earlier papers have estimated that the *critical mass* required is somewhere between 15 percent and 30 percent (see, e.g., Beckwith and Cowell (2007), Bystydzienski (1992), and Studlar and Mcallister (2002)). There is also a consensus in the literature that when the share of women is less than 15 percent, women will have very limited influence in the parliament (Beckwith, 2007). Similar to the critical mass idea, we hypothesize the following:

- **H4**: There is a critical mass (or a critical threshold) of women that is needed for female MPs to speak their mind in the parliament (measured by how differently men and women speak compared to each other). After this critical threshold is exceeded, additional increases in women's descriptive representation do not lead to increased speech differences between male and female MPs.

The above hypothesis (H4) is not necessarily exactly the same concept that is referred to as a critical mass in the literature. Instead of hypothesizing that the whole discussion climate would become more inclusive of female MPs, we only hypothesize that this critical mass would be enough for women to express their views. The hypothesis we formulate here only expects that the *speech differences* between male and female MPs will stop growing at a point where female MPs feel confident to discuss about issues they find important. This

would not necessarily be enough for true representation of women voters positions in actual policies that are made.

3 Institutional context

3.1 The Finnish parliament and plenary sittings

Finland is a parliamentary republic with multiple parties, and this has been the case as long as the Finnish parliament has existed. There are currently (as of December 2022) 9 political parties in the Finnish parliament. The Finnish parliament was founded in 1907, when it replaced the previous system where only the Four Estates of nobility, clergy, bourgeois and peasants were represented. Our paper uses data from a time period starting from the very beginning of the Finnish parliament. During the 110-year time span (1907-2018) we study, there have been large changes affecting the Finnish parliament. Although, for example, women have been elected to the parliament from the very beginning[10], their descriptive representation has risen from the less than 10 % to almost 50 % during the century. Apart from these large demographic shifts that are the topic of this paper, the power of the parliament in decision-making was very limited before the Finnish independence, i.e., during the years 1907-1917: all laws had to be approved by the Russian Emperor (Jyränki and Nousiainen, 2006). During 1915 and 1916 the parliament held no sessions due to Russia increasing their control over Finland during the First World War (Paloheimo, 2007).

Plenary sittings gather MPs on multiple weekdays to give speeches about policy issues they or their parties find important. Plenary sessions in the Finnish parliament are organized usually during Tuesdays, Wednesdays, Thursdays and Fridays. The Speaker of the Parliament decides who gets to speaks. Committee memberships affect who gets to speak first or more easily on specific topics. When group speeches are given, the parliamentary group decides who gives those speeches. Based on discussions with Finnish MPs, it is also perceived to be

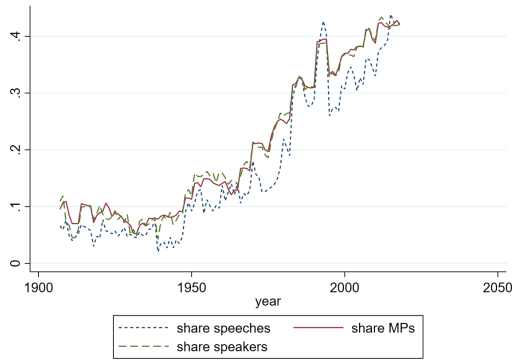
[10]Finland elected the first female MPs in the whole world in 1907.

easier for committee heads, as well as party leaders and long-term MPs, to speak. Speaker slots that have been requested in advance are allocated based on that list.

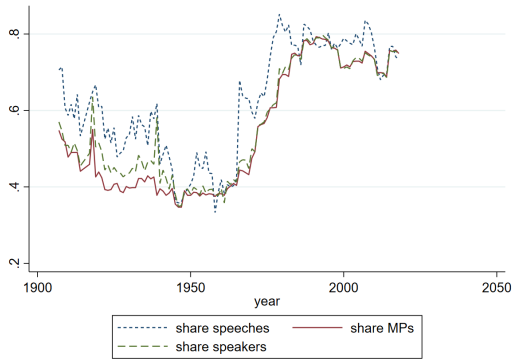
3.2 Descriptive representation and speech frequencies of different groups

As can be seen from Panel (a) of Figure 1, the share of women out of MPs, speakers, and speech was low (under 10 percent) until 1950, but since then has been rising steadily. In 2018, the share of women in the parliament is almost 50 percent. From the same figure, we can also observe that women's share of speeches is a little less than their share of MPs and speakers almost every year, which suggests that women seem to speak less frequently, or they give shorter speeches in the parliament, compared to men. Panel (b) of Figure 1 depicts the shares of university educated MPs. An interesting observation can be made from the figure: university-educated MPs seem to speak more frequently in the parliament than MPs without university education. This disparity persists in the whole observation period, although it seems to have been higher in the earlier years. Similar observation can be made from the share of speech of white-collar workers (Panel d), although in their case, the disparity is only visible in the early years.

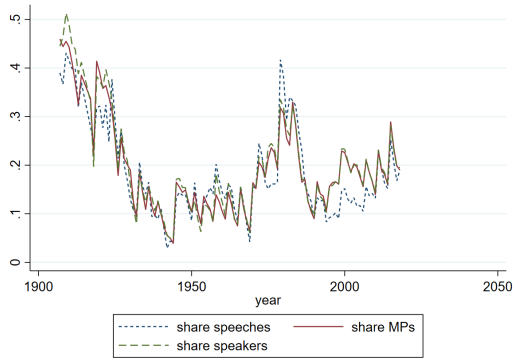
Panel (c) of Figure 1 depicts the shares of MPs under the age of 40. It can be seen from the figure that the share of young (under 40) MPs was high during the first years of the Finnish parliament but declined until the 1950s. After that, the share of young people briefly increased until it peaked in the late 1970s. The peak in late 1970s is almost concurrent with the peak in left-right polarization which happened in 1970s (see Simola et al. (2023)). Panel (e) of Figure 1 shows that the share of MPs from the capital region (Helsinki + Uusimaa) has risen steadily during the observation period, but interestingly, there does not seem to be an increasing trend in their share of speech. Based on the Figure, MPs from the capital region spoke relatively more often in earlier years and less nowadays.



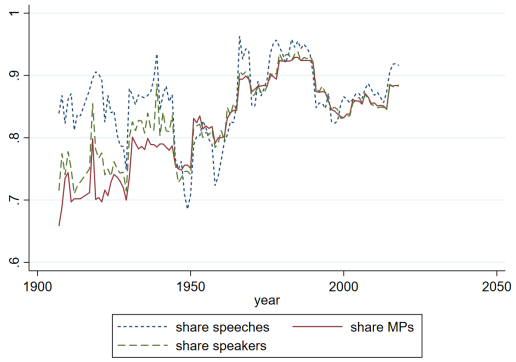
(a) Share women



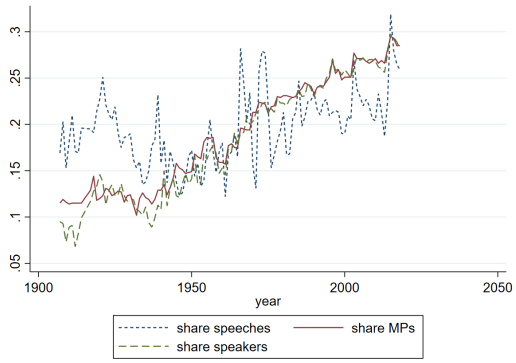
(b) Share university educated



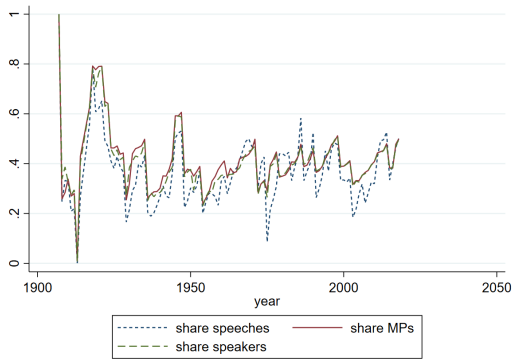
(c) Share under 40 years old



(d) Share in white-collar job



(e) Share from most urban region

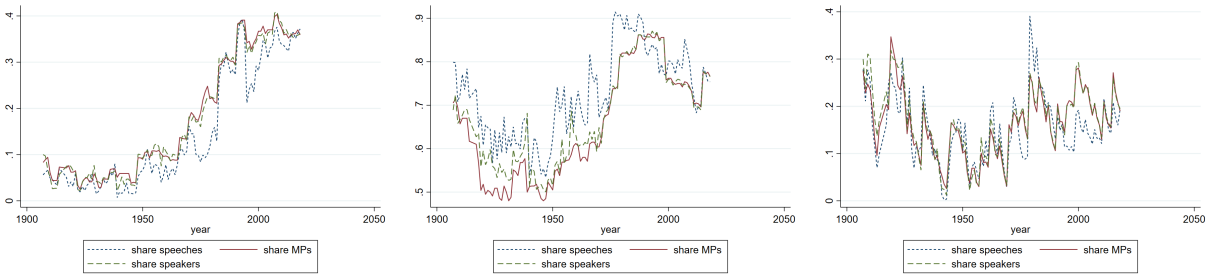


(f) Share first term MPs

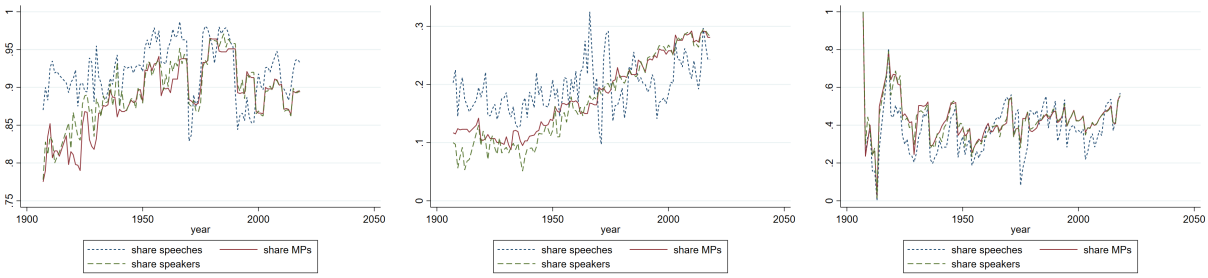
Figure 1: Descriptive figures

Notes. Figure presents different demographic groups' shares out of all MPs, all speakers, and all speeches. Speakers include MPs who speak at least once during the year.

Panel A: Right-wing parties

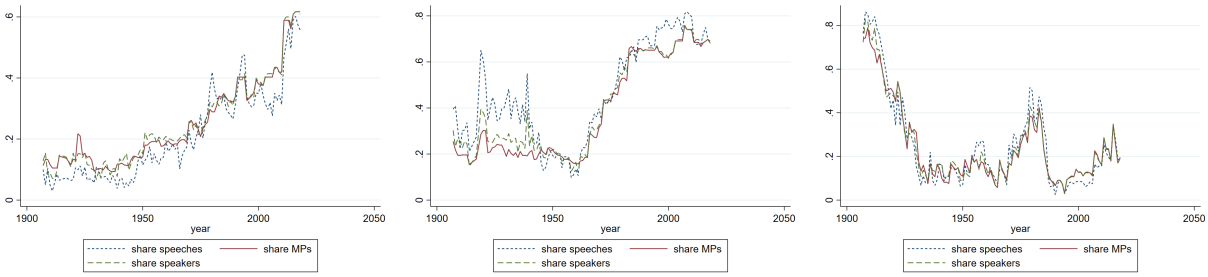


(a) Share women (b) Share university educated (c) Share under 40 years old

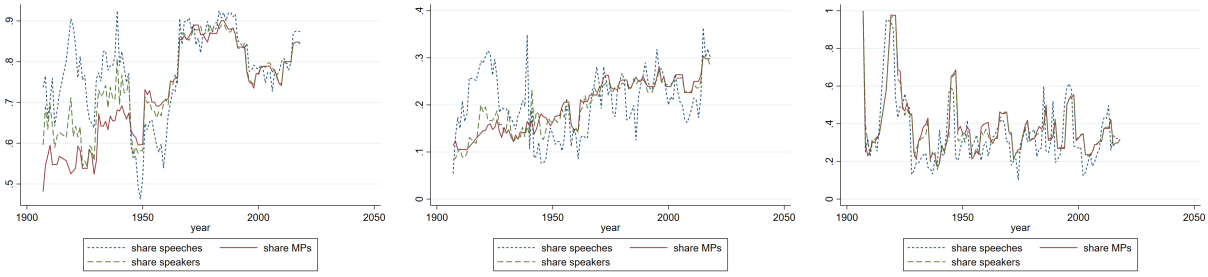


(d) Share in white-collar job (e) Share from most urban region (f) Share first term MPs

Panel B: Left-wing parties



(g) Share women (h) Share university educated (i) Share under 40 years old



(j) Share in white-collar job (k) Share from most urban region (l) Share first term MPs

Figure 2: Descriptive representation in right-wing and left-wing parties

Notes. Figure presents different demographic groups' shares out of all MPs, all speakers, and all speeches within right-wing and left-wing parties. Speakers include MPs who speak at least once during the year.

4 Data

4.0.1 Parliamentary speech data

We use a dataset on parliamentary speeches in the Finnish parliament covering years (sessions) 1907-2018, except for the years when the parliament did not gather (1915, 1916). The dataset is constructed in Simola et al. (2023) who extracts parliamentary speeches from the transcripts of plenary sessions and links them to MP data from the publicly available MP register. The analysis dataset we use is on speaker-session level, and consists of phrase counts for different bigrams (two-word phrases). Details on the dataset and its construction and preprocessing can be found in Simola et al. (2023).

Information about education and occupation was not previously used in Simola et al. (2023). Thus, we added this information to the data from the publicly available MP register. Using these, we are able to identify MPs who have university education and who have white-collar job backgrounds, both of which are attributes that we are interested in when we investigate intra-party speech differences by different characteristics. University educated MPs are defined as those whose education column include words indicating Finnish university degrees or professions that require university education. Occupations are classified to white-collar and blue-collar jobs by hand without using any official classification. See Appendix SI7 in the supporting information (p. 15) for keywords used when constructing university education, white-collar job, and blue-collar job variables.

4.0.2 Data on voter preferences

In order to study the links between descriptive and substantive representation, we need information about the preferences of different groups of voters. To analyze voter preferences, we use publicly available survey datasets from the Finnish Social Science Archieve. The used datasets are the following[11]:

[11] We cannot share these datasets publicly, but they are available for research purposes from the Finnish Social Science Archieve.

- *Finnish Voter Barometer (Puolueiden ajankohtaistutkimus) 1973-1992*
- *Finnish Voter Barometer (Puolueiden ajankohtaistutkimus) 1995-2002*
- *National Election Study (Eduskuntavaalitutkimukset) 2003-2015*

5 Methods

5.1 Measuring intra-party group differences

To measure speech differences along a single bivariate characteristic, such as gender or education level, we use a method introduced in Gentzkow et al. (2019) who applied the method to study partisanship in the United States. In our case, we utilize the *average partisanship* measure introduced in their paper, but instead of studying partisanship, we use it to study differences between demographic groups within parties. Similar approach has been used in Fiva et al. (2023) who study within-party group differences, but from a more contemporary perspective.

The approach we use is based on Gentzkow et al. (2019), and differs from other approaches that have been used in the earlier literature to measure speech divergence between groups, such as Peterson and Spirling (2018). We choose to use the Gentzkow et al. (2019) method, because as discussed in Gentzkow et al. (2019), their method seems to be more robust to noise in the data. For our research question, a structural topic model (STM) would also be one possible approach. Although STM is not our main approach, we present some analyses utilizing a structural topic model in Online Appendix SI10.

We explain below the Gentzkow et al. (2019) approach. As we are interested in group differences based on politician characteristics, the notation used by Fiva et al. (2023) is the most applicable to our case. We approximate multinomial logistic regression with Poisson regressions using R package *distrom*. Following Gentzkow et al. (2019) and Fiva et al. (2023), the phrase choice probabilities in the model of speech are

$$q_{jt}^{D_i(x_{it})} = \frac{e^{u_{ijt}}}{\sum e^{u_{ijt}}} \tag{1}$$

where u_{ijt} is

$$u_{ijt} = \alpha_{jt} + x'_{it}\gamma_{jt} + \phi_{jt}1_{i\in\{D_t=1\}} \tag{2}$$

In equations above, the indicator D_{it} is a dummy variable getting value 1 if an individual has a certain characteristic (e.g. being a woman, or having university education) and 0 otherwise. In our research, we look at group differences based on 6 different background characteristics: gender, university education, living in Uusimaa region, age under 40, and being a first-term MP. Since we want to estimate intra-party group differences, and not just plain differences between the groups, we include party fixed effects in x, similar to Fiva et al. (2023). We do not observe committee memberships yearly, so we cannot include committee fixed effects. The coefficient ϕ_{jt} is the difference in how often phrase j is used in session t by the group we are interested (i.e., for whom $D_{it} = 1$). Following Gentzkow et al. (2019) and Fiva et al. (2023), polarization in time t at covariates x is defined as

$$\boldsymbol{\pi}_t(x) = \frac{1}{2}\boldsymbol{q}^{D_i=1}(\boldsymbol{x})\boldsymbol{\rho}_t(x) + \frac{1}{2}\boldsymbol{q}^{D_i=0}(\boldsymbol{x})(1 - \boldsymbol{\rho}_t(\boldsymbol{x})) \tag{3}$$

where ρ_{jt} is the expected posterior that the observer assigns $D_i = 1$ after hearing the phrase j:

$$\rho_{jt} = \frac{q_{jt}^{D_{it}=1}(x)}{q_{jt}^{D_{it}=0}(x) + q_{jt}^{D_{it}=1}(x)} \tag{4}$$

The polarization measure is the average of the expression in (3) over different values of covariates x:

$$\bar{\pi}_t = \frac{1}{N_t} \sum_i \pi_t(x_{it}) \tag{5}$$

This average polarization measure describes how different speech is between the bivariate groups that are compared. It tells the probability of a neutral observer correctly deducing the bivariate group (e.g. university education status or gender) to which the MP belongs, after hearing just one phrase. The measure gets values between 0.5 and 1, where 0.5 indicates no differences in speech, while the value of 1 would indicate that the groups compared use completely different words.

5.2 Estimation

Because the estimation of the multinomial logistic regression model described above would be computationally challenging, we follow Gentzkow et al. (2019) and estimate the model using the distributed multinomial regression method which approximates multinomial logit with Poisson regressions that can be run in parallel. As the speech data has a very large number of phrases that can be spoken, small sample bias is a big problem in estimation. The measure we use, developed in Gentzkow et al. (2019) takes this into account by including a LASSO penalty to the estimated coefficients of interest. We use their model and R package *distrom* to estimate the model presented above with the LASSO penalty. Following Gentzkow et al. (2019) and Fiva et al. (2023), the estimator with the LASSO penalty is given by minimizing the following negative likelihood function:

$$\sum_t \sum_i [m_{it} exp(\alpha_{jt} + \gamma_{\mathbf{j}}' \mathbf{x_{it}} + \phi_{jt} \mathbb{1}\{L_{it}\}) - c_{ijt}(\alpha_{jt} + \gamma_{\mathbf{j}}' \mathbf{x_{it}} + \phi_{jt} \mathbb{1}\{L_{it}\}) + \lambda_j |\phi_{jt}|] \tag{6}$$

where $\lambda_j|\phi_{jt}|$ is the LASSO penalty imposed. This penalty limits the small sample bias. The negative log likelihood function for the factorial of J Poisson distributions is the sum of all J likelihoods:

$$(7) \quad \sum_j \{\sum_t \sum_i [m_{it} exp(\alpha_{jt} + \gamma_{\mathbf{j}}' \mathbf{x_{it}} + \phi_{jt} \mathbb{1}\{L_{it}\}) - c_{ijt}(\alpha_{jt} + \gamma_{\mathbf{j}}' \mathbf{x_{it}} + \phi_{jt} \mathbb{1}\{L_{it}\}) + \lambda_j |\phi_{jt}|]\}$$

5.3 Confidence intervals

Confidence intervals are calculated by subsampling, i.e., the polarization series is estimated 100 times for different subsamples and the speech difference estimate is re-estimated in each of the subsamples. Subsampling is used instead of more standard bootstrapping methods, because the estimator with LASSO is a non-smooth estimator, and thus only subsampling can be used, as subsampling requires less assumptions than even nonparametric bootstrap (Gentzkow et al., 2019). We conduct the subsampling by resampling 20 % of the data without replacement. Results are robust to changing the subsample size to 30 % and these results are reported in Appendix SI5 in the supporting information (p. 13). The nominal coverage of the confidence intervals is set to 80 %, similarly as in Gentzkow et al. (2019) and Simola et al. (2023). This is to allow easier comparison to these previous studies. We present 95 % confidence intervals in Appendix SI3 in the supporting information (p. 11). The conclusions remain the same when 95 % CIs are used. The confidence intervals are calculated as follows:

$$(8) \qquad 0.5 + \exp[\log(\widehat{\pi}_t - 0.5) - Q^k_{t(11)}/\sqrt{N}], 0.5 + \exp[\log(\widehat{\pi}_t - 0.5) - Q^k_{t(90)}/\sqrt{N}]$$

where $Q^k_{t(p)}$ is the p'th quantile of the distribution for $Q^k_t = \sqrt{N_k} * [\log(\widehat{\pi}_{t,k} - 0.5) - \log(\overline{\widehat{\pi}}_t - 0.5)]$. $\overline{\widehat{\pi}}_t$ is the average of $\widehat{\pi}_{t,k}$ over draws $k = 1, \ldots, 100$.

5.4 Phrase partisanship

In order to study which phrases are most predictive of different background characteristics of MPs, we define *phrase polarization* such that those words that affect our estimate of speech differences the most, are more *polarized*. This is the same definition Gentzkow et al. (2019) use to define phrase polarization in their paper: the polarization of an individual phrase reflects how the expected posterior of assigning a politician to a particular group changes on average if the particular phrase is removed from the vocabulary. Thus, it can get both positive and negative values, depending on which group tends to use that word. The expression for phrase polarization is:

$$\zeta_{jt}(x_{it}) = 0.5 - 0.5 \sum_{k \neq j} \left(\frac{q^L_{kt}(x_{it})}{1 - q^L_{jt}(x_{it})} + \frac{q^R_{kt}(x_{it})}{1 - q^R_{jt}(x_{it})} \right) \rho_{kt}(x_{it})$$

6 Results

6.1 Group differences in parliamentary speech

Figure 2 presents estimates and confidence intervals for intra-party speech differences.[12] In addition, each plot includes a placebo estimate, which is calculated by assigning the group indicator randomly. Standard errors are calculated by subsampling, i.e., the polarization series is estimated 100 times in order to be able to conduct statistical inference.

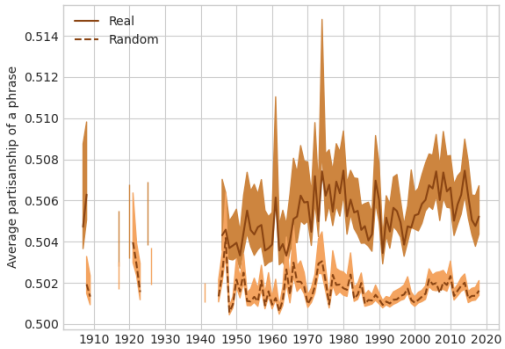
Panel (a) of Figure 2 presents actual and placebo estimates for intra-party speech differences by gender. Because the share of women (see Figure 1, Panel a) was very small in the early 1900s, we are not able to estimate confidence intervals for speech differences in some years during the early 1900s. Because of inadequate number of observations, we omit those years from the analysis. However, since the late 1940s we can estimate the series without gaps. We can clearly see significant speech differences: the confidence intervals of the polarization series for gender do not overlap with those of the placebo estimate. Furthermore, it also seems that polarization by gender has increased slightly over time, since the estimate is higher in 2010s compared to 1940s, for example. The largest continuous period of increasing speech differences happened between the late 1940s and early 1970s. The descriptive representation of women also began to increase during 1950s, so the beginning of the increase in the descriptive representation of women coincides with the beginning of the increase in within-party speech differences between women and men. The size of the estimate is between 0.504 and 0.507, which is quite large, since Fiva et al. (2023) estimate that contemporary gender speech differences are around 0.501-0.503 in Norway. These intra-party estimates are also somewhat larger than between-party (left-wing vs. right-wing) differences estimated in Finland (around 0.502-0.504, see Simola et al. (2023)). This suggests that within-party demographic composition may be even more policy relevant than the between party seat shares.

[12] Results without party fixed effects are reported in Appendix SI4 (Supporting Information, p. 12).

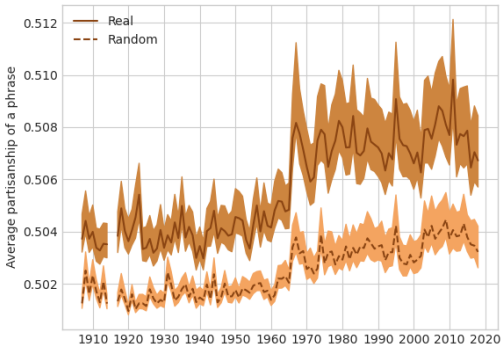
Panel (b) of Figure 2 depicts intra-party speech differences by university education. The figure shows that highly educated MPs clearly speak differently than the other MPs, and have spoken differently during the whole observation period. There is an increase in speech differences based on university education status in the 1970s, approximately coinciding with the increase in the descriptive representation of university educated people in the parliament. The level of speech differences by education level are around 0.504 before this increase and around 0.508 after the increase. Because earlier research has suggested that government/opposition status drives a large part of speech differences between politicians (see Lauderdale and Herzog (2016)), we also show that the results are robust to adding a government opposition dummy to the model (see Appendix SI4, Figure SI5).

Panels (c), (d), (e), and (f) show results by age, white-collar job status, living in capital region, and first term MP status, respectively. We do not find any significant speech differences based on these factors. Of the insignicant characteristics, the also the magnitudes seem small in urbanicity, first-term MPs, and young MPs. For white collar MPs, the magnitude seems large but the method is unable to detect the 'signal' (real speech differences) from the random noise in the data, which we suspect is due to the fact that only a very small number of MPs come from non-whitecollar backgrounds, and we use the same group shares when constructing the random series.

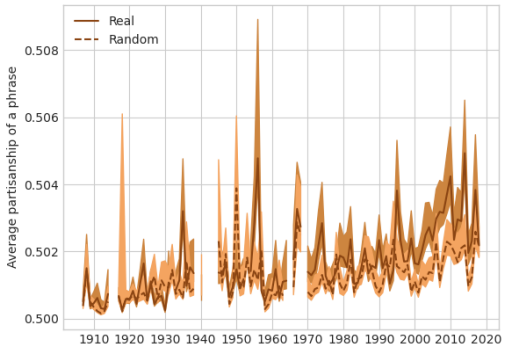
Based on these results, within-party differences seem to be the strongest between the genders and between university and non-university educated MPs: in both of these dimensions, speech differences also seem to have slightly increased with time. Results regarding gender are consistent with the results in Fiva et al. (2023), as gender is one of the most important drivers of intra-party differences also in their contemporary data. However, they did not study intra-party differences between university educated MPs and other MPs.



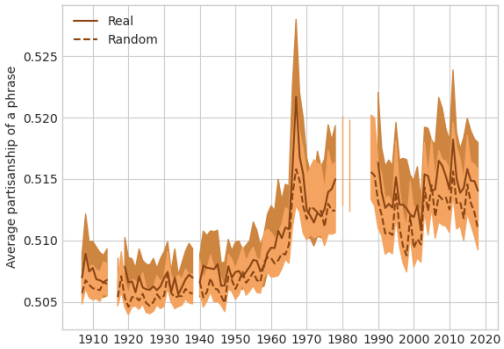
(a) Gender



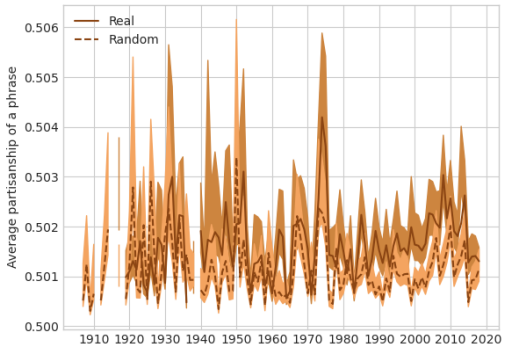
(b) University education



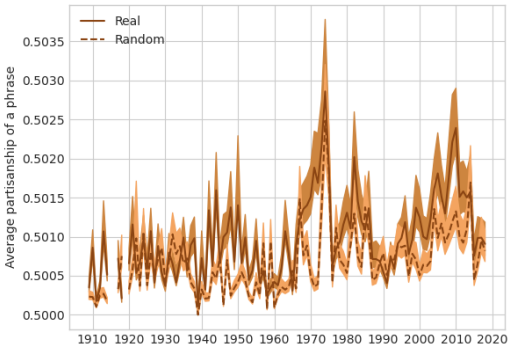
(c) Young (under 40 years old)



(d) White-collar job



(e) Most urban districts (Helsinki and Uusimaa)



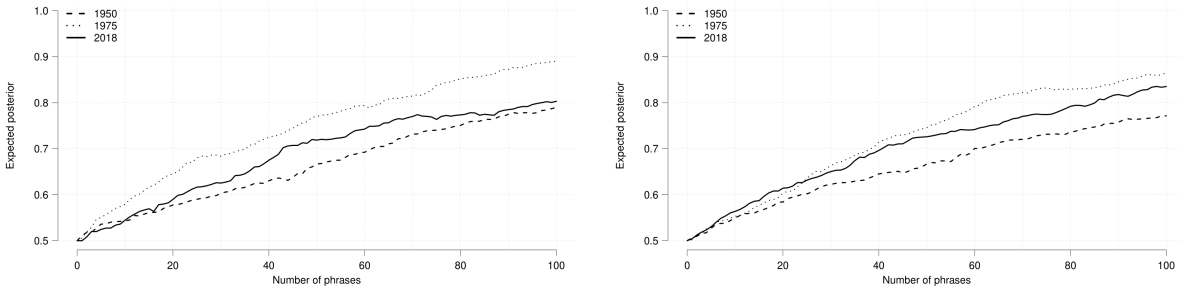
(f) First term MP

Figure 3: Intra-party group differences

Notes. Figure presents intra-party group differences by six different demographic groups. Estimation includes year and party fixed effects. Confidence intervals have 80% nominal coverage. See Appendix SI3 (Supporting Information p. 11) for 95 % nominal coverage CIs.

6.1.1 Magnitude

It may be challenging to infer the magnitude of our findings on speech differences by gender and education based on the main estimate, as the main estimate is the probability of guessing the group right after hearing just *one phrase*. In order to show how the expected posterior of guessing the group right changes when the observer hears more words, we show the expected posterior for up to 100 phrases in Figure 4. 100 phrases is a relatively large amount of speech as our phrases are bigrams, i.e., they contain two words. Unfortunately, we do not have time stamps available in the data and thus cannot estimate, e.g., how many phrases can be spoken during 1 minute. The estimated magnitudes are very sizeable for both gender and education: they are significantly larger compared to magnitudes in Fiva et al. (2023). It can also be observed that for both gender and education, the levels of speech differences in 2018 are higher than they were during the 1950s. Speech differences with respect to both gender and education peaked in 1970s, which was a time when political polarization in general was very high in Finland (Simola et al., 2023).



(a) Gender

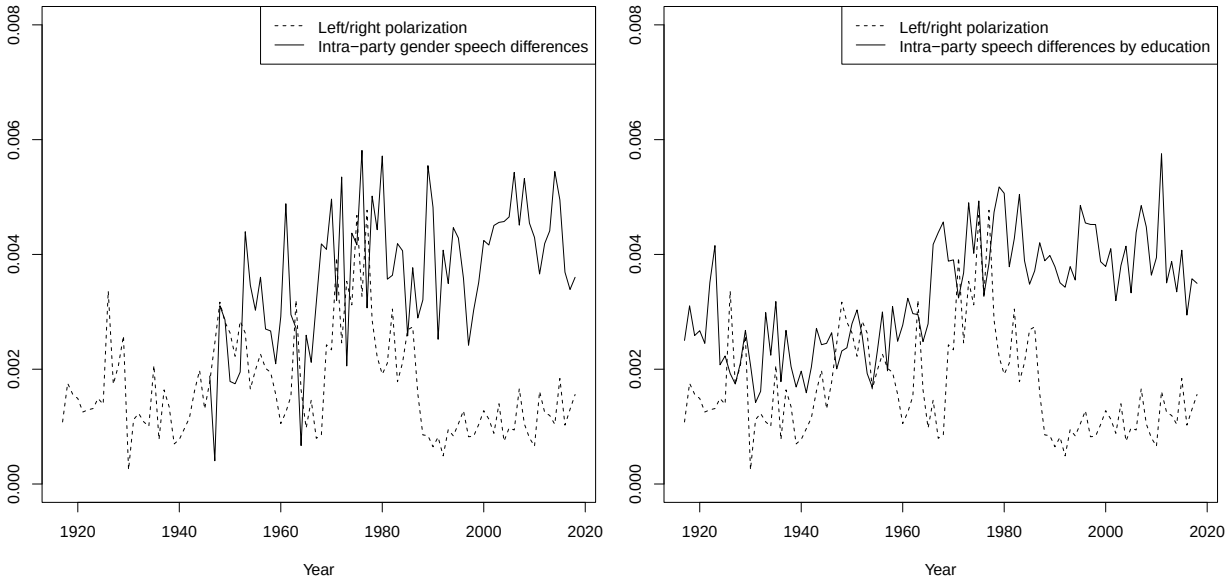
(b) University education

Figure 4: Magnitude of speech differences: expected posterior when neutral observer hears x number of phrases

Notes. Figure presents the expected posterior of guessing the group identity right after hearing up to 100 words. The analysis is conducted similarly as in Gentzkow et al. (2019).

Figure 5 compares the size of the intra-party speech differences to left-right partisanship estimated with the same data. The random series has been substracted from both within-party and left-right time series. The figure shows that within-party speech differences by both

gender and education are larger than between-party (left-right) speech differences in Finland.



(a) Gender (b) University education

Figure 5: Comparison to left-right speech differences

Notes. Figure presents the levels of within-party speech differences compared to left-right speech differences in Finland.

6.2 Topics, specialization and content

6.2.1 Topic classification

We want to define topics in order to study 1.) whether the between-topic or within-topic component of speech differences is larger for speech differences by gender and education, and 2.) which topics have been favored by female MPs and university educated MPs at different times. Our main approach to define topics is the same manual topic classification procedure as also used in Gentzkow et al. (2019). This approach is described below.

When defining the topics, we first look at most polarizing phrases for each group from each year, and classify those bigrams to 18 different topics. Based on these phrases, we search for keywords that the bigrams are comprised of. For example, if we have a phrase such as *työtätekev.luok* (working class) we would use both *työtätekev* and *luok* as keywords when we search for phrases that contain specific keywords. Once we have found all bigrams that contain specific keywords, we drop 'false matches', i.e., bigrams that do not belong to a topic even if they contain a specific keyword. All 18 topics are manually validated in this way. Topics are not varied over time. To mitigate the problem related to using researcher discretion in this approach, we also conduct an unsupervised, structural topic model (STM) analysis in the Online Appendix SI10.

6.2.2 Decomposing speech differences to within-topic and between-topic components

In this section, we examine whether the speech differences found between the genders and the educational groups reflect different views within issues or specialization in specific issues. To achieve this, we decompose speech differences into within-topic and between-topic speech differences in a similar fashion as in Gentzkow et al. (2019). We conduct manual classification of most polarized phrases to topics, and compare speech differences within those topics to speech differences between different topics. Gentzkow et al. (2019) conduct a similar

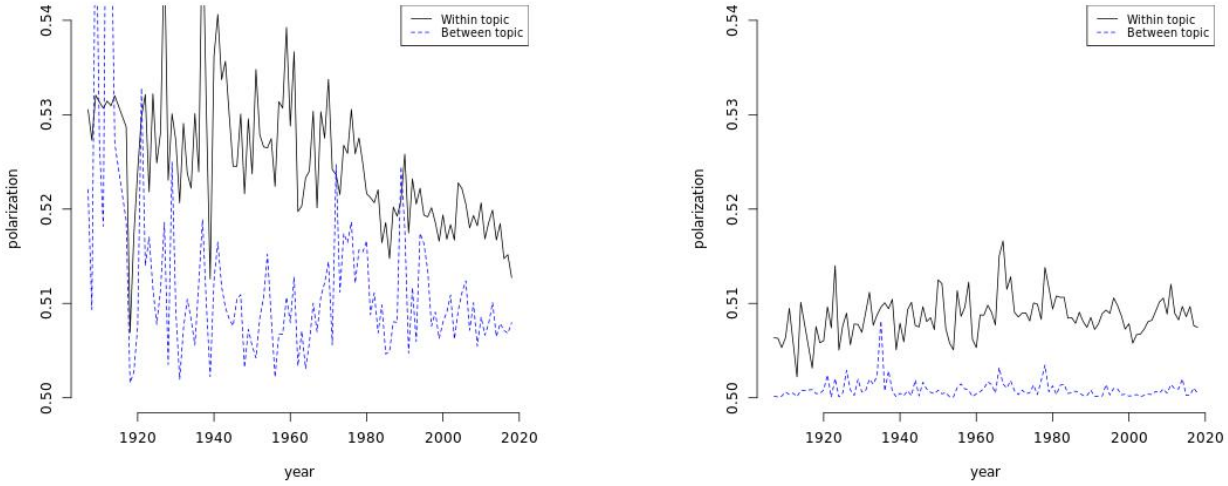
analysis of left-right polarization in the United States and find that estimates of within topic polarization are much larger than those of between-topic polarization.

The measure of *within-topic speech differences* is defined such that in a specific topic, it is the probability of correctly guessing the group to which the MP belongs to after hearing them speak just one phrase, when the vocabulary consists only of words in that specific topic. The average within-topic speech differences are then calculated by weighting each topic by its frequency in the data. In turn, *between-topic speech differences* are measured by the probability of correctly guessing the group to which the speaker belongs to after knowing the topic but not the exact phrase the MP chooses.

The results from the decomposition are presented in Figure 5. For gender, the estimates before the 1940s may not be reliable, as the number of women in the parliament is very low before the late 1940s. We find that within-topic speech differences are larger than between-topic speech differences both for gender and university education. This suggests that different genders and different educational groups use different phrases when they talk about the same topics. This can be suggestively interpreted as indicating that parliamentary speech reveals to some extent individual policy preferences rather than parties only selecting certain MPs to specialize in certain topics.

This does not, however, preclude that different groups would also specialize in topics. Although the between-topic component is much smaller than the within-topic one, there are still large differences in how often different groups talk about different topics, especially between the genders: the estimated within-topic speech differences for gender are quite large at approximately 0.51 on average. Moreover, it seems that the between-topic component is driving the increase in gender speech differences observed during 1950s-1970s. In contrast, for higher educated MPs it is the within-topic differences that are driving the observed increase in the 1960s and 1970s. Regarding specialization, we also find evidence that the share of female MPs out of all MPs who talk about family issues, or the share of university educated MPs out of all MPs who speak about education policy, are much larger than the seat shares

of those demographic groups in the parliament (see figures in Appendix SI1 in Supporting Information 1, p. 3), suggesting that some specialization in topics is occurring.



(a) Gender

(b) University education

Figure 6: Decomposition to within and between topic speech differences

6.2.3 Topics favored by female MPs and university educated MPs

Appendix SI1 in supporting information (p. 3) shows different groups' shares of speakers when a specific topic is discussed. For women, we find that before the 1940s, specialization was quite low, except for three topics: family, children and women's issues. When the share of women in the parliament began to increase in the late 1940s, women began to become over-represented (relative to their seat share) in many policy topics, such as healthcare, social policy, and education, while simultaneously becoming also more over-represented in those three topics (family, children, and women's issues) that they already were talking about before the increase in women's descriptive representation. Women's increased representation is also reflected in how often different topics are discussed in the Parliament: in absolute terms, issues women care about began to increase only after more women were represented in the Parliament. Also in relative terms, some topics women care about, such as healthcare, became more common after women's descriptive representation increased. See Appendix SI10

for the topic phrases mentioned in absolute terms (p. 19) and in relative terms (p. 20).

6.2.4 Most polarized phrases

In Appendix SI2 (supporting information p. 5), we present 5 most polarizing words in four years (1912, 1950, 1971, 2017). These show what types of policy topics different demographic groups have discussed in the parliament. Supporting Information table A.1 shows most polarized words by gender: words related to the rights of children and the elderly, as well as to healthcare and education, are clearly present in the list of most polarized words for women. This is consistent with women's preferences: Chattopadhyay and Duflo (2004) find that village leaders invest more in public goods that are relevant to the needs of their own genders. For men, the list includes more general words as well as economy-related words. In the list of most polarized words by university education status, it can be seen that some education related words, such as *education reform* and *provider of education*, appear in 2017. In earlier years, the words are less education-related, but there are still a number of those such as *parents of students*, *mental work*, and *foreign language*.

We present 8 most polarizing words for each topic in Appendix SI6 in the supporting information (p. 14). These are only estimated for gender and education, since these are the two dimensions where we found statistically significant differences in speech.

The phrases about the elderly were polarizing during 1940-1995 but not anymore, and family words were polarizing up until 1970. Nowadays the healthcare topic contains more polarizing phrases for women. Until 1970 the labor word *sama.palk*, meaning equal pay, was polarizing for women, but polarization of that phrase has since declined. The word *tyon.perh*, which is related to work-life balance, has been polarizing for women in the latest decades. When discussing about crime, the use of word *rikosl.20*, which is the law on sexual violence, is polarized for women. For men, the word *rikosl.10*, the law on the freedom of faith, is polarized.

6.2.5 Comparison to voter preferences

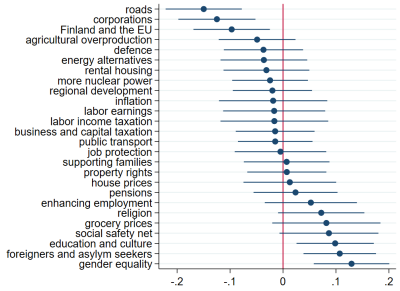
Based on the voters surveys (see Figure 7), women care about gender equality, fighting crime, healthcare, unemployment benefits, care for the elderly, minorities, social safety net in general, as well as education & culture. If we compare these to the topic plots, we can see that the issues women voters care about are closely related to issues that women MPs talk about. First, caring about gender equality was the most predictive factor of the survey respondent being a woman. When we look at the share of women when women's issues are talked about, we see that women have always been over-represented, relative to their seat share, when gender equality is discussed. Moreover, once women's descriptive representation began to increase, women's over-representation when women's issues are talked about increases further.

Regarding the other topics women voters care about, 6 out of 8 are also topics where women MPs specialize: these are healthcare, unemployment benefits, social safety net in general, care for the elderly, and education & culture. Topic plots show women MPs specialize in healthcare, social policy, elderly, education, and culture - which roughly correspond to aforementioned issues women voters care about. Women voters also care about fighting crime, but for crime we find that women MPs talk about crime as often as their share is in the parliament, i.e., they are not over-represented there.

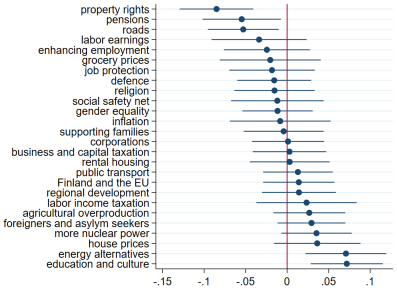
Males care about EU, firms and entrepreneurship, (un)employment, roads, and defence. Indeed, it seems that men MPs are over-represented when foreign policy is discussed, and they like to use words related to the EU (see most polarized phrases). Therefore, it seems that male voters preference for EU issues is an issue that men MPs act for in the parliament. Regarding firms and entrepreneurship, economy and finance topic is favored by male MPs, suggesting that male MPs also care about firm performance and the private sector.

University educated voters mostly care about education and culture, energy, housing, racism, gender equality, firms, the European Union, the environment, drug abuse, and smaller public sector. When we look at topics favored by highly educated MPs, we see that university-educated MPs are over-represented when education policy is discussed, culture, economy and

finance or foreign policy are discussed. Thus, at least in this way, university-educated MPs seem to talk about issues many university educated voters find important.

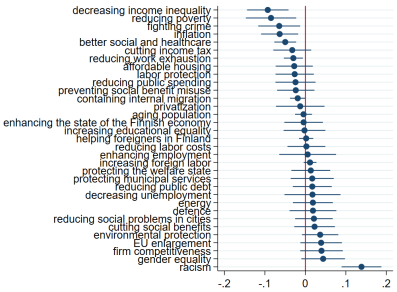


(a) Women voters' preferences (1973-1990)

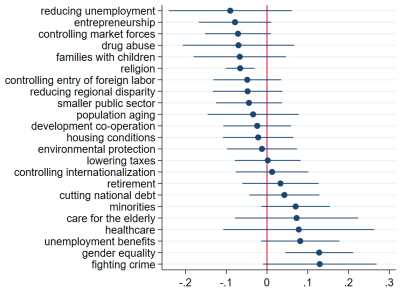


(b) University educated voters' preferences (1973-1990)

(c) Women voters' preferences (1992-2005)



(d) University educated voters' preferences (1992-2005)



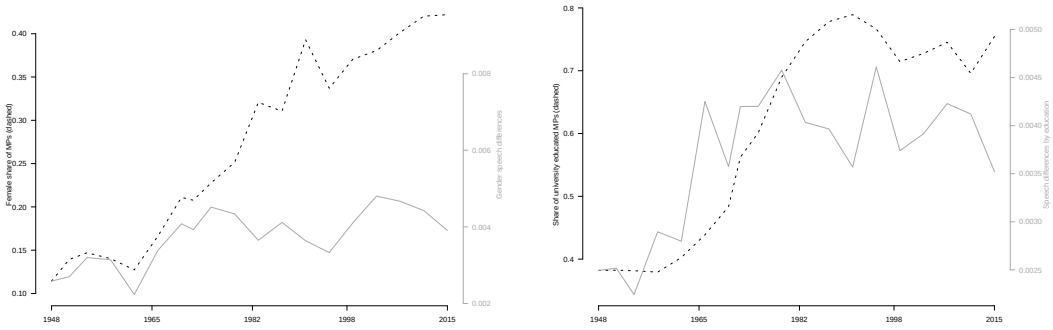
(e) Women voters' preferences (2003-2015)

(f) University educated voters' preferences (2003-2015)

Figure 7: Voters' preferences

Notes. Figure presents regression coefficients from regressions where the outcome variable is the bivariate group characteristic (female in the left hand side panel and university education status in the right-hand side panel) and independent variables are indicators for different policy issues, getting value 1 if the individual finds the issue important.

6.3 Links between descriptive and substantive representation



(a) Gender (b) University education

Figure 8: Speech differences compared to descriptive representation

Notes. Figure compares speech differences by gender (panel a) and by education (panel b) to the descriptive representation of women and university educated MPs, respectively.

This section discusses the link between descriptive and substantive representation. The link is discussed primarily for women, as the concept of substantive representation is originally defined to study female representation, and because the concept is not very suitable to be applied when discussing the representation of university educated MPs. Nevertheless, we still also discuss briefly whether the descriptive representation of university educated MPs is linked to speech patterns and how.

Figure 8 presents the development of gender speech differences together with the evolution of women's decriptive representation. The figure shows that gender speech differences increase during a period when women's descriptive representation increases from 10 % to 20 % but do not further increase after that. Regression results shown in Table 1 and the results shown in Figure 9 paint a similar picture as they suggest the association between women's seat share and gender speech differences is strong before the 1970s but after the decriptive representation of 20 % is exceeded, speech differences seem to have reached a level that stays constant on average, even when women's descriptive representation increases substantially more during later decades. This does not, of course, mean that this relationship would

necessarily be causal, as there are many different societal trends that may be behind both increasing women's descriptive representation as well as changing how women speak in the Parliament. However, the association between these two variables is strikingly strong in the pre-1970s Finland.

Table 1: Association between the descriptive representation of women and speech differences by gender (years aggregated by election terms)

	Outcome: Gender speech differences		
	years 1945-2018	**years 1945-1969**	**years 1970-2018**
Panel A: Levels			
seat share of women	0.0045*** (0.001)	0.0203** (0.007)	0.0003 (0.0016)
Number of observations	19	6	13
Panel B: Changes			
seat share of women	0.0037 (0.0111)	0.08480 (0.0409)	0.0059 (0.0102)
Number of observations	19	6	13

Notes. Tables shows results from a linear regression model where the dependent variable is the level or change in the speech differences between female MPs and male MPs, and the independent variable is the seat share of women. Standard errors in parentheses. Confidence levels: * 0.1 ** 0.05 *** 0.01

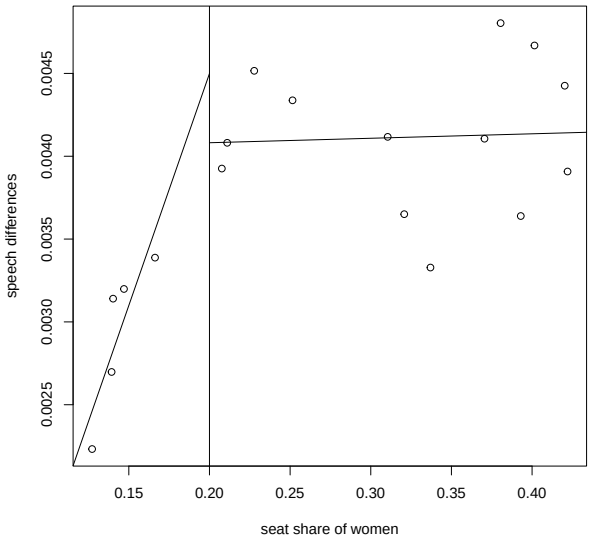


Figure 9: Association between the descriptive representation of women and gender speech differences

Notes. Figure is a scatter plot that has the seat share of female MPs on the x axis and the speech differences between male and female MPs (with the random series substracted) on the y axis. The lines are linear model fits calculated using a regression model where speech differences are predicted by the seat share of the group in question.

As can be seen from Figure 8, speech differences by education are flat when the seat share of university educated is constant, i.e., during the period until 1965, and then again during 1982-2015. Coinciding with the large increase in the seat share of university educated MPs (from 40 % to around 80 %) between years 1965 and 1982, there is a concurrent increase in speech differences by education. This is also visible in regression results in Table 2 showing that the association (in levels) is also statistically significant. This association could stem from university educated MPs gaining a majority in the parliament, or from compositional changes among the university-educated MPs resulting from the large expansion of higher education in Finland during that time.

Table 2: Association between the descriptive representation of university educated MPs and speech differences by university education status

	Outcome: Speech differences, education
Panel A: Levels	
seat share of university-educated	0.0048*** (0.0006)
Number of observations	48
Panel B: Changes	
seat share of university-educated	0.003 (0.003)
Number of observations	47

Notes. Tables shows results from a linear regression model where the dependent variable is the level or change in the speech differences between university educated MPs and other MPs, and the independent variable is the seat share of university educated MPs. Standard errors in parentheses. Confidence levels: * 0.1 ** 0.05 *** 0.01

As it has been suggested that the critical mass concept could be more useful if it was used within parties (Beckwith, 2007), we also want to discuss the link between descriptive representation and speech differences within party blocks (left-wing and right-wing parties). Online Appendix Figure SI6 shows speech differences estimates within parties and Figure SI7 presents analyses otherwise similar to Figure 8, but separately for left-wing and right-wing party blocks. Here, we also include a longer time period even though the estimates before 1950s may be quite unreliable/noisy due to a very low number of women in the Parliment. It can be observed that the change during years 1950-1970 seems to be driven mainly be right-wing parties, which makes sense, as women's seat share in left-wing parties is over 15 % during most years, while in right-wing parties their seat share is much lower.

Behind the results regarding the link between women's descriptive and substantive representation, there can of course be many explanations, only one of which is that descriptive representation would directly affect substantive representation. Our paper argues the two

most likely explanations are *selection*, i.e. changes in MP composition that happen during the same time, or *behavioral change*. We are not able to completely rule out either of these, although we display more supportive evidence for the behavioral channel. Regarding selection, we find evidence of MP composition changing around the time period when the biggest changes in speech patterns occured (see Online Appendix SI13), but the changes in speech do not continue as long as the change in selection does, which does not support the selection argument. Nevertheless, it is still a possibility that selection is the driving force as we cannot separate the effect of selection from any other confounding trends. Regarding behavioral change, when we compare female MPs who were in the parliament both before and after the largest shift in speech differences happened (also shown in Online Appendix SI13), we show that those women began to talk more about healthcare, labor issues and women's issues. This supports further the behavioral change mechanisms, but we cannot make very firm conclusions about behavioral change due to a lack of statistical power.[13]

7 Conclusion

Our paper offers a century-long overview of the evolution of the composition of political parties and that of the speech differences between various within-party group in the Finnish parliament. We document that there exists intra-party group differences in parliamentary speech by gender and education, while other characteristics, such as urbanicity, white collar occupation, or age, do not seem to predict speech patterns in the parliament. Our extensive set of descriptive evidence is also consistent with many earlier papers arguing that female MPs speak differently from male MPs, as well as with evidence that female MPs speak less than men (see Bäck et al. (2014) who argue women MPs speak less because discussions revolving more on 'hard' policy issues) and that the topics women nowadays talk about are different than the topics they used to talk about before (see Hargrave and Blumenau (2022)).

[13]Due to a small number of observations, we are not able to distinguish changes in most of the other topics or in the aggregated "all female-favored topics" variable despite relatively large point estimates.

Our paper also shows that simultaneously as the descriptive representation of women and university educated people, respectively, increased, speech differences by gender and education also increased. For gender, we also find that women began to specialize many topics, such as healthcare and social policy, simultaneously as their descriptive representation increases. The topics women begin to specialize in are ones that also female voters find important based on electoral surveys.

If these changes we observed resulted from a behavioral change, our results would be consistent with the view that when the descriptive representation of female MPs is low, they may not be able to speak their mind in the Parliament, and thus translating descriptive representation of women into substantive representation requires large enough descriptive representation. However, it has to be noted that our findings only indicate that there were simultaneous changes in the descriptive and substantive representation of women. Many other changes may also have occured during the times when these changes happened, so it is likely that many other forces are also at play.

Our results could plausibly be generalized at least to other Nordic countries, since findings in Fiva et al. (2023) regarding contemporary speech differences in Norway are very similar to our estimates in the last decades. Regarding generalizability to other countries, one needs to be cautious, since the Finnish political system is more homogeneous than that of many other countries. The results are supportive of concerns over the lack of representation also in other context than politics. For example, the speakers on top policy panels still average three men for every woman. This male dominance of Europe's most important policy conferences may have ripple effects on policies and, as a result, national laws and European regulations (Grabbe, 2018).

According to publicly available data on women's descriptive representation in other countries during 1970-2010[14] (see Appendix SI9 in the supporting information, p. 18), women's descriptive representation was around 10-15% in countries such as Albania, Azerbaijan,

[14]This data (Vanhanen (2011)) is available from the Finnish Social Science Archieve.

Bangladesh, and Zambia in 2010.[15] During the last decade, women's descriptive representation may have increased in the parliament also in the aforementioned countries. If the critical mass is somewhere between 10% and 15%, which is a possible reading of our results, it would predict a disproportionate increase also in women's substantive representation in these countries. Based on the low gender shares globally our results suggest that women are not represented substantively in a vast number of countries.

[15]The results presented in Appendix SI9 in the supporting information (p. 18) also indicate that many developed countries have had very low descriptive representation until the very latest decades.

References

Bäck, Hanna and Marc Debus. 2019, 8. "When Do Women Speak? A Comparative Analysis of the Role of Gender in Legislative Debates". *Political Studies* 67 (3): 576–596.

Bäck, Hanna, Marc Debus, and Jochen Müller. 2014. "Who Takes the Parliamentary Floor? The Role of Gender in Speech-making in the Swedish Riksdag". *Political Research Quarterly* 67 (3): 504–518.

Beckwith, Karen. 2007, 3. "Numbers and newness: The descriptive and substantive representation of women". *Canadian Journal of Political Science* 40 (1): 27–49.

Beckwith, Karen and Kimberly Cowell. 2007. "Sheer Numbers: Critical Representation Thresholds and Women's Political Representation". *Perspectives on Politics* 5 (3): 553–565.

Berkman, Michael B and Robert E O'Connor. 1993. "Do Women Legislators Matter? Female Legislators and State Abortion Policy". *American Politics Quarterly* 21 (1): 102–124.

Blumenau, Jack. 2019. "The effects of female leadership on women's voice in political debate". *British Journal of Political Science* 51 (2): 750–771.

Bratton, Kathleen A. 2005. "Critical Mass Theory Revisited: The Behavior and Success of Token Women in State Legislatures". *Politics and Gender* 1 (1): 97–125.

Bratton, Kathleen A and Leonard P Ray. 2002. "Descriptive Representation, Policy Outcomes, and Municipal Day-Care Coverage in Norway". *American Journal of Political Science* 46 (2): 428–437.

Bystydzienski, Jill M. 1992. *Women Transforming Politics: Worldwide Strategies for Empowerment*. Bloomington: Indiana University Press.

Caiazza, Amy. 2004. "Does women's representation in elected office lead to women-friendly policy? Analysis of state-level data". *Women and Politics* 26 (1): 35–70.

Catalano, Ana. 2009. "Women Acting for Women? An Analysis of Gender and Debate Participation in the British House of Commons 2005–2007". *Politics and Gender* 5 (1): 45–68.

Chattopadhyay, Raghabendra and Esther Duflo. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India". *Econometrica* 72 (5): 1409–1443.

Chen, Li Ju. 2021, 2. "Female policymakers and educational expenditures: cross-country evidence". *European Journal of Law and Economics* 51 (1): 129–155.

Childs, Sarah and Mona Lena Krook. 2006. "Should Feminists Give Up on Critical Mass? A Contingent Yes". *Politics and Gender* 2 (4): 522–530.

Childs, Sarah and Mona Lena Krook. 2008, 10. "Critical mass theory and women's political representation". *Political Studies* 56 (3): 725–736.

Clayton, Amanda, Cecilia Josefsson, and Vibeke Wang. 2017, 6. "Quotas and women's substantive representation: Evidence from a content analysis of ugandan plenary debates". *Politics and Gender* 13 (2): 276–304.

Cowell-Meyers, Kimberly and Laura Langbein. 2009. "Linking Women's Descriptive and Substantive Representation in the United States". *Politics and Gender* 5 (4): 491–518.

Crowley, Jocelyn Elise. 2004. "When Tokens Matter". *Legislative Studies Quarterly* 29 (1): 109–136.

Dolan, Julie. 1998. "Support for women's interests in the 103rd congress: The distinct impact of congressional women". *Women and Politics* 18 (4): 81–94.

Fiva, Jon H, Oda Nedregård, and Henning Øien. 2023. "Group Identities and Parliamentary Debates". https://www.jon.fiva.no/docs/FivaNedregardOien.pdf.

Gentzkow, Matthew, Jesse M. Shapiro, and Matt Taddy. 2019. "Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech". *Econometrica* 87 (4): 1307–1340.

Grabbe, Heather. 2018. "When women's voices are heard". https://www.politico.eu/article/when-womens-voices-are-heard.

Grey, Sandra. 2002. "Does Size Matter? Critical Mass and New Zealand's Women MPs". *Parliamentary Affairs* 55 : 19–29.

Hargrave, Lotte and Jack Blumenau. 2022. "No Longer Conforming to Stereotypes? Gender, Political Style, and Parliamentary Debate in the UK *". *British Journal of Political Science* 52 (4): 1584–1601.

Hessami, Zohal and Mariana Lopes da Fonseca. 2020, 6. "Female political representation and substantive effects on policies: A literature review". *European Journal of Political Economy* 63 .

Hogan, Robert E. 2008. "Sex and the Statehouse: The Effects of Gender on Legislative Roll-Call Voting n". *Social Science Quarterly* 89 (4).

Hyytinen, Ari, Jaakko Meriläinen, Tuukka Saarimaa, Otto Toivanen, and Janne Tukiainen. 2018. "Public Employees as Politicians: Evidence from Close Elections". *American Political Science Review* 112 (1): 68–81.

Jensen, Jacob, Suresh Naidu, Ethan Kaplan, Laurence Wilse-Samson, David Gergen, Michael Zuckerman, and Arthur Spirling. 2012. "Political Polarization and the Dynamics of Political Language: Evidence from 130 Years of Partisan Speech". *Brookings Papers on Economic Activity*: 1–81.

Jyränki, Jaakko and Antero Nousiainen. 2006. *Eduskunnan muuttuva asema*. Edita.

Lauderdale, Benjamin E. and Alexander Herzog. 2016. "Measuring political positions from legislative speech". *Political Analysis* 24 (3): 374–394.

Lippmann, Quentin. 2022, 3. "Gender and lawmaking in times of quotas". *Journal of Public Economics* 207 .

Meriläinen, Jaakko. 2022. "Political Selection and Economic Policy". *The Economic Journal* 132 (648): 3020–3046.

Norris, Pippa. 1996. "Women Politicians: Transforming Westminster?". *Parliamentary Affairs* 491 (1): 89–102.

Osborn, Tracy and Jeanette Morehouse Mendez. 2010, 1. "Speaking as women: Women and floor speeches in the Senate". *Journal of Women, Politics and Policy* 31 (1): 1–21.

Paloheimo, Heikki. 2007. "Eduskuntavaalit 1907-2003". In H. Paloheimo and A. Ollila (Eds.), *Kansanedustajan työ ja ariki*, pp. 173–369. Suomen eduskunta.

Paxton, Pamela and Melanie Hughes. 2007. *Women, Politics, and Power: A Global Perspective*. 2455 Teller Road, Thousand Oaks California 91320 United States : SAGE Publications, Inc.

Pearson, Kathryn and Logan Dancey. 2011, 12. "Elevating women's voices in congress: Speech participation in the house of representatives". *Political Research Quarterly* 64 (4): 910–923.

Peterson, Andrew and Arthur Spirling. 2018. "Classification Accuracy as a Substantive Quantity of Interest: Measuring Polarization in Westminster Systems". *Political Analysis* 26 (01): 120–128.

Pitkin, Hanna. 1967. *The Concept of Representation*. Berkeley: University of California Press.

Poggione, Sarah. 2004. "Exploring Gender Differences in State Legislators' Policy Preferences". *Political Research Quarterly* 57 (2): 305–314.

Rasmussen, Magnus Bergli and Sirianne Dahlum. 2021. "Educated to Democratize? Education and Elite support for Universal suffrage". https://www.researchgate.net/publication/350618055.

Saint-Germain, Michelle A. 1989. "Does Their Difference Make a Difference? The Impact of Women on Public Policy in the Arizona Legislature*". *Social Science Quarterly* 70 (4).

Schwarz, Daniel, Denise Traber, and Kenneth Benoit. 2017. "Estimating Intra-Party Preferences: Comparing Speeches to Votes". *Political Science Research and Methods* 5 (2): 379–396.

Simola, Salla, Jeremias Nieminen, and Janne Tukiainen. 2023. "A century of partisanship in Finnish political speech". ACE Discussion Paper No. 160. https://ace-economics.fi/kuvat/dp160.pdf.

Studlar, Donley T and Ian Mcallister. 2002. "Does a critical mass exist? A comparative analysis of women's legislative representation since 1950". *European Journal of Political Research* 41 : 233–253.

Suomen Gallup and Helsingin yliopisto. 2007. "Puolueiden ajankohtaistutkimus 1973-1990: yhdistetty aineisto [sähköinen tietoaineisto]. Versio 1.0 (2007-08-16). Yhteiskuntatieteellinen tietoarkisto [jakaja],".

Suomen Gallup and TNS Gallup Oy. 2008. "Puolueiden ajankohtaistutkimus 1992-2005: yhdistetty aineisto [sähköinen tietoaineisto]. Versio 1.1 (2008-01-30). Yhteiskuntatieteellinen tietoarkisto [jakaja]".

Swers, Michele L. 1998. "Are Women More Likely to Vote for Women's Issue Bills than Their Male Colleagues?". *Legislative Studies Quarterly*: 435–448.

Tam, Waikeung. 2017, 10. "Women Representing Women? Evidence from Hong Kong's Semi-Democratic Legislature". *Representation* 53 (3-4): 201–218.

Thomas, Sue and Susan Welch. 1991. "The Impact of Gender on Activities and Priorities of State Legislators". *Western Political Quarterly* 44 (2): 445–456.

Tolbert, Caroline J. and Gertrude A. Steuernagel. 2001. "Women lawmakers, state mandates and women's health". *Women and Politics* 22 (2): 1–39.

Towns, Ann. 2003. "Understanding the Effects of Larger Ratios of Women in National Legislatures: Proportions and Gender Differentiation in Sweden and Norway". *Women and Politics* 25 (1-2): 1–29.

Tremblay, Manon. 1998. "Do Female MPs Substantively Represent Women? A Study of Legislative Behaviour in Canada's 35th Parliament". *Canadian Journal of Political Science* 31 (3): 435–465.

Vaalitutkimuskonsortio, Lauri Karvonen, Heikki Paloheimo, Sami Borg, Kimmo Grönlund, and Elina Kestilä-Kekkonen. 2017. "Eduskuntavaalitutkimukset 2003-2015: yhdistetty

aineisto [sähköinen tietoaineisto]. Versio 1.0 (2017-11-03). Yhteiskuntatieteellinen tietoarkisto [jakaja].”.

Vanhanen, Tatu. 2011. “Naisten edustus kansallisissa parlamenteissa 1970-2010 [sähköinen tietoaineisto]. Versio 4.0 (2011-06-22). Yhteiskuntatieteellinen tietoarkisto [jakaja]. http://urn.fi/urn:nbn:fi:fsd:T-FSD2183”.

Wäckerle, Jens and Bruno Castanho Silva. 2023, 1. “Distinctive Voices: Political Speech, Rhetoric, and the Substantive Representation of Women in European Parliaments”. *Legislative Studies Quarterly*.

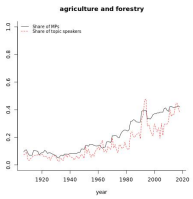
Political representation and the evolution of group differences within parties: Evidence from 110 years of parliamentary speech

Supporting Information

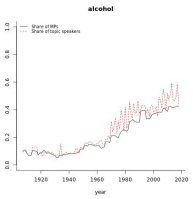
June 5, 2023

Contents

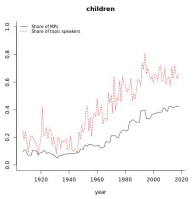
SI1: Specialization in topics (group shares of topic speakers)



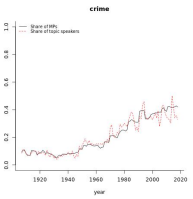
(a) Agriculture and forestry



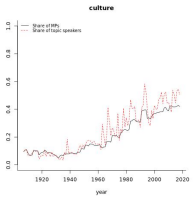
(b) Alcohol



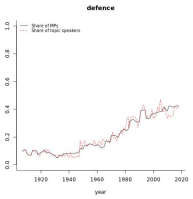
(c) Children



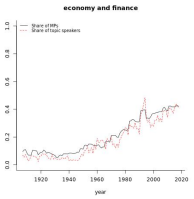
(d) Crime



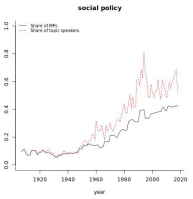
(e) Culture



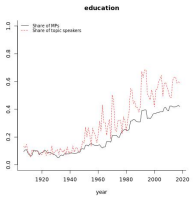
(f) Defence



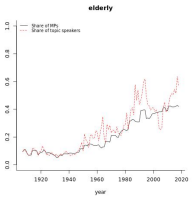
(g) Economy and finance



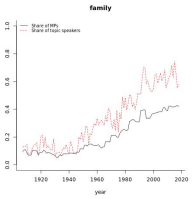
(h) Social policy



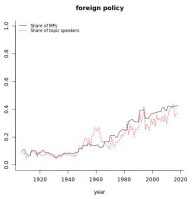
(i) Education



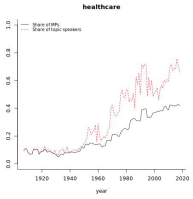
(j) Elderly



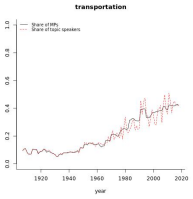
(k) Family



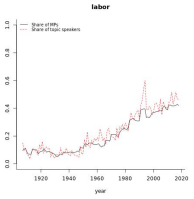
(l) Foreign policy



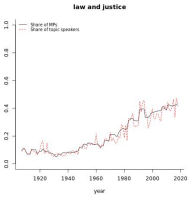
(m) Healthcare



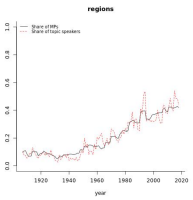
(n) Transportation



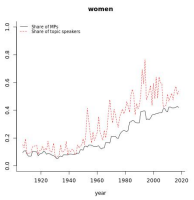
(o) Labor issues



(p) Law and justice



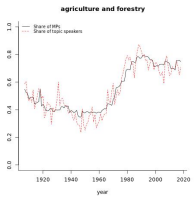
(q) Regions



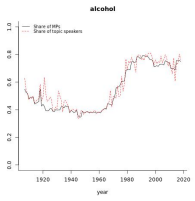
(r) Women

Figure SI1: Specialization (gender): share of female MPs of topic speakers

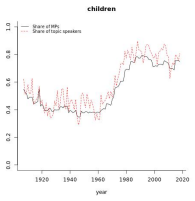
Notes. Figures present women's share of MPs who speak about a specific topic.



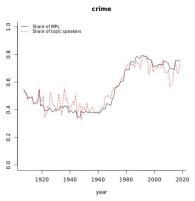
(a) Agriculture and forestry



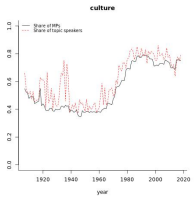
(b) Alcohol



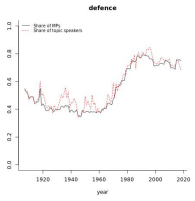
(c) Children



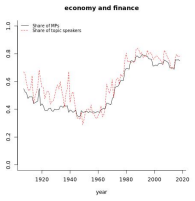
(d) Crime



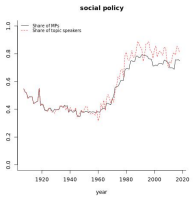
(e) Culture



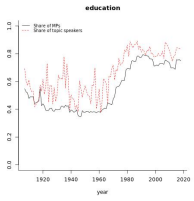
(f) Defence



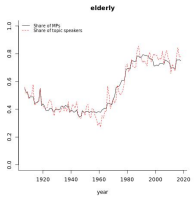
(g) Economy and finance



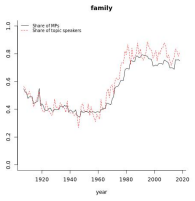
(h) Social policy



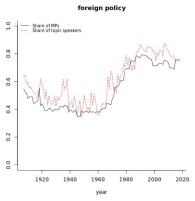
(i) Education



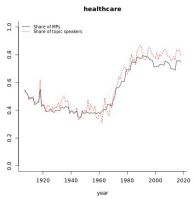
(j) Elderly



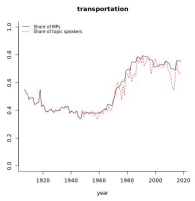
(k) Family



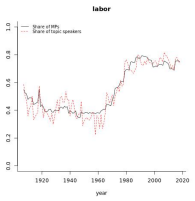
(l) Foreign policy



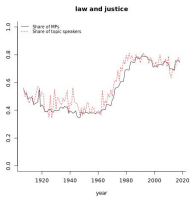
(m) Healthcare



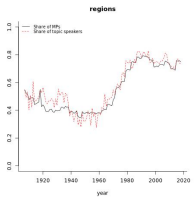
(n) Transportation



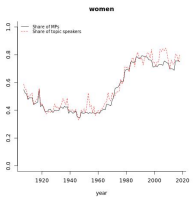
(o) Labor issues



(p) Law and justice



(q) Regions



(r) Women

Figure SI2: Specialization (education): share of university-educated MPs of topic speakers

Notes. Figures present women's share of MPs who speak about a specific topic.

SI2: Most polarized phrases in 1912, 1950, 1971 and 2017

By gender

				1912		
Women	eng.	#W	#M	Men	eng.	#W	#M
syntyn.laps	children born	521	7	täl.kert	this time	309	321
laps.isä	father of a child	515	7	mietintö.liitety	report attached	298	308
aviottom.laps	illegitimate children	220	0	jonku.ver	somewhat	224	232
avioliito.ulkopuol	outside marriage	216	7	tääl.eduskun	here at the parliament	185	191
en.synnyty	not birth	204	0	pyydä.kannat	ask for support	171	177
				1950			
Women	eng.	#W	#M	Men	eng.	#W	#M
soda.aika	wartime	114	39	toime.haltij	official	271	448
suomalais.laps	Finnish children	73	0	vira.toime	tenure description	267	439
helsing.kaupung	City of Helsinki	92	37	valtio.palveluks	in service of the state	32	92
X40.milj	40 million	64	16	vilj.hin	price of grain	34	73
vanhust.asem	position of the elderly	53	13	timot.siemen	seed of timothy	12	45
				1971			
Women	eng.	#W	#M	Men	eng.	#W	#M
huumausain.käytö	use of drugs	217	59	vuon.1970	Year 1970	59	136
valmet.telak	Valmet shipyeard	198	42	suome.pan	Bank of Finland	18	66
helsing.kaupung	City of Helsinki	178	67	tasaval.president	The President	56	92
huumausain.käyttäj	users of drugs	96	12	vuon.1969	year 1969	96	117
kunnollis.eläk	decent pensions	62	2	milj.vuon	millions in a year	5	21
				2017			
Women	eng.	#W	#M	Men	eng.	#W	#M
ammatillis.koulutuks	vocational education	429	147	euroop.union	The European Union	81	161
koulutuks.reform	educational reform	117	27	harm.taloud	grey economy	50	104
sosiaal.terveydenhuolo	healthcare	102	15	talous.kasv	economic growth	41	81
koulutuks.järjestäj	provider of education	62	14	viime.vaalikaud	last term	23	61
kuntouttav.työtoimin	rehabilitative work	49	5	sisäis.turvallisuud	internal security	50	85

			1912				
University	eng.	#U	#NU	No university	eng.	#U	#NU
vuode.1913	year 1913	69	24	viera.kielt	foreign language*	9	59
vast.muutam	only a few	18	0	toise.kappal	another unit	17	40
viera.kiele	foreign language*	24	7	oikea.pyys	right ask	1	23
eduskun.määrä	power of the parliament	16	0	mil.ehdo	-	1	21
eduskun.oikeud	rigths of the parliament	15	0	valtiopäiv.aika	time of the parliament	22	41
			1950				
University	eng.	#U	#NU	No university	eng.	#U	#NU
henkis.työn	mental work	148	44	vira.toime	tenure post	247	492
työn.tekij	worker	85	27	toime.haltij	official	255	487
neuvostoliito.vas	Soviet Union	50	5	timot.siemen	seed of timothy	4	126
soda.aika	wartime	60	38	es.olev	proposed	218	299
rikosl.16	criminal law	29	10	valtio.vira	government agencies	129	204
			1971				
University	eng.	#U	#NU	No university	eng.	#U	#NU
oppil.vanhemp	parents of students	26	1	hinto.maksu	payment of prices	31	79
muun.henkilökun	other personnel	23	1	tulo.menoarvio	141	182	
es.olev	proposed	165	146	oy.yleisradio	YLE Oy	7	46
venäj.kiele	Russian language	18	0	prosent.vuon	percent per year	30	69
vennamo.käyttäm	Used by Vennamo	29	11	yleisradio.ab	YLE Ab	7	38
			2017				
University	eng.	#U	#NU	No university	eng.	#U	#NU
ammatillis.koulutuks	vocational education	275	106	hyvä.as	good thing	139	219
sosiaal.terveydenhuolo	healthcare	86	46	pitä.muist	should remember	36	80
koulutuks.reform	educational reform	58	25	oma.miele	own opinion	12	53
kestäv.kehityks	sustainable development	72	43	jona.päivä	that day	7	47
koulutuks.järjestäj	provider of education	38	18	syytä.muist	reason to remember	8	43

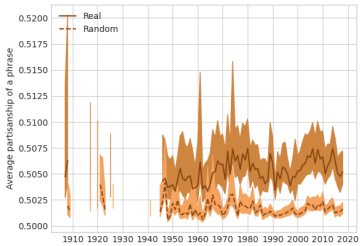
1912							
Uusimaa	eng.	#U	#O	Other	eng.	#U	#O
vuon.1910	year 1910	331	71	täl.kert	this time	324	327
sivu.toise	page other	128	3	mietintö.liitety	report attached	298	302
eduskun.päättäm	decision by the parliament	161	36	jonku.ver	somewhat	222	225
toise.kappal	another unit	116	21	tääl.eduskun	here at the parliament	184	185
nykyis.senaat	current senate	112	22	pyydä.kannat	ask for support	178	180
1950							
Uusimaa	eng.	#U	#O	Other	eng.	#U	#O
henkis.työn	mental work	262	45	toime.haltij	official	306	415
työn.tekij	worker	171	22	vira.toime	tenure post	300	407
samo.myösk	same also	90	22	valtio.vira	goverment agency	114	180
hin.palkkaneuvosto	price and wage council	67	9	valtio.metalliteht	public metal factory	5	62
eduskun.enemmistö	majority of the parliament	92	58	valtio.palveluks	at the service of the state	43	86
1971							
Uusimaa	eng.	#U	#O	Other	eng.	#U	#O
helsing.kaupung	City of Helsinki	164	67	vastattav.esit	must answer to proposed	23	37
huumaav.aine	drug	60	3	helsing.kaupunginvaltuusto	Helsinki City Council	19	33
työvoimapiir.alue	labor force district	32	8	vuode.1970	year 1970	67	76
lain.ede	before the law	36	14	teen.kunnioittav	do honourable	6	11
helsing.hovioikeud	Helsinki Court of Appeal	25	3	koulu.väl	school	2	5
2017							
Uusimaa	eng.	#U	#O	Other	eng.	#U	#O
nost.es	bring up	146	106	ammatillis.koulutuks	vocational education	125	268
agend.2030	Agenda 2030	57	22	koulutuks.reform	educational reform	12	64
euroop.neuvosto	European Council	50	21	koko.aja	all the time	105	152
parlamentaaris.työryhm	Parliamentary task force	43	16	good thing	hyvä.as	130	164
ulko.turvallisuuspolitiik	Foreign Affairs	38	12	Vaasa Central Hospital	vaasa.keskussairaal	4	31

		1912					
White-collar	eng.	#W	#B	Blue-collar	eng.	#W	#B
viera.kielt	foreign language	44	0	kem.tornio	Kemi and Tornio	1	38
asiallin.käsittely	appropriate process	43	0	oikea.pyys	right ask	4	29
vuon.1906	year 1906	31	0	ansaits.tul	deserve to become	1	24
lain.al	under the law	39	14	istunto.ehdot	session proposal	5	24
viera.kiele	foreign language	19	0	kehit.itseä	develop oneself	1	20
		1950					
White-collar	eng.	#W	#B	Blue-collar	eng.	#W	#B
henkis.työn	mental work	110	37	timot.siemen	seed of timothy	18	347
työn.tekij	worker	65	22	harv.asutu	sparsely populated	24	95
kyse.olev	in question	74	56	valtio.vira	government agency	157	210
imatr.kauppal	Imatra township	23	5	valtio.metalliteht	public metal factory	41	83
pienviljelij.liito	Small farmers' union	35	17	maaseudu.asuntoolo	rural living conditions	18	60
		1971					
White-collar	eng.	#W	#B	Blue-collar	eng.	#W	#B
pohjoism.neuvosto	Nordic council	35	2	valmet.telak	Valmet factory	58	186
sotaväe.rikosl	military criminal law	38	17	vuon.1970	Year 1970	103	216
kulutuskysyn.kasvu	increase in consumer demand	17	0	the value of that	8	82	
perh.pienviljelm	small family farm	17	2	yleisradio.ab	YLE Ab	12	78
muun.henkilökun	other personnel	13	0	oy.yleisradio	YLE Oy	13	75
		2017					
White-collar	eng.	#W	#B	Blue-collar	eng.	#W	#B
ammatillis.koulutuks	Vocational education	244	115	yhä.ede	still	7	78
euroop.union	European union	137	96	harm.taloud	grey economy	57	128
euroop.neuvosto	European council	30	9	malm.lentokent	Malmi airport	15	69
ulko.turvallisuuspolitiik	Foreign policy	20	2	hyvä.kolleg	good colleague	18	54
vapa.sivistystyö	liberal adult education	31	13	alko.monopol	alcohol monopoly	16	46

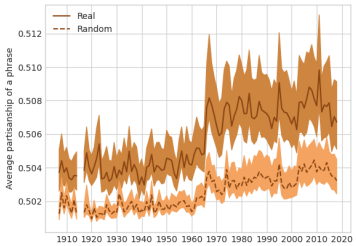
				1912			
Young (under 40)	eng.	#Y	#O	Old (over 40)	eng.	#Y	#O
sama.vuon	same year	98	15	lain.valvon	law enforcement	12	44
eduskun.päättäm	decided by the parliament	65	39	edellis.valtiopäiv	previous parliament	34	49
valtiopäiv.aika	time of the parlieament	39	19	aika.sopiv	quite suitable	15	22
suome.vankilo	Finnish prisons	16	0	tääl.edellis	previous	15	22
vuon.1910	year 1910	103	88	täl.kert	this time	327	327
				1950			
Young (under 40)	eng.	#Y	#O	Old (over 40)	eng.	#Y	#O
kyse.olev	in question	158	61	toime.haltij	official	290	403
vuode.1949	year 1949	155	71	vira.toime	tenure post	314	393
rikosl.16	criminal law 16	86	11	timot.siemen	seed of timothy	14	66
X16.luvu	-	88	14	es.olev	proposed	271	272
luvu.24	-	85	13	täl.kert	this time	260	262
				1971			
Young (under 40)	eng.	#Y	#O	Old (over 40)	eng.	#Y	#O
huumausain.käytö	use of drugs	175	81	liikevaihtovero.lisävero	revenue tax	38	102
dollar.hyväksyt	dollar approved	104	31	helsing.kaupunginvaltuusto	Helsinki City Council	24	32
huumaav.aine	drug	67	3	keskiviiko.2	Wednesday	16	22
vastattav.seuraav	-	196	136	prosent.liikevaihtovero	percent revenue tax	10	14
mukais.merkits	-	49	5	vuode.vuote	year after year	9	13
				2017			
Young (under 40)	eng.	#Y	#O	Old (over 40)	eng.	#Y	#O
ammatillis.koulutuks	vocational education	677	147	euroop.union	European union	79	149
koulutuks.reform	educational reform	173	24	toine.as	other issue	39	82
koulutuks.järjestäj	provider of education	112	18	pitä.muist	should remember	16	52
en.vuode	next year	208	115	erit.hyvä	very good	139	162
toise.ast	secondary school	138	59	tode.palj	very much	24	45

		1912					
First-term MP	eng.	#F	#N	Non-first-term MP	eng.	#F	#N
asiallis.käsittely	appropriate processing	59	9	savo.rada	Savo track2	23	
venäj.kans	with Russia	55	9	jää.suome	stay in Finland	2	6
vuon.1910	year 1910	121	84	esimerk.vaasa	for example Vaasa	3	6
valtiopäiv.aika	time of the parliament	43	23	hakem.apu	seek help	2	3
kun.sisä	-	19	5	täl.kert	this time	326	327
		1950					
First-term MP	eng.	#F	#N	Non-first-term MP	eng.	#F	#N
vira.toime	tenure post	490	333	timot.siemen	seed of timothy	27	83
toime.haltij	official	489	345	vanho.talo	old house	10	34
harv.asutu	sparse population	116	17	valtio.tilintarkastaj	government audit	7	24
valtio.vira	government agency	224	134	vuokr.korotuks	rent increases	17	33
henkis.työn	mental work	146	72	luvu.24	-	8	23
		1971					
First-term MP	eng.	#F	#N	Non-first-term MP	eng.	#F	#N
vastattav.seuraav	answer to next	231	75	es.olev	proposed	120	187
jäsen.vastattav	memeber must answer	240	109	perh.pienviljelm	small family farm	1	41
asianomais.jäsen	members in question	218	111	tasaval.president	The President	60	92
valtioneuvosto.asianomais	-	200	95	vuode.vuote	Year after year	4	28
valmet.telak	Valmet factory	115	34	mark.käytettäv	Marks (money) in use	16	39
		2017					
First-term MP	eng.	#F	#N	Non-first-term MP	eng.	#F	#N
kute.edustaj	like member	137	66	viime.kaude	last term	60	112
erit.hyvä	very good	180	141	voi.sano	can saY	51	92
vaasa.keskussairaal	Vaasa Central Hospital	39	10	euroop.union	European union	122	151
äärimmäis.tärkeä	of paramount importance	67	40	työllisyys.paran	increase in employment	18	35
uusiutuv.energia	renewable energy	64	43	viime.hallituskaud	last government	23	38

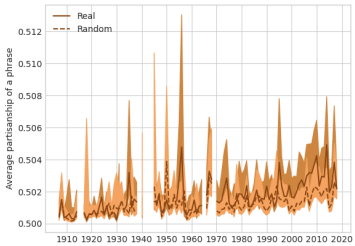
SI3: 95 % nominal coverage CIs



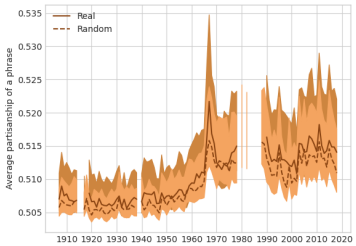
(a) Gender



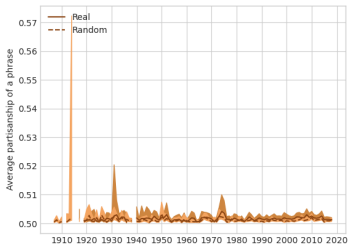
(b) University education



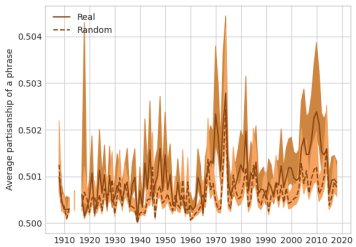
(c) Young (under 40 years old)



(d) White-collar job



(e) Most urban districts: Helsinki and Uusimaa)



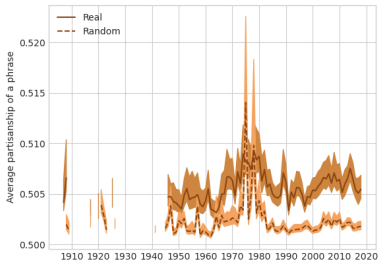
(f) First term MP

Figure SI3: Intra-party group differences

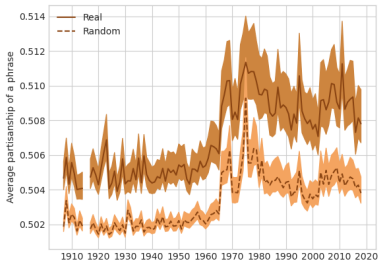
Notes. Figure presents intra-party group differences by six different demographic groups. Estimation includes year and party fixed effects. Confidence intervals have 95% nominal coverage.

SI4: Alternative specifications

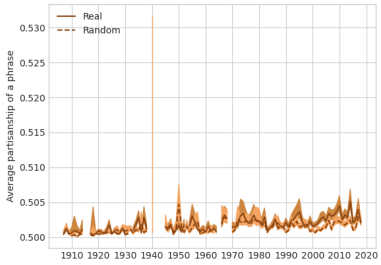
Without party fixed effects



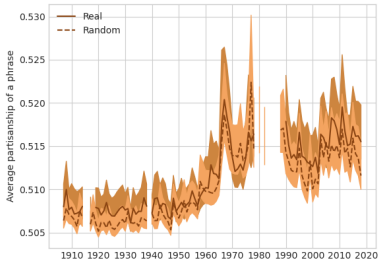
(a) Gender



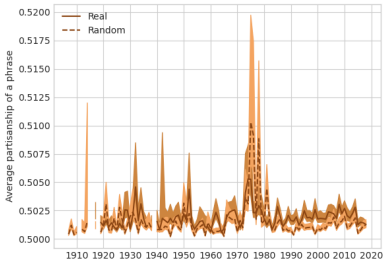
(b) University education



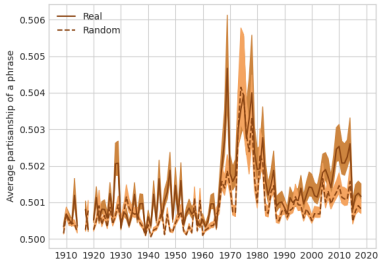
(c) Young vs. old



(d) White-collar vs. blue-collar



(e) Helsinki + Uusimaa

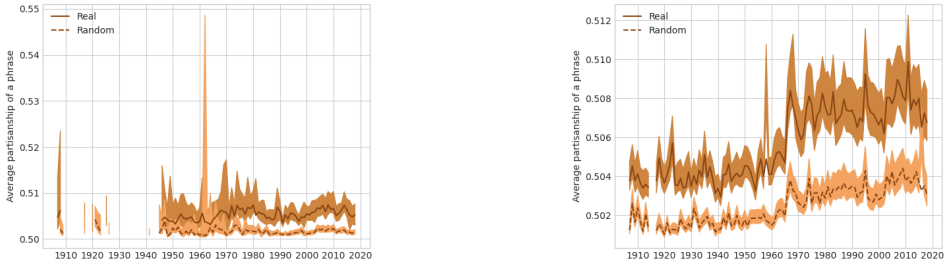


(f) First term MPs

Figure SI4: Intra-party group differences

Notes. Figure presents intra-party speech differences by six different demographic groups. Estimation includes year fixed effects but does not include party FE. Confidence intervals have 80% nominal coverage.

Adding a government-opposition dummy



(a) Gender

(b) University education

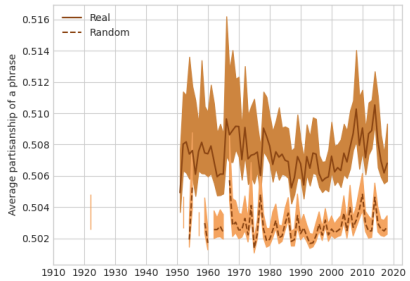
Figure SI5: Intra-party group differences (gender and uni education), govt/opposition dummy included

Notes. Figure presents intra-party speech differences by six different demographic groups. Estimation includes year FE, party FE and government/opposition dummy. Confidence intervals have 80% nominal coverage.

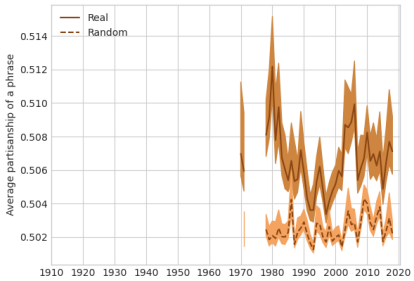
Heterogeneity by party

Figure SI6 shows results where we estimate the speech differences (for gender and education) separately for left-wing parties and right-wing parties. Party fixed effects are still included in estimation. Only years with a sufficient number of observations per group (more than 15) are shown here, but a whole time series is shown in Section 6.3 (Figure 10), where we compare these within party-block speech differences to within-party block descriptive representation. These results show some heterogeneity in speech difference estimates with respect to which party we look at. It can be seen from Figure 4 that gender speech differences used to be large in right-wing parties around 1980 while they are more moderate today. In left-wing-parties, gender speech differences have been more flat. Previous work has found that in the United States, gender differences in political preferences are larger in the right-wing of the political spectrum than in the left-wing (Poggione, 2004). Our results suggest that has been the case in Finland in around 1980, but nowadays gender differences in speech are somewhat larger in left-wing parties. Regarding speech differences between MPs with university education and those without, the speech differences in left-wing parties can be detected from the noise of

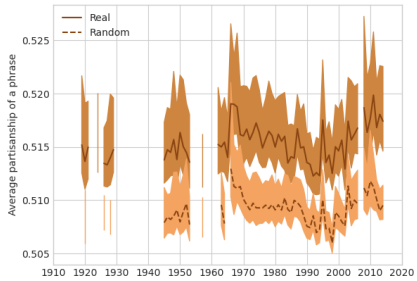
the data, unlike in right-wing parties. Although not detectable from random noise, the size of the estimate is larger in right-wing parties.



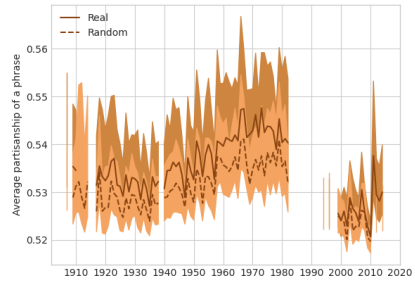
(a) Women, left-wing parties



(b) Women, right-wing parties



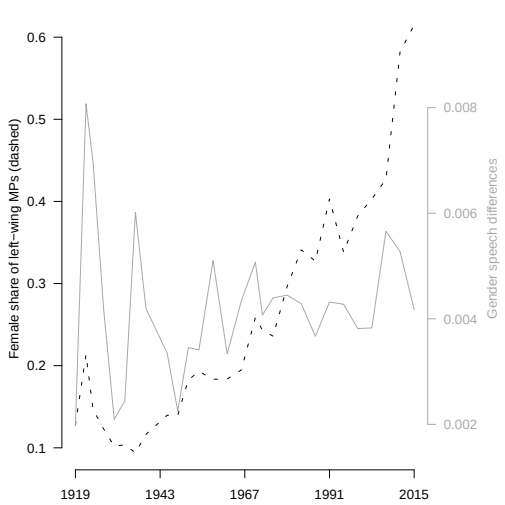
(c) University educated, left-wing parties



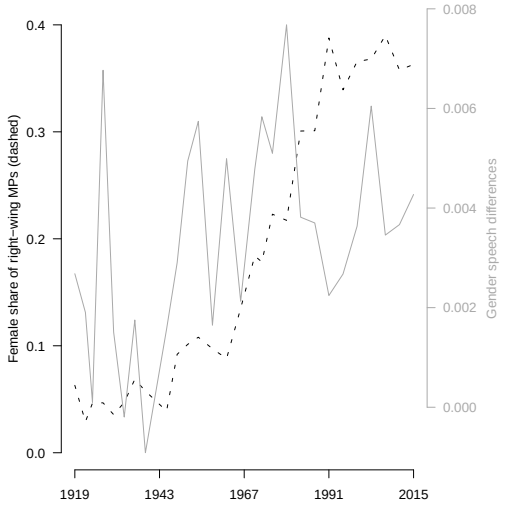
(d) University educated, right-wing parties

Figure SI6: Results separately for left and right wing parties

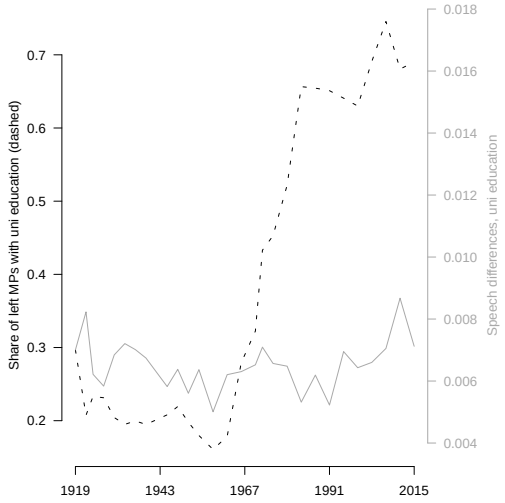
Notes. Figure presents speech differences by gender and university education status separately for left-wing and right-wing parties. Left wing includes Social Democrats and the current Left Alliance and its predecessors. Right wing includes the National Coalition Party, Center Party and Agrarian League. Confidence intervals have 80% nominal coverage.



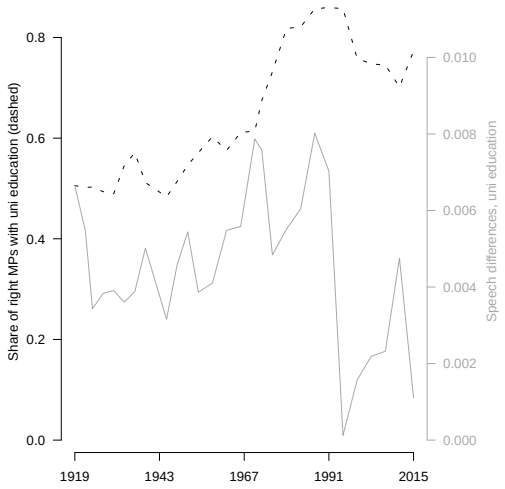
(a) Gender, left-wing



(b) Gender, right-wing



(c) University education, left-wing

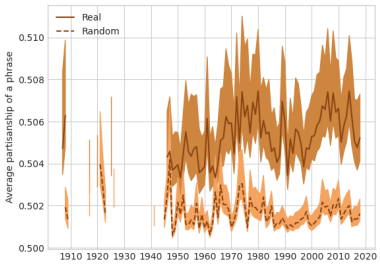


(d) University education, right-wing

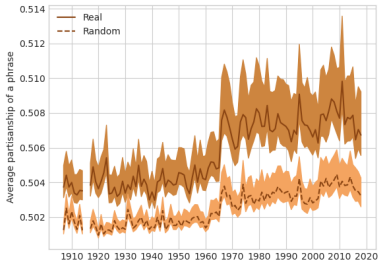
Figure SI7: Links between descriptive representation and speech differences, separately for left-wing and right-wing parties

SI5: Subsampling robustness

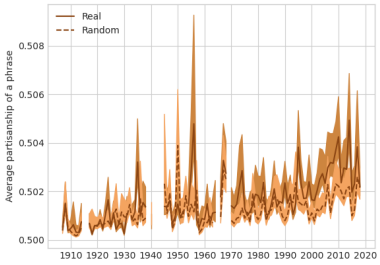
30 % subsample size



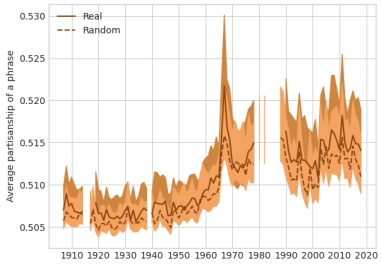
(a) Gender



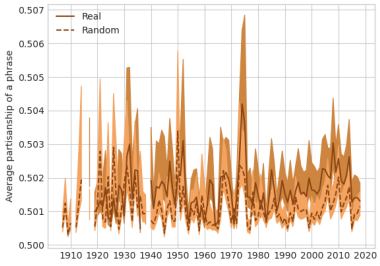
(b) University education



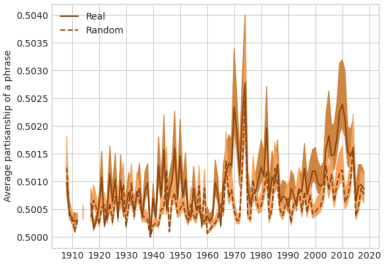
(c) Young vs. old



(d) White-collar vs. blue-collar



(e) Helsinki + Uusimaa

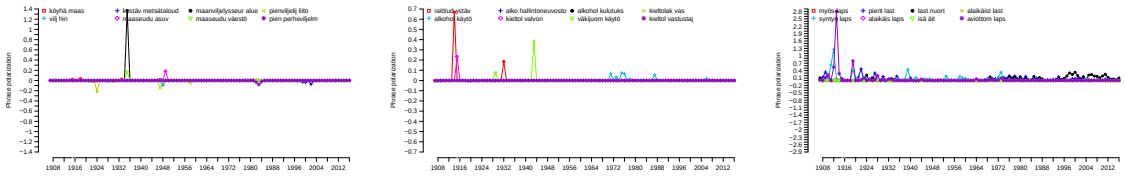


(f) First term MPs

Figure SI8: Intra-party group differences

Notes. Figure presents intra-party speech differences by six different demographic groups. Estimation includes year fixed effects but does not include party FE. Confidence intervals have 80% nominal coverage.

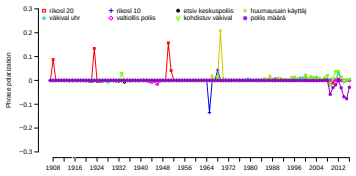
SI6: Polarizing phrases within topics



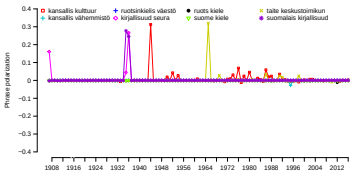
(a) Agriculture and forestry

(b) Alcohol

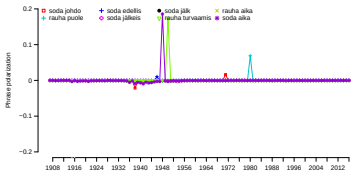
(c) Children



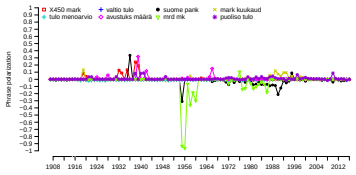
(d) Crime



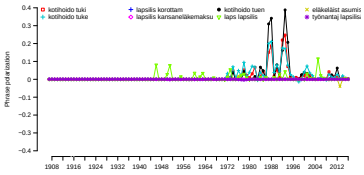
(e) Culture



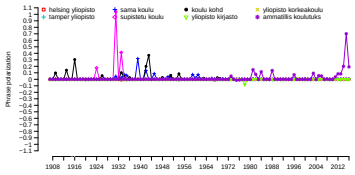
(f) Defence



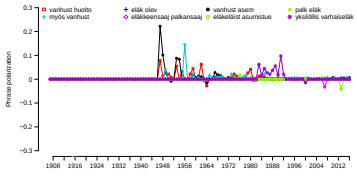
(g) Economy and finance



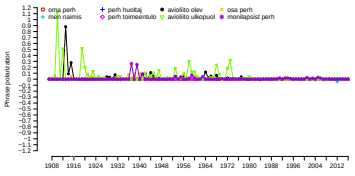
(h) Social policy



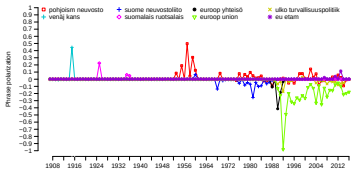
(i) Education



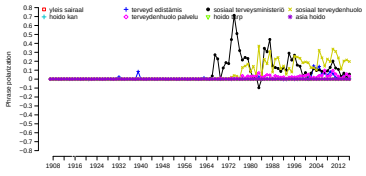
(j) Elderly



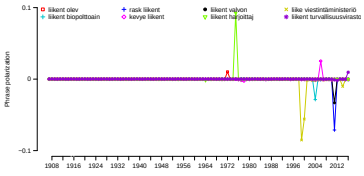
(k) Family



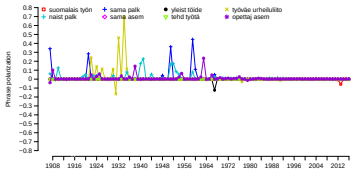
(l) Foreign policy



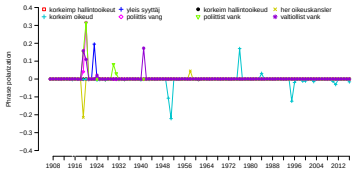
(m) Healthcare



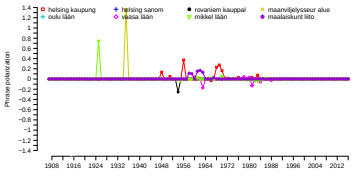
(n) Transportation



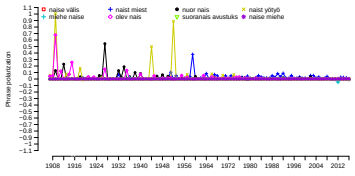
(o) Labor issues



(p) Law and justice



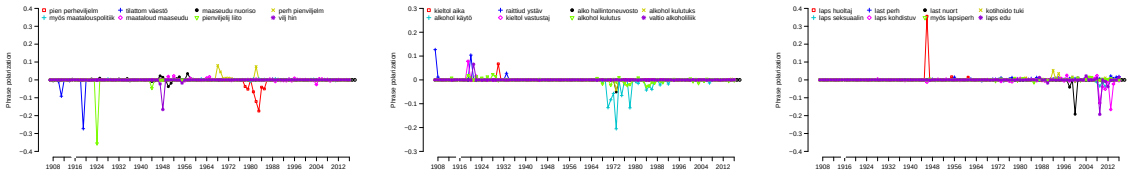
(q) Regions



(r) Women

Figure SI9: Gender: most polarizing phrases within different topics

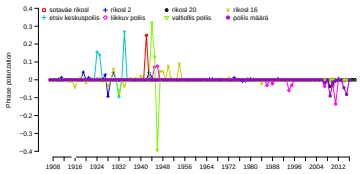
Notes. Figures present 1000 times the estimated phrase polarization.



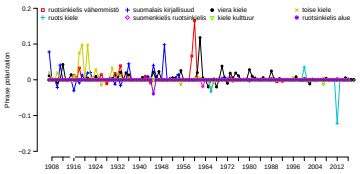
(a) Agriculture and forestry

(b) Alcohol

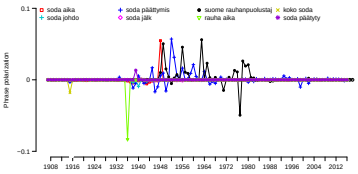
(c) Children



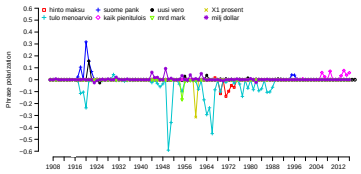
(d) Crime



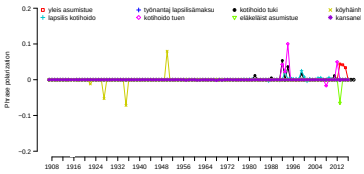
(e) Culture



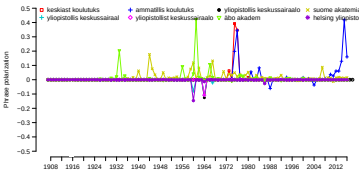
(f) Defence



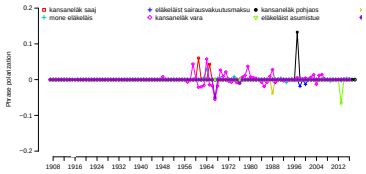
(g) Economy and finance



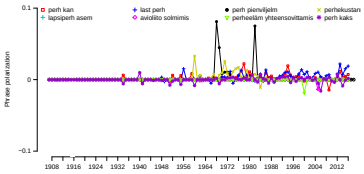
(h) Social policy



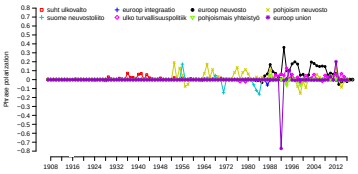
(i) Education



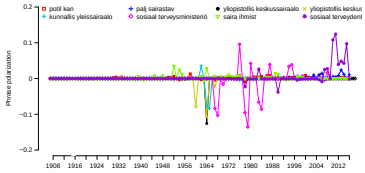
(j) Elderly



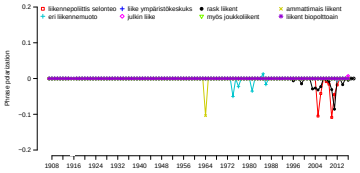
(k) Family



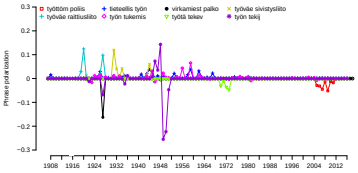
(l) Foreign policy



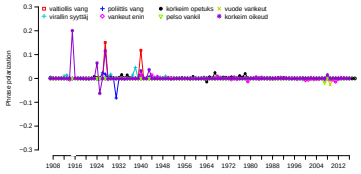
(m) Healthcare



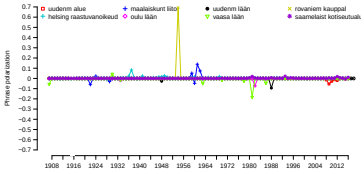
(n) Transportation



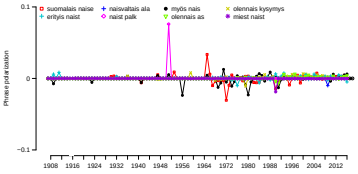
(o) Labor issues



(p) Law and justice



(q) Regions



(r) Women

Figure SI10: Education: most polarizing phrases within different topics

Notes. Figures present 1000 times the estimated phrase polarization.

SI7: Variable creation and sample selection

Adding occupations and education level to the data

Keywords used to identify MPs with university education:
Keywords below are used to identify MPs with university education. If the education field contrains any of the keywords in the list below, the MP will be considered to have university education. The first four keywords ("tohtori", "maisteri", "kandidaatti", "lisensiaatti") relate to general degrees and identify most of the university educated MPs, but some MPs have expressed their education without using these words, thus the need to include other words too in the keyword list.

"tohtori", "maisteri", "kandidaatti", "lisensiaatti", "agronomi", "professori", "notaari", "varatuomari", "ekonomi", "pappisvihkimys", "diplomi-insinööri", "opettaja", "Yhteiskunnallinen korkeakoulu", "proviisori", "farmaseutti", "arkkitehti", "luokanopettaja", "yliopisto", "lastentarhanopettaja", "erityisopettaja", "kotitalousopettaja", "kauppakorkea", "metsänhoitaja", "Bachelor", "Master", "bachelor", "master", "University", "university", "agrologi", "tuomari", "oikeustutkinto, "kameraalitutkinto", "kadettikoulu", "lääkäri", "sanomalehtitutkinto", "tieteiden ylioppilas"

Keywords used to identify MPs with white-collar or blue-collar profession:

Keywords used to identify MPs with white-collar jobs:
"opettaja", "tuomari", "asianajaja", "agronomi", "lisensiaatti", "lääkäri", "toimittaja", "johtaja", "sosionomi", "liikemies", "luennoitsija", "kirjanpitäjä", "maisteri", "kirkkoherra", "rovasti", "insinööri", "sihteeri", "professori", "tohtori", "asiantuntija", "maaherra", "nimismies", "ministeri", "pastori", "neuvos", "kirjailija", "psykiatri", "kandidaatti", "ekonomi", "pormestari", "opiskelija", "vapaaherra", "puhuja", "Kansanedustaja", "suurlähettiläs", "konsuli", "lehtori", "lakimies", "päällikkö", "tarkastaja", "toimisto", "farmaseutti", "asiamies", "presidentti", "myynninedistäjä", "arkkitehti", "piispa", "notaari", "rehtori", "tiedottaja", "konsultti", "kirjuri", "konttoristi", "huvilanomistaja", "virastoapulainen", "analyytikko", "senaattori", "majuri", "kansleri", "aktuaari", "tutkija", "kenraali", "tradenomi", "merkonomi", "työllistäjä", "apteekkari", "kansanedustaja", "kommodori", "prokuristi", "asessori", "kappalainen", "everstiluutnantti", "dosentti", "journalisti", "Professori", "pappi", "runoilija", "toiminnanohjaaja", "liikkeenomistaja", "geronomi", "esimies", "ylioppilas", "yrittäjä", "valtuutettu", "Euroopan komission jäsen"

Keywords used to identify MPs with blue-collar jobs:
"viljelijä", "viljeijä", "vijeliä", "vijelijä", "kirvesmies", "vartija", "vahtimestari", "maalari", "viilaaja", "kauppias", "sairaanhoitaja", "räätäli", "suutari", "posteljooni", "muurari", "perhepäivähoitaja", "myymälänhoitaja", "ompelija", "levyseppä", "luottamusmies", "talonmies", "kultaseppä", "satamavalvoja", "rakennusmestari", "puvustonhoitaja", "seppä", "mylläri", "konemies", "sorvari", "liikunnanohjaaja", "konstaapeli", "vaatturi", "mestari", "veturinkuljettaja", "mäkitupalainen", "ulosottomies", "nuorisotyönohjaaja","työmies", "ylikonstaapeli",

"komisario", "emäntä", "isännöitsijä", "teknikko", "tehdas", "työläinen", "lastenohjaaja", "rakennusurakoitsija", "hoitaja", "asioitsija", "asentaja", "faktori", "agrologi", "viilaaja", "kuljettaja", "kuljetusyrittäjä", "kätilö", "tilanomistaja", "turvallisuusvalvoja","työntekijä", "betoni", "raudoittaja", "sitoja", "sorvaaja", "torppari", "korjausmies", "valaja", "kaavaaja", "fysioterapiayrittäjä", "koneistaja", "muusikko", "palo", "verhoilija", "toimitsija", "kanttori", "urkuri", "asemamies", "kutoja", "leipuri", "korjaaja", "yksityisetsivä", "ehkäisevän päihdetyön ohjaaja", "tilallinen", "kuljettaja", "kirjaltaja", "merikapteeni", "puutarhuri", "autoilija", "rautatieläinen", "harjoittaja", "tarkkaaja", "hieroja", "liikkenharjoittaja", "talollinen", "ajomies", "mekaanikko", "puutavaramies", "koneenkäyttäjä", "näyttelijä", "kokki", "rakennusmies", "laborantti"

Sample selection: parties that are excluded

When we analyze groups other than women, we include all parties in the sample. When we analyze speech differences between the gender, we exclude some parliamentary groups, most of which have been small groups, often consisting of only a single MP. We do this, because many parties have had no women, or a very small number of women MPs in some years. These parties are excluded from the analysis. Most of the excluded parties are small groups, except for the Centre Party (*"Keskustan eduskuntaryhmä"*) which is traditionally a large party. Our results are robust to not dropping any parties: the speech difference estimate would look very similar if we included all parties. The parties that are excluded in our main analyses are the following:

"Alkiolainen keskustaryhmä", "Eduskuntaryhmä Immonen", "Eduskuntaryhmä Nuorsuomalaiset ja Risto Kuisma", "Eduskuntaryhmä Puhjo", "Eduskuntaryhmä Virtanen", "Hannu Suhosen eduskuntaryhmä", "Kansalaispuolueen eduskuntaryhmä", "Kansanpuolue", "Keskustan eduskuntaryhmä", "Kristillisdemokraattinen eduskuntaryhmä", "Kristillisen liiton eduskuntaryhmä", "Liberaalien eduskuntaryhmä", "Liike Nyt -eduskuntaryhmä", "Muutospuolueen eduskuntaryhmä", "Nuorsuomalaisten eduskuntaryhmä", "Remonttiryhmä", "Ruotsalainen vasemmisto", "Ryhmä Erlund", "Sosialidemokraattinen riippumaton eduskuntaryhmä", "Suomalaisen rintaman eduskuntaryhmä", "Suomen kansanpuolue", "Suomen kristillisen työväen liitto", "Suomen pientalonpoikien puolue", "Suomen pienviljelijäin ja maalaiskansan puolue", "Työväen ja pienviljelijäin vaaliliitto", "Vaihtoehto Suomelle -eduskuntaryhmä", "Vapaamielisten liitto", "Vapaiden demokraattien eduskuntaryhmä", "Vasemmistoryhmä", "Vasenryhmän eduskuntaryhmä", "edustaja Väyrynen"

SI8: Altenative definition for urbanicity: 10 biggest cities

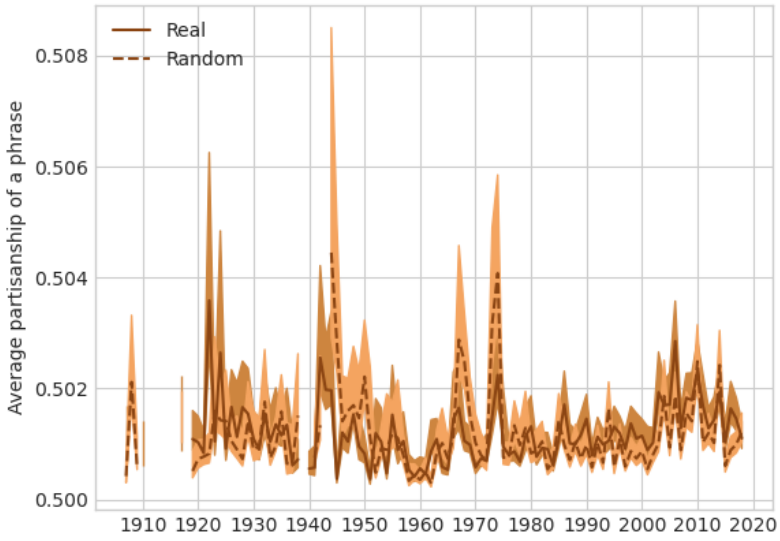


Figure SI11: Urbanicity

Notes. Figure presents speech differences between urban and non-urban politicians, where urbanicity is defined as being born in one of the 10 biggest cities. Confidence intervals have 80% nominal coverage.

SI9: Shares of female MPs and university educated MPs in different committees

We only have data on committee memberships the MPs have had during their whole term in the Parliament so these figures show the share of women who have been in a committee at some point out of all MPs who have been in the committee at some point. In addition, as not all of the committees have always existed, we pool together committees that are similar but have different names in different years.

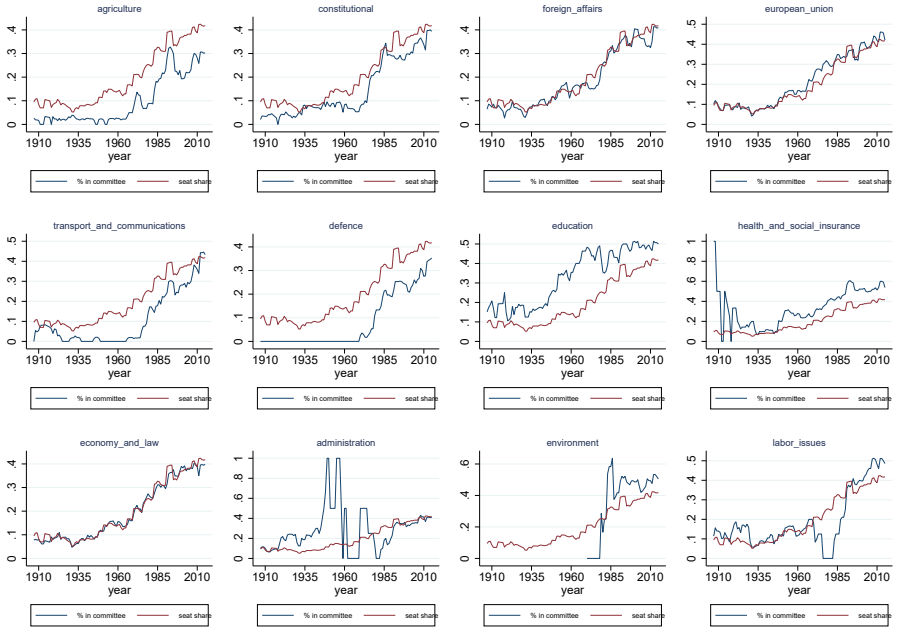


Figure SI12: Share of female MPs in different committees

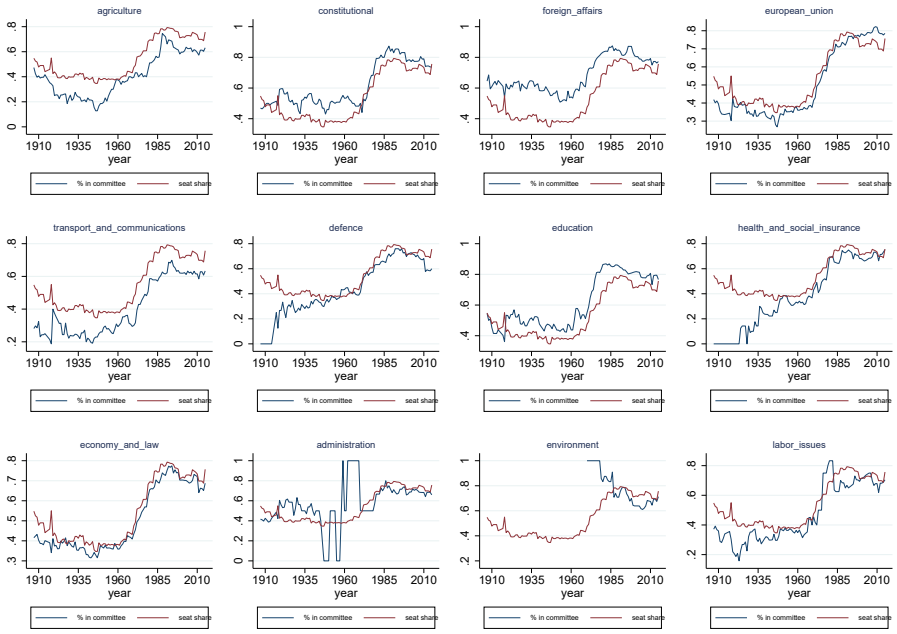


Figure SI13: Share of university educated MPs in different committees

SI10: Structural topic model

Figure SI15 shows topics created by a structural topic model. This approach differs considerably from our main topic analyses, as this a structural topic model is an unsupervised method, i.e., we do not influence in any way what the topics are or which words belong to which topics. Topic models are also not exclusive, i.e., words can belong to multiple topics. Figure SI15 shows the topics created by the structural topic model.

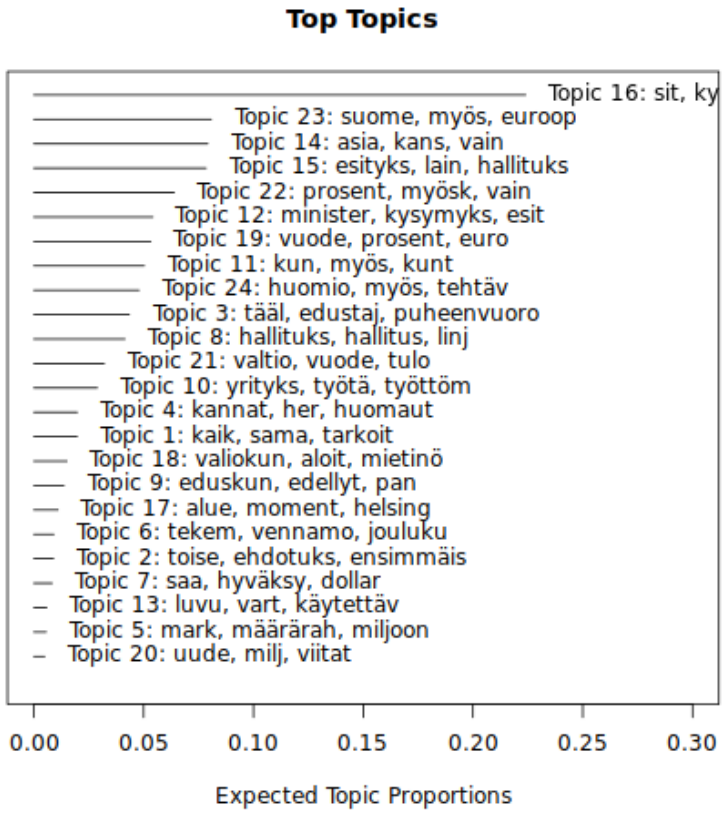


Figure SI14: Topics created by structural topic model

Notes. Figure shows topics created by a structural topic model.

Based on going through top 100 words in each of the 24 topics, we are able to name 18 of the topics generated by the structural topic model. Other topics cannot be classified as they only contain procedural words or all kinds of words without any common theme. Table SI1 shows how we name those 18 topics.

Topic 2: Government proposals
Topic 3: Phrases related to speakers and discussion in the Parliament
Topic 4: Statements
Topic 5: Budget
Topic 8: Government/opposition
Topic 10: Labour markets/employment and entrepreneurship
Topic 11: Healthcare, family, children
Topic 12: Ministries
Topic 14: Democracy
Topic 15: Legislation
Topic 17: Regional policy
Topic 18: Committees
Topic 19: The economy
Topic 20: Public finance
Topic 21: Budget (similar to topic 5)
Topic 22: Agriculture
Topic 23: Foreign policy
Topic 24: Education

Table SI1: Named topics

Figure SI15 below shows the expected topic proportions by year, i.e., how common different topic are in different years. It can be observed that the topic 11, "Healthcare, family, children" start to grow after the 1960s, which is consistent with the results we found with our main topic classification method. This topic 11 is also the most clearly "female" topic according to the Figure **??** which shows which topics are common for female MPs and male MPs. Figure SI17 shows which topics are common for university educated and non-university educated MPs. The latter suggests that university educated MPs care a lot about democracy, and also care about foreign policy and education.

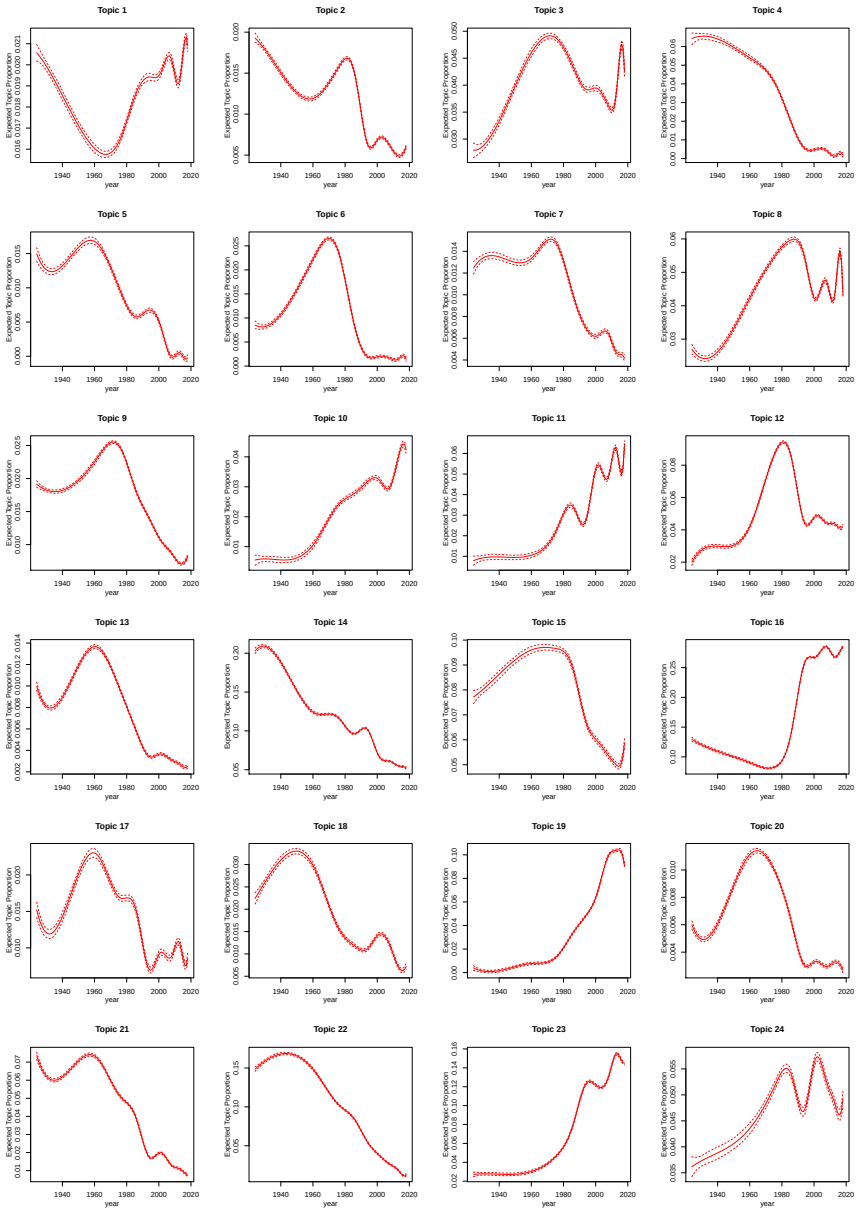


Figure SI15: Expected proportions of topics created by structural topic model

Notes. Figure shows the expected proportions of topics created by a structural topic model.

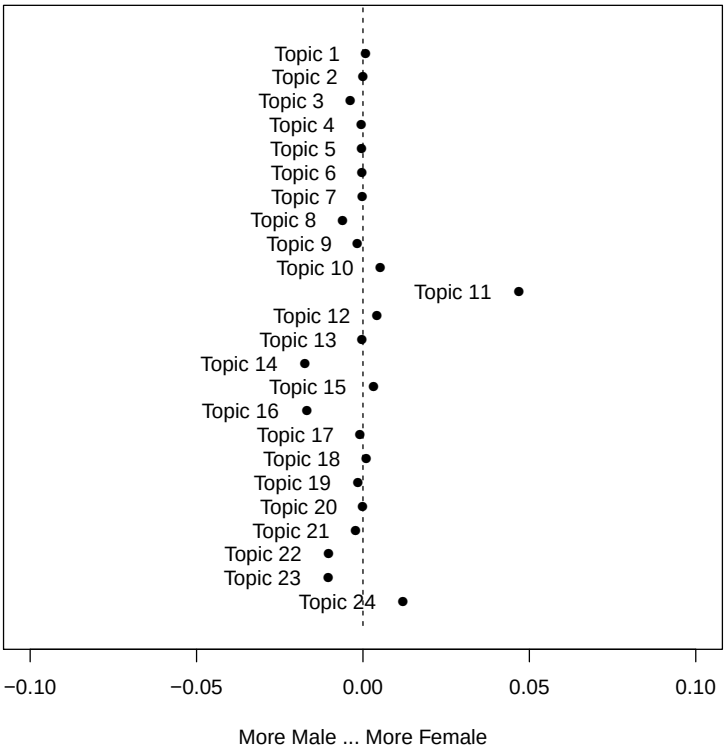


Figure SI16: Topics (STM) favored by gender

Notes. Figure shows which topics (created using a structural topic model) are favored by female MPs and which topics by male MPs.

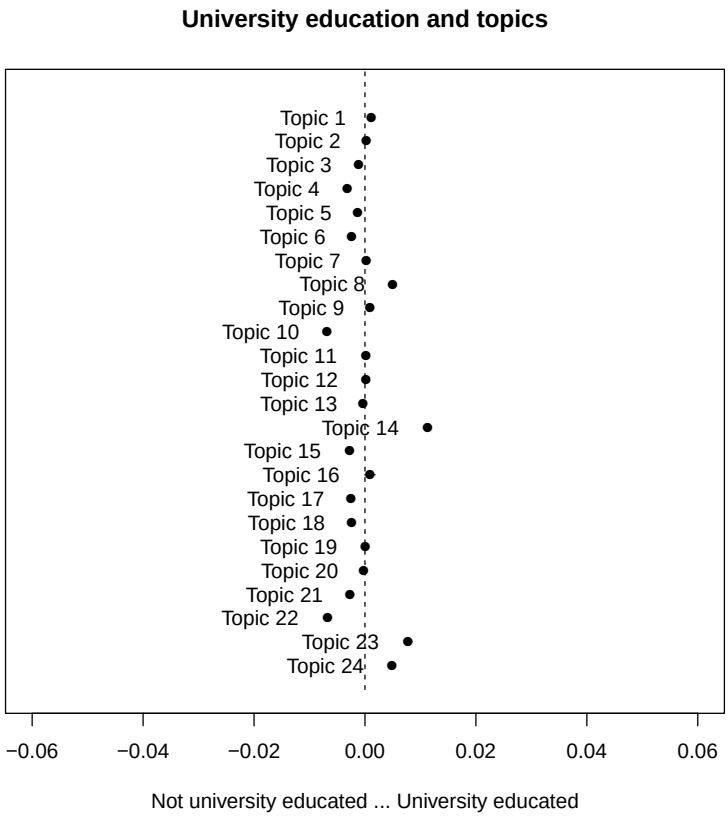


Figure SI17: Topics (STM) favored by university education status

Notes. Figure shows which topics (created by a structural topic model) are favored by highly educated MPs and which by non-highly-educated MPs.

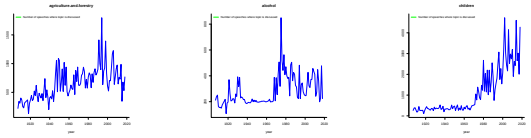
SI11: Evolution of women's seat share in parliaments around the world

Figure SI18: Women's seat shares

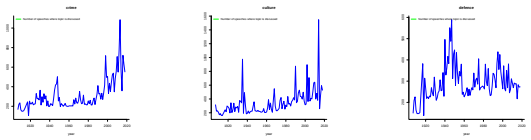
Notes. Figure presents the evolution of women's seat shares in many different countries. Data is from Finnish Social Science Archive.

SI12: How common are different topics

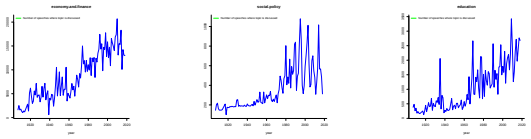
Number of topic phrases mentioned in speeches



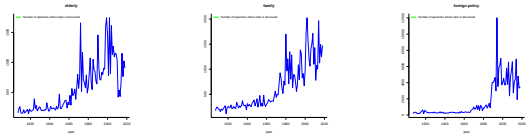
(a) Agriculture and forestry (b) Alcohol (c) Children



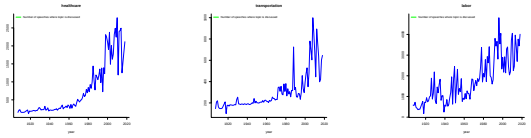
(d) Crime (e) Culture (f) Defence



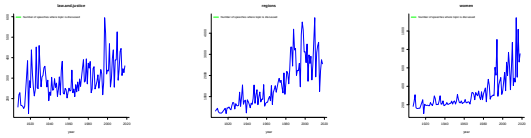
(g) Economy and finance (h) Social policy (i) Education



(j) Elderly (k) Family (l) Foreign policy



(m) Healthcare (n) Transportation (o) Labor issues



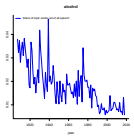
(p) Law and justice (q) Regions (r) Women

Figure SI19: Number of topic phrases mentioned

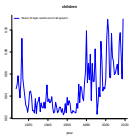
Notes. Figure presents the number of phrases spoken that belong to a specific topic.

Share of topic phrases mentioned in speeches

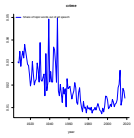
(a) Agriculture and forestry



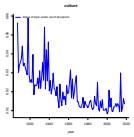
(b) Alcohol



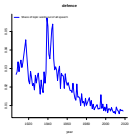
(c) Children



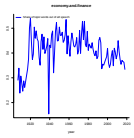
(d) Crime



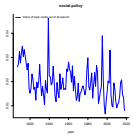
(e) Culture



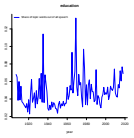
(f) Defence



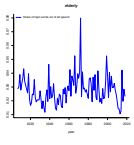
(g) Economy and finance



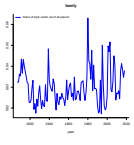
(h) Social policy



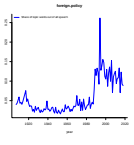
(i) Education



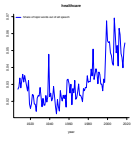
(j) Elderly



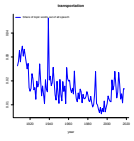
(k) Family



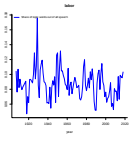
(l) Foreign policy



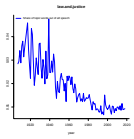
(m) Healthcare



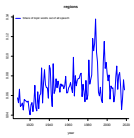
(n) Transportation



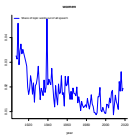
(o) Labor issues



(p) Law and justice



(q) Regions



(r) Women

Figure SI20: Number of topic phrases mentioned

Notes. Figure presents the number of phrases spoken that belong to a specific topic.

SI13: Selection or behavioral change?

Selection analysis

Table SI2: Characteristics of female MPs elected in 1966/1975 (after observed change in speech patterns) compared to female MPs elected in 1958/1962 (before the changes happened)

	Dependent Variable			
	Unieducation	Whitecollar	Under40	Left
(Intercept)	0.4359***	0.8380***	0.1149***	0.6720***
	(0.0190)	(0.0118)	(0.0141)	(0.0172)
Elected66_75	0.1522***	0.0934***	0.0819***	0.0761***
	(0.0249)	(0.0156)	(0.0185)	(0.0226)
R-squared	0.0227	0.0221	0.0120	0.0069
Adjusted R-squared	0.0221	0.0215	0.0114	0.0063
F-statistic	37.27***	35.98***	19.58***	11.39***
Observations	1609	1597	1609	1632

Note: Standard errors in parentheses. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Table SI3: Characteristics of female MPs elected in 1979/1983 compared to female MPs elected in 1966/1975

	Dependent Variable			
	Unieducation	Whitecollar	Under40	Left
(Intercept)	0.5882***	0.9314***	0.1968***	0.7482***
	(0.0151)	(0.0091)	(0.0131)	(0.0150)
Elected79_83	0.1719***	-0.0271*	0.0049	-0.1233***
	(0.0207)	(0.0124)	(0.0179)	(0.0205)
R-squared	0.0338	0.0024	0.000037	0.0174
Adjusted R-squared	0.0333	0.0019	-0.00046	0.0170
F-statistic	69.00***	4.80*	0.07	36.20***
Observations	1976	2004	2016	2042

Note: Standard errors in parentheses. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Analysis of behavioral change (female MPs who are in the Parliament during 1958-1969

Outcome	Coefficient	S.E.	t	p
Agriculture and forestry	-0.15740	0.15753	-0.999	0.329
Alcohol	0.49461	0.31766	1.557	0.134
Children	-0.13240	0.62373	-0.212	0.834
Crime	0.14674	0.12475	1.176	0.253
Culture	0.031267	0.305331	0.102	0.919
Defence	-0.004200	0.075356	-0.056	0.956
Economy and finance	-2.5624	1.8887	-1.357	0.18929
Education	-0.6903	1.7018	-0.406	0.689
Elderly	-0.062874	0.026917	-2.336	0.0295**
Family	1.00677	0.58722	1.714	0.101
Foreign policy	-0.4239	0.9972	-0.425	0.675
Healthcare	1.13813	0.44195	2.575	0.0176**
Labor	0.39694	0.17925	2.214	0.038**
Law and Justice	0.003072	0.064317	0.048	0.962
Regions	0.67150	0.63134	1.064	0.300
Social Policy	-0.48141	0.30820	-1.562	0.1332
Transportation	-0.007368	0.008022	-0.918	0.369
Women	0.63321	0.28646	2.210	0.0383 **

Table SI4: Linear regression, coefficients for female seat share by outcome

	(1) OLS	(2) Fixed Effects
Dependent variable:	Talking about female-favored topics	
Share of female MPs	1.442 (2.145)	1.381 (2.305)
Intercept	0.090 (0.308)	–
Observations	23	23
R^2	0.021	0.032

Table SI5: Talking about female-favored topics

Notes. Table shows the results from a regression where the outcome variable is the share of speech that belongs to female-favored topics (healthcare, education, social policy, elderly, women, family, children, culture)

The **Aboa Centre for Economics (ACE)** is a joint initiative of the economics departments of the Turku School of Economics at the University of Turku and the School of Business and Economics at Åbo Akademi University. ACE was founded in 1998. The aim of the Centre is to coordinate research and education related to economics.

Contact information: Aboa Centre for Economics, Department of Economics, Rehtorinpellonkatu 3, FI-20500 Turku, Finland.

www.ace-economics.fi

ISSN 1796-3133