Simola, Salla; Nieminen, Jeremias; Tukiainen, Janne

Working Paper

A century of partisanship in Finnish political speech

Discussion paper, No. 160

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

Salla Simola, Jeremias Nieminen, and Janne Tukiainen

A century of partisanship in Finnish political speech

Aboa Centre for Economics

Discussion paper No. 160
Turku
May 2023

The Aboa Centre for Economics is a joint initiative of the economics departments of the University of Turku and Åbo Akademi University.

ISSN 1796-3133

Printed in Uniprint
Turku
May 2023

Salla Simola, Jeremias Nieminen, and Janne Tukiainen

A century of partisanship in Finnish political speech

Aboa Centre for Economics
Discussion paper No. 160
May 2023

ABSTRACT

We use novel data to describe the evolution of party differences in parliamentary speech in Finland during 1907–2018. We find a peak in left-right polarization in the 1970s, driven by the extreme left party, and co-occuring with a high prevalance of Soviet Union related phrases, perhaps resulting from Soviet information influencing. The period was also marked with short-lived coalition governments and inefficient policymaking. Moreover, as we find that left-right partisanship fluctuates during the majority of the 20th century, our results show that the levels of polarization currently perceived as high in many countries may not be that exceptional.

JEL Classification: D72, P00

Keywords: text analysis, parliamentary speech, polarization

Contact information

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

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

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

Acknowledgements

We thank Jaakko Meriläinen, Matti Mitrunen, Tuomo Virkola, Andrey Zhukov, Natalia Zinoyeva, Manuel Bagues, Tuomas Pekkarinen, Matti Sarvimäki, Marko Terviö, seminar and workshop participants at Helsinki GSE, Linköping Institute for Analytical Sociology, and KT-päivät in Turku for helpful discussions. We also thank Kimmo Kettunen, Markus Alenius, Kimmo Makkonen, Yunus Emre Sahin, Anil Yildirim, Ozan Yanar, parliament library and the staff at CSC for advice. The computer resources of the Finnish IT Center for Science (CSC) and the FGCI project (Finland) are acknowledged. Simola thanks OP Pohjola Group Research Foundation. All inaccuracies and mistakes are of our own.

1 Introduction

Political polarization is regarded as an important phenomenon by social scientists due to its associations with political efficiency and stability. For example, high levels of polarization may cause gridlocks in decision-making (Jones, 2001; Binder, 2004) or even threaten democracy (Arbatli & Rosenberg, 2021). While the exact definition of polarization is up for scholarly debate (Boxell, Gentzkow, & Shapiro, 2017), the definitions typically incorporate concepts like the 'extent of disagreement' or 'differences in attitudes or ideologies' between groups (DiMaggio, Evans, & Bryson, 1996). A predominant approach for measuring polarization in the political science literature has been to use votes cast in the parliament (roll-call votes) to estimate politicians' political positions and then aggregate distances between positions to a chosen level (e.g. Poole and Rosenthal (1985)). When party discipline is high, these approaches may underestimate the level of polarization. In this paper, we document the extent of differences in *speech* between political parties and party groups in Finland during the last century; a phenomenon that we will call partisanship of speech or simply partisanship in what follows.

Our paper uses 110 years of parliamentary speech data, covering all speeches from a period from 1907 to 2018. This time period covers large shifts in the societal landscape. For example, our data begins before the Finnish independence in 1917. Moreover, during our observation period, there has been a civil war (1918) as well as two wars (1939-1940 and 1941-1944). The period also marked a transition from a poor, agrarian country to a developed, urban society that Finland is today. It also includes a time period during the Cold War sometimes called *finladization*, which was characterized by strong Soviet Union influence in Finnish politics. In addition to the aforementioned changes, demographic shifts in the Parliament were also large, e.g., the seat share of women increasing from 10 percent to almost 50 percent. Our paper studies historical, long-term changes in political polarization. Using data from a long time span enables us to put the current trends in partisanship into a historical context.

In addition to having a historical perspective, this paper complements the currently sparse literature on the partisanship of political speech in a multiparty context, where "multiplicity of oppositions dominates decision making" (Nyholm, 1972). We study partisanship of political speech using minutes from parliamentary meetings as data. We extract text from plenary records from a period of more than one hundred years between 1907–2018, totaling more than 400,000 pages, and build a corpus of all speeches given in the Finnish parliament during the period. After preprocessing these speeches, we construct a vocabulary of all unique subsequent two-word pairs, bigrams. The terms 'bigram' and 'phrase' will be used interchangeably throughout the paper. The data for analysis consists of counts for phrases in the vocabulary on the speaker-year level[1].

The partisanship measure that we employ in this paper, introduced in Gentzkow, Shapiro, and Taddy (2019a), corresponds to the expected posterior probability of correctly guessing the party of the speaker after hearing them speak a single phrase. For this purpose, we first identify phrases that are used disproportionately by a party by contrasting the probability of that phrase in a party's speech against its total probability in that party's and the other party's speech. This phrase-level measure corresponds to the posterior probability of correctly guessing that the speaker comes from a specified party, given that they spoke the phrase. For example, the phrase 'kansalaissod.jälk' ('after the civil war') is spoken solely by speakers coming from the left parties during the 1920s as the right-wing used a different term when they talked about that war[2]. The posterior that the speaker is from a left party after hearing that phrase thus gets a value of 1. The posterior gets a value equal to 0.5 when a phrase forms a similar share in speech on both sides of the split. Second, we compute the yearly measure of partisanship by averaging the posteriors over the two party blocks.

The measure is intuitive but comes with a practical challenge. Phrases spoken only once during the parliamentary year seem extremely partisan since the probability of such phrase conditional on party will be zero for all but one of the parties. Since text data is often inherently extremely sparse and phrases spoken only few times are its dominant feature, such a deficit becomes a first-order concern. We address this sparsity or finite-sample problem using a recipe from Gentzkow et al. (2019a) and Taddy (2015). We substitute the relative frequencies calculated from the

[1]Subsequent studies have used these data to analyze the links between descriptive and substantive representation of different demographic groups within parties (Nieminen, Simola, & Tukiainen, 2023b) as well as the causal effect of media presence on political polarization (Nieminen, Simola, & Tukiainen, 2023a).

[2]Among the right-wing, the term 'vapaussota' (freedom war) was used to reflect the alternative framing for the conflict.

data with predictions from a LASSO-penalized multinomial logistic model. A desirable property of this penalization is that it imposes a threshold for what will be interpreted as a signal of partisan differences in phrase usage, rather than just noise. Using Poisson approximation as in Gentzkow, Shapiro, and Taddy (2019b) makes computing multinomial choice probabilities feasible by facilitating parallel computations of phrase choice probabilities for all phrases.

We study differences in speech 1) between left and right parties, 2) between parties in government and opposition, and 3) between pairs of parties that have been active since the early days of the unicameral parliament. The division between the economic left and the right has long been the fundamental cleavage characterizing political disagreement. We first study partisanship in speech between the left parties and the right parties. We find that differences in speech between the left and right parties fluctuate before the 1990s, peaking around 1910, at the end of the 1920s and in 1950. The sharpest increase takes place in the 1970s, and the high level is sustained until the end of the Cold War. Many of the most partisan phrases of the period concern foreign policy; examples include 'Finland Soviet' ('suome neuvostoliito'), 'German Democratic' ('saks demokraattis'), 'socialist countries' ('sosialistist maide') and references to the Finno-Soviet Agreement of Friendship, Cooperation, and Mutual Assistance of 1948 ('ystävyys.yhteistyö'). Since the 1990s, differences in speech have steadily increased after having decreased from the high levels observed in the 1970s. The levels of today are still far from reaching historically high levels. The current level of partisanship of speech is comparable to the level of the 1930s.

The highest levels of left-right partisanship in the Finnish data are observed in the mid-1970s. Comparing to Gentzkow et al. (2019a) who employ the same method in the U.S., the peak level of partisanship observed in the 1970s in Finland corresponds to the mid-1990s levels in the US. The increase in partisanship from the mid-1960s to 1970s roughly corresponds to the jump in the US from 1990 to mid-1990s. Partisanship seems to be driven by diverging speech of the extreme left; when the extreme left party (SKDL) is excluded, the large peak in left-right polarization in the 1970s flattens out.

One possible explanation for the high levels of polarization observed in the 1970s is Soviet Union information influencing through the extreme left party (SKDL), as SKDL – and especially the Finnish Communist Party which was part of the SKDL – was strongly influenced by the Soviet Union (Andrew & Mitrokhin, 1985). This explanation is supported by our observations that 1.) SKDL drives the 1970s peak in left-right polarization, 2.) Soviet Union related phrases were spoken very frequently in the Parliament in the 1970s, and 3.) the vast majority (70-80 percent) of Soviet Union related phrases were spoken by the SKDL. The results regarding the high prevalence of Soviet Union phrases in the 1960s and 1970s are also in line with the historical narrative of Finland where the entire post-1958 Cold War period, but especially the 1960s and 1970s, goes by the name of Finlndization and is characterized by strong Soviet influence in Finnish internal politics (Arter, 1998). As previous research suggests there are similarities between the information influencing tools used by the USSR and those of the contemporary Russia (Yablokov, 2022), our results may also be relevant to the present day.

Because Finland has a multi-party system, we also measure differences between individual parties to complement the left-right comparison. In those analyses, we measure differences between four parties that have existed since the early years in the data; the National Coalition Party, the Centre Party, the Social Democratic Party and the Communist Party (subsequently the Left Alliance). The Communist Party returned to parliamentary politics in 1944 after being illegal earlier. During more than three decades after this comeback, differences in speech between the Communist Party and the three other parties are large. Thereafter, the differences decrease until the 2000s. As expected, differences are smallest between the Communist Party and the Social Democrats and largest between the Communist Party and the right-leaning National Coalition Party. Differences in speech between the three other parties are small, however with a few distinguishable patterns. Differences in speech between the National Coalition Party and the Centre Party first increase during the 1970s and decrease thereafter. Differences between Social Democrats and the Centre Party, on the other hand, grow larger from the 2000s onwards.

In a multiparty system, a major force that may moderate differences between parties' political positions is the necessity to form government coalitions across party lines. We find that differences in speech between government and opposition emerge only in the 1970s. The differences in government-opposition speech re-emerge after the mid-1990s, exceeding the 1970s level. However, the differences between government-opposition speech today are only half of levels of left-right partisanship during their peak in the 1970s.

Overall, the results regarding the role of the left-right and the government-opposition cleavages are consistent with earlier voting-based studies by Nyholm (1972) and Pajala (2013) in Finland. Pajala (2013) finds evidence of the dominance of left-right division over the government-opposition division in plenary voting from the 1960s to the 1980s; government cohesion in voting is weaker than today. Starting from the 1990s, government votes in complete or nearly complete unison regardless of the parties that form the coalition. The diverging voting behavior of the extreme left is documented in Nyholm (1961) for the 1950s and in Pajala (2013) for the later decades.

Partisanship of political speech in multiparty context is studied by Peterson and Spirling (2018) in the UK and by Lauderdale and Herzog (2016) in Ireland. Peterson and Spirling (2018) use prediction accuracy of machine learning algorithms as the measure of partisanship. They find that partisanship has decreased since the beginning of the 2000s and is currently at levels of the 1960s.[3] Lauderdale and Herzog (2016) show that most variation in political speech in the multiparty system of the Irish Dáil takes place between the government and the opposition, with differences between establishment and anti-establishment parties being another important source of variation. Authors do not examine changes in government-opposition differences over time. Our time frame is also somewhat larger than that of Peterson and Spirling (2018) and much larger than that of Lauderdale and Herzog (2016), which is advantageous. For example, having a time period starting from 1907 allows us to study polarization in the context of a newly independent Finland and during a major national conflict like the Finnish civil war. This would not be possible with a shorter time frame[4].

Studying a different (bipartisan) context, Gentzkow et al. (2019a) find that partisanship of speech in the United States has increased dramatically during the last two decades. According to

[3] As pointed out by Gentzkow et al. (2019a), a falsification test presented in the paper's appendix may question this interpretation, as a series with randomly allocated party labels to speakers exhibits high levels of partisanship and thus signals of large bias in the measure.

[4] Fiva, Nedregård, and Øien (2022) study left-right partisanship in Norway with a shorter time frame beginning in the 1980s. Their paper is thus primarily informative of the recent changes in polarization. The results of their paper are also consistent with our results, as both our paper and theirs observe an increase in partisanship in the recent decades.

the authors, the probability of correctly guessing the party of the speaker based on a minute of speech (around 30 bigrams) has increased from 57 percent in 1989 to 73 percent in 2007. In our data, the probability of guessing the group affiliation right based on hearing 30 bigrams would be around 55 percent nowadays and was somewhat below 70 percent during the peak observed in the 1970s (see Appendix A for these magnitude calculations). Gentzkow et al. (2019a) suggest that the sharp increase in the partisanship of speech results from the professionalization of political speech ("innovation in political persuasion") after the Republican 1994 campaign *Contract with America*, with subsequently growing numbers of communications staff coining new terms to parties' vocabulary with an emotional appeal to the electorate. It is not clear *ex ante* what kind of patterns should be expected in political speech elsewhere. The United States seems to be an outlier in terms of how drastic the changes in affective polarization — antipathy towards party or parties other than one's own — have been during the last two decades (Boxell, Gentzkow, & Shapiro, 2021). Moreover, the US bipartisan political system stands in contrast with the multiparty system in place in Finland. In a multiparty system, political parties form government coalitions across party lines, and policymaking under coalition government necessitates bargaining and compromise (Martin & Vanberg, 2004). Understanding whether these dynamics moderate partisanship of political speech in multiparty systems is of interest on its own.

The splits studied in this paper are in no way a complete or an exhaustive description of the realm of political speech in a multiparty system. Some recent research from multiparty systems suggests that the current political polarization may take place along an alternative axis coined Green/Alternative/Libertarian and Traditional/Authoritarian/Nationalist (GAL-TAN), a trend that may be reconciled with the rise of identity politics (see Besley and Persson (2019) for theoretical discussion). Our result of modest left-right partisanship in the 2000s supports the narrative of the decreasing importance of the left-right dimension. Otherwise, the role of GAL-TAN parties is left open in this paper, as the parties relevant to GAL-TAN scale have only existed during the very latest decades of data we have. Our study aims to open the discussion on partisan differences in political speech over dimensions that are important through most of the history of the Republic of Finland. Historical evidence provided by this paper challenges the narrative that political polarization nowadays would necessarily be exceptionally high. Instead, our results suggest that the values of partisanship fluctuate and higher values have been observed in various points of Finnish history.

The paper proceeds as follows. Section 2 discusses Finland during the last century and describes the institutional background. Section 3 describes data constuction and preprocessing. Section 4 describes the underlying model of speech and the estimation procedure. Section 5 presents our results. Section 7 concludes.

2 Background

2.1 Finland during the last century

Finland gained its independence from Russia in 1917. The societal trend figures shown in Figure 1 show how the Finnish economy has evolved since gaining independence. At the start of its independence, Finland was a poor, rural country without much of a social welfare system. Panel C of Figure 1 shows that the share of public consumption expenditure was less than 10 percent, at the same level where countries such as Indonesia and Tanzania are nowadays.[5]

During 1918, just after Finland gained its independence, a civil war began in Finland. This is visible in the GDP growth figure presented in Panel A of Figure 1. The figure shows that in 1918, there was a largest relative drop in GDP (more than -10 percent) during the whole century, immediately followed by a very large increase (+ 20 percent) in GDP. The Panel A of Figure 1 also shows that GDP growth was rapid in Finland during the period from 1917 to 1970s. During this period, the average growth rate seems to have been around 5 percent per year, generally ranging from 0 percent to 10 percent if the years during the Second World War are excluded. After the 1980s, the growth rate has plummeted and there have been two large declines in GDP, the Finnish recession in the 1990s and the Great Recession that began in 2008.

The Second World War is clearly visible in Panel C, which plots the share of public consumption expenditure out of GDP. This measure gets its largest values in the years when Finland fought

[5] See World Bank (2021): General government final consumption expenditure (percent of GDP). Url: https://data.worldbank.org/indicator/NE.CON.GOVT.ZS, searched 6 March 2023

two wars (1939-1945). This subfigure also shows the big increase in public consumption after the wars. After the wars, the expansion of government expenditures peaked during the 1970s, which was also the time during which income inequality (Gini index) decreased rapidly. The expansion of government expenditures continued until the slump of 1990s and then again in the 2000s. This shows how Finland rapidly expanded its welfare system and the size of the public sector in general.

2.2 Finnish parliament during 1907-2018

The Finnish parliamentary system was subject to significant changes between 1907 and 2018. 1907 marks the beginning of the unicameral Parliament of Finland. A multiparty system with no minimum vote threshold replaced the former legislative assembly, where Four Estates of nobility, clergy, bourgeois and peasants had representation. Nearly 90 percent of the population above the voting age 24 were eligible to vote (Paloheimo, 2007), while in the 1905 elections the share had been around less than one in ten (Jyränki & Nousiainen, 2006). During its first decade, the unicameral parliament served the purpose of discussing societal matters with a fairly comprehensive representation of the electorate. However, the Parliament was severely limited in its decision making, as any bills needed the approval of the Emperor of Russia to be passed (Jyränki & Nousiainen, 2006). The Russian Emperor also convened the Parliament. During the First World War, Russia would exercise strengthened influence over Finland, and no assembly of the Parliament was called in 1915 and 1916 (Paloheimo, 2007).

The role of the opposition as the supervisor and the challenger of the government only originates after independence. The predecessor of the modern government was called the Senate. Senators were selected by the Russian Emperor, and they could also come from outside of the Parliament and were sometimes of Russian origin. The Senate responded to the Emperor, not the Parliament. The comparisons of speech between opposition and government in this paper thus start only in 1917 when the government assumes parliamentary responsibility.

Plenary speeches have been transcribed, first entirely in real time by professional stenographers and later with the help of audio recordings, already since the 1800s. Speech is transcribed with "as few and subtle alterations as possible" necessary for readability. For example, regional versions of personal pronouns are converted to the written standard language, but other regional and social variation in word choices are transcribed as such (Voutilainen, 2017). There is no upper limit for the length of a plenary speech – the right for unrestricted speech is constitutional as long as speaker sticks to the topic. However, certain speech types that have emerged since the mid-1960s are exceptions to this rule. Speeches during question hours, a plenary type introduced in 1966 to animate plenary discussions, are restricted in length to a few minutes. This restriction also applies to debates, which have been a part of the plenary type repertoire since 2012. The Speaker of the Parliament is allowed to exercise speaker selection during these plenary types.[6]

There is a consensus among parliament members that plenary speeches are mainly a means to communicate to the media and the electorate instead of trying to convince other parliament members or influence the content of legislative bills (Pekonen, 2011). The chances of parliament members to reach the public through plenary speeches vary with the publicity given to them by the media. The first radio live broadcast took place in 1926 but regular radio broadcasts started later. The first plenary session was broadcast in television in 1960, but regular televised broadcasts of plenaries only started in the 1980s.[7] Coming to the 2010s, the televised plenaries still reach hundreds of thousands of views.[8]. The causal effects of introducing TV broadcasting of plenary sessions have been analyzed in Nieminen et al. (2023a). Plenary sessions are also streamed online and their transcripts are published on parliament website typically within a couple of hours after a plenary session.

3 Data

This section briefly describes the data used in this paper. For a detailed description of data construction, preprocessing and sample selection, see Appendix A. The main dataset of our paper covers all records of the plenary sessions of the Parliament of Finland (*Eduskunta*) from 1907–2018. Since the Parliament did not gather in 1915 and 1916, the time series has a break for these

[6]https://www.eduskunta.fi/FI/naineduskuntatoimii/julkaisut/Documents/ekj_2+2017.pdf, retrieved 17 May, 2023.

[7]Email exchange, Päivi Erkkilä, The Library of Parliament's Information Service

[8]https://yle.fi/uutiset/3-6083211, https://www.finnpanel.fi/tulokset/tv/kk/ohjryh/viimeisin/ohjryh.html, retrieved April 10, 2020.

years. We perform Optical Character Recognition for data from 1907–2015 page by page using the tesseract OCR engine. Text for 2016–2018 is extracted directly from pdf file metadata. After splitting speech sections to speeches, speakers' names are linked to data from MP register[9] which contains, for example, speaker's party label, their gender, municipality of birth, their electoral districts and electoral terms.

In order to represent speeches as a large data matrix, we apply preprocessing to raw speeches, combine preprocessed words to phrases consisting of two consecutive words (*bigrams*), construct a dictionary from unique bigrams and count the occurrences of bigrams in the dictionary on speaker-year level. The word stems are concatenated to bigrams, combinations of two consecutive stems. Using combinations of consecutive stems introduces context – frequencies for *'työtätekev.luok'* (the stemmed bigram for 'working class') may convey more partisan information than frequencies for 'työtätekev' ('working') and 'luok' ('class') separately. Using two words for the context window size is an arbitrary choice. The common usage of compound words in Finnish to convey multiple meanings would make unigrams, single words, another potential candidate for the vocabulary unit. However, we follow Gentzkow et al. (2019a) in sticking with the bigrams, as bigrams are better than single words to capture meanings. Even though the Finnish language has long compound words, single words still rarely contain as much information as bigrams, i.e., if we used unigrams, we would lose some of the framing. For example, the context captured in the bigram ystävyys.yhteistyö (refers to the 1948 pact with the Soviet Union) is exactly the kind of relevant topic context we want to capture, as opposed to the generic unigrams ystävyys (friendship) or yhteistyö (co-operation).

The vocabulary consist of all unique bigrams uttered as part of parliamentary speeches, given that they exceed certain minimum frequency thresholds. The phrase must be used at least 100 times in total over all speeches, it must be used at least 10 times during at least one parliamentary year, and it must be used by 10 unique speaker-years. The restrictions follow the ones made by Gentzkow et al. (2019a), and come with the benefit of reducing the dimension of the count

[9] Kansanedustajamatrikkeli, retrieved from the Parliament library as a spreadsheet.

matrix.[10] After restrictions, the vocabulary consists of approximately 53,000 phrases.

Using a fixed vocabulary over the whole time period means that any new phrases emerging during the more recent years have less time to reach the total count of 100. Also, phrases popular in the early years and unpopular these days will still be a part of today's vocabulary while this is not the case in the other direction. We also omit a set of frequently appearing but ideologically uninformative procedural phrases and attempt to drop phrases containing speaker or party names. A more detailed description of these restrictions is provided in the Appendix.

The final data used for analysis consists of counts for the number of times each MP used each dictionary phrase during a parliamentary year. This count matrix has a row for each MPs who spoke a positive number of phrases in a year and has in total 19,094 rows (speaker-years) and 53,705 columns (phrases).

4 Model and estimation

4.1 Partisanship measure

Our data consists of counts for the usage of two-word phrases, i.e., word combinations like 'humanitarian crisis', on the MP-year level. To measure partisanship of a single phrase, a natural way to get started are the relative phrase choice probabilities, i.e., the proportion of speech by members of a party that went to the usage of a phrase, and contrast them in some way to the proportion of the phrase usage in other parties' speech.

Define the count of phrase j in speaker i's speech in year t as c_{ijt} and the total amount of speech by speaker i in year t as $m_{it} = \sum_j c_{ijt}$. The proportion of speech going to phrase j in party P in t is then $Pr(j|P,t) = q^P_{jt} = \frac{\sum_{i \in P_t} c_{ijt}}{\sum_{i \in P_t} m_{ijt}}$.

First consider the case where the possible parties are the left party and the right party, $P_{it} \in$

[10] We refer the reader interested in the potential effect of these restrictions on results to Gentzkow et al. (2019a), where their Online Appendix Figure 1 shows little effect of tightening this restriction on the average partisanship series. It is, however, worth noting that a crude stemming method like Porter is likely to behave worse on morphologically rich language like Finnish relative to English. As a result, counts for the same lemma of a word may be split between two vocabulary units (e.g. 'kotihoido tuen', 'kotihoido tuke' as in 'homecare allowance'), and thus the impact of the minimum thresholds could differ from results derived with English.

$\{L, R\}$. Following Gentzkow et al. (2019a), we define the "leftness" of a phrase as

$$\rho_{jt} = \frac{q^L_{jt}}{q^L_{jt} + q^R_{jt}} \tag{1}$$

Based on the probabilities for each party speaking phrase j, we thus compute the probability that a speaker comes from the left party L given that phrase j was spoken. The partisanship measure has the interpretation of the posterior probability that the speaker comes from the left party L given phrase j when the prior probability that the speaker comes from one of two parties is equal for both parties. If a phrase is used only by the left party, this measure will get a value of 1, and if a phrase is only used by the right party, the leftness measure will get a value of 0. The "rightness" measure is the complement of the "leftness measure", $1 - \rho_{jt}$. The measure of overall partisanship of speech in year t is an average of phrase partisanship over phrases and parties.

The underlying model of speech is the following. The vocabulary consists of two-word phrases $j \in \{1, ..., J\}$. Counts for phrases spoken by MP i during parliamentary year t are captured in a $1 \times J$ vector $\mathbf{c_{it}}$. The counts come from a multinomial distribution

$$\mathbf{c_{it}} \sim MN(m_{it}, \mathbf{q_t^{P_i}}(\mathbf{x_{it}})) \tag{2}$$

indexed by the total number of phrases spoken by MP i during parliamentary year t $m_{it} = \sum_j c_{ijt}$ and a vector of phrase choice probabilities $\mathbf{q_t^{P_i}}(\mathbf{x_{it}})$.

The multinomial choice probabilities are expressed in terms of phrase utilities

$$q^{P_i}_{jt}(x_{it}) = \frac{e^{u_{ijt}}}{\sum_k e^{u_{ikt}}} \tag{3}$$

where the utility for speaker i from phrase j in year t is

$$u_{ijt} = \alpha_{jt} + \gamma'_{\mathbf{j}}\mathbf{x_{it}} + \varphi_{jt}\mathbb{1}\{L_{it}\} \tag{4}$$

α_{jt} accounts for differences in popularity of a phrase across years. $\mathbf{x_{it}}$ includes potential confounders that might cause differences in phrase usage across parties for reasons not regarded as partisan, such as government-opposition status of speaker's party. $\mathbb{1}\{L_{it}\}$ indicates whether the speaker belongs to the left party L. It could also be a vector of party indicators with dimension $\#parties - 1$. The phrase choice probabilities within a year only differ due to the speaker's party and their characteristics x_{it}.

Partisanship at x, t is the posterior probability that a neutral observer assigns to speaker's true party after a single phrase of speech and is defined as

$$\pi_t(\mathbf{x}) = 0.5 \cdot \mathbf{q_t^L}(\mathbf{x}) \cdot \rho_\mathbf{t}(\mathbf{x}) + 0.5 \cdot \mathbf{q_t^R}(\mathbf{x}) \cdot (1 - \rho_\mathbf{t}(\mathbf{x})) \tag{5}$$

where $\mathbf{q_t^L}(\mathbf{x}), \mathbf{q_t^R}(\mathbf{x})$ and $\rho_\mathbf{t}(\mathbf{x})$ are vectors with dimension J (= total number of unique phrases). This is the total probability of correctly guessing the party of the speaker with characteristics x when party seat shares are 0.5, or, alternatively, when the observer making the guess knows that speaker comes from either left or right party with equal prior probability. The measure will always be at least 0.5. When speech is not partisan, $\rho_{jt}(x) = 1 - \rho_{jt}(x) = 0.5$. When speech is partisan and $\rho_{jt}(x) > 1 - \rho_{jt}(x)$, also $q_{jt}^L(x) > q_{jt}^R(x)$ by the definition of $\rho_{jt}(x)$ and vice versa. Thus $\rho_{jt}(x)$ values above 0.5 will get a higher weight in averaging than low (below 0.5) values. Note that this would not be true if, instead of 0.5, we used the true party seat shares in averaging.

The final partisanship measure is an average of the measure in (5) over all speakers active in year t.

$$\overline{\pi}_t = \frac{1}{|L_t \cup R_t|} \sum_{i \in |L_t \cup R_t|} \pi_t(\mathbf{x_{it}}) \tag{6}$$

4.2 Estimation

The Gentzkow et al. (2019a) phrase partisanship measure above is intuitive and comes with the benefit of identifying the most partisan phrases at different points in time. The measure mimics the decision-making process of someone familiar with the partisan connotations of language. As an example, in 1923, the phrase 'kansalaissod.jälk' ('after citizenwar') yields phrase partisanship

$\rho^L = 1$ with a count of 22 for left party MPs and 0 for others. On the right side, the war often went by the name of the Freedom War or the Red Rebellion. However, the same computation for 'iha.oikeast' ('really', count 1 vs. 0) also results in $\rho^L = 1$ in 1987. In general, using raw relative frequencies when the choice set is extremely large relative to the amount of speech is problematic. For any party-year combination, the counts for the majority of phrases are 0 or 1. Thus, the partisanship measure will often get the extremely partisan value of 1 just by chance. Note that this problem does not automatically disappear by imposing a minimum threshold for the total count of a phrase — even when the total exceeds 100 in all data, the finite sample problem can demonstrate itself in a single year. Thus, a solution is needed to filter the clearly partisan phrases from phrases that occur with different levels due to randomness, typically having small group-level differences in absolute counts.

Instead of using raw relative frequencies, one can construct the phrase choice probabilities $\hat{\mathbf{q}}_\mathbf{t}^\mathbf{P}$ from the predicted probabilities of the multinomial logistic model (Gentzkow et al., 2019a). Multinomial logistic model is a standard tool for modeling situations where discrete choices are made between multiple mutually exclusive alternatives. The simplest non-trivial multinomial logistic model for our purposes would have year and party-year indicators as predictors for phrase choice. The computations then consist of estimating $\#phrases \times (1 + \#predictors)$ parameters. This is an essential piece of the estimation strategy for controlling the severe finite-sample bias.[11] (Gentzkow et al., 2019a).

However, the estimation of the multinomial logistic model is computationally infeasible given that the dimension of potential choices is in the tens of thousands. Poisson regressions conditional on the log of total phrase count m_{it} can, fortunately, be used to approximate the parameters of the multinomial logistic distribution Taddy (2015). The Poisson regressions are separable across phrases and can thus be run in parallel. Parameter estimates acquired from the Poisson regressions

[11] Computing choice probabilities through multinomial logistic model has various benefits. Importantly for solving the above-mentioned finite sample problem, the model facilitates the usage of regularization, i.e. a method for parameter shrinkage and variable selection. Selection is particularly crucial for party-year indicators that control whether the predicted phrase choice probabilities differ between parties in a given year. As such, they indirectly determine which phrases will show up as partisan. A regression-based method like multinomial logit also allows for controlling for potential confounders, which could include speaker's government-opposition status or the speaker's birthplace that could affect speech because of regional linguistic variation.

are plugged in the formula for multinomial logistic probabilities. Similarly to Gentzkow et al. (2019a), we use this approach to estimate the partisanship series. The estimation procedure is describe more in detail below.

Following Gentzkow et al. (2019a) and based on a result in Taddy (2015) that the likelihood function for a factorial of independent Poisson regressions conditional on $\log(m_{it})$ approximates the likelihood for the conditional multinomial logistic model, we find the parameters for the multinomial choice probabilities by estimating J Poisson models. The advantage of the Poisson model is that, unlike the multinomial logit for which parameters of each choice probability need to be communicated to all terms of the likelihood function during estimation, it is separable across phrases. Thus, parameters of the Poisson distribution for each phrase can be estimated on a different CPU. The Poisson sacrifices the interdependence between phrases – independence of other phrases is assumed for the phrase choice probabilities, meaning that for example, the fact that a speaker just said "dear chair" would not impact their probability of saying "honored miss chairperson" next.

In the approximation, counts are distributed

$$c_{ijt} \sim Poisson(\exp(\mu_{it} + u_{ijt})) \tag{7}$$

where $\mu_{it} = log(m_{it})$. The negative likelihood function to minimize is, with a LASSO term:

$$\sum_t \sum_i [m_{it} exp(\alpha_{jt} + \gamma_{\mathbf{j}}' \mathbf{x_{it}} + \phi_{jt} \mathbb{1}\{L_{it}\}) - c_{ijt}(\alpha_{jt} + \gamma_{\mathbf{j}}' \mathbf{x_{it}} + \phi_{jt} \mathbb{1}\{L_{it}\}) + \lambda_j |\phi_{jt}|] \tag{8}$$

In the simplest regression, $x_{it} := x_t$ only includes year indicators.

The last term, $\lambda_j|\phi_{jt}|$, in the above formula penalizes the objective function for any non-zero ϕ_{jt} coefficient. Since the penalty in this method, known as the Least Absolute Shrinkage and Selection Operator (LASSO), is relative to the absolute value of the coefficient, the penalty has a kink at $\phi_{jt} = 0$. This form of penalty favors sparse solutions in which coefficients are set exactly to zero. Coefficient λ_j controls the strength of the penalty. It defines a threshold for the magnitude of differences in counts that are considered noise, rather than a signal of party differences. A larger λ results in a sparser solution, i.e. more zero ϕ_{jt} coefficients and thus fewer phrases whose usage will show up as partisan.

The value for λ_j is determined by trying multiple potential values and choosing the one that minimizes the Bayesian Information Criterion for the model. The largest potential λ_j value is determined by finding the lambda that sets ϕ_{jt} to zero. The other potential λ_j's are incrementally decreasing values of that largest potential value.

The negative log likelihood function for the factorial of J Poisson distributions is just the sum of the J likelihoods:

$$\sum_j \{\sum_t \sum_i [m_{it} exp(\alpha_{jt} + \gamma'_{\mathbf{j}}\mathbf{x_{it}} + \phi_{jt}\mathbb{1}\{L_{it}\}) - c_{ijt}(\alpha_{jt} + \gamma'_{\mathbf{j}}\mathbf{x_{it}} + \phi_{jt}\mathbb{1}\{L_{it}\}) + \lambda_j|\phi_{jt}|]\} \tag{9}$$

Gentzkow et al. (2019a) also recommend setting a small penalty for coefficients other than $\phi_{\mathbf{t}}$. The penalty enters as ψ the formula below. This helps in model convergence but comes with the downside of restricting the choice of λ. We set the value of $\psi = 10^{-5}$ but also test robustness for setting it to $\psi = 10^{-6}$ in Appendix Figure A6b.

$$\begin{aligned}\sum_j \{\sum_t \sum_i [m_{it} exp(\alpha_{jt} + \mathbf{x_{it}}\gamma_{\mathbf{jt}} + \phi_{jt}1_{i\in L_t}) - c_{ijt}(\alpha_{jt} + \mathbf{x_{it}}\gamma_{\mathbf{jt}} + \phi_{jt}1_{i\in L_t}) \\ + \psi(|\alpha_{jt}| + ||\gamma_{\mathbf{jt}}||_1) + \lambda_j|\phi_{jt}|\}\end{aligned} \tag{10}$$

4.3 Inference

Confidence intervals are constructed by subsampling. Intuitively, the subsampling procedure uses the distance of each yearly subsample estimate from the mean over all 100 yearly subsample estimate to approximate the variability of the yearly estimate. We draw 100 20 percent subsamples of data without replacement and re-estimate the series for each draw k to get $\hat{\pi}_{t,k}$. The confidence intervals are centered at estimate $\hat{\pi}_t$.

They are formed according to the formula

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

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 $k = 1, \ldots, 100$.

4.4 Phrase partisanship

The advantage of building the partisanship measure up from phrase level is the possibility of identifying the most partisan phrases that drive the partisanship measure at different points in time. A 'partisan phrase' has a high impact on the partisanship measure either because it is disproportionately used by the parties (high ρ_j), it is used a lot (high q), or both. A phrase that is used in similar amounts by both parties has low ρ_j, but its removal from the vocabulary may still result in large changes in partisanship measure because the weights for other vocabulary phrases will be scaled up. Put differently, phrase might have a big impact on partisanship also by moderating it.

The definition for partisanship of phrase j at x_{it}, following Gentzkow et al. (2019a), is

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

This is the change in the expected posterior for the event that the speaker comes from a left party at x_{it} when phrase j is removed from the vocabulary. It accounts to removing phrase j from the averaging step and scaling the choice probabilities for other phrases $k \neq j$ up to account for the removal of phrase j. The second term in the expression will be larger than 0.5 when removing the phrase increases the posterior for the event that the speaker is from a left party. Thus, negative values of $\zeta_{jt}(x_{it})$ will be right and positive values left. The absolute value is the magnitude of partisanship at x_{it}. Phrase partisanship $\overline{\zeta}_{jt}$ is the average of the measure over all speakers active in t.

5 Results

We study differences in speech for two-dimensional splits of the political party space. In Section 5.1, we study the traditional left-right dimension and classify parties to left and right parties. On a theoretical level, leftist parties emphasize a more egalitarian distribution of resources relative to other parties in parliament among other goals. In practice, we classify explicitly leftist parties into left parties based on mentions of keywords such as 'left', 'socialist', 'social democrat' or 'workers' in party's or party's predecessor's name.[12] Thus, parties around the center of the dimension are classified as right parties. Some parties in the resulting 'right parties' category are characterized by moderate or sometimes even somewhat leftist economic policies. As such, the label does not match perfectly with the content of the policies of those parties but is just one possible transparent classification. Appendix Figure A15 shows that the seat share of left parties fluctuates between 40 and 50 percent until it decreases to less than 30 percent after the 2000s. The sharp drop in the seat share after the Civil War of 1918 is due to all but one left party MPs having died, exiled to Russia or being trapped in prison camps (Jussila, Hentilä, & Nevakivi, 2009).

To analyze to what extent changes in the set of parties that gain seats in the parliament influence our results, we provide robustness analyses that focus only on parties that exist over a long time span. In those analyses, presented in Section 5.2.1, we study speech by four individual parties against each other. The parties are the National Coalition Party, the Centre Party, the Social Democratic Party and the Communist Party (SSTP, STPV and SKDL, followed by the Left Alliance). The first three of these parties have existed since the beginning of the 20th century and the last since 1918.

In Section 5.2.2, we consider differences in speech between parties in the government and in the opposition, as this is a dimension that has been found in previous literature to be important (Lauderdale & Herzog, 2016). Government coalitions are formed across party lines and vary in size over time. While the last minority government was in force in 1977, minority governments were pretty common in decades preceeding World War II and again in the late 1950s and early 1960s. Appendix Figure A16 shows the prime minister party and the share of parliament seats that parties in government coalition altogether hold at each point in time.

[12] Left parties are Demokraattinen vaihtoehto (Deva), Eduskuntaryhmä Puhjo, Ruotsalainen vasemmisto (RV/RVP), Sosialidemokraattinen eduskuntaryhmä (SDP), Sosialidemokraattisen opposition eduskuntaryhmä, Sosialistinen eduskuntaryhmä "kuutoset", Suomen kansan demokraattisen liiton eduskuntaryhmä (SKDL), Suomen sosialistinen työväenpuolue (SSTP), Työväen ja pienviljelijäin puolue, Työväen ja pienviljelijäin sosialidemokraattinen liitto (TPSL), Työväen ja pienviljelijäin vaaliliitto (STPV), Vasemmistoliiton eduskuntaryhmä, Vasemmistoryhmä and Vasenryhmän eduskuntaryhmä (vr).

The figures below show the evolution of the partisanship measure over time for this set of splits. As a sanity check, we also show results from a permutation test where the party label is randomly assigned to active MPs, holding constant the relative shares of the labels. Partisanship that stays constant and close to 0.5 in the random series validates the results in the real series – then, any changes in the real series are driven by differences in speech between the party groups and not by randomness.

The random series may, however, also exhibit changes. This would suggest that there are drivers other than partisan differences in speech that results in differences in speech; one such driver could be the finite sample bias discussed in Section 4.2. In a regularized series, variability in the random series could signal that the LASSO is not successful in distinguishing signal from noise. On the other hand, reasons other than the failure of regularization could result in unstable random series. Strong manerisms or otherwise very divergent speech by individual representatives will always drive partisanship, no matter which side of a split they occupy. Adding control variables will help in accounting for compositional changes in the parties in terms of gender, government status and region.

5.1 Left-right partisanship

Figure 2 shows the main result of the paper. The left-right partisanship measure is constructed from the penalized choice probabilities. Control variables used for the estimated choice probabilities of this main specification include an indicator for whether the speaker's party is in the government, the speaker's gender, and speaker's region.

The partisanship measure, net of the random series, exhibits high levels before the 1918 Civil War and goes down in the aftermath when all MPs except for one from the losing leftist side have either exiled in Russia, died or been trapped in a prison camp (Jussila et al., 2009). Partisanship again increases during the 1920s, characterized by tensions in the domestic policy and difficulties in parliamentary cooperation, until the passing of Communist laws in 1930. The measure stays flat for the next decade. Partisanship again peaks in 1950. The period from the 1970s to the mid-1980s stands out the most, with left-right reaching it highest levels in mid-1970s. The period starting from 1990 shows relatively stable partisanship. The permutation test, labelled the random series, provides support for a partisan interpretation of the fluctuations in the real series — the random series stays flat throughout the period.

Even the highest levels of left-right partisanship in Finland are well below the US levels of recent years. In the US, the average phrase partisanship rises above 0.51 after 2010, with simulations showing that this corresponds to correctly inferring the party of the speaker with around 73 percent probability after one minute of speech ($\sim$ 33 phrases, Gentzkow et al. (2019a)). In the Finnish data, the highest level of average phrase partisanship is around 0.506 in the mid-1970s. This corresponds to the mid-1990s levels in the US data, when partisanship had already jumped from its 1990 level. The increase in partisanship from the mid-1960s to 1970s roughly corresponds to the jump in the US from 1990 to the mid-1990s, when the 1994 election presumably professionalized the language of politics for good.

Appendix Figure A2 illustrates the impact of covariates in the evolution of the partisanship measure. When no controls are added (Appendix Figure A2a), the random series fluctuates together with the real series. The difference between the permutation tests for the main specification in Figure 2 and in Appendix Figure A2a with no added covariates could signal about speech differences driven by the changing composition of the parliament.

Appendix Figure A2b shows that the evolution in partisanship is not driven by parties' government status. The speech between the government and the opposition varies for non-partisan reasons demonstrated by phrases like 'order of the day' or 'government [budget] cuts' ('yksinkert päiväjärjestyks', 'päiväjärjestyks tehd', 'hallitus leik').

Appendix Figure A2c shows that adding controls for representative's region considerably smoothens the random series. One reason for controlling for region is the dialectical differences in speech between areas. The region covariates indeed seem to have a large impact on the choice probabilities of certain fill words ('sit.tämmöis', 'elik.tääl'), but also on potential agenda phrases like 'Kuusamo Posio' and 'northern people' ('kuusamo posio', 'pohjois ihmis'). Thus, region controls will partly account for differences in speech resulting from agenda differences between representatives coming from different regions. Since analyses aim to describe general agenda differences in speech between parties, not regions, this property will be desirable. Also, controlling for region will not impact the phrase choice probabilities for phrases concerning the rural people in general, as long as such phrases are used by representatives coming from various (rural) regions.

The main specification finally includes controls for the speaker's gender. Controls for gender further smoothen the evolution of the random series. Phrases with large gender coefficients include phrases like 'abortion', 'born outside' and 'work life balance' (raskaud.keskeytyks, ulkopuol.syntyn, perheeläm.yhteensovittamin).

5.1.1 Soviet Union influence through the extreme left party in the 1960s and 1970s

The 1970 hike in partisanship coincides with the pro-Soviet movement ("taistoism", according to the movement's leader Taisto Sinisalo) gaining a foothold within the communist party SKDL. The party was split into two factions with deep disagreement but still functioned as a single parliamentary group. Figure 3 presents left-right partisanship results when SKDL is left out from the analysis. The figure shows that the when SKDL is dropped, the time series flattens and the pronounced levels of partisanship in the 1970s shown in Figure 2 disappear. This suggests that the 1970s peak is mainly driven by the SKDL. When SKDL is dropped, we still observe significant left-right partisanship, but the level of partisanship seems to be quite stable at around 0.502 throughout the whole time period.

SKDL had close ties with the Soviet Communist Party, which was known to employ information influencing and propaganda as "active measures" - tools in Soviet political warfare (Cull, Gatov, Pomerantsev, Applebaum, & Shawcross, 2017). Thus, we wonder if we can detect any signs of Soviet influence in the speech by SKDL that could then have contributed to the central role the party had in driving polarization in the Finnish parliament in the 1970s.

We plot the prevalence of a set of handpicked Soviet Union related phrases[13] in the Finnish Parliament and find that these Soviet-related phrases were most common in the 1970s. Figure 4 shows the series. The peak coincides with peak polarization observed in Figure 2. The largest

[13]These phrases include all bigrams that contain the term 'neuvostoliito' (Soviet Union), as well as other words that are closely related to Finland-Soviet Union relations (such as 'ystävyys.yhteistyö' related to a co-operation pact between the countries). These bigrams are listed in Online Appendix B ('List of Soviet Union related bigrams').

peak ending in around 1985 matches well the fact that Soviet Union information influencing efforts decreased when Gorbatshev became the leader of the country in 1985 (Galeotti, 2019). Thus, our figures would be consistent with Soviet Union information influencing driving the high prevalence of Soviet Union phrases and high polarization before Gorbatchev's reign. The strong correlation between the polarization series and the use of Soviet Union phrases totally disappears at the time of the dissolution of the Soviet Union in the 1990s. Figure 5 shows that the vast majority (around 70 percent) of the Soviet related phrases were spoken by speakears from SKDL.

The association between SKDL and speech related to Soviet Union is also consistent with what is known in the literature about the party. For example, Arter (2022) argues that SKDL was an important channel for the Soviet Union to advance their interests. Research has also suggested that the Soviet Union strongly influenced and financed the Finnish Communist Party (SKP), which was part of the SKDL.[14] During the same time that we observe peak polarization, The National Coalition Party, which was the most vocal opponent of the Finnish Soviet policies with around 20 percent seat share in the 1970s and 1980s, was left outside of government coalitions for 22 years from 1966 to 1987 for "general reasons that anyone considering the position of Finland understands", described by the then Speaker of the Parliament Johannes Virolainen in 1979.

The high prevalence of Soviet Union phrases in the 1960s and 1970s also coincides with Finlandization, a period characterized by strong Soviet Union influence in Finnish politics (Arter, 1998). During the mid-1970s time when left-right polarization was the highest in Finland, Soviet influence was present also in many other countries, including South Africa (Barratt, 1981), Egypt (Dawisha, 1979) and Niger (Ojo, 1985). According to Andrew and Mitrokhin (1985), the 1970s were also a time during which more illegal Soviet agents began working for communists parties in Europe, Asia and Africa.

In addition to the largest spike in the 1970s, there are also two smaller peaks in the polarization series. First, there is an increase in polarization lasting approximately from 1940 to 1948. This coincides with the period that has been later described as "the years of danger" in Finland,

[14] The Soviet Union also favored the party when they interfered with the formation of coalition governments in Finland. For example, according to Andrew and Mitrokhin (1985), the Finnish Communist Party (which was part of the SKDL) was given many key positions in the Finnish government in 1945, via negotiations through a special channel between the Soviet Union and the bourgeois parties.

as the risk of Finland becoming communist was perceived to be high at that time (Rautkallio, 1990). During these years, the SKDL was also popular among the public. In 1946, the SKDL parliamentary group was the largest group in the Finnish Parliament, but they lost a large number of seats in 1948. However, we also observe a short hike in left-right partisanship in the early 1960s, when the Cold War had some of its most tense moments, such as the Berlin Crisis of 1961 and the Cuban Missile Crisis of 1962.

Overall, we show that the all-time high in Finnish polarization coincides with a peak in Soviet-related speech, spoken primarily by the party that is responsible also for the peak in polarization. According to historical writing, SKDL was used as a channel by the Soviet Communist Party to promote Soviet interests in Finnish politics. These factors together propose a possible link between Soviet information influencing and political polarization in Finland.

5.1.2 The impact of the populists

One important actor in the Finnish politics of the 1970s was also the Finnish Rural Party SMP, which gained 18 parliament seats in the "protest elections" of 1970. The party was populist, with a message targeted to the rural population (the "forgotten people"). The party dominates in the number of speeches given throughout the 1970s, with the representatives, having 10 percent of parliament seats, accounting for 30 percent of parliamentary speech. The party was also infamous for constituting an enormous surge in the number of Parliament's legislative bills – so much so that the surge goes by the name of "Vennamo effect" or the "SMP effect" among Finnish political scientists (Pajala, 2010). The party is a part of the right parties in our left-right categorization, but ideologically it is hard to classify along the traditional left-right axis. It could, however, have a considerable impact on the partisanship measure given its high proportion in speeches.

Appendix Figure A3 shows that while constructing the partisanship measure without representatives from SMP lowers year-specific partisanship peaks in the 1970s, it does not alter the overall salience of the decade in the series. After controlling for gender, dialect region, and government status, the differences in speech between the left and the right in the 1970s do not seem to be driven at all by the populist protest party.

5.1.3 The impact of the LASSO

Appendix Figure A4 illustrates the problem of constructing the partisanship measure from raw empirical phrase choice probabilities. In the Figure, phrase choice probabilities are constructed directly from data according to $\widehat{q}^L_{jt} = \frac{c^L_{jt}}{\sum_k c^L_{kt}}$ and $\widehat{q}^R_{jt} = \frac{c^R_{jt}}{\sum_k c^R_{kt}}$, where c^L_{jt} is the count for phrase j in t for speakers in left parties L. Right parties are denoted by R. 'Real' series is based on counts from the data for left and right parties, and the 'random' series is a result of a permutation test based on counts from data where party labels are randomly assigned. The figure shows how the real and random series move in parallel throughout the time interval. Random noise dominates any signal in the data to the extent that there is no noticeable difference neither between the levels nor the trends in the series. Accounting for finite sample bias completely changes the story as demonstrated earlier in Figure 2.

Appendix Figures A5a and A5b give further idea about the impact of the LASSO. These figures show that a small share, around 1 percent, of phrases used in a given year have a positive party-year coefficient. Appendix Figure A5a shows that, like the amount of speech, also the number of phrases used increases over time. Appendix Figure A5a shows that the number of phrases giving rise to partisanship measure π_t via phrase partisanship ϕ_{jt} similarly increases over time broadly at the same rate as phrases with nonzero counts, except for a couple of recent years in the data where the share of partisan phrases increases.

As mentioned in Section 4.2, a penalty imposed on the covariates other than the party-year indicators helps in faster convergence of the model, but comes at the cost of restricting the choice of penalty parameters λ for the coefficients of the party-year indicators. Appendix Figure A6 illustrates the impact of changing the penalty parameter ψ. The partisanship in the series is 'no controls'; however, ψ imposes a cost on year indicators that are part of all of the partisanship models estimated in this paper, though not explicitly mentioned. A lower ψ restricts the choice of λ less. Thus, the penalization is stronger and both the random and the real series are smoother in Appendix Figure A6b than in Appendix Figure A6a.

5.1.4 Phrase partisanship

Table A4 shows five phrases with highest ζ_j in the main specification with government party, gender, and region controls. The table also shows predicted counts for these phrases per 100,000 phrases. For comparison, the total phrase count after preprocessing and pruning the phrase set is around 50,000 in 1907, after which it increases and crosses 100,000 yearly phrases in 1917, and amounts to around 500,000 phrases in 2018. Table shows an overview of the partisan phrases driving variation in π_t for every tenth year in the data.

The most partisan phrases include divisive topics from price increases in the 1950s to the European integration in the 1990s. Appendix Figure A14 shows the evolution of partisanship for some of the most partisan phrases of each decade. 'university of helsinki' ('helsing yliopisto') is a phrase with high right partisanship, peaking in mid-1930s and again around 1970. The right-wing student movement of the 1930s, headed by the right-wing Academic Karelia Society (Akateeminen Karjala-Seura AKS), demanded that the teaching language in universities should be changed from Swedish to Finnish (Jussila et al., 2009). 'constitution 14' (hallitusmuodo.14) also relates to the language question; Section 14 of the 1919 Constitution lays down the national languages of Finland, Finnish and Swedish. In the 1940s, left parties make a disproportionate number of references to the state police ('valtiollis poliis'). Under the governance of the post-WWII Minister of the Interior Yrjö Leino from the Communist Party, the organization previously devoted to anti-communist activities took a new turn, and its staff was largely replaced by new members with connections to the extreme left. The usage of the phrases 'big capital advantage' ('suurpääom etu') and 'finland soviet' ('suome neuvostoliito') increases in the 1970s.

In general, Marxist terminology ('bigcapital.advantage', bourgeois) is primarily a part of the vocabulary of the left, and the issues of the war veterans ('sotie veteraan', 'ylimääräis rintamalis') are disproportionately discussed by the right. 1970s phrases are charaterized by content relating to the foreign policy. Phrases 'renewable energy' ('uusiutuv energia') and 'asylum seekers' ('turvapaik hakij') appear among the most partisan phrase of the 2000s. This pattern may be interpreted as evidence of the surging importance of the GAL-TAN dimension, since neither of these issues are not distinctively left or right.[15] Appendix Tables A5 – A16 list the 5 most partisan phrases for every year in data.

5.2 Other between-party comparisons

5.2.1 Differences in speech between individual parties: Pairwise comparisons

Next, we compare partisanship of speech between four parties that have existed since the early years of the unicameral parliament: the Social Democratic Party, the National Coalition Party (the Finnish Party until 1918), the Centre Party (Agrarian League until 1965) and the Communist Party (SSTP, STPV and SKDL, the Left Alliance after discontinuation ofthe Communist Party). The National Coalition Party is typically positioned on the right end of the left-right axis, the Centre Party in the center and Social Democratic Party on the left.[16] The Communist Party represents the extreme left. We construct the following series without control variables to avoid "overfitting" due to overly small cell sizes.

Appendix Figures A7a, A7c and A7e show that when measuring differences in speech between these more atomic units, the level of the random series is not at 0.5 and fluctuates together with the real series. This might suggest that the penalized estimator is not successful in distinguishing signal from noise. However, looking at net-of-random variation in the figures, the differences in speech between the National Coalition Party and the Centre party seem to be narrowing down during the 1980s. Also, the net-of-random partisanship the between the Social Democrats and Centre Party increases after a pretty stable gap until the mid-2000s. The differences between these parties could be undermined when the parties are on the same side of the government-opposition split: the Social Democrats and the Centre Party are part of the same government coalition from the mid-1960s to the mid-1980s. From the mid-2000s until 2019, the parties are on the opposite sides of the government-opposition split. The National Coalition Party is in the opposition for 22 consecutive years from 1966 to 1987.

Partisanship seems to be driven by diverging speech of the extreme left. The Communist

[15] The phrases are characteristic of right parties in our classification potentially because the Green League and the Finns Party are not explicitly left in our criteria and are thus classified into the 'right parties' group.

[16] e.g. https://www.hs.fi/politiikka/art-2000006026006.html

Party returned to parliamentary politics in 1944 after being illegal earlier. During more than three decades after the parliamentary comeback, differences in speech between the Communist Party and the three other parties are large. Thereafter, the differences decrease until the 2000s. As expected, differences are smallest between the Communist Party and the Social Democrats and largest between the Communist Party and the National Coalition Party. Differences in speech between the three other parties are small, however with a few distinguishable patterns. Differences in speech between the National Coalition Party and the Centre Party first increase during the 1970s and decrease thereafter. Differences between Social Democrats and the Centre Party, on the other hand, grow larger from the 2000s onward.

5.2.2 Differences in speech between the government and the opposition

Studying government - opposition differences in speech over time is also of interest. Earlier studies by Nyholm (1972) and Pajala (2013) show that parties often vote according to government-opposition line and that the government-opposition dimension seems to dominate the left-right dimension in the politics of the 2000s. It is interesting to see whether differences in speech follow similar patterns. The series starts in 1917, when the government assumes parliamentary responsibility and data becomes available.

The random series in Appendix Figure A8 exhibits considerable variation, again suggesting difficulties in filtering signal from noise in data. Net of random variation in partisanship measure, the differences between government-opposition speech are high during the seventies and rise again starting in the mid-1990s. The level of government-opposition partisanship today is slightly larger than left-right partisanship.

5.3 Polarization and other societal phenomena

Earlier literature has found that polarization correlates positively with a smaller government (Lindqvist & Östling, 2010) and income inequality (Acosta, Pena, & Saalfeld, 2020; Grechyna, 2016), and negatively with trust in government (Grechyna, 2016). High levels of polarization have also been associated with legislative gridlock (Jones, 2001).

To understand more about the societal dynamics that interplay with polarization, we plot time series of various societal indicators along with the polarization series in Figure 6. The outcomes we compare to the polarization of parliamentary speech include voter polarization, number of extremely long speeches (filibustering), law proposals/bills by MPs, length of coalition governments, laws passed and government law proposals. Based on the series, polarization of parliamentary speech seems to be linked to higher voter polarization, shorter length of coalition governments and more inefficient policymaking (more bills, less laws).

6 Conclusions

This paper documents the evolution of differences in speech between left and right parties, between government and opposition parties, and between the four parties that were active in the Finnish parliament since the introduction of the unicameral parliament in 1907. We find that the recent increases in polarization are nothing unusual in the modern history of Finland. In Finland, the highest levels of speech partisanship are documented in the 1970s.

The relationship with the superpower Soviet Union seems to play a role in speech partisanship throughout the 1970s – the stemmed version of 'Finland Soviet Union' appears among the five most left phrases in four years both during the 1970s and 1980s. Moreover, we find that the share of all phrases where the Soviet Union is mentioned is the highest during the 1970s. The party the drives both the peak in polarization and the prevalance of Soviet Union related phrases is the extreme-left party SKDL, which is a predecessor of the current Left Party and at that time the home to the Finnish Communist Party. One interpretation of these results is that Soviet Union information influencing may have played a role in increasing left-right divides in Finland in the 1970s, or at the very least, that Soviet Union related matters were behind most of the high left-right divides observed. As the 1970s peak in polarization also co-occured with short-lived government coalitions and less efficient policymaking (more bills, less laws), it looks like Soviet Union may have succeeded in creating instability in Finland in the 1970s.

Our results suggesting Soviet information influencing may have played a role in increasing polarization and decreasing stability are very relevant also to the present world situation, as there are many papers arguing that the present Russia uses very similar tools that Soviet Union used. For example, the contemporary Russian information influencing aims to weaken the trust in government in foreign countries, for example by promoting various conspiracy theories (Yablokov, 2022). According to Yablokov (2022), there are striking similarities between the information influencing by contemporary Russia and the information influencing campaigns in the Soviet Union.[17] Erlich and Garner (2023) find that those with partisan or ethnolinguistic ties to Russia are more prone to believing Russian misinformation. This is similar to our results in a sense that it was a specific party (SKDL) with ties to the USSR that drove the polarization observed in the 1960s and 1970s.

Popular writing often suggests hate speech as an emerging trend in the arenas of public speech. It is worth noting that the partisanship measure will not capture phenomena such as across-the-line harshening of political rhetoric, if all parties toughen their rhetoric to a similar extent simultaneously. Even though such phenomena could be related to affective polarization, i.e. antipathy towards parties other than one's own party, it seems that hostile speech or hate speech would clearly be a phenomenon distinct from partisanship in speech.

This paper provides a comprehensive, century-long overview of the history of parliamentary speech in Finland. This paper thus offers a long within-country analysis of polarization in a European country, complementing papers such as Boxell et al. (2021) that have studied polarization in cross-country settings. The goal of this paper is to use our unique historical data and paint a long narrative of polarization over the course of over 100 years of history of Finland. Causal explanations for the reasons behind polarization are beyond the scope of this paper, but we show time series data of several co-occuring phenomena, which may be used as pointers for further examinations of the topic. Other potential extensions could study the role of publicity in spurring partisan speech, which has been touched in Nieminen et al. (2023a) with regards to the effect of TV, but more research could be done relating to the effects of other forms of media on parliamentary speech. To complement the depiction of partisanship provided in this study, more partitions of the party space could be studied. Also the influence of Soviet propaganda in speech could be

[17] Yablokov (2022) discusses some examples from the Soviet Union era, such as the instance of USSR promoting the conspiracy theory of AIDS having been developed in U.S. laboratories. In the contemporary world, Yablokov (2022) mentions that Russia has been involved in promoting conspiracy theories related to the COVID-19 pandemic.

examined further.

References

Acosta, C. P., Pena, J. C., Saalfeld, T. (2020). *Inequality, macroeconomic performance and political polarization a panel analysis of 20 advanced democracies.*

Andrew, C., Mitrokhin, V. (1985). *The sword and the shield - the mitrokhin archive and the secret history of the kgb.*

Arbatli, E., Rosenberg, D. (2021). United we stand, divided we rule: how political polarization erodes democracy. *Democratization*, *28*, 285-307. doi: 10.1080/13510347.2020.1818068

Arter, D. (1998). Kekkonen and the 'dark age' of finlandised politics? *Irish Studies in International Affairs*, *9*, 39-49.

Arter, D. (2022). From finlandisation and post-finlandisation to the end of finlandisation? finland's road to a nato application. *European Security*. doi: 10.1080/09662839.2022.2113062

Barratt, J. (1981). *The soviet union and southern africa.* Suid-Afrikaanse Instituut van Internasionale Aangeleenthede = South African Institute of International Affairs.

Besley, T., Persson, T. (2019). The Rise of Identity Politics.

Binder, S. A. (2004). *Stalemate: Causes and consequences of legislative gridlock.* Brookings Institution Press.

Boxell, L., Gentzkow, M., Shapiro, J. (2021). Cross-country trends in affective polarization. *The Review of Economics and Statistics*, 1-60.

Boxell, L., Gentzkow, M., Shapiro, J. M. (2017). Greater internet use is not associated with faster growth in political polarization among us demographic groups. *Proceedings of the National Academy of Sciences*, *114*(40), 10612–10617.

Cull, N., Gatov, V., Pomerantsev, P., Applebaum, A., Shawcross, A. (2017). *Soviet subversion, disinformation and propaganda: How the west fought against it an analytic history, with lessons for the present.*

Dawisha, K. (1979). *Soviet foreign policy towards egypt.* Palgrave Macmillan London.

DiMaggio, P., Evans, J., Bryson, B. (1996). Have American's social attitudes become more polarized? *American Journal of Sociology*, *102*(3), 690–755.

Erlich, A., Garner, C. (2023, 1). Is pro-kremlin disinformation effective? evidence from ukraine. *International Journal of Press/Politics*, *28*, 5-28. doi: 10.1177/19401612211045221

Fiva, J. H., Nedregård, O., Øien, H. (2022). *Polarization in party-centered environments: Evidence from parliamentary debates.* Retrieved from https://www.jon.fiva.no/docs/FivaNedregardOien.pdf

Galeotti, M. (2019). *Active measures: Russia's covert geopolitical operations.* Retrieved from https://www.marshallcenter.org/en/publications/security-insights/active-measures-russias-covert-geopolitical-operations-0

Gentzkow, M., Shapiro, J. M., Taddy, M. (2019a). Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech. *Econometrica*, *87*(4), 1307-1340. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA16566 doi: 10.3982/ECTA16566

Gentzkow, M., Shapiro, J. M., Taddy, M. (2019b). Measuring group differences in high-dimensional choices: Method and application to congressional speech. *Econometrica*, *87*, 1307-1340. doi: 10.3982/ecta16566

Grechyna, D. (2016, 7). On the determinants of political polarization. *Economics Letters*, *144*, 10-14. doi: 10.1016/j.econlet.2016.04.018

Jones, D. R. (2001). Party polarization and legislative gridlock. *Political Research Quarterly*, *54*(1), 125–141.

Jussila, O., Hentilä, S., Nevakivi, J. (2009). *Suomen poliittinen historia 1809-2009.* WSOY Oppimateriaalit.

Jyränki, A., Nousiainen, J. (2006). *Eduskunnan muuttuva asema.* Edita.

Lauderdale, B. E., Herzog, A. (2016). Measuring political positions from legislative speech. *Political Analysis*, *24*(3), 374–394.

Lindqvist, E., Östling, R. (2010, 8). Political polarization and the size of government. *American Political Science Review*, *104*, 543-565. doi: 10.1017/S0003055410000262

Martin, L. W., Vanberg, G. (2004). Policing the bargain: Coalition government and parliamentary scrutiny. *American Journal of Political Science*, *48*(1), 13–27.

Nieminen, J., Simola, S., Tukiainen, J. (2023a). *Does the presence of tv cameras affect polarization in parliamentary speech?*

Nieminen, J., Simola, S., Tukiainen, J. (2023b). *Evolution of within-party group differences: Evidence from 110 years of parliamentary speech.*

Norvig, P. (2007). How to write a spelling corrector. *Online at: http://norvig. com/spell-correct. html*.

Nyholm, P. (1961). *Suomen eduskuntaryhmien koheesio.*

Nyholm, P. (1972). Parliament, Government and Multi-Dimensional Party Relations in Finland. *Commentationes Scientiarum Socialium*, *2*.

Ojo, O. J. B. (1985). The soviet union and nigeria: The quest for influence 1967-1983. *Africa Spectrum*, *20*, 167-189.

Pajala, A. (2010). Plenary votes and voting cohesion in eduskunta: A descriptive analysis of the 1991-2006 parliamentary years. *Acta politica aboensia B*, *2*.

Pajala, A. (2013). Government vs opposition voting in the Finnish parliament Eduskunta since World War II. *European Journal of Government and Economics*, *2*(1), 41–58.

Paloheimo, H. (2007). Eduskuntavaalit 1907–2003. *In Ollila, Anne ja Paloheimo, Heikki: Kansanedustajan työ ja arki. Suomen eduskunta*, *100*, 173–369.

Pekonen, K. (2011). *Puhe eduskunnassa.* Vastapaino.

Peterson, A., Spirling, A. (2018). Classification accuracy as a substantive quantity of interest: Measuring polarization in Westminster systems. *Political Analysis*, *26*(1), 120–128.

Poole, K. T., Rosenthal, H. (1985). A spatial model for legislative roll call analysis. *American Journal of Political Science*, 357–384.

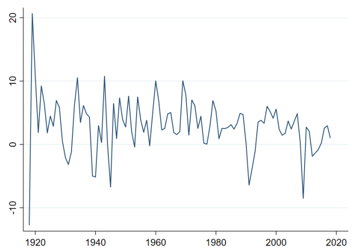
Rautkallio, H. (1990). *Paasikivi vai kekkonen, suomi lännestä nähtynä 1945-1956.* Tammi.

Taddy, M. (2015). Distributed multinomial regression. *The Annals of Applied Statistics*, *9*(3), 1394–1414.

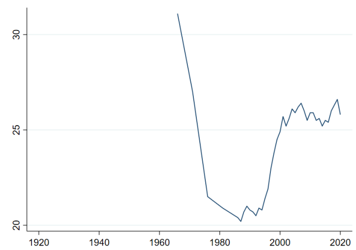
Voutilainen, E. R. J. (2017). The regulation of linguistic quality in the official speech-to-text reports of the Finnish Parliament. *CoMe: Studies on Communication and Linguistic and Cultural Mediation*, 61–73.

Yablokov, I. (2022, 6). Russian disinformation finds fertile ground in the west. *Nature Human Behaviour*, *6*, 766-767. doi: 10.1038/s41562-022-01399-3

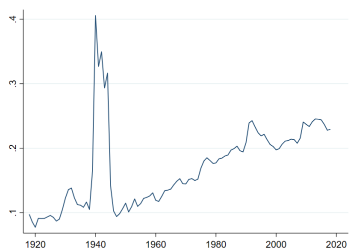
Figures



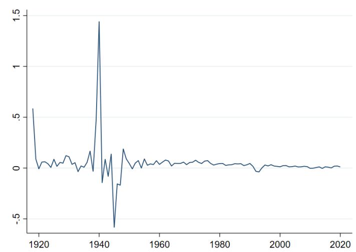
(a) GDP growth



(b) Gini index



(c) Public consumption expenditure/GDP

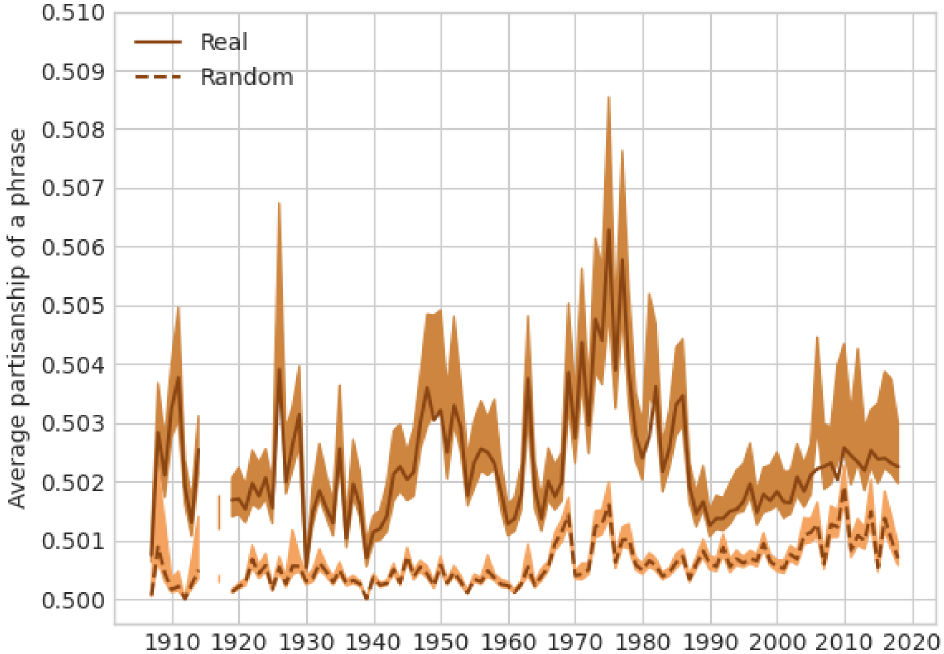


(d) Yearly change in public consumption expenditure

Figure 1: The Finnish economy during 1917-2018

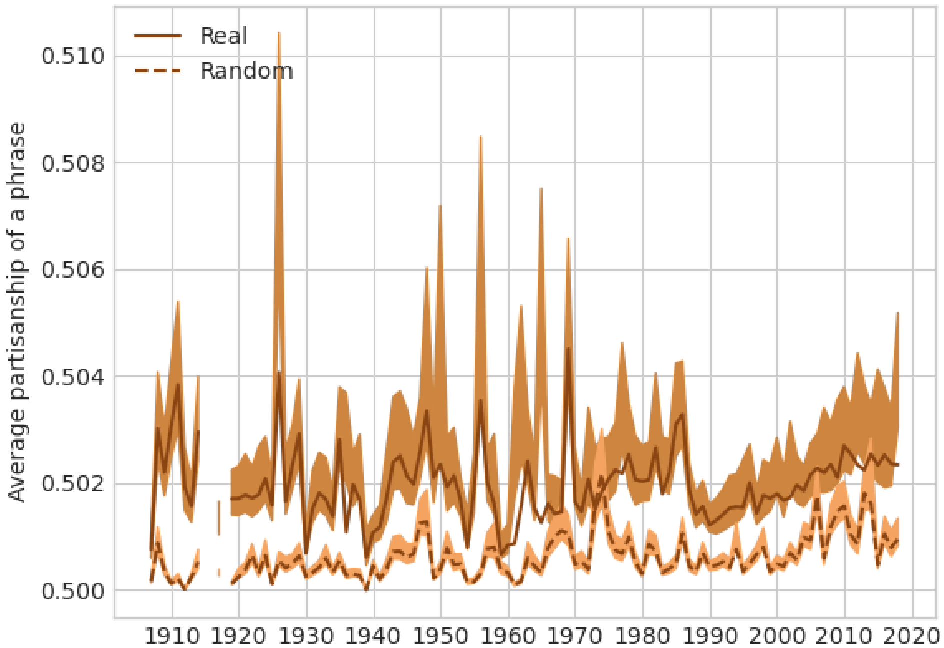
Notes. Figures show how various economic time series have developed in Finland during years 1917-2018.

Figure 2: Left-right partisanship, controls for government status, speaker gender and region



Notes: Results are based on regressions with controls for government status of speaker's party, speaker's gender and the region of the speaker. Confidence intervals are based on subsampling and have 80 percent nominal coverage. The confidence intervals are computed via subsampling and centered around the estimate. The series breaks in 1915–1916, when the parliament did not gather, and in 1918 and 1939, when the number of speakers from left parties was below 30 (1 and 24, respectively).

Figure 3: Left-right partisanship without the extreme left party (SKDL), controls for government status, speaker gender and region



Notes: Results are based on regressions with controls for government status of speaker's party, speaker's gender and the region of the speaker. Confidence intervals are based on subsampling and have 80 percent nominal coverage. The confidence intervals are computed via subsampling and centered around the estimate. The series breaks in 1915–1916, when the parliament did not gather, and in 1918 and 1939, when the number of speakers from left parties was below 30 (1 and 24, respectively).

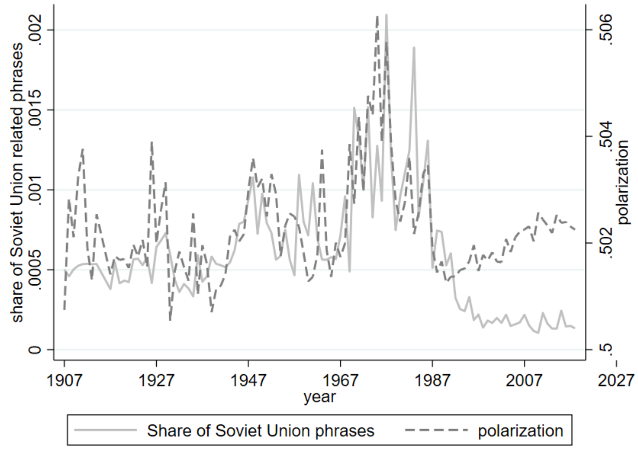
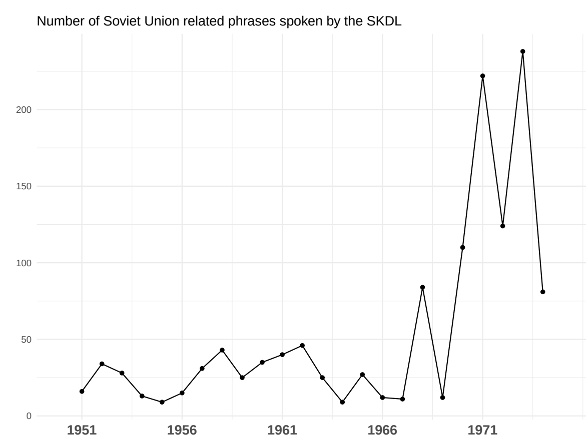
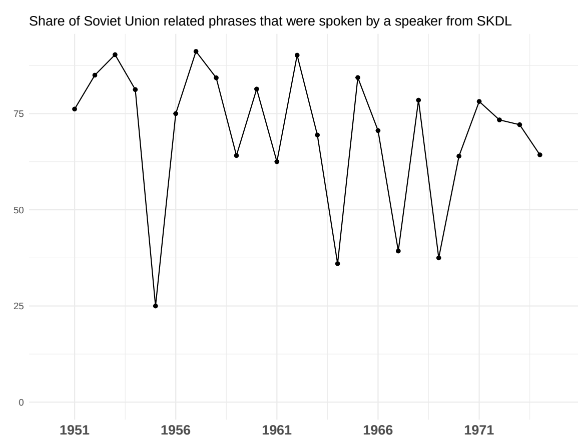


Figure 4: Phrases related to the Soviet Union

Notes: Figure shows in the same figure i.) the share of phrases related to the Soviet Union, and ii.) left-right polarization series from which we have subtracted the 'random' series. Soviet Union related phrases include all phrases containing the phrase 'Soviet Union' in some form, as well as other hand-picked Soviet-related words such as words related to the 'friendship and co-operation pact' (Finno-Soviet Treaty of 1948) between Finland and the Soviet Union. These bigrams are listed in Online Appendix B ('List of Soviet Union related bigrams').



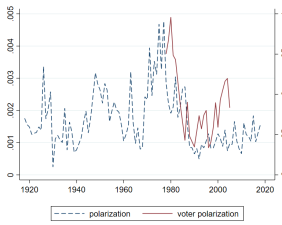
(a) Number of Soviet Union phrases used by the SKDL



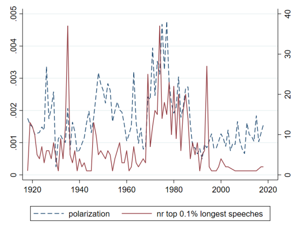
(b) Share of SKLD of Soviet Union phrases spoken

Figure 5: SKDL and Soviet Union phrases

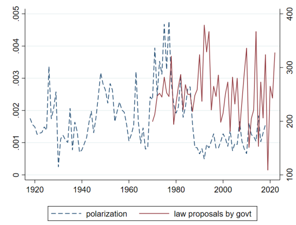
Notes: Panel A of this figure shows the number of Soviet Union phrases used by the SKDL during a period between 1950 and 1975. Panel B shows the yearly share of SKLD of Soviet Union phrases spoken in the Parliament.



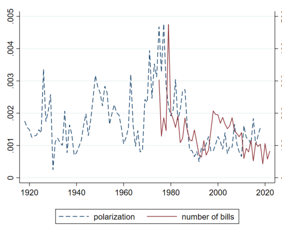
(a) Voter polarization



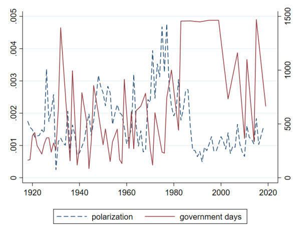
(b) Number of long (top 0.1 percent longest) speeches



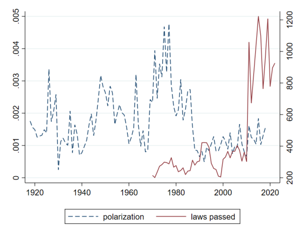
(c) Law proposals by the government



(d) Number of bills by MPs



(e) Length of coalition governments



(f) Laws passed

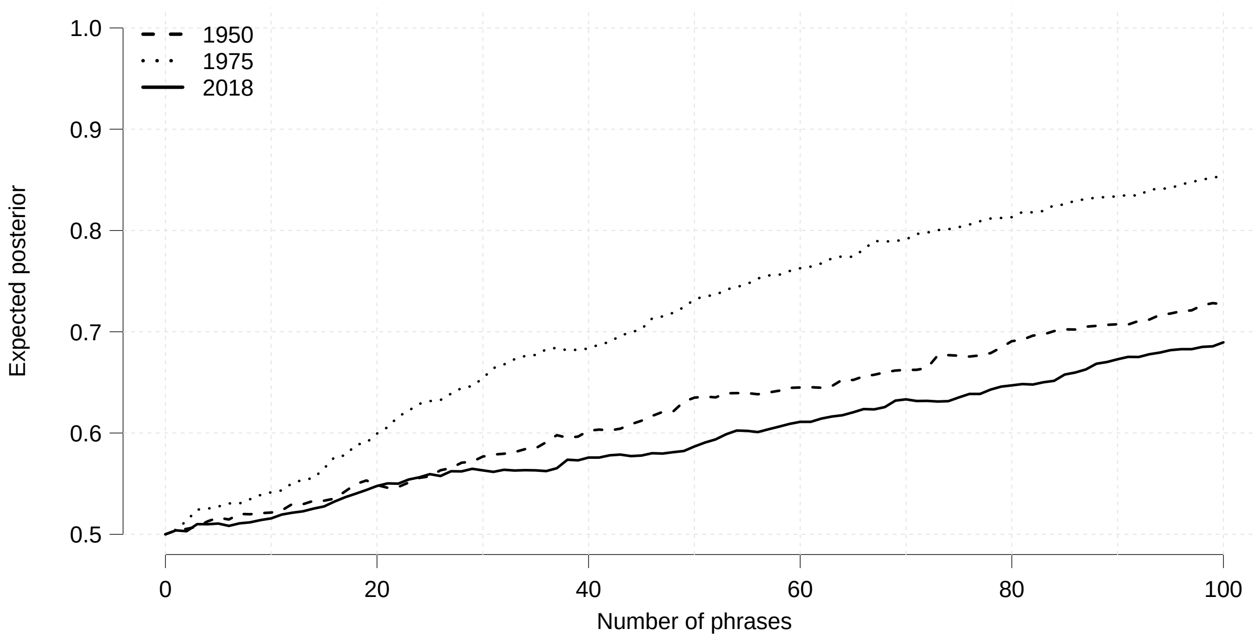
Figure 6: Polarization of parliamentary speech and other societal trends

Notes. Figures compare the development of left-right polarization to that of other societal trends. Voter polarization is defined as the standard deviation of a survey question asking where voters place themselves on the left-right scale.

Online Appendices

Online Appendix A: Polarization results

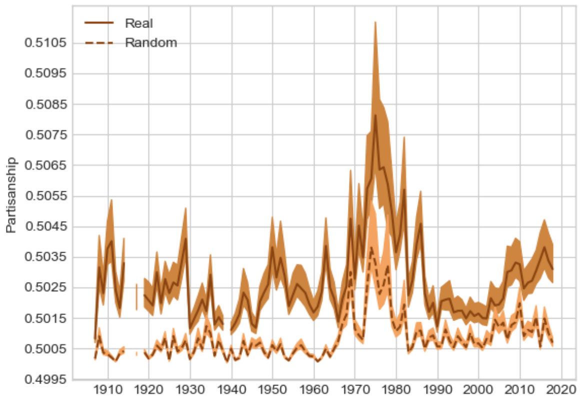
Figure A1: Magnitude of polarization when more than 1 phrase is heard



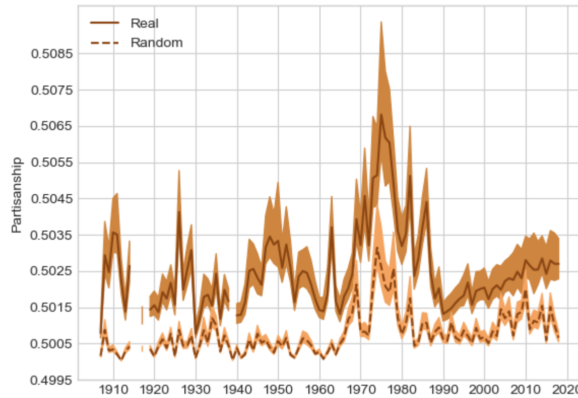
Notes: Figure shows the expected posterior of guessing the group identity right after hearing up to 100 words. The analysis is conducted similarly as in Gentzkow et al. (2019b).

Figure A2: Left-right partisanship: Impact of covariates

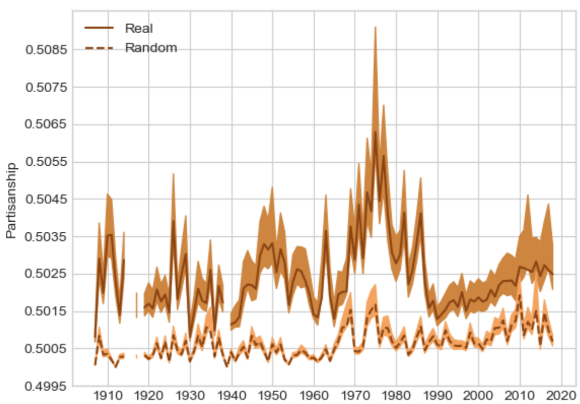
(a) No covariates



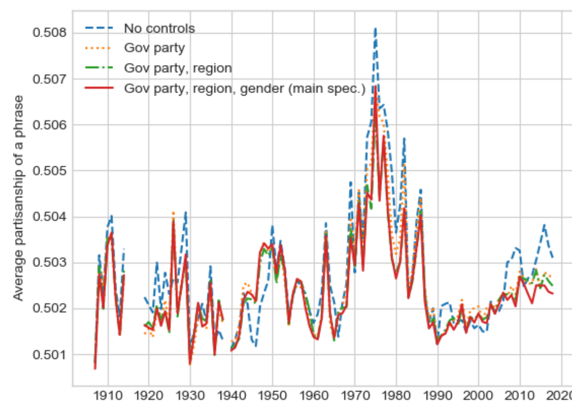
(b) Covariates: indicator for gov party



(c) Covariates: indicator for gov party and region



(d) Comparison of covariate sets

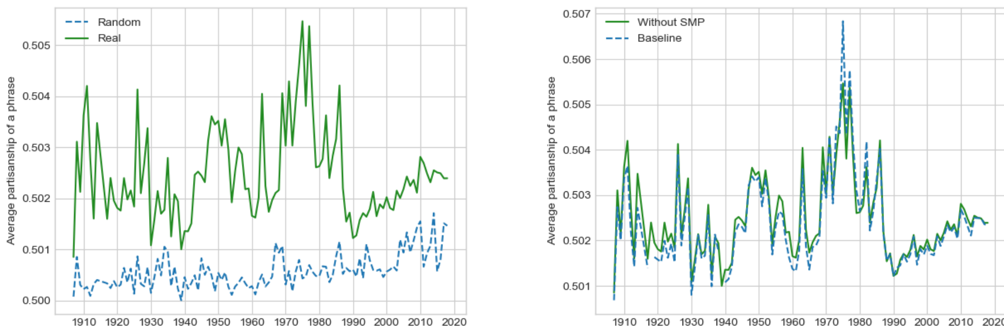


Notes: Figures A2a-A2c show the real and random series for average partisanship for different sets of covariates. Figure A2d shows a comparison of the three series varying in covariates and the main specification that additionally controls for speaker's gender. Confidence intervals in Figures A2a-A2c are computed via subsampling, are centered around the estimate and have 80 percent nominal coverage. The series breaks in 1915–1916, when the parliament did not gather, and in 1918 and 1939, when the number of speakers from left parties was below 30 (1 and 24, respectively).

Figure A3: Left-right partisanship without SMP

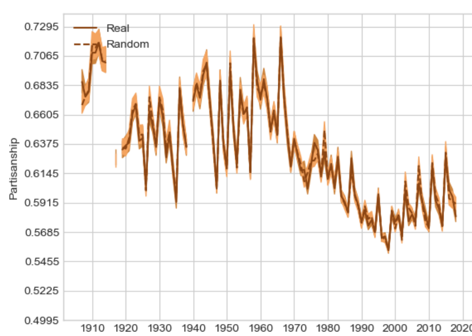
(a) Real series without SMP and random falsification series

(b) Left partisanship with and without SMP



Notes: Figure A3a shows the real and random series from estimation without SMP. Controls for government status, gender and dialect region are added. Figure A3b shows the same series plotted against the baseline series (left-right, controls for government status, gender and dialect region). The series break in 1915–1916, when the parliament did not gather.

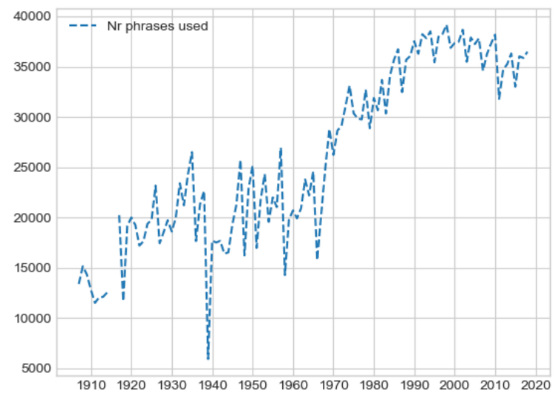
Figure A4: Left-right partisanship calculated from empirical phrase choice probabilities



Notes: Results are based on calculations where partisanship measure is constructed from empirical phrase choice probabilities $q_j = c_j / \sum_k c_k$. Confidence intervals are based on subsampling and have 80 percent nominal coverage. The confidence intervals are computed via subsampling and centered around the estimate. The series breaks in 1915–1916, when the parliament did not gather, and in 1918 and 1939, when the number of speakers from left parties was below 30 (1 and 24, respectively).

Figure A5: Phrase usage and partisanship

(a) Number of phrases in vocabulary with nonzero counts c_{jt}



(b) Number of phrases with nonzero coefficients ϕ_{jt} for party-year indicators

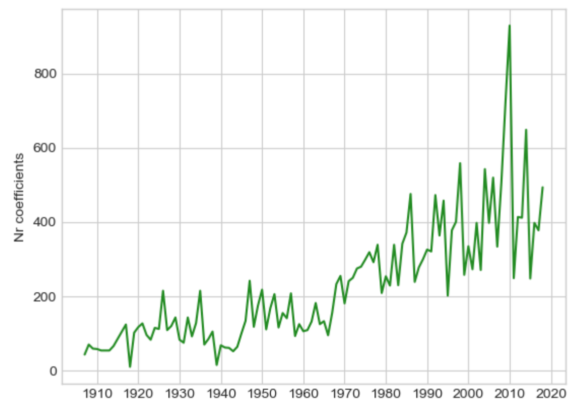
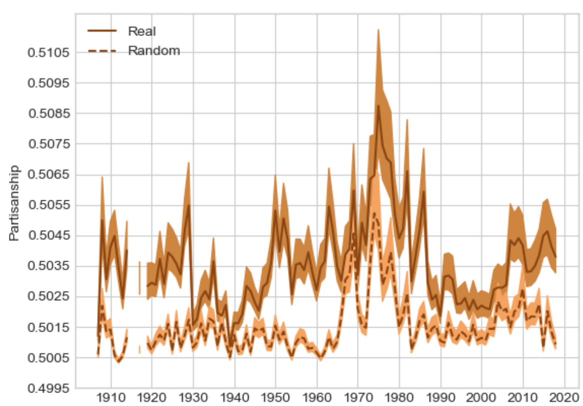
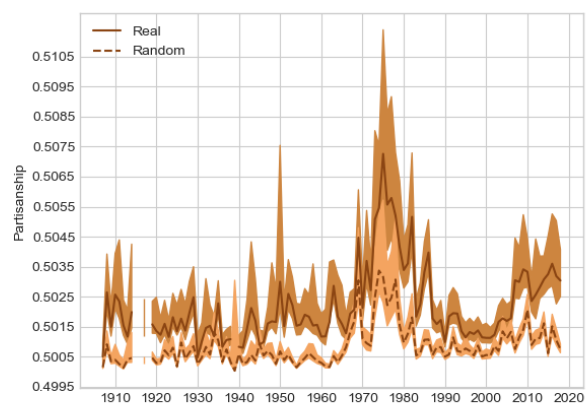


Figure A6: Left-right partisanship: the impact of ψ

(a) $\psi = 10^{-4}$



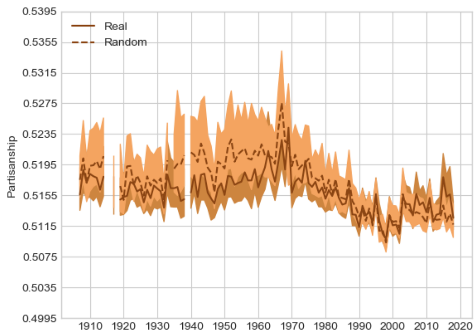
(b) $\psi = 10^{-6}$



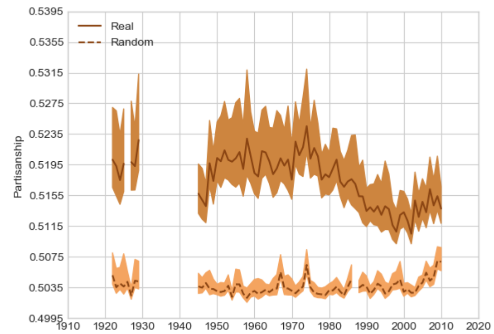
Notes: Results are based on regressions with controls only for year indicators and party-year indicators. Confidence intervals are based on subsampling and have 80 percent nominal coverage. The confidence intervals are computed via subsampling and centered around the estimate. The series breaks in 1915–1916, when the parliament did not gather, and in 1918 and 1939, when the number of speakers from left parties was below 30 (1 and 24, respectively).

Figure A7: Partisanship, pairwise comparisons

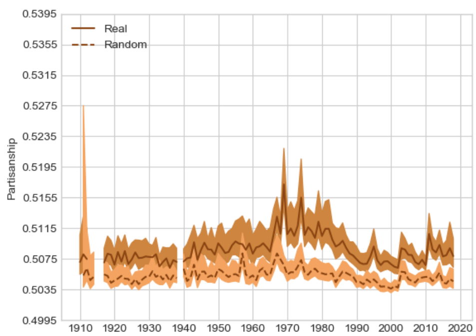
(a) Social Democrats and National Coalition Party



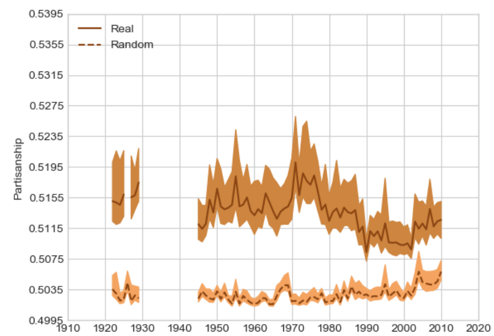
(b) Extreme left and the National Coalition Party



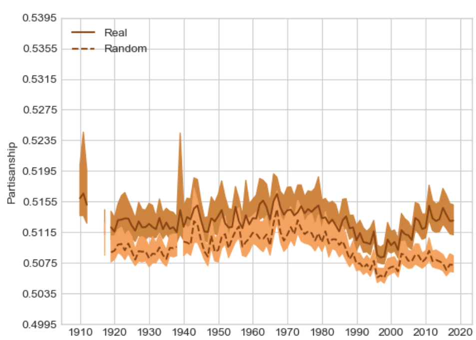
(c) National Coalition Party and Centre Party



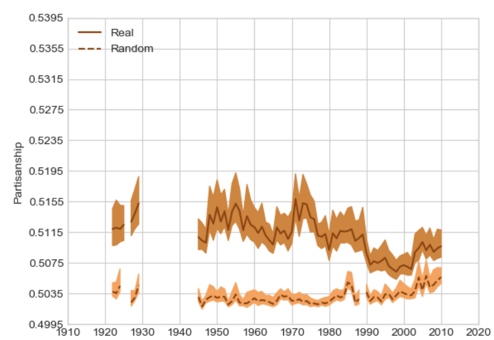
(d) Extreme left and the Centre Party



(e) Social Democrats and Centre Party

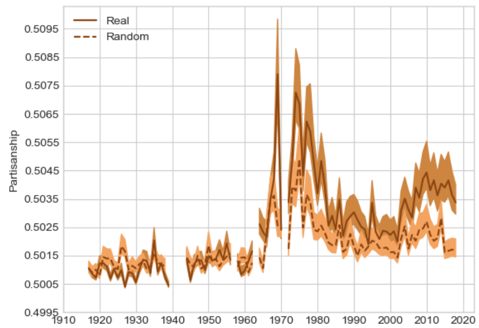


(f) Extreme left and the Social Democrats



Notes: Figures show differences in speech between parties indicated in the subcaption. Results are based on regressions with no other speaker characteristics than speaker's party. Gaps in the time series are for years for which the number of MPs from one of the two parties was below 15. Confidence intervals are based on subsampling and have 80 percent nominal coverage. The confidence intervals are computed via subsampling and centered around the estimate.

Figure A8: Government-opposition partisanship, controls for gender and dialect region



Notes: Figures show partisanship between government and opposition parties. Results are based on regressions with no other speaker characteristics than speaker's party. Confidence intervals are based on subsampling and have 80 percent nominal coverage. The confidence intervals are computed via subsampling and centered around the estimate. The time series break in 1939–1943, when the number of MPs in the opposition was below 30 (strong majority government), and in 1957, 1963 and 1971, when the number of MPs in government was below 30 (governments by non-partisan officials; *virkamieshallitus*).

Online Appendix B: Data and preprocessing

Data

The data covers all records of the plenary sessions of the Parliament of Finland (*Eduskunta*) from 1907–2018. Since the Parliament did not gather in 1915 and 1916, the time series has a break for these years. We perform Optical Character Recognition for data from 1907–2015 page by page using the tesseract OCR engine. Text for 2016–2018 is extracted directly from pdf file metadata.

The processing of text data begins with the retrieval of speech sections from the records. Our automated script finds the beginnings and ends of speech sections based on character string searches. Speech sections typically start with titles *'Keskustelu:'* (discussion), *'Yleiskeskustelu:'* (general discussion) or a phrase containing one of them ('Discussion continues'). Since the image-to-text conversion made with the OCR program is imperfect, the script allows for some common misspellings (e.g. *'Keskustelu;'*, *'Keskustelu :'*). MPs also speak during the plenary sessions to make procedural questions or comments, to ask for corrections to an accidental faulty vote or to make announcements. Such speeches are not recorded under discussion tags but under a different subtitle (e.g. *'Puheenvuoron saatuaan lausui'*). Since these speeches are not attempting to convey speaker ideology, the automated script only targets discussion sections and ignores other speech types.

Speech sections are split to speeches based on speaker tags. The format of these tags changes several times over the years. Some examples of the tags are *'Ed. E. S. Yrjö-Koskinen:'* and *'Ed. Procopé:'* for pre-2000s, *'Jaana Ylä-Mononen /kesk (vastauspuheenvuoro):'* for 2000–2015 and *'16.04 Ulkoasiainministeri Timo Soini (vastauspuheenvuoro):'* for 2015 onwards. For pre-2000s, speaker initials are used if two or more active MPs have the same last name.

Figure A17 shows that the share of speeches successfully retrieved by the automated script is well above 90 percent for most decades in an audit of a randomly selected subset of transcripts. The recall dips for the transcripts from the second, third, and fourth decade of the 1900s. The numbers are based on a manual audit of 10 randomly selected transcripts from a randomly selected year for each decade. Because the audit was conducted in October 2019 and the script has been updated and improved since, the numbers may be interpreted as a reasonable estimate of the lower bound for the coverage of the data.

After splitting speech sections to speeches, speakers' names are linked to data from MP register[18] which contains, for example, speaker's party label, their gender, municipality of birth, their electoral districts and electoral terms. Some of the former minister "professionals" are not affiliated with any party and are also not included in the data set. We acquire their information from the parliament website[19]. The final MP register consists of a total of 2,528 MPs.

Linking is complicated by speaker name changes and, again, by OCR misspellings. Marriage is an important reason for the former, and is mostly addressed by going through the non-matched speaker names manually. To account for OCR misspellings, we adapt the spelling corrector in Norvig (2007) for an extended alphabet. MP register data can be used to construct a pool of correctly spelled MP names active in each parliamentary year. If a match for a retrieved speaker name is found in the pool of active MPs, no spelling correction is applied. If no match is found for the exact spelling, the spelling corrector tries to match candidate corrections for the retrieved speaker name. The first tier of candidate corrections consists of modifications of the retrieved name that are one edit (deletion, insertion, transposition, replacement) away from the exact spelling. The second tier of modifications consists of corrections with a minimum edit distance of two. To avoid false positives, we do not consider candidate corrections further than two edits away from the retrieved spelling.

For example, a common misspelling for Representative Procopé in the OCR'd text is 'Procop6'. Matching based on the exact spelling of the name in the OCR'd text will be unsuccessful. The true name Procopé is, however, one replacement away from the misspelled name and thus included in the first tier of candidate corrections. Due to spelling corrector, 'Procop6' will successfully link to the MP register data. A speaker match is found for 97 percent of speeches in the data.

In general, all speeches transcribed under discussion sections are included in the analysis. Speech by the Deputy Speakers or the Speaker of the Parliament is excluded because it primarily

---

[18] Kansanedustajamatrikkeli, retrieved from the Parliament library as a spreadsheet.

[19] https://valtioneuvosto.fi/tietoa/historiaa/hallitukset-ja-ministerit/raportti/-/r/v5s/henkilo.nimi

guides the proceedings of the Parliament. Speech by the MP from the autonomous Åland province is also omitted as it primarily takes place in Swedish. If no match for speaker name is found or if the speaker is not referred to by name (e.g. 'Puhuja:'), speech is not included in the analysis. Since Swedish is the second official language of Finland, some parliamentary speeches are given in Swedish. We perform automatic language detection on speech level and leave out speeches that are classified as Swedish. However, this still leaves some Swedish speech in the data because many speeches contain sections both in Swedish and Finnish. Around 20 percent of speeches are primarily in Swedish at the beginning of the time series, and the share goes down to less than .5 percent in 2018. Speakers with missing party labels are excluded from the analysis. Those speakers who switch parties during the parliamentary year are assigned to their party in the beginning of the parliamentary year.

Figure A18a shows that the number of speeches given in the Parliament increases over time. The cyclical pattern in the number of speeches reflects the varying length of the parliamentary year. In particular, the plenary season lasts longer in election years. There are a couple of important changes in the plenary format during the time period which also affects the amount of speech. In 1967, in an effort to make make plenary discussions more engaging, a session type called government question hour was introduced. In these sessions, opposition MPs could pose questions to the government about current matters. In 2012, debate speeches were introduced. Debates are fast-paced discussions between two parliament members coming from different parties. Debates have a prespecified topic, and each speech can last a maximum of one minute. The shortening of an average speech shows up in the two figures, with the total number of speeches going up but with no clear trend in the average number of phrases spoken.

Figure A18b shows that the average number of phrases spoken by an MP also increases over time. Thus, the increase in the total number of speeches is not driven by the same amount of speech split to multiple shorter speeches.

Preprocessing

In order to represent speeches as a large data matrix, we apply preprocessing to raw speeches, combine preprocessed words to phrases consisting of two consecutive words (*bigrams*), construct a dictionary from unique bigrams and count the occurrences of bigrams in the dictionary on speaker-year level.

In the preprocessing step, the text is normalized, i.e. certain manipulations are performed on the text to collapse certain word forms to a single normal form. When an MP is speaking, additions, comments and interruptions by other MPs are transcribed in the plenary minutes in parentheses. To simplify preprocessing, we discard these parenthetical additions. We replace certain non-alphanumeric symbols (€, \$, %, §) with their literal form and replace all other non-alphanumeric characters, including any punctuation, with a space. Then, the text is converted to lowercase and split into words based on spacing. We remove certain very common words (*stop words*) that may not carry much partisan information. Appendix 6 lists all removed stop words. Finally, we use the Porter2 stemmer to chop off word suffices and to find the stem of a word, i.e. the part of a word that contains its main meaning.

The word stems are concatenated to bigrams, combinations of two consecutive stems. Using combinations of consecutive stems introduces context – frequencies for *'työtätekev.luok'* (the stemmed bigram for 'working class') may convey more partisan information than frequencies for 'työtätekev' ('working') and 'luok' ('class') separately. Using two words for the context window size is an arbitrary choice. The common usage of compound words in Finnish to convey multiple meanings would make unigrams, single words, another potential candidate for the vocabulary unit. However, we follow Gentzkow et al. (2019a) in sticking with the bigrams. This comes at the cost of a larger dimension of the data matrix.

The vocabulary consist of all unique bigrams uttered as part of parliamentary speeches, given that they exceed certain minimum frequency thresholds. The phrase must be used at least 100 times in total over all speeches, it must be used at least 10 times during at least one parliamentary year, and it must be used by 10 unique speaker-years. The restrictions follow the ones made by Gentzkow et al. (2019a), and come with the benefit of reducing the dimension of the count matrix.[20] After restrictions, the vocabulary consists of approximately 53,000 phrases.

Using a fixed vocabulary over the whole time period means that any new phrases emerging during the more recent years have less time to reach the total count of 100. Also, phrases popular in the early years and unpopular these days will still be a part of today's vocabulary while this is not the case in the other direction.

We also omit a set of frequently appearing but ideologically uninformative procedural phrases and attempt to drop phrases containing speaker or party names. A more detailed description of these restrictions is provided in the Appendix.

The final data used for analysis consists of counts for the number of times each MP used each dictionary phrase during a parliamentary year. This count matrix **C** has a row for each MPs who spoke a positive number of phrases in a year and has in total 19,094 rows (speaker-years) and 53,705 columns (phrases). For comparison, the number of phrases is around one-tenth, the number of years around 70 percent and the number of unique speakers around one-fourth of the corresponding numbers in the US data used by Gentzkow et al. (2019a).

Details on data construction

Plenary records are retrieved as pdf files from Open Data Service by the Parliament of Finland. Pdf's until 2000 combine multiple plenary records in one file, while files from 2000s contain one plenary record each. Optical Character Recognition is performed page by page using tesseract OCR engine for data from 1907–2015. Text for 2016–2018 is extracted from pdf file metadata using PDF2text program.

In addition to the following list of stopwords, I remove phrases including 1) representative names, 2) party names, 3) addresses of the chairperson ("Arvoisa puhemies"), 4) months and 5) a list of other procedural phrases.

[20]We refer the reader interested in the potential effect of these restrictions on results to Gentzkow et al. (2019a), where their Online Appendix Figure 1 shows the effect of tightening this restriction on the average partisanship series. It is, however, worth noting that a crude stemming method like Porter is likely to behave worse on morphologically rich language like Finnish relative to English. As a result, counts for the same lemma of a word may be split between two vocabulary units (e.g. 'kotihoido tuen', 'kotihoido tuke' as in 'homecare allowance'), and thus the impact of the minimum thresholds could differ from results derived with English.

List of stopwords

ei	johon	keiltä	meistä	niille	näistä	siihen	teitä
eivät	joiden	keinä	meitä	niillä	näitä	siinä	tuo
emme	joihin	keissä	mihin	niiltä	nämä	siitä	tuohon
en	joiksi	keistä	miksi	niin	ole	siksi	tuoksi
et	joilla	keitä	mikä	niinä	olemme	sille	tuolla
ette	joille	keneen	mille	niissä	olen	sillä	tuolle
että	joilta	keneksi	millä	niistä	olet	siltä	tuolta
he	joina	kenelle	miltä	niitä	olette	sinua	tuon
heidän	joissa	kenellä	minkä	noiden	oli	sinulla	tuona
heidät	joista	keneltä	minua	noihin	olimme	sinulle	tuossa
heihin	joita	kenen	minulla	noiksi	olin	sinulta	tuosta
heille	joka	kenenä	minulle	noilla	olisi	sinun	tuotä
heillä	joksi	kenessä	minulta	noille	olisimme	sinussa	tähän
heiltä	jolla	kenestä	minun	noilta	olisin	sinusta	täksi
heissä	jolle	kenet	minussa	noin	olisit	sinut	tälle
heistä	jolta	ketkä	minusta	noina	olisitte	sinuun	tältä
heitä	jona	ketä	minut	noissa	olisivat	sinä	tämä
hän	jonka	koska	minuun	noista	olit	sitä	tämän
häneen	jos	kuin	minä	noita	olitte	tai	tänä
hänelle	jossa	kuka	missä	nuo	olivat	tallä	tässä
hänellä	josta	kun	mistä	nyt	olla	te	tästä
häneltä	jota	me	mitkä	näiden	olleet	teidän	tätä
hänen	jotka	meidän	mitä	näihin	ollut	teidät	vaan
hänessä	kanssa	meidät	mukaan	näiksi	on	teihin	vai
hänestä	keiden	meihin	mutta	näille	ovat	teille	vaikka
hänet	keihin	meille	ne	näillä	poikki	teillä	yli
häntä	keiksi	meillä	niiden	näiltä	se	teiltä	
itse	keille	meiltä	niihin	näinä	sekä	teissä	
ja	keillä	meissä	niiksi	näissä	sen	teistä	

List of Soviet Union related bigrams

entis.neuvostoliito

esim.neuvostoliito

esimerk.neuvostoliito

kommunistis.puolue

muide.sosialistist

myös.neuvostoliito

neuvostoliito.alue

neuvostoliito.kaup

neuvostoliito.kaupa

neuvostoliito.muide

neuvostoliito.puole

neuvostoliito.suht

neuvostoliito.suome

neuvostoliito.taho

neuvostoliito.taloudellis

neuvostoliito.tapahtuv

neuvostoliito.väl

neuvostoliito.välin

neuvostoliito.välis

neuvostoliito.vas

sosialistist.maide

suht.neuvostoliito

suome.kommunistis

suome.neuvostoliito

tuont.neuvostoliito

vient.neuvostoliito

yhteistyö.avunantosopimuks

yhteistyö.neuvostoliito

ystävyys.avunantosopimuks

ystävyys.yhteistyö

Online Appendix C: Most polarized phrases

Table 3.5. 5 most left and right phrases based on $\widehat{\zeta}$, real series, main specification: 1900s

1907 Left	#L	#R	Right	#L	#R
X8 tunt	62	19	suome pan	62	102
syvä rivi	50	11	suome pank	72	96
X8tuntis työpäiv	49	15	porvarillis edustaj	1	4
kans syvä	39	10	työtätekev väestö	1	4
kans pohjakerrost	39	12	kans kaht	2	4
1908 Left	#L	#R	Right	#L	#R
kans näyttämö	233	78	alem kansakoulu	76	208
porvarillis edustaj	104	2	suomalais puolue	99	174
täl istuntokaud	120	47	ylem kansakoulu	39	86
porvarillis jäsen	80	16	suome eduskun	66	105
yhä ede	167	108	X12 viiko	12	39
1909 Left	#L	#R	Right	#L	#R
naist yötyö	161	29	suome eduskun	100	172
porvarillis edustaj	82	3	suomalais puolue	43	96
pitk työpäiv	91	33	suome pank	121	149
X8tuntis työpäiv	72	18	suome pan	142	165
ensimäis lukemis	76	23	täl kert	440	444

Notes: Table shows the five most left and right phrases and their predicted counts for 100,000 phrases. These phrases have the highest impact on the partisanship measure π_t. The coefficients come from a regression with controls for government status of speaker's party, gender and speaker's dialect region. Table shows partisan phrases for every year of the decade available in data.

Table 3.6. 5 most left and right phrases based on $\widehat{\zeta}$, real series, main specification: 1910s

1910 Left	#L	#R	Right	#L	#R	1917 Left	#L	#R	Right	#L	#R
tilattom väestö	661	180	alem kansakoulu	17	118	viime kesä	137	52	venäj valtio	21	64
hyvä her	103	16	suome rautat	27	125	korkeim oikeud	112	64	ehdot pykäl	17	51
työtätekev väestö	78	0	ylem kansakoulu	44	141	porvaristo taho	52	10	hallituks edustaj	38	63
porvarillin enemmistö	71	0	suome koskev	14	61	porvarillis edustaj	41	3	maataviljelev väestö	11	35
porvar taho	70	4	X4000 mark	23	59	oikeisto taho	76	41	mietintö liitety	66	88
1911 Left	#L	#R	Right	#L	#R	**1918 Left**	#L	#R	Right	#L	#R
porvarillin enemmistö	187	0	suome pan	192	529						
porvarillis edustaj	129	4	tilattom maalaisväestö	36	202						
her porvar	96	4	suome pank	144	286						
yhdistyks jäsen	103	15	suomalais puolue	25	37						
porvar taho	66	0	suurt tulo	1	11						
1912 Left	#L	#R	Right	#L	#R	**1919 Left**	#L	#R	Right	#L	#R
hyvä her	149	17	työväestö keskuud	1	8	valkois suome	140	29	vahvistetu tulo	3	32
tilattom väestö	233	132	eduskun määrä	6	10	porvaristo taho	60	6	suome valtakun	20	44
porvarillin enemmistö	66	5	täl kert	317	321	suome työväenluok	52	0	suome pank	31	54
liitety ensimäis	74	19	mietintö liitety	296	299	kansalaissod jälk	43	2	välikysymyks tekij	52	73
her porvar	50	0	kem tornio	2	5	suome porvaristo	45	4	ruots kans	5	18
1913 Left	#L	#R	Right	#L	#R	**1920 Left**	#L	#R	Right	#L	#R
tilattom väestö	183	90	suome pan	177	221	tulo menoarvio	589	432	milj kilo	33	98
porvar taho	75	0	mark menoer	20	25	kuluv vuode	141	73	suome pan	67	131
porvarillin enemmistö	62	0	täl kert	321	324	porvarillin enemmistö	39	0	kannat tehty	12	45
mietintö liitety	219	168	miljoon mark	292	294	porvarillis edustaj	36	0	ehdot eduskun	62	92
porvaristo taho	47	4	viime vuon	251	253	valkois suome	33	0	lyhytaikais luoto	15	45
1914 Left	#L	#R	Right	#L	#R	**1921 Left**	#L	#R	Right	#L	#R
toise kappal	140	15	maataviljelev väestö	26	66	tilattom väestö	184	48	kansainliito neuvosto	9	47
her porvar	74	0	käsityks eduskun	13	51	porvaristo taho	55	2	suome pank	38	66
omistav luok	87	13	vuode 1914	12	47	vuode 1919	103	57	suome pan	69	89
suome porvaristo	70	0	täl kert	382	387	porvarillis edustaj	41	2	tasaval president	78	91
tilattom väestö	326	265	kem tornio	2	7	X13 kuukaud	42	8	suomalais kirjallisuud	6	17

Notes: Table shows the five most left and right phrases and their predicted counts for 100,000 phrases. These phrases have the highest impact on the partisanship measure π_t. The coefficients come from a regression with controls for government status of speaker's party, gender and speaker's dialect region. Table shows partisan phrases for every year of the decade available in data. Note that parliament did not gather in 1915 and 1916.

Table 3.7. 5 most left and right phrases based on $\widehat{\zeta}$, real series, main specification: 1920s

1920 Left	#L	#R	Right	#L	#R	**1925** Left	#L	#R	Right	#L	#R
tulo menoarvio	589	432	milj kilo	33	98	täl kert	500	426	suome pan	23	62
kuluv vuode	141	73	suome pan	67	131	porvarillin enemmistö	34	2	harv asutu	28	63
porvarillin enemmistö	39	0	kannat tehty	12	45	poliittis vang	29	0	valtio virkatalo	16	36
porvarillis edustaj	36	0	ehdot eduskun	62	92	porvaristo taho	29	0	kommunistis puolue	13	30
valkois suome	33	0	lyhytaikais luoto	15	45	yhdy kannattam	110	83	venäj valtio	7	18
1921 Left	#L	#R	Right	#L	#R	**1926** Left	#L	#R	Right	#L	#R
tilattom väestö	184	48	kansainliito neuvosto	9	47	yhdy kannattam	158	80	en vuode	95	164
porvaristo taho	55	2	suome pank	38	66	porvaristo taho	81	3	her schauma	2	66
vuode 1919	103	57	suome pan	69	89	täl kert	425	352	supistetu kansakoulu	15	68
porvarillis edustaj	41	2	tasaval president	78	91	työväe urheiluliito	64	7	hallituks ohjelm	70	121
X13 kuukaud	42	8	suomalais kirjallisuud	6	17	etsiv keskuspoliis	96	41	harv asutu	20	63
1922 Left	#L	#R	Right	#L	#R	**1927** Left	#L	#R	Right	#L	#R
tulo menoarvio	365	293	her schauma	12	64	työtätekev väestö	94	0	valtio haltu	25	60
työtätekev luoka	66	0	lepääm jätety	23	54	porvarillin enemmistö	56	0	harv asutu	22	49
poliittist vank	57	7	tasaval president	60	86	porvaristo taho	62	15	milj kilo	24	47
porvaristo taho	45	0	hallinnollis asetuks	12	37	työväe urheiluliito	48	15	tie vesirakennust	14	34
poliittis vang	44	4	suome pan	118	134	maan työtätekev	33	0	asia järjestäm	10	19
1923 Left	#L	#R	Right	#L	#R	**1928** Left	#L	#R	Right	#L	#R
kansalaissod jälk	52	0	suome pan	83	174	asevelvollist palvelusaj	114	22	kuusamo kuolajärv	21	55
virk palvelusmiest	66	24	yliopisto opettaj	11	61	työtätekev väestö	78	3	tasaval president	69	102
porvaristo taho	40	0	järjestysmuodo perust	57	104	yhdy kannattam	114	52	valtio vira	21	43
porvarillin enemmistö	40	3	uusi vero	11	43	työtätekev luok	46	0	kajaan kihlakun	8	28
puhtaast poliittis	40	3	suome pank	77	96	palvelusaj lyhentäm	56	12	vira toimenhaltij	25	34
1924 Left	#L	#R	Right	#L	#R	**1929** Left	#L	#R	Right	#L	#R
etsiv keskuspoliis	124	20	valtio haltu	34	70	täl kert	629	452	vira toimenhaltij	97	190
työväe urheiluliito	115	37	milj kilo	9	44	työtätekev luoka	57	0	valtio vira	65	111
porvaristo taho	69	3	valtio koulu	19	38	poliittist vank	64	13	vuon 1928	18	53
työväe sivistysliito	69	17	kajaan kihlakun	13	30	y m	237	187	naimis olev	6	36
työväe edustaj	37	0	erinomais suure	8	23	työtätekev väestö	68	22	kommunistis puolue	23	46

Notes: Table shows the five most left and right phrases and their predicted counts for 100,000 phrases. These phrases have the highest impact on the partisanship measure π_t. The coefficients come from a regression with controls for government status of speaker's party, gender and speaker's dialect region. Table shows partisan phrases for every year of the decade available in data.

Table 3.8. 5 most left and right phrases based on $\widehat{\zeta}$, real series, main specification: 1930s

1930 Left	#L	#R	Right	#L	#R	**1935** Left	#L	#R	Right	#L	#R
porvarillis edustaj	25	3	kansanliik vas	1	55	milj mark	327	218	helsing yliopisto	47	246
porvaristo taho	20	0	työtätekev väestö	10	27	miljoon mark	137	98	järjestysmuodo perust	13	81
porvarillis piire	22	3	vuote verrat	4	20	vuode 1936	94	59	ylimääräis valtiopäiv	94	151
töide järjestämis	14	0	lain hyv	4	18	y m	106	71	suomalais yliopisto	17	71
porvarillis sanomalehd	13	0	työtätekev kans	0	13	ammatillis järjestö	31	6	valtio yliopisto	7	60
1931 Left	#L	#R	Right	#L	#R	**1936** Left	#L	#R	Right	#L	#R
milj kilo	93	11	suome pan	97	264	työväe urheiluliito	261	182	korkeim oikeud	41	71
kunt yhteistoimin	30	4	supistetu kansakoulu	74	188	milj kilo	51	8	suome voimistelu	48	77
miljoon mark	278	254	suome pank	77	182	poliittist vank	47	9	itsenäisyyt vapaut	7	34
porvarillis piire	23	0	maalaiskunt liito	29	80	poliittis vang	33	0	voimistelu urheiluliito	51	78
valtio viljavarasto	49	28	last huoltaj	9	50	työläist palko	36	3	itsenäisyyd vapaud	17	39
1932 Left	#L	#R	Right	#L	#R	**1937** Left	#L	#R	Right	#L	#R
milj mark	552	467	lapua liik	51	252	miljoon mark	339	175	hallitusmuodo 14	18	176
tulo menoarvio	205	136	suome pan	56	206	milj mark	470	363	sotaväe päällikö	16	126
kuluv vuode	150	87	kansanliik vas	1	94	köyhäinhoido vara	57	6	ruotsinkiel opetu	13	50
en vuode	193	135	eteläis vaalipiir	4	61	X2 mark	49	19	valtio yliopisto	12	28
milj kilo	67	18	suome pank	67	122	ammatillis järjestö	30	0	suomalais yliopisto	11	26
1933 Left	#L	#R	Right	#L	#R	**1938** Left	#L	#R	Right	#L	#R
milj mark	861	587	hallitusmuodo 14	26	92	miljoon mark	269	111	tasaval suojelulak	26	85
työväe urheiluliito	272	98	hyvä her	9	33	milj mark	651	558	milj kilo	35	76
poliittis vang	86	5	ammatillis järjestö	15	24	teknillis opisto	59	5	lain hyv	10	26
poliittist vank	68	0	j annal	13	21	valtiollis poliis	60	17	kansanliik vas	10	23
vuon 1932	89	28	vasemmisto keskust	0	9	ammatillis järjestö	34	2	X35 pen	6	18
1934 Left	#L	#R	Right	#L	#R	**1939** Left	#L	#R	Right	#L	#R
milj mark	736	427	kansanliik vas	7	75						
valtio mets	126	28	työväe urheiluliito	136	203						
etsiv keskuspoliis	110	34	lapua liik	64	131						
kysymyks olev	235	188	suome pan	45	87						
miljoon mark	188	158	j annal	13	25						

Notes: Table shows the five most left and right phrases and their predicted counts for 100,000 phrases. These phrases have the highest impact on the partisanship measure π_t. The coefficients come from a regression with controls for government status of speaker's party, gender and speaker's dialect region. Table shows partisan phrases for every year of the decade available in data.

Table 3.9. 5 most left and right phrases based on $\widehat{\zeta}$, real series, main specification: 1940s

1940 Left	#L	#R	Right	#L	#R	**1945** Left	#L	#R	Right	#L	#R
milj mark	443	384	korvauks oikeutetu	62	122	valtiollis poliis	303	68	timot siemen	28	86
uus vero	47	14	uude tila	8	50	valtiollin poli	102	19	yleis kulutuks	32	61
julkis sana	47	22	uus tilo	36	77	esit lähet	72	8	milj kilo	15	33
työläist palko	27	2	perjant 3	23	60	X35 dollar	58	10	yhtiö mets	3	20
ammatillis järjestö	20	0	uus olo	21	55	muuttamis sisältäv	58	11	lain toimeenpano	21	36
1941 Left	#L	#R	Right	#L	#R	**1946** Left	#L	#R	Right	#L	#R
hinto nousu	151	56	korvauks oikeutetu	15	135	valtiollis poliis	287	77	korsholm lään	7	77
vuode 1939	104	51	kansanliik vas	1	25	valtiollin poli	107	33	yleis kulutuks	19	72
ansiottom arvonnousu	74	36	X29 dollar	14	38	vuode 1947	150	96	vuon 1945	33	78
valtiollis poliis	31	5	kaik työläis	1	5	neuvostoliito vas	48	7	lään perustamis	9	47
maalaiskunt liito	41	23	käsittely ehdot	5	8	työväe edustaj	33	4	milj kilo	20	55
1942 Left	#L	#R	Right	#L	#R	**1947** Left	#L	#R	Right	#L	#R
toime haltij	490	149	palautetu alue	169	235	toime haltij	335	175	miljard mark	190	301
vira toime	488	189	korvauks oikeutetu	37	73	vira toime	329	184	valtio mets	34	79
omaisuuks omistaj	39	7	vuon 1941	40	71	valtiollis poliis	180	45	korsholm lään	13	46
palveluks olev	86	56	suome pan	48	70	hin palkkaneuvosto	123	33	milj kilo	16	47
palk eläv	23	0	miljoon mark	167	179	työtätekev kans	58	0	milj kuutiometr	8	36
1943 Left	#L	#R	Right	#L	#R	**1948** Left	#L	#R	Right	#L	#R
toime haltij	458	152	korvauks oikeutetu	20	149	valtiollis poliis	697	96	lään perustamis	14	95
vira toime	448	147	tuotantokustannuks vas-taav	27	52	valtiollin poli	214	32	korsholm lään	32	94
valtio vira	328	70	valtakun etu	13	38	miljard mark	393	266	henkis työn	16	67
palkkausluok olev	46	12	yhtiö mets	4	22	liikkuv poliis	90	15	pyöreä puutavar	9	45
työväestö keskuud	27	0	tuot hin	33	38	suome neuvostoliito	62	22	maataloustuot hin	15	35
1944 Left	#L	#R	Right	#L	#R	**1949** Left	#L	#R	Right	#L	#R
vira toime	318	103	yhtiö mets	4	129	milj mark	822	556	posio kun	0	45
toime haltij	303	102	om mets	7	44	pienviljelij liito	106	18	henkis työn	40	77
valtio vira	214	56	valtio mets	42	75	toime haltij	172	96	maataloud asem	20	40
valtiollis poliis	75	5	puolustusvoim käytö	11	32	työläist palko	77	12	täl haava	8	23
sotaväe rikosl	109	49	korvauks oikeutetu	19	37	pekkal hallituks	79	19	kelo 17	2	13

Notes: Table shows the five most left and right phrases and their predicted counts for 100,000 phrases. These phrases have the highest impact on the partisanship measure π_t. The coefficients come from a regression with controls for government status of speaker's party, gender and speaker's dialect region. Table shows partisan phrases for every year of the decade available in data.

Table 3.10. 5 most left and right phrases based on $\widehat{\zeta}$, real series, main specification: 1950s

1950 Left	#L	#R	Right	#L	#R	1955 Left	#L	#R	Right	#L	#R
fagerholm hallituks	104	30	henkis työn	25	205	miljard mark	487	425	maataloud asem	20	103
hinto nousu	117	43	työn tekij	25	100	työläist palko	48	0	tuomio hallituks	43	114
fagerholm hallitus	75	18	useam yhd	5	33	mark vuode	98	52	tuotantokustannuks vas-taav	14	41
vuode 1949	102	47	ensimmäis laps	13	39	ammatillis järjestö	44	2	maataloud harjoittaj	49	73
timot siemen	90	37	ruumiillis työn	10	30	pyöreä puutavar	48	8	maataloustuot hin	37	60
1951 Left	#L	#R	Right	#L	#R	**1956 Left**	#L	#R	Right	#L	#R
miljard mark	835	422	henkis työn	48	162	rovaniem kauppal	194	45	vira toime	59	159
pyöreä puun	293	101	valtio vira	18	83	milj mk	554	453	toime haltij	57	155
vuon 1949	91	26	toime haltij	52	95	milj mark	410	309	sak johto	7	32
vuode 1948	85	36	vira toime	52	92	miljard mark	653	561	valtio vira	23	44
pyöreä puutavar	372	334	poliittis realiteet	6	34	yhdistyn kansakunt	71	13	utsjoe kun	13	31
1952 Left	#L	#R	Right	#L	#R	**1957 Left**	#L	#R	Right	#L	#R
milj mark	860	656	vira toime	150	384	milj mark	358	187	pyöreä puutavar	20	69
pyöreä puutavar	131	47	toime haltij	135	329	hinto korotuks	57	10	toime haltij	47	95
kuluv vuode	213	150	valtio vira	31	217	milj mk	604	564	valtio oppikoulu	14	61
kans etu	53	4	korkeim oikeud	57	122	työläist pienviljelij	35	0	vira toime	51	96
hinto nousu	73	25	virko toim	43	103	kans elintaso	45	16	voi oike	16	60
1953 Left	#L	#R	Right	#L	#R	**1958 Left**	#L	#R	Right	#L	#R
työtätekev kans	49	0	valtio meno	169	272	työvoimapiir alue	53	0	milj rupl	26	209
kuluv vuode	142	101	vira toime	36	135	kyme lään	83	44	pyöreä puutavar	4	106
milj mark	523	488	toime haltij	33	119	olev luvu	85	49	vanh lais	11	63
milj mk	459	427	yliopisto apteek	23	77	milj mark	475	439	van lain	25	77
kans etu	34	6	virko toim	17	69	hinto korotuks	88	57	mrd mark	315	342
1954 Left	#L	#R	Right	#L	#R	**1959 Left**	#L	#R	Right	#L	#R
luvu 3	130	43	uude käsiteltäv	8	14	milj mk	864	528	suome pan	86	128
olev luvu	84	22	valtio vira	32	37	milj mark	368	293	pyöreä puutavar	6	47
ehdot luvu	85	46	milj mark	490	495	työläist pienviljelij	30	0	puhe olev	67	98
yhtyn paperiteht	49	13	miljard mark	451	455	pohjoism neuvosto	49	20	X12 mrd	16	34
lisäm tarp	58	23	milj mk	374	378	työtätekev kans	28	0	X6 mrd	38	54

Notes: Table shows the five most left and right phrases and their predicted counts for 100,000 phrases. These phrases have the highest impact on the partisanship measure π_t. The coefficients come from a regression with controls for government status of speaker's party, gender and speaker's dialect region. Table shows partisan phrases for every year of the decade available in data.

Table 3.11. 5 most left and right phrases based on $\widehat{\zeta}$, real series, main specification: 1960s

1960 Left	#L	#R	Right	#L	#R	**1965** Left	#L	#R	Right	#L	#R
rautatiehallituks pääjohtaj	71	21	yleis työmarkkino	4	45	porvarillin enemmistö	85	2	milj mk	493	594
r virtas	55	11	pyöreä puutavar	6	31	tulo menoarvio	219	154	kunnallis keskikoulu	6	46
pakollis vakuutuks	57	24	milj kilo	8	25	eduskun porvarillin	47	0	yliopisto korkeakoulu	46	83
sosialistist maide	26	0	porvarillin enemmistö	3	14	porvarillis enemmistö	29	0	oulu yliopisto	87	120
milj mk	611	587	valtiopäiväjärjestyks 67	15	22	olev luvu	45	20	milj nmk	23	54
1961 Left	#L	#R	Right	#L	#R	**1966** Left	#L	#R	Right	#L	#R
sosialistist maide	23	0	valtio viljavarasto	52	92	oma asuno	91	46	jyväskyl yliopisto	18	72
ammatillis järjestö	22	0	suome lääkäriliito	18	58	liikent harjoittaj	45	0	margariin hin	37	90
sosdem puolue	23	4	milj kg	22	34	armeij meno	45	2	margariin hint	10	53
kansandemokraat taho	23	4	sosiaalis sivistyksellis	0	7	X66 prosent	56	18	milj kilo	7	18
asumiskustannust alentamis	18	0	os nimenom	1	5	kans etu	40	4	liikkuv poliis	13	23
1962 Left	#L	#R	Right	#L	#R	**1967** Left	#L	#R	Right	#L	#R
milj mark	314	239	X2 dollar	51	109	valtio varo	106	76	valtio hankintakeskuks	5	55
porvarillin enemmistö	77	6	valtio hankintakeskuks	3	53	nuore työntekij	28	5	taite keskustoimikun	0	47
X1 prosent	188	150	ruotsinkielis vähemmistö	15	59	alais olev	27	5	oy yleisradio	12	57
helsing yliopistollis	55	18	milj nykymark	6	31	hinto korotuks	41	21	seuraav kysymyks	51	95
maataloud hintalak	44	8	auto moottoripyörävero	16	40	teknillis korkeakoulu	46	28	esit seuraav	65	109
1963 Left	#L	#R	Right	#L	#R	**1968** Left	#L	#R	Right	#L	#R
hinto korotuks	139	16	milj mk	152	281	milj mark	303	265	peruskoulu yläast	6	68
porvarillin enemmistö	97	5	korkeim hallintooikeud	23	150	kuluv vuode	119	88	maatilataloud kehittämisrahasto	7	50
hinto nousu	120	41	äbo akadem	41	161	pysyv työpaiko	67	40	seuraav kysymyks	49	86
vanh mark	96	37	turu yliopisto	24	80	uude pelo	28	5	kaik oppil	13	40
eduskun porvarillin	50	2	milj nmk	43	92	ammatillis järjestö	22	1	viera kiele	34	58
1964 Left	#L	#R	Right	#L	#R	**1969** Left	#L	#R	Right	#L	#R
porvarillin enemmistö	76	3	savo selu	16	156	milj mk	357	291	milj kg	32	222
margariin hin	92	33	milj kilo	10	42	eduskun aika	89	54	helsing yliopisto	91	261
vuon 1963	116	79	v mattil	27	58	vuode 1967	89	56	sisäis hallino	20	115
hinto korotuks	77	41	tois kotim	16	45	en vuode	195	163	yliopisto korkeakoulu	31	124
olev luvu	54	20	viera kiele	21	50	vuon 1967	92	62	korkeakoulu hallino	17	97

Notes: Table shows the five most left and right phrases and their predicted counts for 100,000 phrases. These phrases have the highest impact on the partisanship measure π_t. The coefficients come from a regression with controls for government status of speaker's party, gender and speaker's dialect region. Table shows partisan phrases for every year of the decade available in data.

Table 3.12. 5 most left and right phrases based on $\widehat{\zeta}$, real series, main specification: 1970s

1970 Left	#L	#R	Right	#L	#R	**1975** Left	#L	#R	Right	#L	#R
suome neuvostoliito	139	57	esit kunnioitae	77	184	milj mk	765	451	edellyt hallituks	33	113
ystävyys yhteistyö	76	11	kunnioitae seuraav	97	171	milj mark	413	220	tääl eduskun	187	249
kans etu	65	6	pyydä kunnioittav	26	87	hinto maksu	119	17	saada vast	3	49
sosialistist maide	47	4	kysymyks esit	74	129	sors virolais	124	30	asia yhteyd	30	70
saks demokraattis	45	7	kysymyks toimenpit	43	91	valtio tulovero	89	10	maatilataloud ke-hittämisrahasto	25	62
1971 Left	#L	#R	Right	#L	#R	**1976** Left	#L	#R	Right	#L	#R
milj mk	510	386	liikevaihtovero lisävero	77	131	milj mk	781	499	edellyt hallituks	35	101
hinto maksu	142	23	edellyt hallitus	126	170	milj mark	342	229	hyväksyttäv seuraav	15	70
hinto korotuks	136	27	huumausain käytö	71	112	kysymyks esit	162	64	tääl eduskun	155	191
sosialistist maide	111	6	voim tode	15	44	en vuode	213	147	koko kans	38	73
muide sosialistist	45	2	edellyt hallituks	30	50	kun toimial	85	22	tasaval president	47	82
1972 Left	#L	#R	Right	#L	#R	**1977** Left	#L	#R	Right	#L	#R
sosialistist maide	73	5	tasaval president	211	268	suome neuvostoliito	178	71	keskiast koulutuks	173	258
tamper yliopisto	99	32	oma asuno	77	128	työttöm määrä	110	17	kunnioitae seuraav	57	126
milj mk	377	321	täl kohd	20	52	latinalais amerik	98	5	tasaval president	87	144
kysymyks esit	117	67	ehdot perustelu	24	48	taloudellis yhteistyö	93	10	mark säästö	14	57
eduskun hyväksym	70	30	tääl eduskun	184	201	sosialistist maide	82	3	edellyt hallituks	40	75
1973 Left	#L	#R	Right	#L	#R	**1978** Left	#L	#R	Right	#L	#R
suome neuvostoliito	166	22	tääl eduskun	201	321	milj mark	306	220	kysymyks esit	44	119
yleis yhtäläis	120	10	ede todettav	10	41	suome neuvostoliito	69	22	orientoiv vaihe	14	80
sosialistist maide	119	19	ääni periaat	14	45	toimenpit hallitus	83	36	maatilataloud ke-hittämisrahasto	13	71
hinto maksu	105	24	edellyt hallituks	39	60	hinto nousu	55	12	kunnioitae seuraav	44	99
yhtäläis äänioikeud	88	10	ehdot perustelu	34	54	korkeim oikeud	54	13	asia yhteyd	51	102
1974 Left	#L	#R	Right	#L	#R	**1979** Left	#L	#R	Right	#L	#R
milj mark	340	239	korvaav koulu	97	178	milj mark	400	253	kunnioitae seuraav	100	185
pieni keskituloist	88	1	tääl eduskun	184	263	työnantaj lap-silisämaksu	70	13	esit kunnioitae	78	142
valtio tulovero	83	6	edellyt hallituks	37	87	en vuode	430	375	edellyt hallituks	26	85
mietintö liitety	111	39	yksityis oppikoulu	29	76	työväe urheiluliito	33	3	maatilataloud ke-hittämisrahasto	5	49
hinto maksu	84	15	nest oyn	30	71	työttöm toimeentulo-turv	30	2	oma osa	85	123

Notes: Table shows the five most left and right phrases and their predicted counts for 100,000 phrases. These phrases have the highest impact on the partisanship measure π_t. The coefficients come from a regression with controls for government status of speaker's party, gender and speaker's dialect region. Table shows partisan phrases for every year of the decade available in data.

Table 3.13. 5 most left and right phrases based on $\widehat{\zeta}$, real series, main specification: 1980s

1980 Left	#L	#R	Right	#L	#R	**1985** Left	#L	#R	Right	#L	#R
työnantaj lap-silisämaksu	92	12	kunnioitae seuraav	58	170	kysymyks seuraav	84	17	sotie veteraan	3	47
en vuode	293	216	kotimais energia	65	127	suome neuvostoliito	75	14	esit kunnioitae	48	89
työttöm määrä	42	10	esit kunnioitae	79	133	etel afrik	72	15	pien perheviljelm	24	64
kysymyks toimenpit	59	29	sotie veteraan	3	57	X55 vuode	73	20	rint mal	20	44
toimenpit hallitus	77	48	kotimaist polttoain	14	51	sosiaal terveydenhuolo	101	57	ylimääräis rintamalis	2	24
1981 Left	#L	#R	Right	#L	#R	**1986** Left	#L	#R	Right	#L	#R
valtio yritystoimin	86	5	kunnioitae seuraav	106	266	demokraattis vaihtoehdo	112	4	esit kunnioitae	28	127
työnantaj lap-silisämaksu	62	3	edellyt hallitus	57	111	kysymyks seuraav	98	20	ylimääräis rintamalis	5	38
milj mark	276	225	edellyt hallituks	35	77	suome neuvostoliito	96	20	valtiontaloud tarkas-tusvirasto	14	46
suome neuvostoliito	71	21	esit kunnioitae	98	131	milj mark	270	204	uus lään	60	91
en vuode	279	234	sotie veteraan	4	37	neuvostoliito kaupa	51	9	puheenvuoro johdo	64	93
1982 Left	#L	#R	Right	#L	#R	**1987** Left	#L	#R	Right	#L	#R
milj mark	404	170	edellyt hallitus	24	84	demokraattis vaihtoehdo	56	3	edellis hallituks	97	173
milj mk	543	354	sotie veteraan	6	61	suome pan	68	26	en vuode	319	359
toimenpit hallitus	130	43	edellyt hallituks	29	70	työttöm ol	37	8	viime vaalikaud	30	69
aiko ryhty	111	38	ehdot perustelu	13	54	X55 vuote	31	2	taust vast	39	72
vuode 1983	193	121	kotimais energia	25	65	työeläm uudistamis	49	25	vuode 1988	96	127
1983 Left	#L	#R	Right	#L	#R	**1988** Left	#L	#R	Right	#L	#R
suome neuvostoliito	189	122	milj mk	577	729	suome pan	96	52	en vuode	197	276
valtio lainoittam	74	16	koto tapahtuv	22	55	demokraattis vaihtoehdo	43	3	edellis hallituks	42	87
post telelaitoks	68	13	sotie veteraan	3	33	euroop neuvosto	94	57	milj mk	349	378
kunnioitae seuraav	144	107	vuode 1983	89	116	suome pank	52	26	en vuon	158	181
sosiaal terveydenhuolo	83	47	vaasa lään	34	61	erityis painav	34	10	maatilataloud ke-hittämisrahasto	4	23
1984 Left	#L	#R	Right	#L	#R	**1989** Left	#L	#R	Right	#L	#R
harm raha	77	5	milj mk	551	684	suome pan	116	53	kotihoido tuen	24	76
toimenpit hallitus	100	67	X5 prosent	54	108	demokraattis vaihtoehdo	34	1	en vuode	165	214
raha markkino	34	3	esit kunnioitae	87	124	suome pank	65	35	miljoon mark	253	300
kysymyks seuraav	68	37	kotihoido tuen	16	49	pasil konepaj	31	6	hallituks aika	52	95
X55 vuode	40	10	pien perheviljelm	20	47	vaihtotas vaje	82	61	lap korkeakoulu	22	62

Notes: Table shows the five most left and right phrases and their predicted counts for 100,000 phrases. These phrases have the highest impact on the partisanship measure π_t. The coefficients come from a regression with controls for government status of speaker's party, gender and speaker's dialect region. Table shows partisan phrases for every year of the decade available in data.

Table 3.14. 5 most left and right phrases based on $\widehat{\zeta}$, real series, main specification: 1990s

1990 Left	#L	#R	Right	#L	#R	**1995** Left	#L	#R	Right	#L	#R
suome pan	79	39	kotihoido tuen	30	56	euroop neuvosto	206	55	kotihoido tuen	77	162
euroop neuvosto	130	98	julkis hallino	41	62	euroop union	303	204	miljoon mark	293	369
suome pank	53	31	ylimääräis rintamalis	5	25	viime kaude	127	76	miljard mark	298	373
pitä sisä	38	17	ruots kiele	41	58	kansaneläk pohjaos	73	23	laajennetu rauhanturva-toimin	42	74
nuorisoast koulutuks	36	17	halua tode	87	102	talous rahaliito	57	17	o ojal	16	41
1991 Left	#L	#R	Right	#L	#R	**1996** Left	#L	#R	Right	#L	#R
miljoon mark	387	314	euroop yhteisö	185	343	euroop union	352	204	miljoon mark	205	279
valtio mets	44	7	o ojal	17	42	euroop neuvosto	73	15	miljard mark	175	249
yksilöllis varhaiseläk	33	2	ker kaikia	56	77	ansiosidonnais työttömyysturv	45	15	sähkö tuotano	15	58
sosiaal terveydenhuolo	57	29	sisäis devalvaatio	6	21	mark kuukaud	59	31	ammatillis koulutuks	29	65
m laukkan	46	21	valtio paikallishallino	6	17	pohjoism neuvosto	41	13	kotihoido tuen	50	80
1992 Left	#L	#R	Right	#L	#R	**1997** Left	#L	#R	Right	#L	#R
suome pan	160	109	sama miel	173	209	euroop union	327	227	sukupuo olev	14	37
suome pank	129	91	ker kaikia	54	79	euroop neuvosto	73	29	sama sukupuo	14	36
miljoon mark	300	263	energia säästö	16	41	X7 prosent	47	15	kolmant vaihe	40	60
aho hallituks	64	30	miljard mark	365	389	mark kuukaud	43	12	miljoon mark	275	291
hallituspuolue edustaj	65	31	taloudellis tilant	62	82	aho hallituks	87	59	tot kai	96	111
1993 Left	#L	#R	Right	#L	#R	**1998** Left	#L	#R	Right	#L	#R
miljoon mark	439	255	ker kaikia	54	107	euroop neuvosto	168	57	miljoon mark	279	343
en vuon	171	89	o ojal	29	58	euroop union	313	223	tot kai	113	146
kiinnit huomio	181	135	kotimais energia	38	57	kansaneläk pohjaos	91	24	vars hyv	32	56
euroop union	112	67	välillis työvoimakustan-nuks	16	33	aho hallituks	95	45	täytyy tode	22	41
en vuode	180	144	välillist työvoimakus-tannust	10	27	suome pan	56	28	miljard mark	238	255
1994 Left	#L	#R	Right	#L	#R	**1999** Left	#L	#R	Right	#L	#R
miljoon mark	291	223	euroop union	772	1013	euroop neuvosto	286	121	miljoon mark	434	484
en vuode	164	119	miljard mark	250	313	euroop union	264	170	en vuode	181	207
kiinnit huomio	166	121	neuvoaantav kansanäänestyks	22	67	kansaneläk pohjaos	158	90	uude vuosituhan	15	34
sosiaal terveydenhuolo	45	17	X109 keskiviiko	20	59	aho hallituks	102	53	erit hyvä	103	122
ulko turvallisuuspolitiik	113	87	keskiviiko 2111994	28	66	maksu korotuks	67	25	vuode budjet	61	79

Notes: Table shows the five most left and right phrases and their predicted counts for 100,000 phrases. These phrases have the highest impact on the partisanship measure π_t. The coefficients come from a regression with controls for government status of speaker's party, gender and speaker's dialect region. Table shows partisan phrases for every year of the decade available in data.

Table 3.15. 5 most left and right phrases based on $\widehat{\zeta}$, real series, main specification: 2000s

2000 Left	#L	#R	Right	#L	#R	**2005** Left	#L	#R	Right	#L	#R
euroop neuvosto	209	70	miljard mark	160	253	euroop neuvosto	155	24	kiinnit huomio	144	197
euroop union	214	133	miljoon mark	305	378	euroop union	255	156	erit hyvä	90	131
koko aja	142	100	turvapaik hakij	30	68	varallisuusvero poisto	60	25	erit tärkeä	76	115
mark kuukaud	79	46	en vuode	171	204	X7 prosent	55	27	perintö lahjavero	8	44
asem olev	57	28	valtakunnallis aluei-denkäyttötavoit	22	51	iha oikeast	31	5	laittom lako	11	40
2001 Left	#L	#R	Right	#L	#R	**2006** Left	#L	#R	Right	#L	#R
euroop union	316	158	miljard mark	192	283	euroop union	445	240	erit hyvä	78	132
euroop neuvosto	163	35	last nuort	45	89	euroop neuvosto	200	66	kiinnit huomio	125	166
kansaneläk pohjaos	49	13	miljoon mark	307	343	yk mandaat	91	22	toi es	23	59
euroop unio	63	29	erit hyvä	76	101	toise ast	107	52	miljoon euro	372	402
ansiosidonnais työttömyysturv	36	3	kotihoido tuen	13	38	euroop unio	100	46	vanhas hallituks	18	48
2002 Left	#L	#R	Right	#L	#R	**2007** Left	#L	#R	Right	#L	#R
euroop union	195	118	miljoon euro	210	286	sosiaal terveydenhuolo	218	90	miljoon euro	760	823
kansaneläk pohjaos	93	35	miljard mark	63	100	stora enso	84	27	euroop union	77	131
euroop neuvosto	77	27	en vuode	126	161	julkis palvelu	108	56	en vuon	157	209
pohjaos leikkauks	47	8	vuosaar satam	21	55	en vaalej	104	54	erit hyvä	118	169
koko aja	139	104	erit hyvä	83	113	ensimmäis laps	51	9	omaishoido tuen	5	43
2003 Left	#L	#R	Right	#L	#R	**2008** Left	#L	#R	Right	#L	#R
euroop neuvosto	145	18	kotihoido tuen	9	43	sosiaal terveydenhuolo	151	54	uusiutuv energia	36	97
euroop union	283	200	täl vaalikaud	13	43	stora enso	98	33	miljoon euro	521	580
pohjoism neuvosto	77	15	erit hyvä	74	97	euroop neuvosto	97	47	erit hyvä	131	171
kaik pienitulois	51	13	kansallis terveyshank	13	34	lissabon sopimuks	111	64	en vuon	109	144
kaik köyhim	34	3	annel jäätteenmäe	16	37	harm taloud	66	26	omaishoido tuen	6	37
2004 Left	#L	#R	Right	#L	#R	**2009** Left	#L	#R	Right	#L	#R
euroop union	338	215	erit hyvä	83	116	euroop neuvosto	97	46	erit hyvä	96	164
pohjoism neuvosto	92	33	täl vaalikaud	15	46	miljoon euro	479	434	uusiutuv energia	31	71
miljoon euro	550	514	vanhas hallituks	22	46	harm taloud	74	33	en vuon	128	156
sehä tarkoit	35	9	esa lahtel	41	62	hyvä kolleg	39	11	erit hyv	42	60
iha oikeast	27	6	uus työpaiko	37	56	sosiaal terveydenhuolo	64	37	kiinnit huomio	130	146

Notes: Table shows the five most left and right phrases and their predicted counts for 100,000 phrases. These phrases have the highest impact on the partisanship measure π_t. The coefficients come from a regression with controls for government status of speaker's party, gender and speaker's dialect region. Table shows partisan phrases for every year of the decade available in data.

Table 3.16. 5 most left and right phrases based on $\widehat{\zeta}$, real series, main specification: 2010s

2010 Left	#L	#R	Right	#L	#R	**2018** Left	#L	#R	Right	#L	#R
harm taloud	226	145	uusiutuv energia	68	186	toise ast	267	152	tot kai	123	194
miljoon euro	398	337	erit hyvä	110	170	julkis taloud	132	66	ärad talm	4	40
euroop neuvosto	106	46	euroop union	175	212	ammatillis koulutuks	127	83	esityks ehdot	24	61
sinivihr hallitus	46	4	erit hyv	30	65	yhä ede	37	7	erit hyvä	137	169
hyvä kolleg	46	12	ilm muuta	46	72	talouspolitiik arvioin-tineuvosto	30	2	täl vaalikaud	61	89
2011 Left	#L	#R	Right	#L	#R	**2016** Left	#L	#R	Right	#L	#R
harm taloud	300	139	miljoon euro	659	767	viime kaude	218	145	miljoon euro	393	513
viime kaude	194	108	uusiutuv energia	19	103	ammatillis koulutuks	140	90	tot kai	123	172
edellis hallituks	142	75	vihreä taloud	1	20	harm taloud	129	79	uusiutuv energia	14	51
X8 miljard	78	15	sama sukupuo	12	28	en vaalej	73	26	miljard euro	126	159
kaik pienitulois	68	18	sukupuo olev	11	26	palkansaaj eläkkeensaaj	43	2	miehe naise	9	38
2012 Left	#L	#R	Right	#L	#R	**2017** Left	#L	#R	Right	#L	#R
nuort yhteiskuntataku	132	64	miljoon euro	538	611	kasvu kuulu	79	11	miljoon euro	328	458
viime kaude	107	48	erit hyvä	106	141	ammatillis koulutuks	274	208	erit hyvä	96	173
harm taloud	124	70	uusiutuv energia	15	46	harm taloud	99	48	laittom maas	3	44
työmarkkinatue tarve-harkin	56	7	miljard euro	212	241	hallituspuolue edustaj	77	33	ärad talm	6	36
oma osa	112	64	vihreä taloud	3	22	viime kaude	116	74	erit tärkeä	90	115
2013 Left	#L	#R	Right	#L	#R	**2018** Left	#L	#R	Right	#L	#R
harm taloud	135	49	uusiutuv energia	30	74	toise ast	267	152	tot kai	123	194
toise ast	87	47	erit hyvä	130	161	julkis taloud	132	66	ärad talm	4	40
en vuode	162	124	liikkuv poliis	65	92	ammatillis koulutuks	127	83	esityks ehdot	24	61
viime kaude	66	33	kunt tehtäv	21	48	yhä ede	37	7	erit hyvä	137	169
tot kai	158	126	miljoon euro	462	489	talouspolitiik arvioin-tineuvosto	30	2	täl vaalikaud	61	89
2014 Left	#L	#R	Right	#L	#R	**2019** Left	#L	#R	Right	#L	#R
harm taloud	105	36	uusiutuv energia	25	86						
toise ast	100	47	ruots kiele	76	118						
euroop union	213	167	täl vaalikaud	32	64						
sosiaal terveydenhuolo	88	62	sukupuo olev	49	80						
miljoon euro	487	468	miehe naise	19	49						

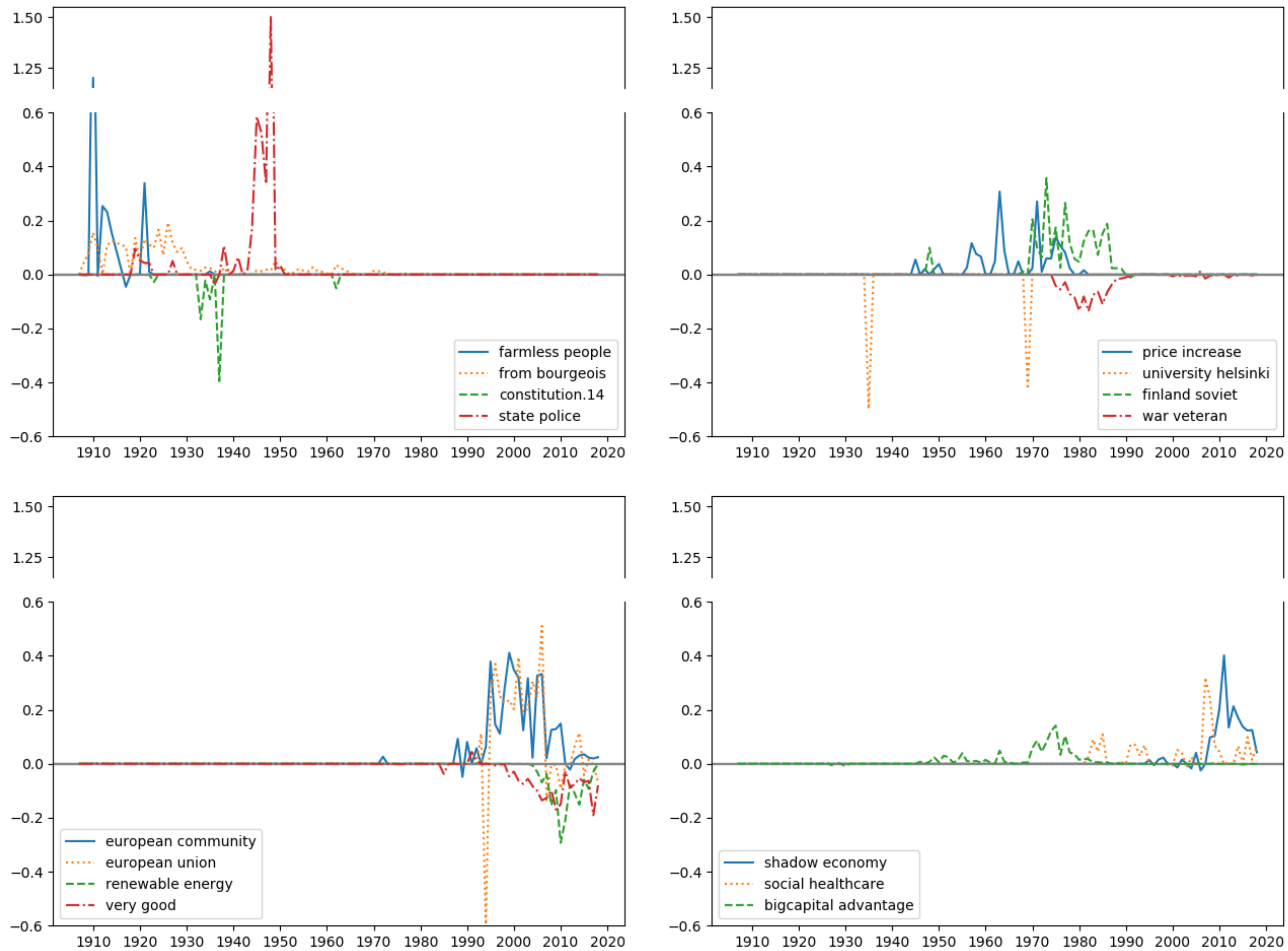
Notes: Table shows the five most left and right phrases and their predicted counts for 100,000 phrases. These phrases have the highest impact on the partisanship measure π_t. The coefficients come from a regression with controls for government status of speaker's party, gender and speaker's dialect region. Table shows partisan phrases for every year of the decade available in data.

Table 3.4. 5 most left and right phrases based on $\widehat{\zeta}$, controls for government status, gender and region (main specification)

1910 Left	#L	#R	Right	#L	#R	1920 Left	#L	#R	Right	#L	#R
tilattom väestö	661	180	alem kansakoulu	17	118	tulo menoarvio	589	432	milj kilo	33	98
hyvä her	103	16	suome rautat	27	125	kuluv vuode	141	73	suome pan	67	131
työtätekev väestö	78	0	ylem kansakoulu	44	141	porvarillin enemmistö	39	0	kannat tehty	12	45
porvarillin enemmistö	71	0	suome koskev	14	61	porvarillis edustaj	36	0	ehdot eduskun	62	92
porvar taho	70	4	X4000 mark	23	59	valkois suome	33	0	lyhytaikais luoto	15	45
1930 Left	#L	#R	Right	#L	#R	**1940 Left**	#L	#R	Right	#L	#R
porvarillis edustaj	25	3	kansanliik vas	1	55	milj mark	443	384	korvauks oikeutetu	62	122
porvaristo taho	20	0	työtätekev väestö	10	27	uus vero	47	14	uude tila	8	50
porvarillis piire	22	3	vuote verrat	4	20	julkis sana	47	22	uus tilo	36	77
töide järjestämis	14	0	lain hyv	4	18	työläist palko	27	2	perjant 3	23	60
porvarillis sanomalehd	13	0	työtätekev kans	0	13	ammatillis järjestö	20	0	uus olo	21	55
1950 Left	#L	#R	Right	#L	#R	**1960 Left**	#L	#R	Right	#L	#R
fagerholm hallituks	104	30	henkis työn	25	205	rautatiehallituks pääjohtaj	71	21	yleis työmarkkino	4	45
hinto nousu	117	43	työn tekij	25	100	r virtas	55	11	pyöreä puutavar	6	31
fagerholm hallitus	75	18	useam yhd	5	33	pakollis vakuutuks	57	24	milj kilo	8	25
vuode 1949	102	47	ensimmäis laps	13	39	sosialistist maide	26	0	porvarillin enemmistö	3	14
timot siemen	90	37	ruumiillis työn	10	30	milj mk	611	587	valtiopäiväjärjestyks 67	15	22
1970 Left	#L	#R	Right	#L	#R	**1980 Left**	#L	#R	Right	#L	#R
suome neuvostoliito	139	57	esit kunnioitae	77	184	työnantaj lapsilisämaksu	92	12	kunnioitae seuraav	58	170
ystävyys yhteistyö	76	11	kunnioitae seuraav	97	171	en vuode	293	216	kotimais energia	65	127
kans etu	65	6	pyydä kunnioittav	26	87	työttöm määrä	42	10	esit kunnioitae	79	133
sosialistist maide	47	4	kysymyks esit	74	129	kysymyks toimenpit	59	29	sotie veteraan	3	57
saks demokraattis	45	7	kysymyks toimenpit	43	91	toimenpit hallitus	77	48	kotimaist polttoain	14	51
1990 Left	#L	#R	Right	#L	#R	**2000 Left**	#L	#R	Right	#L	#R
suome pan	79	39	kotihoido tuen	30	56	euroop neuvosto	209	70	miljard mark	160	253
euroop neuvosto	130	98	julkis hallino	41	62	euroop union	214	133	miljoon mark	305	378
suome pank	53	31	ylimääräis rintamalis	5	25	koko aja	142	100	turvapaik hakij	30	68
pitä sisä	38	17	ruots kiele	41	58	mark kuukaud	79	46	en vuode	171	204
nuorisoast koulutuks	36	17	halua tode	87	102	asem olev	57	28	valtakunnallis alueidenkäyttötavoit	22	51
2010 Left	#L	#R	Right	#L	#R	**2018 Left**	#L	#R	Right	#L	#R
harm taloud	226	145	uusiutuv energia	68	186	toise ast	267	152	tot kai	123	194
miljoon euro	398	337	erit hyvä	110	170	julkis taloud	132	66	ärad talm	4	40
euroop neuvosto	106	46	euroop union	175	212	ammatillis koulutuks	127	83	esityks ehdot	24	61
sinivihr hallitus	46	4	erit hyv	30	65	yhä ede	37	7	erit hyvä	137	169
hyvä kolleg	46	12	ilm muuta	46	72	talouspolitiik arvioinneuvosto	30	2	täl vaalikaud	61	89

Notes: Table shows the five most partisan phrases and their predicted counts for 100,000 phrases from a series with real labels (left and right parties). These phrases have the highest impact on the partisanship measure through both ρ_{jt} and q_{jt}. The coefficients come from the main specification with controls for government status, gender and region. Table shows partisan phrases for every 10 years and the last year in data.

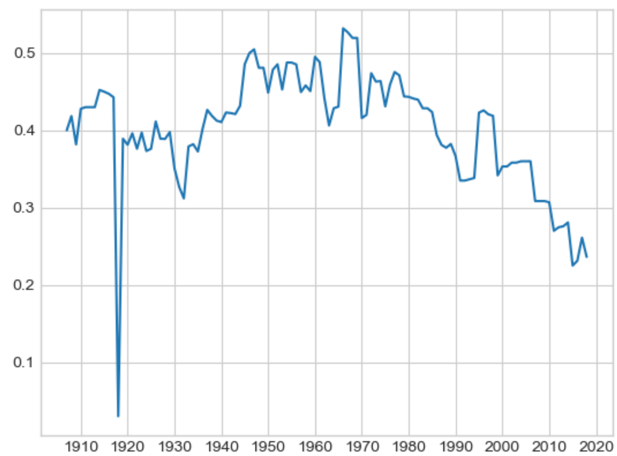
Figure A14: Left-right partisanship ζ_{jt} of selected phrases



Notes: Figure shows phrase partisanship ζ_{jt}, multiplied by 1000, for phrases selected from among the top 3 most partisan phrases of each decade. Positive values are 'left' phrases, negative values 'right' phrases. The ranking of phrases is based on the absolute value of the average ζ_j for the decade. The vertical axis is broken to mitigate the impact of outliers on the clarity of exposition. In the lack of better translation, 'valtiollis.poliis' is translated to 'state police' and 'hallitusmuodo.14' to 'constitution 14'.

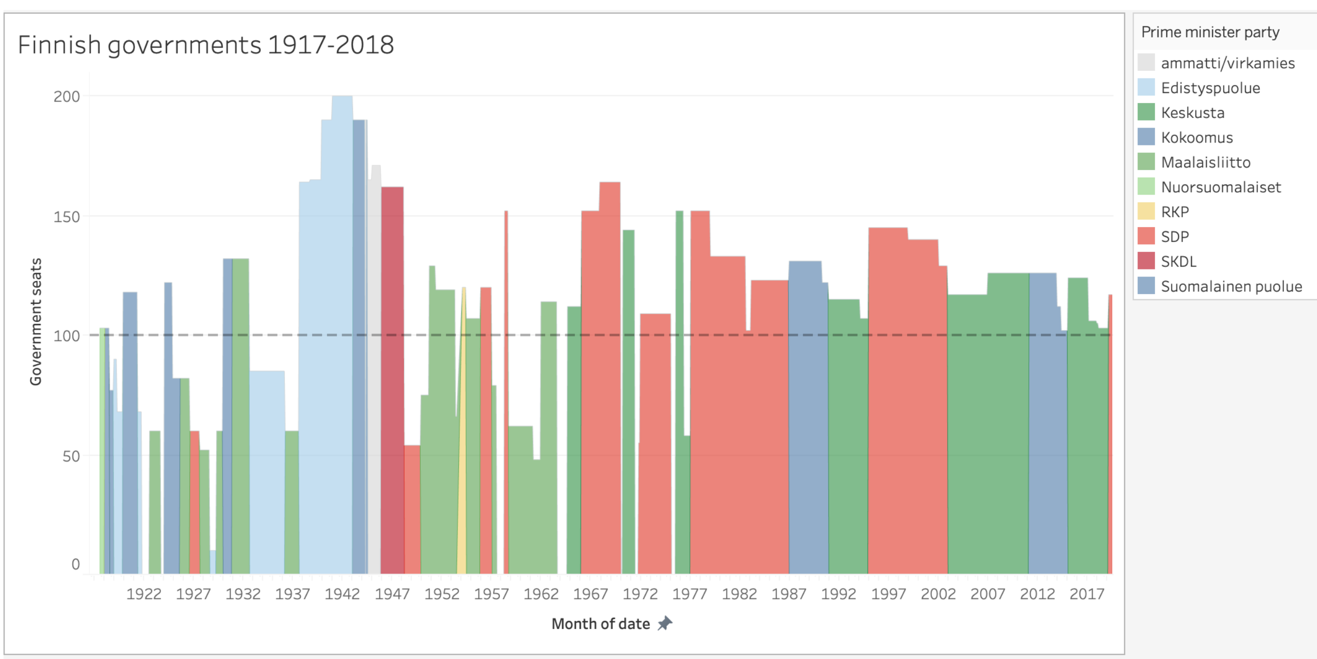
Online Appendix D: Descriptive figures and tables

Figure A15: Seat share of left parties in the Parliament



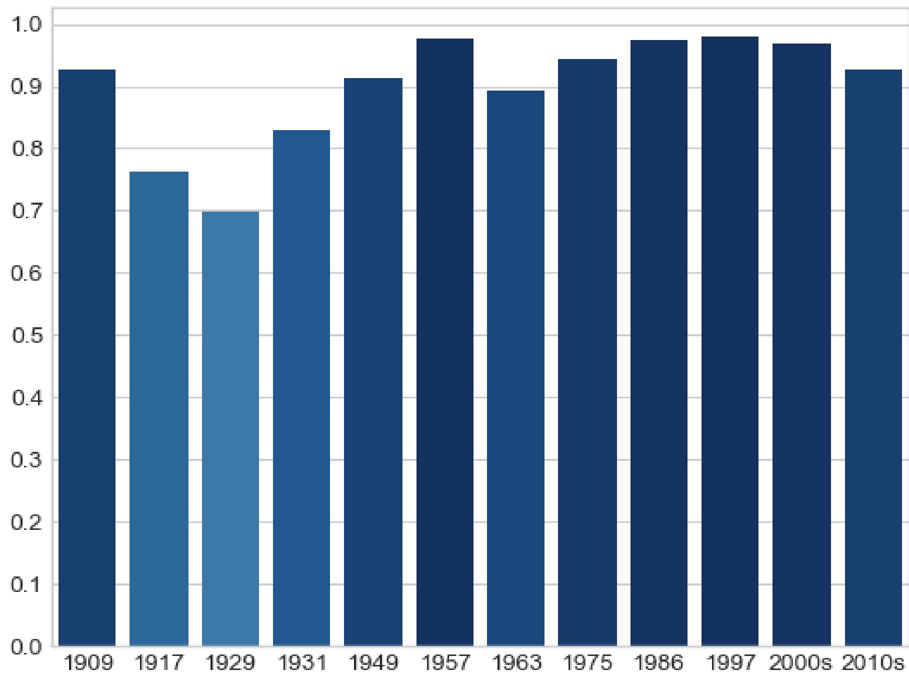
Notes: Figure by author. Calculations are based on MPs active during the last three months of the calendar year.

Figure A16: Government seat share, 1907–2018



Notes: Figure by author.

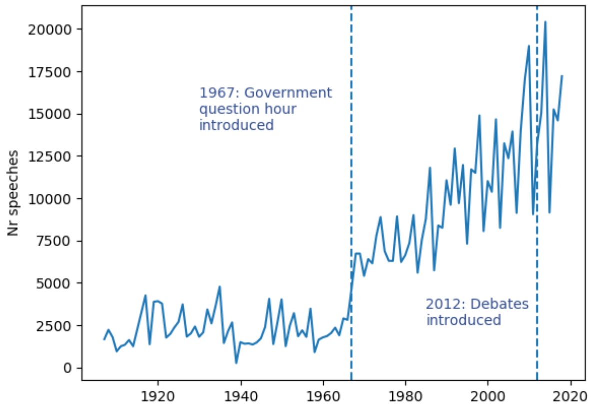
Figure A17: Share of transcribed speeches successfully retrieved



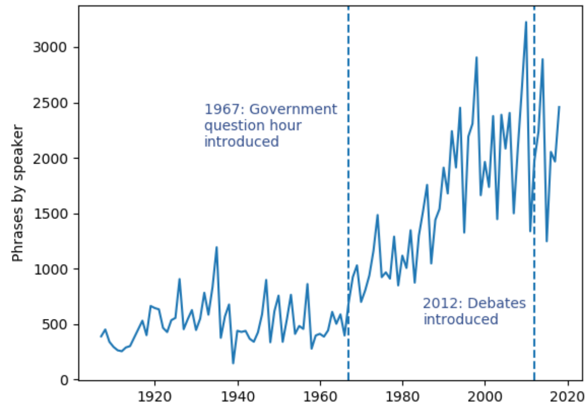
Notes: The figure shows the share of all parliamentary speeches successfully retrieved by the automated parsing script by the author. The number is based on a manual audit of 10 randomly selected parliamentary transcripts from one randomly selected year in each decade. The manual audit was conducted in October, 2019.

Figure A18: Amount of speech over time

(a) Total number of speeches given in a year

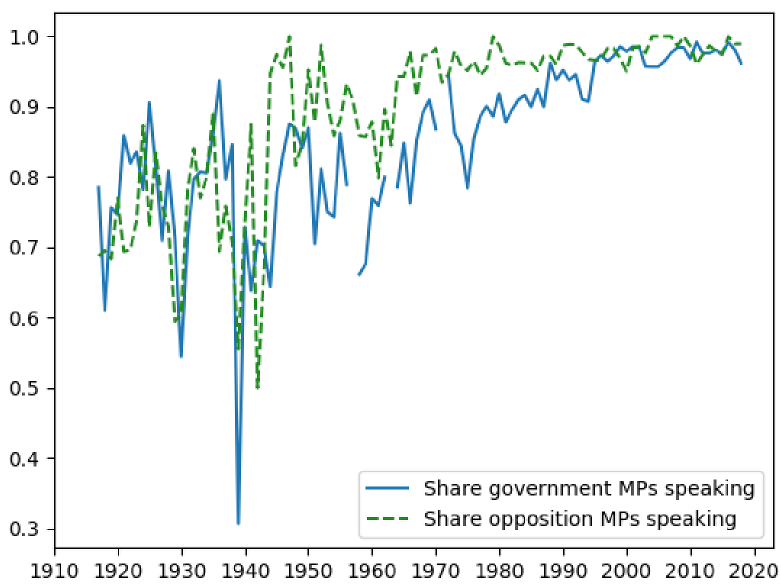


(b) Average number of phrases spoken in a year



Notes: Figure A18a shows the number of Finnish and Swedish speeches retrieved for each parliamentary year. Figure A18b shows the average number of phrases spoken by an MP during the parliamentary year.

Figure A19: Share of government and opposition MPs speaking, 1907–2018



Notes: Figure A19 shows the share of government and opposition MPs speaking during the parliamentary year. Breaks in the time series are for years with governments by non-partisan officials (virkamieshallitus) where number of government MPs is zero.

Table 3.1. Share of MPs and speeches by party

	MPs	Speeches			MPs	Speeches	
Year 1910	Share	Share	Per MP	Year 1920	Share	Share	Per MP
Social Democratic Party	0.43	0.34	2.9	Social Democratic Party	0.4	0.37	13.4
Finnish Party	0.21	0.24	4.0	Agrarian League	0.22	0.29	19.3
Young Finnish Party	0.14	0.26	6.6	National Progressive Party	0.15	0.15	14.6
Swedish People's Party	0.13	0.02	0.6	National Coalition Party	0.15	0.16	15.7
Agrarian League	0.09	0.13	5.4	Swedish People's Party	0.1	0.01	1.6
Christian Workers' Union of Finland	0.01	0.01	9.0	Young Finnish Party	0.01	0.0	0.0
				Liberal Swedish Party	0.01	0.0	7.0
				Christian Workers' Union of Finland	0.01	0.01	14.5
	MPs	Speeches			MPs	Speeches	
Year 1930	Share	Share	Per MP	Year 1940	Share	Share	Per MP
Agrarian League	0.41	0.3	5.5	Social Democratic Party	0.42	0.41	6.1
Social Democratic Party	0.4	0.41	7.6	Agrarian League	0.28	0.31	6.9
National Coalition Party	0.26	0.19	5.3	National Coalition Party	0.12	0.16	8.1
Swedish People's Party	0.16	0.04	1.7	Swedish People's Party	0.1	0.03	1.9
STPV	0.1	0.0	0.1	Patriotic People's Movement (IKL)	0.04	0.05	9.4
National Progressive Party	0.07	0.05	5.0	National Progressive Party	0.04	0.03	5.5
Liberal Swedish Party	0.01	0.0	5.0	Small Farmers' Party (SPP)	0.01	0.01	3.5
Small Farmers' Party (SPP)	0.01	0.0	6.0				
	MPs	Speeches			MPs	Speeches	
Year 1950	Share	Share	Per MP	Year 1960	Share	Share	Per MP
Agrarian League	0.29	0.23	13.6	Agrarian League	0.26	0.22	6.5
Social Democratic Party	0.28	0.24	14.9	SKDL	0.25	0.34	10.7
SKDL	0.18	0.32	29.9	Social Democratic Party	0.18	0.13	5.5
National Coalition Party	0.17	0.18	18.7	National Coalition Party	0.14	0.17	9.0
Swedish People's Party	0.07	0.01	2.2	Swedish People's Party	0.07	0.01	1.5
National Progressive Party	0.02	0.01	11.8	TPSL	0.07	0.06	5.9
Liberals	0.01	0.0	17.0	Liberals	0.03	0.04	9.8
Finnish People's Party	0.01	0.01	21.0	Finnish People's Party	0.01	0.02	12.0
	MPs	Speeches			MPs	Speeches	
Year 1970	Share	Share	Per MP	Year 1980	Share	Share	Per MP
Social Democratic Party	0.27	0.16	15.0	Social Democratic Party	0.27	0.11	12.3
National Coalition Party	0.18	0.21	29.5	National Coalition Party	0.23	0.16	21.1
Centre Party	0.18	0.12	16.7	Centre Party	0.18	0.09	15.0
SKDL	0.18	0.12	17.2	SKDL	0.18	0.17	29.2
Finnish Rural Party (SMP)	0.09	0.3	85.2	Christian Democrats	0.05	0.08	49.6
Swedish People's Party	0.06	0.04	16.9	Swedish People's Party	0.04	0.02	14.0
Liberals	0.04	0.04	21.9	Finnish Rural Party (SMP)	0.03	0.34	351.8
Christian Democrats	0.01	0.01	67.0	Liberals	0.02	0.04	67.0
	MPs	Speeches			MPs	Speeches	
Year 1990	Share	Share	Per MP	Year 2000	Share	Share	Per MP
National Coalition Party	0.28	0.17	33.0	Social Democratic Party	0.26	0.25	49.1
Social Democratic Party	0.28	0.18	34.6	National Coalition Party	0.23	0.2	43.8
Centre Party	0.2	0.2	53.3	Centre Party	0.23	0.26	54.8
Left Alliance	0.1	0.19	103.6	Left Alliance	0.1	0.13	67.7
Swedish People's Party	0.07	0.04	33.5	Swedish People's Party	0.06	0.03	25.5
Finnish Rural Party (SMP)	0.04	0.1	136.0	Green League	0.06	0.05	50.1
Christian Democrats	0.03	0.04	84.2	Christian Democrats	0.05	0.07	67.1
Green League	0.02	0.07	178.5	Sulo Aittoniemi	0.01	0.0	1.0
	MPs	Speeches			MPs	Speeches	
Year 2010	Share	Share	Per MP	Year 2018	Share	Share	Per MP
National Coalition Party	0.27	0.24	77.9	Centre Party	0.24	0.22	64.9
Centre Party	0.25	0.24	82.2	National Coalition Party	0.19	0.16	62.7
Social Democratic Party	0.23	0.2	76.2	Social Democratic Party	0.17	0.2	80.9
Left Alliance	0.09	0.13	134.8	Finns Party	0.09	0.11	92.5
Green League	0.07	0.1	128.4	Blue Reform	0.09	0.07	59.3
Christian Democrats	0.04	0.04	101.3	Green League	0.07	0.08	73.6
Swedish People's Party	0.04	0.01	23.4	Left Alliance	0.06	0.08	98.0
Finns Party	0.03	0.03	96.5	Swedish People's Party	0.04	0.02	39.1

Notes: STPV = Socialist Workers' and Smallholders' Party. TPSL = Social Democratic Union of Workers and Smallholders. SKDL = Finnish People's Democratic League. Table shows statistics for every 10 years and the last year in the data and for 8 parties with the highest number of seats in the Parliament.

Table 3.2. Summary statistics: Left parties, gender and government

	Left parties			Females			Government parties		
	Share MPs	Share speakers	Share speeches	Share MPs	Share speakers	Share speeches	Share MPs	Share speakers	Share speeches
1910	0.43	0.41	0.29	0.09	0.06	0.05	-	-	-
1920	0.39	0.4	0.36	0.09	0.1	0.04	0.59	0.58	0.64
1925	0.39	0.39	0.41	0.09	0.09	0.06	0.42	0.47	0.35
1930	0.36	0.38	0.38	0.07	0.05	0.05	0.64	0.62	0.61
1935	0.38	0.38	0.27	0.07	0.07	0.05	0.52	0.51	0.57
1940	0.42	0.4	0.37	0.08	0.07	0.04	0.91	0.91	0.91
1945	0.51	0.48	0.48	0.09	0.08	0.04	0.81	0.78	0.78
1950	0.45	0.47	0.52	0.11	0.12	0.09	0.38	0.36	0.31
1955	0.49	0.47	0.46	0.15	0.15	0.11	0.54	0.54	0.34
1960	0.5	0.5	0.5	0.14	0.16	0.14	0.32	0.29	0.28
1965	0.43	0.46	0.6	0.13	0.14	0.15	0.56	0.53	0.39
1970	0.44	0.45	0.28	0.21	0.21	0.17	0.71	0.68	0.47
1975	0.46	0.45	0.24	0.2	0.18	0.13	0.7	0.66	0.3
1980	0.45	0.43	0.27	0.25	0.26	0.22	0.61	0.59	0.3
1985	0.42	0.43	0.41	0.32	0.32	0.31	0.6	0.58	0.43
1990	0.38	0.37	0.37	0.31	0.31	0.29	0.62	0.61	0.39
1995	0.43	0.43	0.39	0.34	0.33	0.26	0.7	0.7	0.53
2000	0.36	0.36	0.38	0.37	0.37	0.31	0.7	0.7	0.66
2005	0.36	0.37	0.34	0.38	0.38	0.33	0.58	0.57	0.38
2010	0.31	0.31	0.33	0.39	0.39	0.33	0.63	0.62	0.59
2015	0.23	0.23	0.33	0.42	0.41	0.44	0.62	0.62	0.51
2018	0.23	0.24	0.28	0.42	0.42	0.42	0.53	0.52	0.46

Notes: Table shows summary statistics for three overlapping categorizations: left parties, females and government parties. 'Share MPs' shows the share of parliament members that belong to the category. 'Share speakers' shows the share of speakers (MPs with positive amount of speech) that belong to the category. 'Share speeches' shows the share of all speeches in a year given by an MP that belongs to the category. Table shows statistics for every 5 years and the last year in data. Year 1915 is missing because the Parliament did not gather.

Table 3.3. Summary statistics: Regions

	Helsinki and Uusimaa		Southwest Finland		Häme & Pirkanmaa		Southern Ostrobothnia		Oulu & Lapland		Savo		Southeast Finland	
	Share MPs	Share speeches	Share MPs	Share speeches	Share MPs	Share speeches	Share MPs	Share speeches	Share MPs	Share speeches	Share MPs	Share speeches	Share MPs	Share speeches
1910	0.11	0.18	0.17	0.14	0.11	0.1	0.16	0.14	0.1	0.13	0.2	0.14	0.14	0.17
1920	0.12	0.22	0.15	0.13	0.1	0.11	0.15	0.1	0.1	0.1	0.19	0.11	0.16	0.22
1925	0.12	0.22	0.17	0.19	0.1	0.1	0.15	0.12	0.1	0.1	0.18	0.13	0.16	0.15
1930	0.12	0.19	0.16	0.12	0.1	0.13	0.15	0.12	0.11	0.12	0.2	0.17	0.16	0.16
1935	0.12	0.14	0.15	0.12	0.11	0.11	0.14	0.16	0.11	0.14	0.18	0.2	0.17	0.14
1940	0.13	0.16	0.13	0.13	0.09	0.13	0.14	0.15	0.12	0.07	0.19	0.14	0.2	0.21
1945	0.16	0.13	0.14	0.16	0.1	0.11	0.14	0.16	0.13	0.09	0.16	0.14	0.16	0.21
1950	0.15	0.16	0.13	0.14	0.1	0.14	0.15	0.14	0.12	0.12	0.17	0.13	0.16	0.16
1955	0.18	0.16	0.15	0.16	0.11	0.17	0.14	0.15	0.12	0.12	0.18	0.13	0.1	0.1
1960	0.16	0.16	0.16	0.16	0.13	0.14	0.15	0.15	0.13	0.14	0.18	0.14	0.09	0.1
1965	0.17	0.16	0.15	0.16	0.12	0.15	0.15	0.12	0.13	0.11	0.18	0.18	0.09	0.12
1970	0.21	0.15	0.14	0.14	0.13	0.12	0.12	0.09	0.13	0.14	0.18	0.17	0.08	0.18
1975	0.21	0.22	0.15	0.1	0.15	0.12	0.1	0.05	0.13	0.21	0.18	0.1	0.07	0.19
1980	0.23	0.21	0.14	0.2	0.14	0.08	0.09	0.09	0.12	0.14	0.18	0.09	0.08	0.17
1985	0.23	0.23	0.15	0.15	0.14	0.17	0.1	0.06	0.13	0.14	0.17	0.18	0.07	0.06
1990	0.23	0.23	0.16	0.15	0.14	0.19	0.09	0.07	0.13	0.15	0.17	0.15	0.07	0.06
1995	0.25	0.21	0.15	0.12	0.14	0.16	0.09	0.08	0.13	0.17	0.18	0.2	0.07	0.07
2000	0.25	0.19	0.14	0.17	0.15	0.12	0.09	0.07	0.13	0.14	0.18	0.21	0.07	0.1
2005	0.27	0.23	0.13	0.1	0.16	0.16	0.09	0.11	0.13	0.12	0.17	0.17	0.06	0.1
2010	0.27	0.2	0.14	0.12	0.15	0.13	0.09	0.08	0.12	0.17	0.16	0.2	0.06	0.08
2015	0.29	0.27	0.13	0.18	0.16	0.13	0.08	0.08	0.12	0.1	0.14	0.17	0.07	0.06
2018	0.28	0.24	0.14	0.15	0.16	0.14	0.08	0.14	0.13	0.09	0.14	0.17	0.07	0.07

Notes: 'Share MPs' shows the share of parliament members coming from an electoral district that belongs to the region. 'Share speeches' shows the share of speeches by parliament members from the region. Missing category is omitted and consists of a maximum of 2 percent of MPs. Table shows statistics for every 5 years and the last year in data. Year 1915 is missing since the Parliament did not gather.

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