Kevin Foley

December 3, 2024

Rhetoric and Representation: Party Discipline, Electoral Competitiveness and Rhetorical Alignment with Voter Preferences

Abstract

Despite being central to theories of how democracies function, the relationship between voters and elected representatives is poorly understood, particularly outside of voter participation in regular elections. Studies of political representation and responsiveness seek to understand how the actions of elected officials reflect the preferences of their voters. Departing from existing literature that focuses on policy output, legislative voting or ideological orientation, this paper connects the policy preferences of tens of thousands of individual voters with the public statements of hundreds of elected officials over three decades. It does this by using large language models as "few-shot" classifiers, and, with these tools, building a novel data set that links survey questions to speeches that endorse the same view as (one side of) the question. The paper concludes with an analysis of social and structural factors that predict voter aligment or misalignment with the political speech of their representatives, and finds that the dynamics of national presidential campaigns and national party organization play a large and not well-understood role in the representation of voters by their elected representatives.

1. Introduction

Language is a powerful tool in politics, but its role in understanding political outcomes and political behavior is still not well understood. Political strategy can drive rhetoric, and dramatic shifts in political language can play a role both in short term electoral success and long-term development of party's ideological positioning. This paper takes advantage of increasingly capable language modeling tools to explore the relationship between public attitudes and political rhetoric on a large scale. It does this by linking the views expressed by voters in public opinion surveys to the things politicians who represent them say in floor speeches in the House and Senate floor speeches. Methodologically, the paper combines few-shot prompting techniques (Brown et al. 2020) and recent advances in efficient computation (Kwon et al. 2023) to construct a pair of highly specific document classification models and applies these models at scale to tag hundreds of thousands of Congressional speech transcripts with the identifiers of recent public opinion survey questions; having done this, the second half of the paper analyzes social and structural factors associated with higher and lower degrees of alignment between voter opinion and political rhetoric.

Substantively, this analysis finds that the strongest predictor of a rehetorical representation is party membership. Specifically, in comparison with a Democratic voter represented by a local Democratic official from the same state (in the case of senators) or congressional district (in the case of representatives), a Republican voter represented by a Republican official is about 26 percent less less likely to see their beliefs reflected in House and Senate floor speeches by officials representing their district, and 13 percent less likely even when the representative speaking is a Republican. The second largest predictor of alignment is state-level competitiveness in presidential elections, suggesting either that party interests in presidential races may condition political language in ways that local voters find unpalatable, or that the saturation of swing state media markets with campaign advertisements plays a role in how voters relate to their own representatives. These two effects are an order of magnitude larger than the next-most substantial effects identified in this analysis, namely strength of partisanship education level. Taken together, these results paint a picture in which political language, while in general responsive to the interests of voters across many backgrounds, may also be influenced by structural features associated with national party organization and competition at the expense of some voters who do not see their views reflected in the things their elected officials say.

2. Data and model inference

This study relies on two external data sets. The first is a collection of documents downloaded and parsed from the Congressional Record consisting of 418,350 Speeches House and Senate floor speeches made between 1994 and 2022. Of these, the inference process described below labeled 158,096 (37.8%) unique speeches as relevant to at least one issue question from the most recent American National Election Study (ANES) prior to or during the same year as the speech was given. Speeches are timestamped with year, month and day and associated with a unique identifier for each member from the Congressional Biographical Directory (bioguide.congress.gov).

Public opinion data is taken from the American National Election Study's cumulative data file, which contains a large number of questions, many of which have been repeated over decades, going back to 1948. This study focused on questions from the "issues" section of the study, consisting of 50 broadly worded questions on a range of social and political issues. Variable VCF9223 from the ANES cumulative file, for example, tracks responses to the question "How likely is it that recent immigration levels will take jobs away from people already here?" A full list of variables used in this study will be included in an appendix.

Additionally, the DIME dataset (Bonica 2023) which associates FEC data on campaign financing and district-level democratic vote share in the most recent election with biographical identifiers for members of Congress, was used for the downstream analysis but was not used in the process of joining public opinion data to Congressional speeches.

Preparing the data

In order to facilitate comparison to the speeches of public officials, survey questions were rephrased into declarative "propositions" intended to capture the substance of the question and relable survey responses as either agreeing with the proposition, disagreeing with the proposition (neither / don't know / missing response codes were dropped from the data). Figure 1 provides some examples of this mapping:

Variable	**Question**	**Proposition**	**Response codes**	**Mapping**
VCF0879a	Do you think the number of immigrants from foreign countries who are permitted to come to the United States to live should be increased a little, increased a lot, decreased a little, decreased a lot, or left the same as it is now?	The number of immigrants from foreign countries who are permitted to come to the United States to live should be increased.	1: Increased 3: Same as now 5: Decreased 8: DK 9: NA; no Post IW	1. agree 3. null 5. disagree 8. null 9. null
VCF0809	Some people feel that the government in Washington should see to it that every person has a job and a good standard of living. Others think the government should just let each person get ahead on his/their own. Where would you place yourself on this scale, or haven't you thought much about this?	The government should ensure that every person has a job and a good standard of living.	1: Government see to job and good standard of living ... 7: Government let each person get ahead on his own	1. agree 2. agree 3. agree 4. null 5. disagree 6. disagree 7. disagree
VCF0867	Some people say that because of past discrimination blacks should be given preference in hiring and promotion. Others say that such preference in hiring and promotion of blacks is wrong because it gives blacks advantages they haven't earned. What about your opinion– are you for or against preferential hiring and promotion of blacks?	There should be preferential hiring and promotion of blacks to correct past discrimination.	1: For 5: Against 8: DK; 1990-1994: refused; 1996 and later: other 9: NA; Form A (1986); form B (1990); no Post IW	1. agree 2. disagree 8. null 9. null

VCF9236	Do you favor or oppose the death penalty for persons convicted of murder?	The death penalty should be applied to persons convicted of murder.	1: Favor, 2: Oppose -8: DK; depends (VOL) -9: RF; NA; no post data INAP: Inap. question not used	1: agree 2: disagree -8: null -9: null INAP: null

Figure 1 - Examples of survey question rephrasing

Model inference

Once mapped, all available questions from the most recent ANES study (looking back up to three years, but in most cases less than two years) were paired with each of the survey statements. This results in a very large number of direct comparisons between survey questions (rephrased as propositions) and floor speeches. In total, there were 10,079,931 such comparisons. In order to ensure high quality evaluations of relatedness while controlling cost and inference time, inference took place in two stages, a relevance stage and a position matching stage. In each stage, an iterative process was used to tune prompts for the classification task that yielded the best results on a test set of document / variable pairs. Both tasks used a multi-shot prompting approach (Brown et al. 2020) following the basic format of the multiple choice selection task used in the MMLU evaluation (Hendrycks et al. 2021), and prompts were long (2,767 tokens with 8 shots for the relevance task and 6,004 tokens with 13 shots for the matching task).

Inference was accelerated using paged attention prefix caching (Kwon 2023) and by carefully sorting the order of the inference runs to ensure that the attention maps for prompts and speeches were only calculated once and then reused in future rounds. Figure 2 below demonstrates how the caching approach was used to dramatically reduce compute costs. Thus, the total amount of inference required was far less than what would be required if the texts were sent to an API charging by the token. Nevertheless, cost and energy use are still factors to keep in mind for this kind of analysis. Overall, the inference used 250 hours of running time on an Nvidia A10G GPU, which costs a minimum of around $100 at current prices for spot instances and requires about the same energy use as keeping a refrigerator plugged in for 10 days.[1]

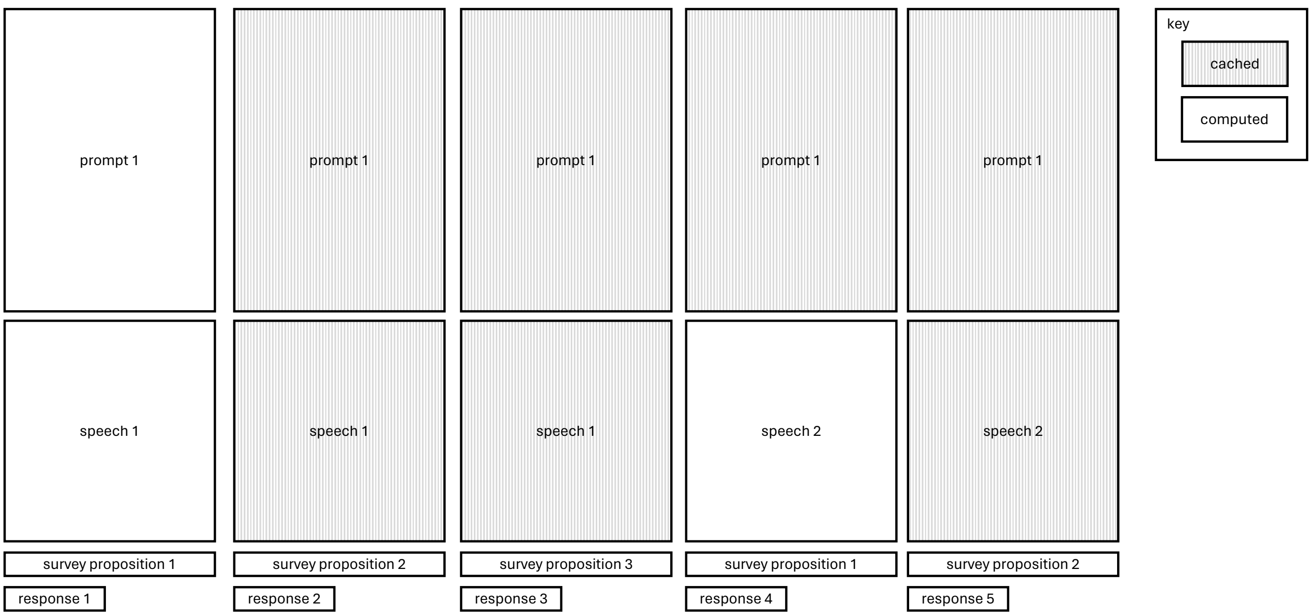

Figure 2 - Cached inference with paged attention allows very long prompts to be used without repeating past computations. The long prompt, around 7,000 tokens in the second stage, was only computed once in each batch of 4,000 records; speeches, which can also be very long, were only computed once despite being matched to proposition sentences from multiple survey variables. This scheme was followed in both stages of llm inference.

Speech to survey variable relevance coding

[1] This cost estimate is just rough guide, and probably understimates the real cost due to model time-to-launch latency, installing Python packages on a cloud compute instance, making mistakes and needing to start over, and so on.

In the relevance stage, a lightweight pretrained language model (Meta's Llama 2 3.2 1b-instruct) was used to classify floor speechs as either relevant or not relevant to the survey variable, reducing the number of comparisons that needed to be passed to a larger model for the matching task.[2] The relevance step yielded 158,096 floor speeches coded as relevant to one or more variables from the most recent ANES study and a total of 326,649 relevant speech-variable pairs, around 3 percent of the original candidate pool. The prompt included examples like the one shown in figure 1, and were encouraged to get the model to return text beginning with the character **1** (relevant) or **2** (not relevant).

```
<|start_header_id|>user<|end_header_id|>
Below is a piece of text selected from the speech of an elected
official in the United States, followed by a proposition sentence
associated with some position on a social, political, economic or
cultural issue. Please indicate whether the text is relevant to
the proposition. In other words, does the text either support,
contradict, or otherwise directly relate to the proposition?
Please follow the same format as shown above.
<statement_text>
Social security is a vital part of what makes our nation strong.
</statement_text>
<prop_text>
Our economy has improved over the last 12 months.
</prop_text><|eot_id|>
<|start_header_id|>assistant<|end_header_id|>
2. not relevant (not about social security or welfare)<|eot_id|>
```

Figure 3 - one of the examples used in the relevance task prompt.

[2] Relatively small open-weight language modelswere used for all LLM inference tasks in this project over much larger and faster commercially available APIs due to concerns about replicability, transparency and durability of commercial language model APIs (Barrie et al. n.d.). Code to reproduce the LLM prompting and inference will be included in an appendix, although due to the nature of batched inferencing with attention caching, LLM classification results for individual examples may not reproduce precisely.

The examples provided in the prompt included short explanations of why a speech was or was not relevant to the sentence. This was helpful for debugging and may have steered the model towards better-reasoned outputs, the classification taks only required the first token. Any response with a higher score on the **1** token (preceding every "relevant" response in the prompt) compared to the **2** token (preceding every "not relevant" response) was coded as relevant. This is the approach used in some language model evaluation tasks like MMLU, and it is helpful because even where the model responds with something other than **1** or **2**, the relative importance assigned to these two labels – even when they are not returned – can serve as a useful binary or multiple choice classifier.[3]

Speech to survey variable position mapping

The second stage of inference was restricted to responses coded as "relevant" by the classifier in the first step. For this stage, a larger model (Meta's Llama 2 3.1 8b) was used after some experimentation showed poor results from the smaller 1b model. The prompt for this task was more than twice as long as the prompt for the relevance task, and an example (one of 13 "shots") from the prompt is shown in figure 2. In this prompt, there were four possible response tokens in a multiple choice setting: **1) shares_voter_view**, **2) disagrees_with_voter**, **3) takes_no_position**, and **4) unrelated_speech**. The use of underscores and labels in the prompt here is totally arbitrary and was just the result of multiple rounds of prompt tuning that improved the accuracy of the model's responses. As with the earlier prompt, the relative label scores on the

[3] For more detail on this please see Fourrier (2023) for an explainer on how multiple-choice LLM evaluation scores are calculated.

first response token were used directly as a kind of multiple-choice classifier; the “codes” **shares_voter_view**, **disagrees_with_voter**, **takes_no_position** and **unrelated_speech** were not used in the classification task.[4]

```
<|start_header_id|>user<|end_header_id|>
Below is a piece of text selected from the speech of an elected official,
followed by a sentence expressing a belief or view held by a voter.
After reading the speech, please indicate whether the official shares the
view of the voter, disagrees with the view of the voter, takes no
position, or if the speech is unrelated to the voter's view.
- Respond with the appropriate code and provide a brief explanation.
- Take a literal approach and do not make any inferences or assumptions
beyond what is in the text of the speech and the proposition or directly
implied. This is politics, so officials will frequently agree or disagree
with the views of voters. There are also many speeches that are unrelated
to voters' views.
Follow the same format.
<speech_text>
This legislation adds $7  billion to the President's budget for military
spending, and adds money  above the amount spent last year. We simply
cannot restore any significant amount of the huge  reductions in
education, in housing, in environmental protection unless  this bill is
brought under financial control.
</speech_text>
Based on what can be understood from this speech, does the official share
the voter's view that "The level of federal spending on the military
should be decreased."?
<|eot_id|>
<|start_header_id|>assistant<|end_header_id|>
1. shares_voter_view. (the official criticizes a bill for increasing
military spending, so the official seems to share the voter's view that
federal military spending should be decreased.)<|eot_id|>
```

Figure 2 One of the examples used in the 13-shot position matching task prompt.

4 Interestingly, earlier versions of this prompt suffered from a common problem: a very high error rate when coding speeches where the official is criticizing or otherwise urging a no vote on a bill. Very often, prompts that used phrases like “agreement” or “disagreement” would result in a “disagreement” code due to the tone and language in the speech, and unrelated to the relationship between the speech and the proposition sentence associated with the survey variable. Switching to prompt language about “sharing the voter’s view on …” resulted in greatly improved performance on the classification task.

Pruning the full dataset

The inference process described above matched 326,649 political speeches to an average of about ten recent ANES survey questions each, yielding more than three million pairs of speeches and responses. These were then joined with the individual survey responses, recoded as described above, and a new variable **rhetorical representation** was generated to indicated whether the views of an individual respondent to an ANES survey were reflected in the speech of an elected official. The join was restricted to only (**response**, **variable**, **speech**) triplets where the respondent lives in the congressional district of the speaker (or in the state, in the case of the Senate). The final dataset for analysis consists of 22 million data points, each of which connects a single survey response given from 1994 to 2020 to a political speech given from 1994 to 2022, and indicates whether that survey respondent was represented in the official's speech.

3. Analysis

A key question in studies of policy representation has been whether or not political actions and outcomes reflect the desires of the broader electorate or if they instead reflect the interest of elites. These studies have shown mixed results, with research showing is worth asking to what extent these findings can be expected to translate to political rhetoric. Rhetoric is complicated and does not map directly to policy outcomes, and in fact at times it may be used to take a strong position on an issue that may not necessarily be reflected in the vote at hand. While the speeches made in Congress can be assumed to represent policy intentions at least some of the time, elected officials are skilled at speaking to multiple audiences at once, and may be signaling to voters or colleagues about ideas and intentions that bear little connection to the issue under discussion.

While it is beyond the scope of this paper to draw conclusions about the relationship between rhetoric and policy, we can start by looking at what factors contribute to more or less rhetorical alignment between voters and their elected representatives, and compare it to what we would expect from the policy representation literature.

First, we set up a weighted linear regression model where the outcome variable is **rhetorical representation**, the variable derived from the LLM inference step above. Each record in the data consists of an individual survey response that is mapped to a political speech made in the House or Senate by a representative (in either chamber) of the survey respondent. Data is clustered at the respondent level, and the respondent effects may be exaggerated due to this clustering, the model was fit using cluster-robust sandwich estimator. An increase of 1 in the dependent variable corresponds to a switch from a respondents' views not being shared by their representative to the opposite, i.e. a speech that reflects agreement on the specific issue question the respondent was asked.

Because income, education, race and gender have all been included in past studies on policy representation, they are included here along with age as predictors in the model. Income is derived from a five category percentile range provided in the ANES and recoded to three levels: low-income (0 to 16th percentile), middle income (17th to 96th percentile), and high income (96th to 99th percentile). Low income and high income are included in the model as indicator variables and middle income, representing the majority of respondents, is left out as a reference category. Age is also coded in the ANES data in ranges. To prevent the introduction of bias by coding unequal ranges as levels, age is filled in the data with a random uniform number somewhere in the specified age range. For example, a respondent with age coded as 17 to 20 might be coded in the regression data as 17, 18, 19 or 20 with equal probability.

Party identification in the ANES cumulative file is on a seven point scale from Strongly Democratic to Strongly Republican, with independents at 4, non-identifying leaners at 3 and 5 and strong partisans at 1 and 7. This was recoded to an indicator variable **republican = 1** and two additional indicators, **strong partisan = 1** and **weak partisan = 1** representing 1 and 7, with independents and partisans with no strong identity as uncoded reference levels. Respondents with a bachelor's degree and above are coded with **college = 1**, and speeches from the Senate, which can be very different from the House in rhetoric and tone, are also included with in indicator **senate = 1**.

It is also plausible that voters do not drive political rhetoric, and that factors outside of voter preferences could play a role in generating misalignment between voter attitudes and political speech. To test this, four structural variables were included. First, outside funding from political action committees is included and shown per 100k for better interpretability (**pac 100k**). Party funding is typically smaller than PAC funding but indicates importance of a race or a candidate to the national party and is also included on the same scale (**party 100k**). Two indicators of strategic pressures on rhetoric are also included: **safe seat**, a number from 0 to 1 indicating the probability of a politician being re-election based on results from the last six (House) or two (Senate) elections, is included to test if elected officials moderate their rhetoric due to strategies involving speaking to interests of independents and swing voters. Finally, a continuous measure of **electoral college risk** reflects the uncertainty of the official's home state in the previous presidential election. This is calculated simply with the variance formula for a proportion

$$\mathrm{Var}(\,p\,) = \frac{p \cdot (1-p)}{n}$$

(where p equals the state's democratic vote share in the previous presidential election) rescaled so the maximum value (at 0.5) is equal to 1. This vote share data is calculated from Dave Leip's county-level vote share data and aggregated to the state level (Leip n.d.).

To better illustrate the differences between the parties, the model is first estimated over all voters and officials in and then broken down by slices of party affiliation. In figure 4 below, the first column shows the results of the full model, followed by (2) a subset of the data where voters and their representatives are both Republicans (or Republican-leaning, in the case of voters). Column (3) shows the same subset for Democrats, (4) shows Republican voters (regardless of the party of the official making the speech), (5) shows the same subset for Democrats, and the last column shows the effects for the subset of data consisting of voters and officials from the same party. Columns (2) through (5) do not include independent voters or politicians in the analysis. The full model also estimates the interaction effect of a **republican = 1** indicator with a **republican official** = 1 indicator, but these terms and their interactions are left out of the subsetted models.

Surprisingly, the single greatest predictor of non-agreement with official speech is membership in, or identification with, the Republican Party. This is true in the full model both for republican voters in general, republican elected officials in general, and also on the joint effect of the interaction between Republican voters who are represented by Republican speakers (the joint effect of all three terms is -0.130 at $P<=0.001$).

The second highest-magnitude (negative) predictor of agreement with representative's speech is being in a state that is at play in the electoral college. A one-unit increase in electoral college risk (going from a completely safe state to a state that had a very close result in the most recent presidential election) results in a twenty-five percentage point reduction in rhetorical representation. Looking at the subset models, it is clear that this effect is most pronounced for Democratic and Democrat-leaning voters. There are multiple possible explanations for this phenomenon, including greater voter exposure to political advertising, efforts by elected officials to persuade voters from the other party through more conservative or liberal rhetoric, or pressure from the national party to avoid extreme positions in a state where elections may have important national political stakes. There are similar dynamics, and similar possible explanations, for voters and official living in or representing safe seats. Because this study only looks at pairs of voters and officials who share a state or district, it is hard to draw a contrast here between nationally vs. locally popular rhetoric, but in general it appears, maybe unsurprisingly, that elected officials in safe seats say things their voters tend to agree with. Similarly not surprising is that strong partisans see their views reflected more in political speech of their representatives, but it is interesting also to see that weak partisans (party-leaning independents) are more likely to see their views represented than other types of partisan attachment.

Social and demographic predictors play a less pronounced role, and it is not clear if the observed effects are due to correlation with Republican party membership or leaning. Education is correlated positively with more rhetorical representation, while the **white = 1** and **male = 1** indicators have the opposite effect, but these could be statistical artifacts due to correlation with party membership and the party effects noted above.

The final set of non-results relate to the role of money, wealth and outside finance on political rhetoric. These results are interesting to the extent that they line up with some existing research findings. First, survey respondents in the top income bracket (96th percentile or above in self-reported household income) do not see a significant effect in either direction in terms of their alignment with representative's political rhetoric. This is consistent with research that has found similar results regarding the relationship between income and policy responsiveness (e.g. Ura and Ellis 2008), but runs counter to the expectations of cross-national studies that find greater policy misalignment across income levels in relatively unequal societies (Rosset et al. 2013), as well as to studies that have found a relationship between income and roll call votes (Bartels 2008) and policy output (Gilens 2005). As far as the role of corporate interests and campaign contributions, these also do not have a significant effect on rhetorical alignment with voters, a finding that is consistent with research in the wake of the 2010 *Citizens United* ruling that sought and largely failed to identify a measurable impact of corporate campaign financing on legislative behavior (Bonica 2016).

	Dependent variable: Rhetorical representation					
	1. Joint	2. Both Rep	3. Both Dem	4. Rep voter	5. Dem voter	6. Same party
Intercept	1.003***	0.795***	0.859***	0.718***	1.120***	1.065***
	(0.060)	(0.086)	(0.087)	(0.104)	(0.078)	(0.004)
Electoral College Risk	-0.252***	-0.085	-0.192*	-0.048	-0.499***	-0.393***
	(0.061)	(0.092)	(0.086)	(0.105)	(0.078)	(0.004)
Safe seat	0.029**	0.005	0.056***	0.001	0.080***	0.056***
	(0.010)	(0.019)	(0.013)	(0.016)	(0.012)	(0.001)
PAC funding (100k)	0.000*	-0.000	0.001***	0.001**	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Party funding (100k)	0.001	0.010	0.005	-0.016***	0.014***	0.008***
	(0.002)	(0.010)	(0.003)	(0.005)	(0.003)	(0.000)
Age (decade)	-0.006***	0.003	-0.007***	-0.006**	-0.006***	-0.004***
	(0.001)	(0.003)	(0.002)	(0.002)	(0.002)	(0.000)
White non-Hispanic	-0.033***	0.001	-0.009	-0.055***	-0.016*	-0.038***
	(0.006)	(0.013)	(0.008)	(0.013)	(0.007)	(0.000)
Low income	0.008	-0.009	-0.000	0.062***	-0.018	0.001**
	(0.008)	(0.021)	(0.010)	(0.015)	(0.010)	(0.000)
High income	0.005	-0.009	-0.013	0.018	0.001	-0.007***
	(0.011)	(0.022)	(0.013)	(0.019)	(0.013)	(0.001)
College and above	0.038***	0.039***	0.072***	-0.011	0.069***	0.073***
	(0.005)	(0.011)	(0.007)	(0.010)	(0.007)	(0.000)
Male	-0.020***	0.004	0.003	-0.037***	0.007	-0.013***
	(0.005)	(0.011)	(0.007)	(0.009)	(0.007)	(0.000)
Senate	0.026***	0.011	0.031***	0.017*	0.029***	0.027***
	(0.004)	(0.009)	(0.006)	(0.008)	(0.005)	(0.000)
Strong partisan	0.041***	0.014	0.084***	-0.068***	0.071***	0.073***
	(0.006)	(0.014)	(0.009)	(0.011)	(0.008)	(0.000)
Weak partisan	0.034***	-0.004	0.020	-0.019	0.019*	0.022***
	(0.007)	(0.015)	(0.011)	(0.011)	(0.010)	(0.000)
Years since 1994	-0.000*	0.000	0.000**	-0.000	-0.000	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Republican voter	-0.264***					
	(0.007)					
Republican official	-0.216***					
	(0.006)					
Rep. official x Rep. voter	0.350***					
	(0.011)					
Observations	21,618,880	1,741,460	9,993,320	7,272,100	11,978,462	11,745,428
R^2	0.112	0.057	0.070	0.066	0.056	0.060
Note:						*p<0.05; **p<0.01; ***p<0.001

Figure 4 - Rhetorical representation by survey respondents' House and Senate representatives. (weighted least squares). Fixed-effect controls for survey variable ID are not shown. Cluster-robust standard error was applied at the level of individual survey respondents.

4. Discussion

How do senators and representatives reflect the views of their constituents? Linking survey responses to congressional speech transcripts can shed new light on questions of political representation. Existing literature focusing on policy responsiveness has demonstrated some connections between income and policy outcomes, while others have noticed a somewhat surprising lack of evidence of any such connections and have called for more nuanced analysis of different dimensions of responsiveness and representation. By shifting the focus from policy output and legislative voting to speech, this study highlights a different kind of representation with uncertain implications.

One potential value of speech as a political tool, compared to policy, is that it is more visible to voters, particularly in the contemporary information environment dominated by national cable news, social media, and the slow death since the 1990s of local newspapers, television and radio stations. Politicians may be able to pursue strategies that use language as the primary area of responsiveness to voters while pursuing policy agendas that may be less favorable to the same voters who agree with their public statements.

A second key difference between language and policy is that, while every policy arguably has winners and losers, the same is not generally true for speech. Presenting different messages to different audiences is a critical skill in modern political communications, and the best practitioners can even manage to do this in the same utterance. This flexibility of political language leaves open the possibility that language might be more meaningful and useful as a tool for winning political support than policy output, while at the same time having no tangible impact on the policy outcomes voters actually care about.

5. Works cited

Barrie, Christopher, Alexis Palmer and Arthur Spirling, "Replication for Language Models: Problems, Principles and Best Practices for Political Science." Working paper presented at the American Political Science Association Annual Meeting, September 2024.

Bartels , Larry M. 2008. *Unequal Democracy: The Political Economy of the New Gilded Age*, New York : Russell Sage Foundation.

Bonica, A. "Avenues of Influence: On the Political Expenditures of Corporations and Their Directors and Executives." *Business and Politics*, vol. 18, no. 4, Dec. 2016, pp. 367-94.

Bonica, Adam. "Database on Ideology, Money in Politics, and Elections: Public Version 3.1." Stanford University Libraries, 2023.

Brown, Tom B., et al. 2020. "Language Models are Few-Shot Learners." *Advances in Neural Information Processing Systems* 33: 1877-1901.

Enns, Peter K., and Christopher Wlezien, eds. *Who gets represented?*. Russell Sage Foundation, 2011.

Fourrier, Clémentine, et al. 2023."What's Going on with the Open LLM Leaderboard?", https://huggingface.co/blog/open-llm-leaderboard-mmlu.

Gilens , Martin. 2005. Inequality and Democratic Responsiveness . *Public Opinion Quarterly*, 69 (5): 778-96.

Leip, David. Dave Leip's Atlas of U.S. Presidential Elections. **http://uselectionatlas.org**. US President General - State and County Level.

Hendrycks, Dan, et al. "Measuring Massive Multitask Language Understanding." Proceedings of the International Conference on Learning Representations (ICLR), 2021.

Kwon, Woosuk, et al. "Efficient Memory Management for Large Language Model Serving with PagedAttention." Proceedings of the 29th Symposium on Operating Systems Principles, Association for Computing Machinery, 2023, pp. 611-626.

Rosset, Jan, Nathalie Giger, and Julian Bernauer. 2013. “More Money, Fewer Problems? Cross-Level Effects of Economic Deprivation on Political Representation.” *West European Politics* 36 (4): 817–35.

Ura , Joseph Daniel and Ellis , Christopher R. 2008 . Income, Preferences, and the Dynamics of Policy Responsiveness. *PS: Political Science & Politics*, 41 (4) : 785-94.