



Evolution of women's descriptive and substantive representation: Evidence from 70 years of parliamentary debates

Jeremias Nieminen^{a, b}, Ilona Lahdelma^c, Salla Simola^d, Janne Tukiainen^{a, *} 

^a Department of Economics, Turku School of Economics, University of Turku, Turku, Varsinais-Suomi, Finland

^b Labour Institute for Economic Research LABORE, Helsinki, Uusimaa, Finland

^c Department of Social Sciences, Carlos III University, Madrid, Spain

^d Wolt, Stockholm, Sweden

ARTICLE INFO

Keywords:

Critical mass
Parliamentary speech
Text analysis
Women's representation

ABSTRACT

While differences in how men and women speak in parliament are well established, we still lack evidence encompassing a longer time span. Using Finnish parliamentary speech data from 1948 to 2018—a period spanning from token female presence to near-parity—we find that while women have always spoken differently than men, the differences widen as women's share of seats surpasses roughly 10–20% in the 1960s. Crucially, women already serving in parliament at that point also shift toward topics more closely aligned with “women's issues”, suggesting the change reflects a collective contextual shift rather than the arrival of new cohorts. Moreover, tenure length is not correlated with how women speak, further indicating that changes in women's behavior relate more to contextual factors than to individual career trajectories.

1. Introduction

In recent decades, many countries have seen large changes in the demographic composition of their parliaments as more women are elected.¹ These changes have in turn fueled a debate about whether female members of parliament behave differently from their male peers when controlling for party differences. Indeed, it has been argued that in addition to being more involved in matters directly affecting women (Chattopadhyay and Duflo, 2004; Caiazza, 2004; Swers, 1998) women are engaged with a different set of issues such as family, education, and social policies, whereas men are typically more involved in financial and foreign policies (Bäck and Debus, 2019; Mendelberg et al., 2014; Catalano, 2009).

Research on women's political representation is necessarily time-limited: descriptive representation can only be studied from the point at which women gained the right to stand for election, while substantive representation requires recorded speech or behavioral data. In practice, most existing studies rely on data from the United States and Western Europe covering the late 1980s to the early 2020s. However, these data represent a time span in which female representation is no longer novel and nascent, but rather established, reaping the benefits of early female political pioneers. Most studies of gendered parliamentary speech draw on data from the late 1980s onwards—a period in which women had already been present in Western legislatures for several decades. This sampling window obscures the earlier era in which women were a genuine novelty in parliament, their presence was modest, and the institutions they entered had been shaped exclusively by men.

* Corresponding author.

Email addresses: jeremias.nieminen@utu.fi (J. Nieminen), ilahdelm@clio.uc3m.es (I. Lahdelma), salla.mari.simola@gmail.com (S. Simola), janne.tukiainen@utu.fi (J. Tukiainen).

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

In contrast to the existing literature, we use the interesting case of Finland, which was the first self-governing state to grant women both the right to vote and run for office in 1906 and has recorded parliamentary speeches ever since. In addition to the unusually long timeline, this case is also a “clean slate” in which, unlike in almost all other cases, women did not have to fight against a more patriarchal establishment, but were there in the creation of the institutions from the very beginning. Moreover, the institutional setting of proportional representation, present from the very first election, has been shown to increase the chances of female candidates getting elected. This allows us to track how women’s presence, representation, and political context intertwined over the years. Finally, the open-list proportional representation system with a personal vote creates an institutional incentive for women to listen to and actively represent women in their constituencies in order to earn their votes. The long time horizon allows us to disentangle which features of women MPs’ speech were present from the beginning, which emerged later with contextual changes, and how stable gender-based differences in political speech are overall.

Whether women’s distinctive issue focus constitutes a limitation or a strength of descriptive representation remains contested. One strand of the literature argues that if women disproportionately engage with topics that are less central to mainstream legislative agendas, the case for gender quotas is weakened (Fernandes et al., 2023). An alternative view holds that women’s representation is valuable precisely because it elevates issues that would otherwise receive insufficient attention (Celis et al., 2008; Wängnerud, 2009).

It has also been debated whether women use systematically different ways of speaking in parliament than men (Wäckerle and Castanho Silva, 2023; Yu, 2013) and whether this difference diminishes over time (Castanho Silva et al., 2024; Hargrave and Blumenau, 2022). Moreover, it has been argued that these differences may only come into existence once a certain threshold of female MPs (critical mass) has been reached (Dahlerup, 2006a). The view that women and men consistently differ in legislative behavior has been challenged by research showing that as women gain tenure and career stakes rise, they increasingly adopt behavioral patterns similar to their male colleagues (Bailer et al., 2022).

The long time dimension of women’s political representation in this paper serves as an optimal setting to examine how women’s political representation “organically” unfolded without institutional incentives and interventions and to see how different milestones such as reaching the critical mass of women representatives (e.g., 10–20%) or adopting welfare state institutions coincided with changes in women’s parliamentary behavior. In addition, we can track how women who had longer tenures in the parliament changed over the years as the context and institutions evolved. Further questions we can answer with the long time horizon of the data are when and to what extent women represent other women, and if differences in a) speech styles and b) topics discussed between men and women have been changing over time. The research question of this paper is twofold: first, do gender-based speech differences change with critical mass, and if so in what direction? Second, do these changes reflect individual career trajectories or collective contextual shifts—that is, do women change as individuals over their careers, or does the context change around them?

To summarize our main results, we observe that, first, women have always spoken differently than men in the Finnish Parliament. This difference is larger and less volatile than the difference in speech between left and right wing parties (Simola et al., 2025a). Second, however, the speech differences increase from the 1960s onward when women’s seat share surpassed the critical mass of 10–20% consistent with a nonlinear relationship between descriptive and substantive representation of women. Crucially, the period of 1950–1970 is also the period of significant expansion of the welfare state, and we can see from the topics women preferred to speak about that women were overrepresented in speeches about welfare topics and that the creation of these institutions also coincided with a steady increase in women MPs. Third, women who had already been serving as MPs also began to speak increasingly differently than men during this period—suggesting that the shift reflects a contextual change affecting incumbents, and is not simply an artifact of new cohorts of women entering parliament. That is, women MPs appear to talk more about topics that women voters care about.

This finding complements the tenure result reported below: while the incumbent analysis rules out a purely compositional explanation, the absence of a tenure effect provides little support for an individual career-trajectory explanation. Together, they are consistent with a contextual account in which the shift is tied to the critical mass transition rather than the arrival of new MPs or the accumulation of individual experience.

Fourth, after the 1970s, the differences plateau. However, differences in speech and topic choices persist. Therefore, gender differences in behavior remain detectable even in a system with high descriptive representation of women.

Finally, in line with this, we find that women’s career progress does not correlate with changes in the composition of topics that women speak about. Our key contribution is to show using speech data spanning over 70 years that women and men speak differently, and when women MPs change their behavior it relates more to contextual factors of overall women representation and changes in voter demand, than to individual women MPs’ characteristics such as tenure.

2. Theoretical background and previous literature

The concepts of descriptive and substantive representation were introduced by Pitkin (1967), who argued that it is important to focus on assessing whether a group’s views are advocated for in the parliament, i.e., substantive representation, instead of focusing only on a group’s seat share in the parliament, i.e., descriptive representation. Since then, a large literature on women’s political representation has studied the links between women’s descriptive and substantive representation. According to Cowell-Meyers and Langbein (2009), the literature measures substantive representation by comparing stated priorities of female MPs to those of female voters, by examining policy proposals, voting, and committee behavior, or by estimating the effect of women’s seat share on actual policy outcomes (see Hessami and da Fonseca (2020) for a comprehensive review). Instead, we measure substantive representation in parliamentary speech, as parliamentary speech data can be used to analyze for whom representatives act, which is the original definition of substantive representation. We note that our speech-based measure captures one specific dimension of substantive representation: issue emphasis and framing in parliamentary debates. Substantive representation of women can also occur through

male MPs advocating for women's priorities, through committee work, bill sponsorship, or policy outcomes. Furthermore, women MPs who focus on broader issues may still represent women through the perspectives they bring. Our topic analysis complements the aggregate measure by providing more detail about what men and women speak about.²

In doing so, we align with the literature on female politicians as strategic actors who want to optimize their re-election chances during their parliamentary mandates and do so by appealing to fellow women in the electorate (Herrnson et al., 2003; Adams et al., 2005). There is a growing field of literature arguing that women might opt out of representing other women (Höhmnn, 2019; Kläy et al., 2025; Itzkovitch-Malka and Oshri, 2024; Shim, 2022) for strategic reasons to increase their chances of re-election and for securing more important positions in government. However, following the logic of Carey and Shugart (1995), in the Finnish context each candidate has a personal constituency and the elected candidates are working toward getting re-elected by the same or very similar people to those who voted for them in the initial round. Moreover, plenary speeches provide MPs with a public and recorded venue for communicating issue priorities, whether constituents encounter them directly, through transcripts, or through media coverage like radio or televised debates. A candidate changing the topics they talk about mid-election cycle is risky behavior and is not likely to lead to significant gains, as each candidate is effectively competing with the others for different demographics. We take this particularity of the Finnish context to signal that a rational politician would a) choose the constituency they represent and b) make use of their parliamentary speeches to signal to their core constituents that they are indeed representing them.

Women's parliamentary speech has been increasingly the target of research mainly because of the advent of text-as-data methods that make it relatively easy to compare speech patterns across time and space. This research has demonstrated that women not only speak about different issues than men (Mendelberg et al., 2014; Lowande et al., 2019; Blaxill and Beelen, 2016), but also that women speak *differently* than men, too, in terms of tone (Adams et al., 2023; Dietrich et al., 2019) and choice of words (Yu, 2013).

Several complementary mechanisms explain why women speak differently than men even when controlling for party affiliation. First, women bring distinct lived experiences and policy priorities to parliament, shaped by gendered socialization and labor market trajectories that make them more likely to prioritize issues such as childcare, healthcare, and education (Thomas and Welch, 1991; Norris and Lovenduski, 1995). Second, role congruity theory predicts that women in leadership face social pressure to speak on communal and caring topics: deviating too far from gender-stereotypical issue domains risks backlash from both voters and colleagues, constraining women's speech even when individual preferences might differ (Eagly and Karau, 2002). Third, and particularly relevant in the Finnish context, the open-list proportional representation system creates a direct electoral incentive to specialize. Following Carey and Shugart (1995), each candidate must cultivate a personal vote by appealing to a distinct constituency. For many women MPs, female voters represent precisely that constituency, and this makes issue specialization in women's and welfare topics a rational electoral strategy rather than simply a reflection of personal preference. These three mechanisms are not mutually exclusive: socialization shapes initial issue orientations, electoral incentives reinforce and sustain them, and the threat of backlash penalizes deviation. Together they predict persistent gender differences in speech that are stable across careers—consistent with our finding that tenure does not erode women's distinctive topic choices.

However, a growing strand of literature is beginning to question whether these differences are permanent, with Castanho Silva et al. (2024) and Hargrave and Blumenau (2022) stating that these differences are disappearing, with women increasingly speaking same way as men in the parliament. This allows us to hypothesize that speech differences between male and female MPs are not necessarily constant but might be prone to changes over time and depend on the larger societal context. A further question our data can address is whether men respond to increasing female representation by specializing in different topics (Höhmnn, 2020).

The observation in the above cited studies that women and men are increasingly similar in their way of speaking might be due to several reasons. For example, due to a need to conform to the masculine way of communicating at the workplace (Castanho Silva et al., 2024), or because women no longer see the need to act like women in parliament (Blaxill and Beelen, 2016). This phenomenon might also be triggered by women having increasingly long parliamentary tenure. Indeed, previous research has shown that the longer a female MP serves, the less she represents women (Bailer et al., 2022). Whichever the reason, there seems to be a time-component to all these explanations, as speech norms change across time and context, and longer female tenures can only exist in contexts with a longer history of female political representation. We argue that this longer time dimension has not been tested so far with a dataset spanning the earlier years of female political representation.

Moreover, the contested concept of critical mass (Dahlerup, 1988; Mansbridge, 1999; Bratton, 2005; Beckwith and Cowell-Meyers, 2007; Childs and Krook, 2006)—according to which women's collective legislative behavior shifts once they pass a threshold share of parliamentary seats—also has an underexplored temporal dimension. The critical mass literature argues for a threshold range of 10–30%, with no single value having received scientific consensus, and with the usefulness of the concept itself having been questioned (Dahlerup, 2006b; Childs and Krook, 2006; Mansbridge, 1999). We operationalize critical mass in a data-driven way as the period when women's seat share in the Finnish Parliament crossed the 10–20% range in the span of a decade, which we observe empirically in the 1960s.

Despite this contested status, the intuition behind a threshold effect has a clear origin. The argument for why a numerical threshold should matter traces back to Kanter (1977), who argued that minorities below approximately 15% of a group function as tokens. This

² A separate question is whether greater women's representation translates into measurable welfare outcomes. Grier et al. (2025), using a cross-country event-study of 52 large sustained increases in female legislative representation during 1985–2023, find no significant effects on a broad set of well-being indicators (maternal mortality, female education, female labor force participation, female property rights, female civil liberties, and the Gender Development Index). On the other hand, Hessami and da Fonseca (2020) reviews the literature and identifies several studies finding positive causal effects of women's legislative presence on policy outcomes in developing countries, especially regarding education and health provision.

means that they are individually visible but collectively unable to shift group norms or build coalitions. Applying this to parliamentary politics, as women cross this or some other threshold—of the various thresholds proposed in the literature, 30% gained the most policy traction (Dahlerup, 2006b)—the token dynamic can be expected to weaken through several distinct mechanisms. Institutionally, a larger female contingent allows women MPs to form cross-party coalitions, secure committee positions on welfare-relevant portfolios, and accumulate the procedural influence needed to advance their priorities (Beckwith and Cowell-Meyers, 2007; Grey, 2002). Psychologically, reduced token pressure may increase willingness to speak distinctively rather than conform to masculine parliamentary norms (Kanter, 1977; Mendelberg et al., 2014). Electorally, a growing share of women MPs may normalize “women’s issues” as mainstream legislative topics rather than niche concerns, reducing the reputational risk of specializing in them (Carey and Shugart, 1995; Fulton et al., 2006). Socially, shifting norms inside the chamber may make it easier for women to speak on previously marginalized topics (Blumenau, 2019; Mendelberg et al., 2014).

Our evidence is best positioned to speak to the electoral and social mechanisms: we observe what women speak about and when those choices change relative to the critical mass transition, including for women already serving before the threshold was crossed. We cannot directly test the institutional mechanism, as we lack yearly committee assignment data, nor the psychological mechanism, as we do not observe individual confidence or risk tolerance. We therefore treat the critical mass transition as a contextual shift and remain agnostic about which specific mechanism drives the behavioral change we document.

We are able to test with our data how women who served during the transition to critical mass changed their speech behavior once the critical mass was met. This enables us to see to what extent the concept of critical mass merely changes the composition of the legislature by bringing in new MPs with new ideas or if the increased representation of women also induced already serving female MPs to change their parliamentary behavior.

3. Institutional context

The theoretical mechanisms outlined above, namely socialization, role congruity, and electoral incentives, all predict that speech differences will be sensitive to the share of women in parliament, and that the critical mass transition should be a particularly informative period to observe. Finland offers an unusually clean test of these predictions. Women have been elected to parliament since its founding in 1907, the proportional representation system has been stable throughout, and the data span the full trajectory from women’s novelty in parliament to near-parity—including the precise decades when the welfare state was constructed. This combination of institutional stability and long time horizon makes it possible to observe whether behavioral change around the critical mass transition reflects individual careers or collective context, and whether the patterns documented in more recent data from other countries hold in the earlier, formative period as well.

The Finnish parliament was founded in 1907 when it replaced the previous system where only the Four Estates of nobility, clergy, bourgeoisie, and peasants were represented. In Finland, women have been elected to parliament since its founding in 1907. Over the 20th century, their seat share rose from below 10% to nearly 50%—one of the steepest and most sustained increases in female descriptive representation in any democracy.

Finland has a multi-party political system and open-list proportional representation elections. There have been no significant changes to the election system during the existence of the Finnish Parliament (Ylisalo et al., 2012). Plenary sittings in the Finnish Parliament are usually organized four times a week. The role of the plenary sittings is mainly for politicians to communicate their views to the public and not necessarily to impact policies or other politicians’ views (Pekonen, 2011). In the early decades, parliamentary proceedings reached the public primarily through printed transcripts and newspaper reports. Radio broadcasting of sessions began selectively in the 1930s. Regular television coverage started in the 1960s–1970s, with Thursday question hours becoming a televised institution (Simola et al., 2025a; Nieminen et al., 2024). Since the 2000s, all plenary sessions have been webcast online. Importantly, our analysis does not require that speeches are the primary communication channel between MPs and constituents; rather, plenary speeches serve as a revealed-preference measure of what MPs choose to emphasize in a formal, public, and recorded setting. The Speaker of the Parliament decides who gets to speak. Committee memberships³ affect who gets to speak first or more easily on specific topics. Group speeches (*ryhmäpuheenvuoro*) are speeches delivered on behalf of a parliamentary group (*eduskuntaryhmä*), typically during major debates such as budget sessions or government statements, and represent the official position of the group rather than the individual MP. The parliamentary group, typically organized along party lines, decides who delivers these speeches. Based on discussions with Finnish MPs, it is also perceived to be easier for committee heads, as well as party leaders and long-term MPs, to speak.

Fig. 1 shows that women’s share of speeches is consistently below their share of MPs, suggesting selection may work against women. If so, our results may understate the extent to which women choose to emphasize different topics, as women who do speak may face additional selection pressures. We acknowledge this as a limitation.

Regarding external validity, Finland offers a uniquely long time series covering the entire trajectory from novelty to normalization of women’s political participation. The proportional representation system, present since the first elections, makes Finland comparable to other Nordic and Western European countries with similar institutions. Indeed, Fiva et al. (2026) apply the same methodology to Norway during 1981–2021 and find relatively similar results regarding the magnitude of speech differences between men and women. However, we acknowledge that the specific connection between women’s speech and welfare state formation is context-dependent: countries with different welfare state trajectories, gender quotas, or majoritarian electoral systems may exhibit different patterns.

³ We do not have yearly information about committee memberships but do have very imperfect data regarding committee memberships MPs held (but not when they held them). Fig. E1 in Online Appendix utilizes these.

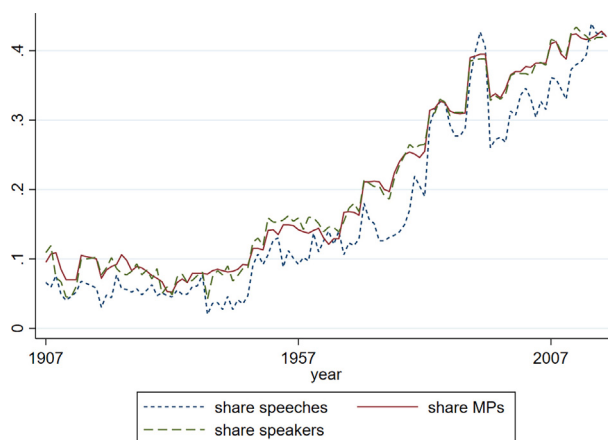


Fig. 1. Descriptive representation of women, 1907–2018. Notes. Figure shows women's share of all MPs, all speakers (who speak at least once), and all speeches.

We focus on parliamentary speeches rather than votes or other legislative behavior for two reasons. First, in the Finnish multi-party system with strong party discipline, roll-call votes are largely determined by party positions and reveal little about individual MPs' priorities. There is also previous evidence in the literature that using roll-call votes to measure, e.g., political polarization can be misleading because of party discipline (Schwarz et al., 2017). Speeches, by contrast, allow MPs to signal their personal priorities and communicate with constituents even when their votes are constrained by party lines. Second, the speech record covers the entire period since 1907, whereas systematic roll-call data are not available for the full period. We acknowledge that speeches are a less consequential form of legislative behavior than votes, bill sponsorship, or committee work: an MP may emphasize an issue without directly shaping policy. However, precisely because speeches serve a communicative function (Pekonen, 2011), they are well suited for measuring issue emphasis and substantive representation as defined by Pitkin (1967).

4. Empirical methodology

We use a dataset on parliamentary speeches in the Finnish parliament, which is available for 1907–2018. However, we focus mainly on the period 1948–2018 as the earlier years had too few women for our main estimator to be used. We use only years that have a minimum of 15 speakers per group. From 1948 onward, every year has at least 22 female speakers (mean 45.4, mean female share 29.8%). Eight pre-1948 years fall below this threshold. See Online Appendix Table O1 for yearly sample sizes and Fig. P1 for a robustness check using a lower threshold of 5 speakers, which results in very large CIs for those additional years that would then be included. Simola et al. (2025b) describes the dataset. The analysis dataset we use in our main analysis is on the speaker-session (year) level, and consists of phrase counts for different bigrams (two-word phrases). In addition, to analyze voter preferences, we use publicly available survey datasets from the Finnish Social Science Archive. These include *Finnish Voter Barometer 1973–1992*, *Finnish Voter Barometer 1995–2002* and *National Election Study 2003–2015* (see Suomen Gallup and TNS Gallup Oy (2008); Suomen Gallup and Helsingin yliopisto (2007); Vaalitutkimuskonsortio et al. (2017)).

To measure speech differences between men and women, we use a method introduced in Gentzkow et al. (2019). They applied this method to study partisanship in the United States. In our case, we utilize the *average partisanship* measure introduced in their paper, but instead of studying differences between parties, we use it to study differences between *women and men*, controlling for party affiliation. A similar approach was used by Fiva et al. (2026), who studied within-party group differences, but from a more contemporary perspective.

In short, the Gentzkow et al. (2019) measure applied to our case gives the probability that a neutral observer correctly deduces the gender of the MP after knowing just one phrase, i.e., bigram they spoke. We also show how the measure changes when we estimate it for knowing more than 1 bigram, but we do not have timestamps for the speeches and thus cannot estimate what, e.g., hearing 1 minute of speech would mean in terms of the number of phrases used. The Gentzkow et al. (2019) measure always gets values between 0.5 and 1, where 0.5 indicates no differences in speech, while the value of 1 would indicate that men and women use completely different phrases. Confidence intervals are calculated using subsampling (see Online Appendix for details). The method is validated by estimating a placebo series where the gender of the MPs is randomly drawn. This placebo should produce estimates close to but slightly above 0.5 due to finite-sample bias in the high-dimensional phrase space, as documented in previous studies using this method (Gentzkow et al., 2019; Fiva et al., 2026). The relevant benchmark for genuine speech differences is therefore whether the real estimates systematically exceed the placebo. They do: in our main period (1948–2018) the real series exceeds the placebo in every year, with a mean real-minus-placebo difference of 0.0037 (SD = 0.0010, min = 0.0013, max = 0.0058). The full year-by-year comparison of placebo and real estimates is reported in Online Appendix K (Table K1).

We explain below the Gentzkow et al. (2019) approach mathematically (see a somewhat more detailed explanation in Online Appendix K). Following Gentzkow et al. (2019) and Fiva et al. (2026), the phrase choice probabilities in the model of speech are

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

where u_{ijt} is

$$u_{ijt} = \alpha_{jt} + x'_{it}\gamma_{jt} + \phi_{jt}1_{i \in \{D_i=1\}} \quad (2)$$

In the equations above, the indicator D_{it} is a dummy variable taking the value 1 if the MP is a woman and 0 otherwise. Since we want to estimate intra-party group differences, and not just plain differences between the groups, we include party fixed effects in x , similar to Fiva et al. (2026). We do not observe committee memberships annually, so we cannot include committee fixed effects. The coefficient ϕ_{jt} represents the difference in how often phrase j is used in session t by the group we are interested in (i.e., for whom $D_{it} = 1$). Following Gentzkow et al. (2019) and Fiva et al. (2026), polarization at time t for covariates x is defined as

$$\pi_t(x) = \frac{1}{2}q^{D_i=1}(x)\rho_t(x) + \frac{1}{2}q^{D_i=0}(x)(1 - \rho_t(x)) \quad (3)$$

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

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

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

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

This average polarization measure describes how different speech is between the binary groups that are being compared.

Because the estimation of the multinomial logistic regression model described above would be computationally challenging, we follow Gentzkow et al. (2019) and estimate the model using the distributed multinomial regression method which approximates multinomial logit with Poisson regressions that can be run in parallel. As the speech data has a very large number of phrases that can be spoken, small sample bias is a serious problem in estimation. The measure we use, developed in Gentzkow et al. (2019) takes this into account by including a LASSO penalty to the estimated coefficients of interest. We use their model and R package *distrom* to estimate the model presented above with the LASSO penalty. Following Gentzkow et al. (2019) and Fiva et al. (2026), the estimator with the LASSO penalty is given by minimizing the negative log likelihood function for the factorial of J Poisson distributions, i.e., the sum of all J likelihoods:

$$\sum_j \left\{ \sum_t \sum_i [m_{it} \exp(\alpha_{jt} + x_{it}\gamma_{jt} + \phi_{jt}1_{i \in \{D_i=1\}}) - c_{ijt}(\alpha_{jt} + x_{it}\gamma_{jt} + \phi_{jt}1_{i \in \{D_i=1\}}) + \psi(|\alpha_{jt}| + \|\gamma_{jt}\|_1) + \lambda_j |\phi_{jt}|] \right\} \quad (6)$$

where $\lambda_j |\phi_{jt}|$ is the LASSO penalty imposed on the coefficients of interest. This penalty limits the small sample bias. The second penalty term, ψ , is a small fixed cost applied to the remaining parameters, i.e., the phrase-session intercepts α_{jt} and the covariate coefficients γ_{jt} . Gentzkow et al. (2019) employ it for numerical reasons: it stabilizes the optimization and improves convergence, at the cost of narrowing the range of admissible values for λ . We set $\psi = 10^{-5}$.

We note that while recent advances in text-as-data methods, particularly large language models (LLMs), have expanded the toolkit for analyzing political text (Benoit et al., 2026), we believe the Gentzkow et al. (2019) approach is well suited for our research question. Benoit et al. (2026) demonstrate that LLM ensembles can replicate expert survey estimates of party policy positions from manifestos, achieving correlations of 0.87–0.92 with expert benchmarks. First, their approach is designed for holistic position estimation, reading an entire document and scoring it on a policy dimension, rather than for measuring speech differences between groups at the phrase level. LLM-based approaches summarize documents holistically and cannot trace group differences to specific phrases. Second, our approach is exactly reproducible given the same data and code, whereas LLM outputs can vary across model versions and configurations. Third, the placebo series and subsampling confidence intervals provide well-understood statistical validation, for which comparable procedures in LLM-based text analysis are less established. The method follows an established lineage (Gentzkow et al., 2019; Fiva et al., 2026) and remains part of the standard toolkit for measuring speech differences between groups. However, we do utilize an LLM-based analysis in the topic analyses as a robustness check for our main topic analyses (see next subsection).

4.1. Topic analyses

While the Gentzkow et al. (2019) method is the main approach we use to analyze speech data, we also conduct some additional topic analyses described more thoroughly in the Online Appendix. Our main approach to defining topics follows the same manual classification procedure as Gentzkow et al. (2019). We first identify the most common phrases (bigrams) used distinctively by women and men in each year. Based on the keywords these bigrams contain, we classify them into topical categories. We then search the full vocabulary for additional bigrams containing topic-relevant stems. The result is 18 topic categories reflecting the major policy domains in Finnish parliamentary debate: healthcare, children, education, economy and finance, foreign policy, family, labor, elderly, regions, social policy, culture, agriculture, defense, crime, law and justice, alcohol, transportation, and women's issues. Each bigram is assigned to at most one topic, and many bigrams remain unclassified ("non-topic phrases"). These manually defined topics are then used as outcomes in some regressions described in the Results section.

In addition, we show in the Online Appendix some analyses where instead of using manually defined topics, we estimate topics using a structural topic model (STM) with unigrams and try to assess whether there are similar trends in the prevalence of female-favored topics in the STM results compared to results estimated using our manually defined topics. As a further robustness check,

we also performed an independent, data-driven topic classification using an LLM (Claude Opus 4.6): we collected all unique bigrams from the top-5 gender-polarizing phrases across all years (1948–2018) and classified them into inductively derived categories without reference to our manual scheme. This data-driven classification replicates the core finding: female-associated bigrams cluster into welfare-state topics, while male-associated bigrams cluster into economy and institutional topics. See Online Appendix R for full details and comparison.

5. Results

5.1. Women as a distinct parliamentary group

Our first research question asks whether gender-based speech differences change with critical mass, and if so in what direction. We address it in two steps: we first establish that persistent speech differences exist throughout the period, and then in [Section 5.3](#) examine whether those differences intensify around the critical mass transition.

[Fig. 1](#) shows the evolution of women's descriptive representation in the Parliament. It shows that the share of women MPs, speakers, and speeches was low (under 10 percent) until 1950, and it has been rising steadily since then. By 2019, the share of women in the Parliament was almost 50 percent. It also shows women's share of speeches is less than their share of MPs and speakers almost every year, suggesting that women MPs speak less than men MPs in the Parliament.

[Fig. 2](#) shows our main results, i.e., speech differences between men and women MPs, calculated using the [Gentzkow et al. \(2019\)](#) method. Panel A of [Fig. 2](#) shows our main estimates and confidence intervals for intra-party speech differences between men and women.⁴ In addition, the plot includes a placebo estimate. Because the share of women (see [Fig. 1](#)) was very small in the early 1900s, we are not able to estimate confidence intervals for speech differences during the early 1900s. Therefore, we omit the years before 1948 from this main analysis.

Looking at Panel A of [Fig. 2](#), we can clearly see significant speech differences between men and women during the whole period: the confidence intervals of the polarization series for gender do not overlap with those of the placebo estimate. Furthermore, it also seems that polarization by gender has increased slightly over time, as the estimate is higher in the 2010s compared to the 1950s, for example. The longest continuous increase in speech differences happened between the late 1940s and early 1970s. Overall, the size of the estimate varies between 0.504 and 0.507 over the long period. For comparison, [Fiva et al. \(2026\)](#) estimate speech differences by gender that vary between 0.503 and 0.505 in Norway during 1981–2021. Interestingly, using the same data and time period as we use, [Simola et al. \(2025a\)](#) report smaller estimates (0.501–0.506) for differences between left and right wing parties. That is, differences based on gender are larger than ideological differences.

It may be challenging to infer the magnitude of our findings on speech differences by gender based on the main estimate shown in Panel A, as that represents the probability of guessing the group correctly after hearing just one bigram. In order to show how the expected posterior of guessing the group correctly changes when the observer hears more words, we present the expected posterior for up to 100 phrases (during 3 selected years) in Panel B. For example, when the observer hears, e.g., 60 phrases, the probability of correctly guessing the gender would be around 75% in 2018 based on these results.

5.2. Topics, specialization and content

Having established that women speak differently than men throughout the period, we now turn to the content of those differences, specifically, whether they reflect distinct topic choices, distinct ways of discussing shared topics, or both.

Next, we examine whether the speech differences found between genders reflect different views within issues or specialization in specific issues. To do this, we use manually defined topics, constructed similarly to [Gentzkow et al. \(2019\)](#). Panel A of [Fig. 3](#) shows the speech differences between men and women if we limit the estimation to those phrases that belong to some of these manually defined topics. In the same figure, we also show speech differences estimated for those phrases that do not belong to any topics. The estimated overall speech differences for the “topic sample” are larger than the overall differences estimated in [Fig. 2](#), but the trend looks similar.

The results of the actual decomposition are presented in Panel B of [Fig. 3](#). These components use only the phrases that belong to topics. The decomposition suggests that both within-topic and between-topic components are sizable. The within-topic speech differences are larger than between-topic speech differences, but unlike the between-topic differences, the within-topic differences have a decreasing trend. It also seems that the between-topic component, i.e., specialization in topics, is driving the increase in gender speech differences observed during the 1950s–1970s. One striking pattern is that within-topic polarization consistently exceeds between-topic polarization. This means that gender differences in speech are driven more by how men and women phrase their arguments within shared topics than by which topics they choose to discuss. This pattern suggests that gender differences in parliamentary speech go beyond topic specialization to include distinct rhetorical styles and framing choices. What is noteworthy is that the 1950s–1970s were the decades of the active construction of the set of institutions comprising the Finnish welfare state.

We present additional topic analyses in the Online Appendix. Appendix A shows different groups' shares of speakers when a specific topic is discussed. Although our main speech-difference measure cannot be constructed for years before 1948 (due to insufficient female speakers), we report topic-share trends for all available years, including the pre-1948 period. Before the year when our observation period in the main analyses starts (1948), women did not strongly specialize in many topics, except in family, children,

⁴ Results without party fixed effects are reported in Online Appendix C.

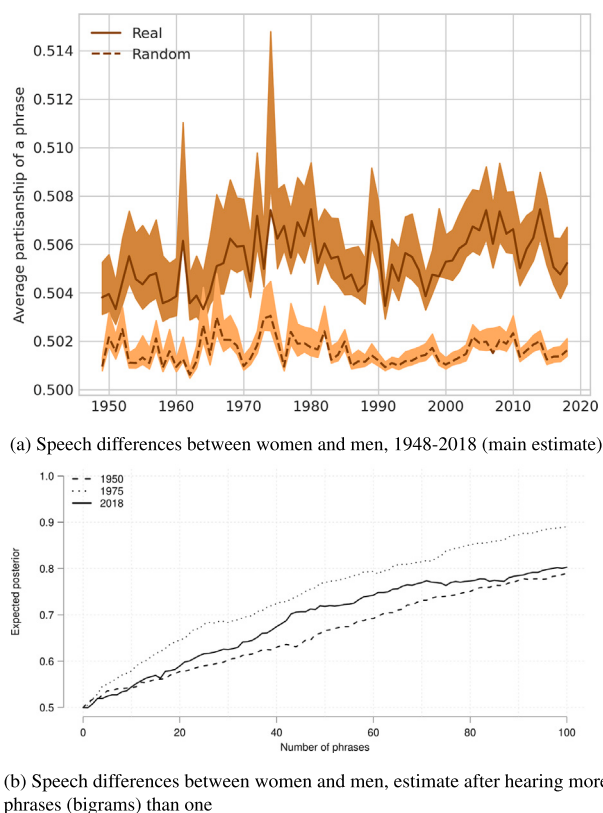


Fig. 2. Evolution of speech differences between men and women MPs: main results. *Notes.* Panel A presents speech differences between genders after a neutral observer knows 1 phrase (main estimate). Panel B shows how the probability of correctly guessing the gender changes after hearing more phrases. The estimation of speech differences includes year and party fixed effects. Confidence intervals have 80% nominal coverage as in Gentzkow et al. (2019). See Online Appendix B for 95% nominal coverage CIs. Speech differences are estimated for the years 1948-2018.

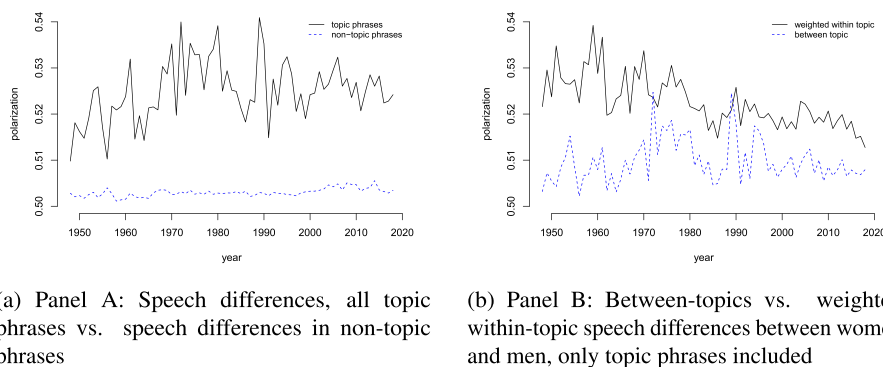


Fig. 3. Decomposition of between-topic and weighted within-topic speech differences between women and men. *Notes.* Panel A compares speech differences estimated using only phrases that belong to manually defined topics (“topic phrases”) with speech differences estimated using all remaining phrases (“non-topic phrases”). Panel B decomposes the topic-phrase speech differences into between-topic and within-topic components. Following Gentzkow et al. (2019), between-topic average partisanship is defined as the probability that a neutral observer would correctly assign a speaker’s gender after learning only which manually coded topic the speaker’s phrase belongs to. Within-topic partisanship measures the probability of correctly assigning gender when the speaker is restricted to using phrases from a single topic. The aggregate within-topic measure is the weighted mean of these topic-level values, with weights given by the relative frequency of each topic across all sessions. That within-topic polarization exceeds between-topic polarization indicates that gender differences in parliamentary speech are driven more by how men and women talk about shared topics (word choice and framing) than by which topics they choose to discuss.

and women’s issues. As the share of women in parliament increased beginning from the late 1940s, they became over-represented in policy areas such as healthcare, social policy, and education, while also reinforcing their earlier focus on topics like family, children, and women’s issues. This increased representation coincided with changes in parliamentary discussions: in absolute terms, issues women prioritized gained prominence only as more women entered parliament, and in relative terms, topics like healthcare

became more frequent. Appendix H details these topic mentions in absolute and relative terms. Regarding the role of tenure length in influencing topics, we find that topic shares of female MPs are not associated with tenure (see Online Appendix Fig. M1).

To sum up the results so far, women are overrepresented in parliamentary debates about education, health care, social policy, and family related issues, and this difference peaked during the era when the most significant expansions of redistributive institutions in these areas took place. This is consistent with the finding in Svaleryd (2009) that greater women's representation is associated with different public spending patterns, particularly in areas such as education and social services. Online Appendix Table Q1 presents the top gender-distinguishing phrases by decade with English translations, confirming these substantive patterns at the phrase level.

Figure L1 in the Online Appendix indicates that since 1973, public opinion surveys show women voters care about the same topics as women MPs, suggesting that women MPs were advocating for issues that they knew were valued among their electorate. However, opinion surveys reveal a long-term decline in topics women used to care about in earlier decades, with gender equality being the only distinctly female issue in later polls. One interpretation is that women voters entrusted women representatives with building the institutions that mattered to them; once these institutions were established, the number of distinctly female issues declined. In other words, this suggests that the growing importance of welfare state related topics such as healthcare, childcare and parental leave among female voters meant that they voted for women who advocated for these issues in the parliament but once these issues were achieved women no longer had to talk about these issues to gain votes. Accordingly, between topic differences had a growing trend until the 1970s after which they stabilized, as seen in Fig. 3.

5.3. Critical mass as a potential turning point

We now address both research questions jointly. The first asks whether speech differences shift around the critical mass transition; the second asks whether any such shift reflects individual career trajectories or a collective contextual change affecting all serving women MPs simultaneously.

Based on Panel A of the main results figure (Fig. 2), speech differences are nowadays higher than they were at the beginning of our observation period (1948). However, the increase in speech differences seems to have happened somewhere between the years 1960 and 1970, after which the level of speech differences between men and women has stayed the same on average. As can be seen from Fig. 1, the 1960s is also a decade when the seat share of women began to increase more strongly, surpassing 10–20% of seats and starting a clear upward trend toward near parity in the 2000s. This coincidence of rising speech differences and surpassing the 10–20% range is consistent with the prediction that women's collective legislative behavior shifts as their numbers move out of the token zone, and it is this transition period that we use as our empirical window for studying behavioral change.

Online Appendix Fig. I1 shows our previously estimated gender speech differences (after subtracting the placebo series) alongside the evolution of women's descriptive representation (seat share). The figure illustrates that gender speech differences increase during the period when women's seat share moves through the 10–20% range, and do not further increase after that—consistent with the interpretation that the behavioral shift coincides with the transition out of the token zone rather than with continued growth in descriptive representation thereafter.⁵

Regression results (shown in Online Appendix Table I1) and the results shown in Fig. I2 paint a similar picture, as they suggest that the association between women's seat share and gender speech differences is strong before the 1970s, but after women's seat share moved through the 10–20% range, speech differences seem to have reached a level that remains constant on average, even when women's descriptive representation increases substantially more during later decades. This plateau coincides with the peak of topic differences around the 1970s, after which the differences somewhat subside. One possible interpretation is that once the welfare state institutions women had collectively advocated for were gradually established, the shared cross-party agenda that had united women MPs weakened, reducing the between-topic differences observed in earlier decades. During this period, women MPs were disproportionately engaged in parliamentary debates on key welfare state reforms, including the national health insurance system (*sairausvakuutuslaki*, 1963), the comprehensive school reform (*peruskoulu-uudistus*, 1968), and the Day Care Act (*päivähoitolaki*, 1973), which established universal municipal childcare. Our phrase-level data confirm that female-associated bigrams during the 1960s–1970s include terms directly related to these reforms (e.g., *sosiaaliterveysministeriö* [Ministry of Social Affairs and Health], *last.päivähoito* [children's daycare]; see Appendix Table Q1). While we cannot establish that women MPs caused these reforms, their speech patterns show they were disproportionately engaged in debating them. After this period, growth in women's descriptive representation was not associated with further increases in gender differences in speeches.

The question of whether the sheer number of women MPs makes a difference in the quality of female representation is a contested one. Wängnerud (2009), in their review piece, conclude that sheer numbers of women are not enough to make substantive change unless women are also given concrete positions in which they can advocate for their preferred policies. However, in general, the more women there are in parliament, the more likely it is that a greater number of women also make it to positions of substantive power in the legislature. Mendelberg et al. (2014) show that in deliberative settings where interruptions are harmful, being a minority is associated with reduced influence over decision making. Applying this logic to the Finnish parliamentary context—where short speeches are given and speakers can support or challenge each other's contributions—a larger female contingent may have allowed women MPs to build momentum by reinforcing each other's arguments. In the absence of supportive MPs who share the objectives

⁵ As one might worry whether the speech changes observed during this transition period are driven by idiosyncratic contributions from a few prominent women MPs, we also estimate in Online Appendix Fig. N1 the polarization series when dropping the female MP who speaks the most. The increasing trend in speech differences before the 1970s remains, after which the level of speech differences does not change on average.

and experiences of a woman speaker, it is harder to build consensus and easier to be outnumbered by a majority male opposition. This suggests that the transition observed in the 1960s, when women's seat share moved through the 10–20% range, may have provided a sufficient base of supportive voices for women MPs to collectively advance issues that mattered to their female electorate.

To examine whether incumbents changed their parliamentary behavior during this transition period, we estimate OLS regressions using the same sample of women who served both before and after the 1960s transition. For this sample, we first estimate regressions of the following type (showing coefficients for female seat share in Table 1):

$$topicshare_{it} = \alpha + \beta \cdot femaleseatshare_t + \varepsilon_{i,t} \quad (7)$$

where $topicshare_{it}$ is the share of MP i 's speech in electoral term t that belongs to a given manually defined topic (e.g., healthcare, economy), measured as the count of bigrams in that topic divided by the MP's total bigram count. The variable $femaleseatshare_t$ is the proportion of parliamentary seats held by women in electoral term t . The unit of observation is a female MP–electoral term pair, and the sample is restricted to women who served in the parliament both before and after the critical mass was reached (i.e., continuously during the 1958–1970 period). The coefficient β captures the association between a one-percentage-point increase in women's seat share and the corresponding change in an MP's topic share (both measured in percentage points). We estimate separate regressions for each of the 18 topics.

We also show results from a regression that includes the placebo group from later years and estimate the regression using a model similar to a difference-in-differences. However, the idea is to compare the actual and placebo groups (see Table 2). In this model, we examine two separate groups of women MPs each of whom served three parliamentary terms at different points in time. The treatment indicator equals 1 for the group of women MPs during whose time in parliament the critical-mass threshold was exceeded. The post variable takes the value 1 for observations from the last electoral term for the respective group. The regression model is now:

$$topicshare_{it} = \alpha + \beta \cdot post_t + \gamma \cdot treat_i + \delta \cdot post_t \times treat_i + \varepsilon_{i,t} \quad (8)$$

where $post_t$ is a dummy equal to 1 for the last electoral term of the respective group, $treat_i$ is a dummy equal to 1 for the group of women during whose parliamentary service the critical-mass threshold was exceeded, and δ is the difference-in-differences coefficient of interest. The identifying assumption is that, absent the critical mass being reached, the treated and placebo groups would have experienced similar changes in topic shares across electoral terms.

We show results from model 1 in Table 1. When we compare female MPs who were in the parliament both before and after the largest shift in speech differences occurred, we show that those women began to talk more about healthcare and women's issues, and less about economy and finance topics (see Table 1). This suggests that moving through the 10–20% range may have led women MPs to devote more of their speech to some of the topics that are the most valued by women voters, suggesting that indeed the years 1950–1970 were a pivotal period in *how* and *about what* women started to speak. We show results from model 2 (which includes a placebo group) in Table 2. Those results show that the aforementioned findings—especially the increase in speaking about “women's issues” and less about economy and finance topics—are robust to including a placebo group.

We also estimate similar behavioral change analyses for male MPs, reported in the Online Appendix (Tables I4 and I5). Unlike our main Gentzkow et al. (2019) estimates, these analyses help shed light on whether the speech changes observed during the 1960s transition were driven by women, men, or both. The results suggest a pattern of gendered specialization: as women shifted toward women's issues and away from economy and finance topics, men shifted away from social policy—consistent with the specialization hypothesis documented by Höhmann (2020). Other changes for men are less robust, varying depending on whether a placebo group is included.

Table 1
Linear regression coefficients for female seat share by outcome.

Outcome	Coefficient	S.E.	<i>t</i>	<i>p</i>
Agriculture and forestry	−0.104	0.163	−0.64	0.53
Alcohol	0.962	0.434	2.22	0.04**
Children	−0.070	0.644	−0.11	0.91
Crime	0.109	0.094	1.16	0.26
Culture	0.038	0.293	0.13	0.90
Defence	0.033	0.092	0.35	0.73
Economy and finance	−2.951	1.538	−1.92	0.07*
Education	−0.917	1.690	−0.54	0.59
Elderly	−0.338	0.877	−0.39	0.70
Family	0.707	0.525	1.34	0.19
Foreign policy	−0.314	0.783	−0.40	0.69
Healthcare	1.288	0.514	2.50	0.02**
Labor	1.034	1.035	1.00	0.33
Law and justice	0.001	0.053	0.01	0.99
Regions	0.489	0.538	0.91	0.37
Social policy	−0.452	0.361	−1.25	0.22
Transportation	−0.007	0.007	−0.92	0.37
Women	0.492	0.239	2.06	0.05*

Notes. Table shows regression estimates where the outcomes are the share of different topics out of all speeches. The explanatory variable is female seat share. Signif. codes: **** 0.01 *** 0.05 ** 0.1.

Table 2
Linear regression coefficients for post:treatment by outcome.

Outcome	Coefficient	S.E.	<i>t</i>	<i>p</i>
Agriculture and Forestry	−0.002	0.012	−0.19	0.85
Alcohol	0.013	0.019	0.69	0.49
Children	0.045	0.038	1.19	0.24
Crime	0.008	0.010	0.84	0.40
Culture	−0.001	0.010	−0.14	0.88
Defence	0.000	0.004	0.05	0.96
Economy and Finance	−0.210	0.067	−3.13	0.00***
Education	0.033	0.074	0.45	0.66
Elderly	0.027	0.029	0.94	0.35
Family	0.051	0.021	2.43	0.02**
Foreign Policy	−0.031	0.040	−0.77	0.45
Healthcare	0.042	0.027	1.56	0.13
Labor	−0.007	0.051	−0.13	0.90
Law and Justice	−0.003	0.003	−1.00	0.32
Regions	0.006	0.037	0.15	0.88
Social Policy	−0.003	0.016	−0.21	0.83
Transportation	0.000	0.003	0.12	0.90
Women	0.033	0.014	2.38	0.02**

Notes. Table shows regression coefficients for post:treatment. The outcomes are shares of different topics out of all speeches. Signif. codes: '***' 0.01 '**' 0.05 '*' 0.1.

We take the results shown in Table 1 to mean that the years 1950–1970 did not only bring a numerical increase of women in the parliament, but that this increase coincided with a period in which women voters' preferences were increasingly visible. The decisive increase in women MPs in parliament in the 1960s possibly meant that there was a critical number of supportive voices backing women MPs speeches. According to the logic presented in Wängnerud (2009), the sheer numbers plausibly translated into increased committee membership on topics that women voters and politicians advocated for. These results are consistent with the view that descriptive representation coincided with changes in substantive representation. The need for welfare state institutions was so universal among women that the momentum gained between 1950 and 1970 clearly stands out as a period in which gendered speech and topic differences and the increase in women MPs coincide in a way that later is no longer the case. This could potentially be interpreted as how shared experiences of women can create momentum and meaningful changes in institutions.

We compare trends in selection based on background variables during the 1960s transition to those in later periods in the Online Appendix. Table I2 shows how selection changed during the transition period while Table I3 shows changes in selection in later years. These indicate that changes in selection in terms of education level were similar in both periods, but other types of selection were not, meaning that changes in the composition of MPs could drive some of the observed patterns and should be kept in mind when interpreting the results. We cannot disentangle the simultaneous factors of the 1960s transition and the creation of welfare state institutions, and the direction of causality cannot be established. Nevertheless, the evidence is consistent with the interpretation that the years 1950–1970 represented a formative period for women's parliamentary representation, during which women developed a distinctive legislative voice, their presence increased, and their speech reflected the policy priorities of their female electorate.

These patterns plateaued after the 1970s but remained persistent and detectable.

5.4. Robustness: data-driven topic classification using LLM

The following two sub-sections assess whether the topic patterns documented above are robust to the choice of classification scheme.

As discussed in Section 4.1, we constructed an independent data-driven topic classification by collecting all 291 unique bigrams that appeared among the top-5 gender-polarizing phrases in any year during 1948–2018 and classifying them into 18 inductively derived categories without reference to our manual scheme, using an LLM (Claude Opus 4.6). We use this alternative classification to replicate all topic-level analyses reported above. Full details and results are in Online Appendix sections R–V.

A clarification on the two schemes is in order: both are built from the same starting point, i.e., from the bigrams that appeared among the top-5 most gender-polarizing phrases in any year during 1948–2018. For the manual scheme, we follow Gentzkow et al. (2019): we assign these top bigrams to 18 categories based on their keywords, expand each category via stem-matching across the full bigram vocabulary, and review the assignments manually, which yields the 5667-bigram scheme. For the data-driven scheme, we skip the vocabulary-expansion step because the LLM (Claude Opus 4.6) was run on a local computer and cannot process the full bag-of-words vocabulary (analyses using the whole vocabulary are run on a remote computing cluster). In this alternative approach, the LLM classifies the 291 “top-5 bigrams” directly into 18 topics and names the topics without input from us.

The within/between decomposition using data-driven topics (Online Appendix R, Fig. R1) reproduces the core pattern from Fig. 3: within-topic polarization exceeds between-topic polarization in the majority of years, and both components exhibit the same decreasing trend observed under the manual classification. The decomposition is noisier, as expected given the smaller vocabulary (291 vs. 5667 bigrams), but the qualitative conclusions remain unchanged.

Table S1 replicates the OLS topic regressions from Table 1. Social and healthcare (coefficient 1.918, $p < 0.01$) and gender equality (0.827, $p < 0.05$) are the topics most strongly associated with female seat share, while economy-related topics carry negative but

insignificant coefficients, mirroring the pattern found with manual topics. The DiD specification (Table T1, replicating Table 2) confirms behavioral change for the same topics: social and healthcare (0.113, $p < 0.05$) and gender equality (0.043, $p < 0.05$). The male placebo (Table V1) shows no significant effects at the 5% level for any data-driven topic, consistent with the manual-topic results.

Finally, Fig. U1 replicates the topic speaker share plots from Online Appendix A and confirms that women are overrepresented among speakers on welfare-state topics (healthcare, children and daycare, education, family) and are represented roughly proportionally on economy and foreign policy topics. Taken together, these results indicate that our findings are robust to the choice of topic classification scheme.

5.5. Additional topic analyses using a structural topic model

We also present additional topic analyses using a different method, a structural topic model, in Online Appendix F. Those results show that topic 11, “Healthcare, family, children” starts to grow after the 1960s, consistent with the results we found using our main topic classification method. That topic is also the most clearly “female” topic according to Online Appendix Fig. F3 which shows which topics are common for female MPs and male MPs.

6. Conclusions

This paper asks two questions about the long-run evolution of gendered parliamentary speech. First, do gender-based speech differences change with the critical mass transition, and if so in what direction? Second, do any such changes reflect individual career trajectories or collective contextual shifts—that is, do women change as individuals over their careers, or does the context change around them?

Our seven-decade overview of the Finnish Parliament yields clear answers to both. On the first question, speech differences not only exist throughout the period but increase as women’s seat share surpasses the 10–20% critical mass threshold in the 1960s, and plateau thereafter. On the second, the shift is collective rather than individual: women who were already serving as MPs before the threshold was crossed also changed their speech during this period, and tenure length is not associated with the composition of topics women discuss. This points to a contextual rather than a career explanation. Taken together, the finding that already-serving women changed their speech, and the finding that career progression itself does not explain topic changes, rule out both the compositional and the individual-career explanations, leaving the contextual account as the most coherent interpretation of the pattern.

Confirming a long line of research (Castanho Silva et al., 2024; Hargrave and Blumenau, 2022; Dietrich et al., 2019; Yu, 2013), we find that women have always spoken less and differently than men. However, as the traditional critical mass of women’s descriptive representation was surpassed, these differences increased, and women also began to speak more and speak increasingly about different topics from men. This difference later plateaued and while differences still exist with regard to the choice of words (within topic differences), the between-topic differences in our times are less important than in the 1950s–1970s. This is relevant because we begin to note a decrease in differences for the same years as most of the text-as-data papers about gendered speech cover in the Western world. In other words, the decades that we know most about regarding speech differences between men and women form part of an era of decline in gendered speech differences, compared to the formative years covering 1950–1970, during which women found their distinctive political agency and showed clear differences not only in the way they spoke but also in *what* they chose to talk about. Therefore the formerly more pronounced differences have so far gone unnoticed, as has the role women’s speech may have played in the creation of the welfare state.

By shifting the perspective backward in history, we find that in the early years of female political representation women not only spoke differently than men, but they also specialized in topics that are traditional women’s issues. As their numbers increased, women expanded their topics to cover education, social policy, and health care, which suggests that women not only had an interest in welfare state institutions, but their increased presence in parliament coincided with the expansion of such institutions. This is also in line with the topics female voters cared about in the same period. As the issues were gradually resolved and female voters started to shift their focus to the same issues as male MPs, women’s distinct policy topics also declined. This is in line with the findings of Kläy et al. (2025) who find that policy responsiveness between men and women is only distinct when social policy is at stake. This seems to have created a feedback loop in which a) women specialized in topics mostly related to redistribution b) prioritized the same topics as their women voters and c) increased in numbers in the parliament. It would be interesting to explore in future research in other countries if the establishment of different welfare-state institutions coincides with specialization between men and women and if the observed decline in differences between female and male speeches coincides with enacted institutional reforms, as in Finland.

Once the institutional reforms were achieved, the differences seem to have subsided, suggesting that the speech differences are mostly driven by different specializations during topics in 1950–1970. The subsequent decrease, which has continued according to recent research, may be attributed either to women career politicians who have ceased to see themselves as representatives of women but rather as representatives of their parties (Bailer et al., 2022), or to a diminishing pressure on female politicians to conform to stereotypically ‘feminine’ styles in recent years (Blumenau, 2019). In addition to these explanations, we propose a new path to investigate the sources of increasing or decreasing differences over time: institutional demands and reforms that unite women MPs across parties. Our data suggest that once a critical number of women entered parliament, there was enough support to collectively advocate for the institutions women voters cared about. And once these were gradually accomplished there was a notable decline in female topics. The 1950s–1970s simultaneously saw the critical mass transition, the expansion of the welfare state, shifts in party competition, and changes in the media environment and public opinion, and our descriptive design cannot disentangle these concurrent forces. However, most existing research does not cover the years when crucial welfare state institutions were created after

the Second World War, so, while our evidence here falls short of establishing causality and remains largely descriptive, the correlation is interesting enough to invite future studies to examine more closely the role institutional demands and reforms play in explaining the ebb and flow of speech differences between men and women over time.

CRediT authorship contribution statement

Jeremias Nieminen: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ilona Lahdelma:** Writing – review & editing, Writing – original draft. **Salla Simola:** Methodology, Data curation, Conceptualization. **Janne Tukiainen:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Previous versions

A working paper version of this paper has been circulated as Aboa Center for Economics Discussion Paper No. 161.

Funding statement

Tukiainen: 101045239—INTRAPOL—ERC-2021-COG. Views and opinions expressed are only those of the authors, however, and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

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

Appendix A. Supplementary data

Supplementary data for this article can be found online at doi:[10.1016/j.ejpoleco.2026.102895](https://doi.org/10.1016/j.ejpoleco.2026.102895).

Data availability

Main datasets will be made available upon publication, but voter surveys from the Finnish Social Science Archive cannot be published.

References

- Adams, J., Bracken, D., Gidron, N., Horne, W., O'Brien, D.Z., Senk, K., 2023. Can't we all just get along? How women MPs can ameliorate affective polarization in Western publics. *Am. Polit. Sci. Rev.* 117 (1), 318–324.
- Adams, J.F., Merrill III, S., Grofman, B., 2005. *A Unified Theory of Party Competition: A Cross-National Analysis Integrating Spatial and Behavioral Factors*. Cambridge University Press, Cambridge.
- Bäck, H., Debus, M., 8 2019. When do women speak? A comparative analysis of the role of gender in legislative debates. *Polit. Stud.* 67 (3), 576–596.
- Bailer, S., Breunig, C., Giger, N., Wüst, A.M., 2022. The diminishing value of representing the disadvantaged: between group representation and individual career paths. *Br. J. Polit. Sci.* 52 (2), 535–552.
- Beckwith, K., Cowell-Meyers, K., 2007. Sheer numbers: critical representation thresholds and women's political representation. *Perspect. Polit.* 5 (3), 553–565.
- Benoit, K., De Marchi, S., Laver, C., Laver, M., Ma, J., 2026. Using large language models to analyze political texts through natural language understanding. *Am. J. Political Sci.* (Forthcoming).
- Blaxill, L., Beelen, K., 07 2016. A feminized language of democracy? The representation of women at Westminster since 1945. *Mod. Brit. Hist.* 27 (3), 412–449.
- Blumenau, J., 2019. The effects of female leadership on women's voice in political debate. *Br. J. Polit. Sci.* 51 (2), 750–771.
- Bratton, K.A., 2005. Critical mass theory revisited: the behavior and success of token women in state legislatures. *Pol. Gen.* 1 (1), 97–125.
- Caiazza, A., 2009. Does women's representation in elected office lead to women-friendly policy? Analysis of state-level data. *Women Polit.* 26 (1), 35–70.
- Carey, J.M., Shugart, M.S., 1995. Incentives to cultivate a personal vote: a rank ordering of electoral formulas. *Elect. Stud.* 14 (4), 417–439.
- Castanho Silva, B., Pullan, D., Wäckerle, J., 2024. Blending in or standing out? Gendered political communication in 24 democracies. *Am. J. Polit. Sci.*, 1–16.
- Catalano, A., 2009. Women acting for women? An analysis of gender and debate participation in the British house of commons 2005–2007. *Pol. Gen.* 5 (1), 45–68.
- Celis, K., Childs, S., Kantola, J., Krook, M.L., 2008. Rethinking women's substantive representation. *Representation* 44 (2), 99–110.
- Chattopadhyay, R., Duflo, E., 2004. Women as policy makers: evidence from a randomized policy experiment in India. *Econometrica* 72 (5), 1409–1443.
- Childs, S., Krook, M.L., 2006. Should feminists give up on critical mass? A contingent Yes. *Pol. Gen.* 2 (4), 522–530.
- Cowell-Meyers, K., Langbein, L., 2009. Linking women's descriptive and substantive representation in the United States. *Polit. Gen.* 5 (4), 491–518.
- Dahlerup, D., 1988. From a small to a large minority: women in scandinavian politics. *Scand. Polit. Stud.* 11 (4), 275–298.
- Dahlerup, D., 2006a. Do women represent women? Rethinking the "critical mass" debate. *Pol. Gen.* 2 (4), 491–530.
- Dahlerup, D., 2006b. The story of the theory of critical mass. *Polit. Gen.* 2 (4), 511–522.
- Dietrich, B.J., Hayes, M., O'Brien, D.Z., 2019. Pitch perfect: vocal pitch and the emotional intensity of congressional speech. *Am. Polit. Sci. Rev.* 113 (4), 941–962.
- Eagly, A.H., Karau, S.J., 2002. Role congruity theory of prejudice toward female leaders. *Psychol. Rev.* 109 (3), 573–598.
- Fernandes, J.M., Lopes da Fonseca, M., Won, M., 2023. Closing the gender gap in legislative debates: the role of gender quotas. *Polit. Behav.* 45 (3), 897–921.

- Fiva, J., Nedregård, O., Øien, H., 2026. Group identities and parliamentary debates. Forthcoming, *The Journal of Politics*.
- Fulton, S.A., Maestas, C.D., Maisel, L.S., Stone, W.J., 2006. The sense of a woman: gender, ambition, and the decision to run for congress. *Polit. Res. Q.* 59 (1), 143–152.
- Gentzkow, M., Shapiro, J.M., Taddy, M., 2019. Measuring group differences in high-dimensional choices: method and application to congressional speech. *Econometrica* 87 (4), 1307–1340.
- Grey, S., 2002. Does size matter? Critical mass and New Zealand's women MPs. *Parliam. Aff.* 55, 19–29.
- Grier, R., Grier, K., Muhoza, F., 2025. The effect of increased women's legislative representation on women's well-being. *Eur. J. Polit. Econ.* 89, 102735.
- Hargrave, L., Blumenau, J., 2022. No longer conforming to stereotypes? Gender, political style, and parliamentary debate in the UK *. *Br. J. Polit. Sci.* 52 (4), 1584–1601.
- Herrnson, P.S., Celeste, J., Atiya, L., Stokes, K., 2003. Women running as women: candidate gender, campaign issues, and voter-targeting strategies. *J. Polit.* 65 (1), 244–255.
- Hessami, Z., da Fonseca, M.L., 2020. Female political representation and substantive effects on policies: a literature review. *Eur. J. Polit. Econ.* 63, 101896.
- Höhmman, D., 2019. When do female MPs represent women's interests? Electoral systems and the legislative behavior of women. *Polit. Res. Q.* 73 (4), 834–847.
- Höhmman, D., 2020. When do men represent women's interests in parliament? How the presence of women in parliament affects the legislative behavior of Male politicians. *Swiss Polit. Sci. Rev.* 26 (1), 31–50.
- Itzkovitch-Malka, R., Oshri, O., 2024. The weight on her shoulders: marginalization of women legislators in parliaments and substantive representation of women. *Br. J. Polit. Sci.* 54 (4), 1340–1361.
- Kanter, R.M., 1977. Some effects of proportions on group life: skewed sex ratios and responses to token women. *Am. J. Sociol.* 82 (5), 965–990.
- Kläy, Y., Eichenberger, R., Portmann, M., Stadelmann, D., 2025. Substantive representation of women: empirical evidence. *Br. J. Polit. Sci.* 55, e32.
- Lowande, K., Bishin, B., Butler, D., Coronel, J., Macdonald, J., Merolla, J., Newman, B., Shah, S., 2019. Descriptive and substantive representation in congress: evidence from 80,000 congressional inquiries. *Am. J. Polit. Sci.* 63 (3), 644–659.
- Mansbridge, J., 1999. Should blacks represent blacks and women represent women? A contingent yes. *J. Polit.* 61 (3), 628–657.
- Mendelberg, T., Karpowitz, C.F., Goedert, N., 2014. Does descriptive representation facilitate women's distinctive voice? How gender composition and decision rules affect deliberation. *Am. J. Polit. Sci.* 58 (2), 291–306.
- Mendelberg, T., Karpowitz, C.F., Oliphant, J.B., 2014. Gender inequality in deliberation: unpacking the black box of interaction. *Perspect. Polit.* 12 (1), 18–44.
- Nieminen, J., Simola, S., Tukiainen, J., 2024. Effects of increased transparency on political divides and MP behavior: evidence from televised question hours in the Finnish Parliament. *Legis. Stud. Q.* 49 (3), 585–616.
- Norris, P., Lovenduski, J., 1995. *Political Recruitment: Gender, Race and Class in the British Parliament*. Cambridge University Press, Cambridge.
- Pekonen, K., 2011. *Puhe Eduskunnassa*. Vastapaino.
- Pitkin, H., 1967. *The Concept of Representation*. University of California Press, Berkeley.
- Schwarz, D., Traber, D., Benoit, K., 2017. Estimating intra-party preferences: comparing speeches to votes. *Polit. Sci. Res. Methods* 5 (2), 379–396.
- Shim, J., 10 2022. Substantive representation of women and policy-vote trade-offs: does supporting women's issue bills decrease a legislator's chance of reelection? *J. Legis. Stud.* 28 (4), 533–553.
- Simola, S., Nieminen, J., Tukiainen, J., 2025a. A century of partisanship in Finnish political speech. *J. Hist. Polit. Econ.* 5 (2), 239–265.
- Simola, S., Nieminen, J., Tukiainen, J., 2025b. Finnish parliamentary speeches dataset. *Sci. Data* 12, 1063.
- Suomen Gallup and Helsingin yliopisto, 2007. Puolueiden ajankohtaistutkimus 1973-1990: yhdistetty aineisto [sähköinen tietoaaineisto]. Versio 1.0 (2007-08-16). Yhteiskuntatieteellinen tietoaarkisto [jakaja].
- Suomen Gallup, TNS Gallup Oy, 2008. Puolueiden ajankohtaistutkimus 1992-2005: yhdistetty aineisto [sähköinen tietoaaineisto]. Versio 1.1 (2008-01-30). Yhteiskuntatieteellinen tietoaarkisto [jakaja].
- Svaleryd, H., 2009. Women's representation and public spending. *Eur. J. Polit. Econ.* 25 (2), 186–198.
- Swers, M.L., 1998. Are women more likely to vote for women's issue bills than their Male colleagues? *Legis. Stud. Q.* 23 (3), 435–448.
- Thomas, S., Welch, S., 1991. The impact of gender on activities and priorities of state legislators. *West. Polit. Q.* 44 (2), 445–456.
- Vaalitutkimuskonsortio, Karvonen, L., Paloheimo, H., Borg, S., Grönlund, K., Kestilä-Kekkonen, E., 2017. Eduskuntavaalitutkimukset 2003–2015: yhdistetty aineisto [sähköinen tietoaaineisto]. Versio 1.0 (2017-11-03). Yhteiskuntatieteellinen tietoaarkisto [jakaja].
- Wäckerle, J., Castanho Silva, B., 2023. Distinctive voices: political speech, rhetoric, and the substantive representation of women in European parliaments. *Legis. Stud. Q.* 48 (4), 797–831.
- Wängnerud, L., 2009. Women in parliaments: descriptive and substantive representation. *Annu. Rev. Polit. Sci.* 12, 51–69.
- Ylisalo, J., Maunula, M., Helin, J., 2012. Vaalijärjestelmämuutoksen kivinen tie: suomen vaalijärjestelmän kehitys suuresta eduskuntaudistuksesta nykypäivään. *Politiikka* 54 (4), 249–268.
- Yu, B., 2013. Language and gender in congressional speech. *Lit. Linguist. Comput.* 29 (1), 118–132.