Talking Policy...or Not:
How Context Shapes when Members of Congress Discuss Issues

Kelsey Shoub, University of North Carolina at Chapel Hill[1]

A common complaint made by members of the minority party in Congress during and after the passage of many major pieces of legislation is that there is not enough public deliberation about the bills. However, members of the majority characterize these statements as complaints without foundation of the minority party. This prompts the questions: do some issue areas see greater discussion than others; if so, why do some issue areas see more discussion than others? I propose that the broader context — the amount of importance the public places on a bill and the amount of attention the area garners in the media — influences how much the parties talk about a general issue area. To test this, I introduce a new data set of the amount of attention parties in each chamber place different policies on the chamber floors in Congress using the Congressional Record. To identify what issues are discussed and whether that discussion is policy oriented, I use a combination of supervised and unsupervised machine learning techniques. Using this data set, I show that as the public places a greater importance on the issue area and as the media gives that area more attention then the parties discuss the policy area more on the chamber floors. Additionally, I show that the Senate discusses policy more than the House, and that the Republican party on average talks about policy less than the Democratic policy.

Paper prepared for presentation at the Visions in Methodology conference held at The Ohio State University on May 7-9, 2018.

[1]Ph.D Candidate, Department of Political Science, University of North Carolina at Chapel Hill, 374b Hamilton Hall, CB 3265, Chapel Hill NC 27599, shoub@live.unc.edu, `http://shoub.web.unc.edu`.

Representative Kevin Brady officially introduced what would come to be known as The Tax Cuts and Jobs Act on November 2, 2017. Within a month both chambers passed the bill. Within two months both chambers had passed the conference version of the bill, and President Trump had signed it. While the construction of the bill took months, public knowledge of its contents was extremely limited before passage. Additionally, while many television news shows spent hours discussing the bill and newspapers devoted numerous stories to it, very little meaningful debate on the introduced bill could occur because of the constrained time frame.

Along with many of his fellow Senators, Senator Durbin (D-IL) tweeted his displeasure about the process—especially about the lack of deliberation, openness about its content, and lack of information about the projected impact. Upon the Senate passing the initial bill, he tweeted "On the floor of the world‘s greatest deliberative body, @SenateGOP rammed through a bill without even reading it."[2] In referencing the old tag line of the Senate, he brought into sharp relief the lack of deliberation that occurred when considering the bill. Then, after the three-day conference session to rectify the House and Senate versions of the bill, occurring only two weeks after the Senate had passed its version, Senator Durbin took to twitter again to question the process. On December 19th, he tweeted: "Why are the GOP rushing through their partisan #GOPTaxScam? Because if the bill were to receive proper scrutiny, it would be revealed for what it isa monumental give-away to corporate special-interests, campaign donors, and the wealthiest Americans." Once again he decried the lack of deliberation, and he bluntly questioned the motives behind stifling discussion and waiting for the official CBO scores.[3] With his tweet, he claimed that with more information the tax bill would have died. Instead it passed and will go into effect for FY2018.

Was there truth to Senator Durbin's complaints? Did Republicans stifle deliberation

[2]Other tweets included: "Let‘s talk about how today Senate Republicans will force a vote on a bill that neither of them has read." and "Trying to review the #GOPTaxScam but they are making hand-written changes to brand new text as we speak can anyone else read this?"

[3]Official CBO scores and projections came out on December 15th, 2017. These arrived in time to possibly sway decisions in the House. However, voting never occurred again in the Senate.

and information in order to pass the bill? Senate majority leader Mitch McConnell (R-KY) publicly disagreed with these narratives. In defense of the process and in response to Democrats in the Senate, he quipped "You complain about process when you're losing."[4] In short, he characterized the problem as one of politics rather than process or policy.

However, these criticisms have been made more generally and over a longer period. There have been many public calls by the media and members of Congress to increase the amount of debate and deliberation, to wait for more information before voting, and time to process given information before voting. These concerns are born out in the numbers in both chambers. According to Curry (2015), 52.1% of important legislation was made public between 0 and 3 days before being acted upon in the House between the 106th and 111th congress. Of this 29.5% of bills saw a layover of 24 hours or less. On average, this is not enough time to read, process, and gather information necessary to make informed decisions or engage in rigorous debate (either in public or in private) about the bill. On the Senate side, the number of speeches given on the floor has decreased over time, which indicates that the amount of discussion may be decreasing. Since the 104th congress, the number of speeches logged in the Congressional record has fallen from almost 18,000 speeches given in the 104th congress (1995/1996) to 13,725 speeches given in the 114th (2015/2016). Each of these statistics indicates that discussion about bills may be limited and the that the discussion that occurs may be poor. An additional complicating factor to these calls, is that discussion may be less robust for some general issue areas than other areas.

While practitioners and watch dogs have raised concerns the amount — or lack thereof — of policy discussion in congress, it has been understudied within our discipline. This prompts the central question for this chapter: when and why are some issues publicly discussed on the House and Senate floors while others are not? Previous research provides clues and a foundation for an answer to this question. Specifically, I draw on and build from studies on who gives congressional floor speeches, what influences other types of floor be-

[4]For a sample news story about the process, see: `http://abcnews.go.com/Politics/republicans-defend-tax-bill-process-transparent-similar-complaints/story?id=51627463`.

haviors, and the role of information on the floor. Building from this, I develop expectations concerning how much policy discussion on the chamber floor about a general area occurs given interactions between the parties, party status in congress, and the outside context.

One challenge to answering this question and testing my expectations is that the requisite data is costly to collect and construct. A two fold problem presents itself: first, the general policy area discussed needs to be identified for each document (here congressional speech); and second, whether and how much of a speech is on policy rather than attempting to shape procedure or facilitating the discussion on the floor to move forward. By using a mixture of supervised and unsupervised machine learning techniques, I identify what policies are discussed and how often they are discussed in the Congressional Record, and show one way that the costs associated with this type of research can be lowered.

1 Studying the Chamber Floor in the House & Senate

Three generally disjointed literatures have turned to the Congressional floor to study member behavior, party influence, and legislative outcomes. While they occasionally examine similar processes and specific outcomes, they each introduce distinct questions and adopt different approaches. These are: studies on who speaks on the floor, the influence of institutions on behavior and outcomes, and the role of information on outcomes.

Of these three, the most proximate to speech on the floor is also the most isolated in its treatment of the House and Senate: research on who speaks on the floor of each chamber and why. Three common findings have emerged from this research. First, two types of speakers are more likely to deliver speeches than anyone else: the leadership in each party (majority/minority leaders, committee chairs, ranking committee members, etc.) and outsiders (ex. extremists, junior members, etc.) (Maltzman & Sigelman 1996, Morris 2001, Pearson & Dancey 2011). This means that those who toe the party line but are not in positions of power are less likely to speak. Second, parties significantly structure one-minute speeches by both providing pre-written speeches for members and handing out talking points for members to center their speeches (Harris 2005). Further, parties are reactive to each other within general policy areas (Hughes 2016). Third, speakers aim to increase their personal diffuse support through their speeches (Hill & Hurley 2002, Proksch & Slapin 2014). These findings are extended in a broader study of parliamentary debate in multiple countries by Proksch & Slapin (2014). They posit and show that the institutions governing parliamentary speech, party unity, and need for a personal (rather than party) vote informs who talks and how often. Additionally, they treat speeches as a vehicle for members of a chamber to stake out policy positions rather than as a tool to persuade or change the outcome of a bill. This is a common assumption throughout the more specific research on speech in Congress.

However, typically these studies do not expand beyond one specific type of speech be it one-minute speeches in the House (Maltzman & Sigelman 1996, Hughes 2016, Morris 2001, Harris 2005) or symbolic speech in the Senate (Hill & Hurley 2002). Additionally, aside from the Hughes (2016) study, the focus has been on the frequency of use of any kind of speech rather than a focus on substantive policy discussion by policy or on the use and discussion of procedure during discussion. Finally, while these studies typically assume that floor speeches are symbolic and used to simply transmit policy positions, research on the role of information in debate seems to run contrary to this assumption. One limiting factor in these studies is the high cost of collecting and building the datasets needed, because humans have been relied upon to hand code all or large proportions of each set of speeches studied. In this chapter, I seek to contribute to this literature in three ways: expand the types of speech to be considered, focus on attention to policy and procedure rather than general attention, and relax the assumption that speech is simply symbolic. To facilitate this analysis, I use a mixture of supervised and unsupervised machine learning techniques.

Rather than focusing on what is said and who says it on the House and Senate floors, the second literature focused on chamber floor proceedings. First, studies have shown that the use of different procedures change as the context outside of Congress changes, as the level of polarization changes, and degree of party strength changes. Procedures studied have ranged from a focus on the use of the filibuster and holds in the Senate (Howard & Roberts 2015, Koger 2010, Smith 2014) to the role and use of amendments in both chambers (Smith 1989, Sinclair 2016) to the ability of House leadership to forcibly shape legislation (Cox & McCubbins 2007, Rohde 2010, Curry 2015). Notably, many researchers have documented the centralization of the decision making process and increasing use of procedural warfare in both chambers since the 1970s (Smith 1989, Sinclair 2016, Mann & Ornstein 2006). Further, not only does their use change but the institutions themselves may also change (Jeong, Lowry, Miller & Sened 2014, Roberts 2010). Second, studies have shown that the degree of polarization and procedural changes are endogenous to the process: while the use of different procedures changes in response to increasing polarization, strategic decisions and procedures can also increase polarization in the chamber (Theriault & Rohde 2011, Roberts & Smith 2003). These two points together have underscored the key role of party and a party's status in the chamber on their behavior, the strategies they adopt, and their ability to effect change. However, while much has been learned about the behavior of individual members and parties in each chamber through this literature, many of these lessons have not been used to explain when, how, and what members and parties talk about.

Finally, bridging these two literatures about what happens on the House and Senate floors, are studies on the role of information—how members limit or increase the amount of information about policies and bills available using procedure. In this conception of the House and Senate floors, they are the last place where something can happen to alter the final outcome by the injection of new information (Riker 1986, Austen-Smith 1990, Curry 2015, Esterling 2009). As a result, those seeking to ensure the passage of a bill attempt to limit the flow of information, while those seeking to ensure the failure of a bill inundate those watching with information in an attempt to sway them. Two distinguishing features to this literature are that: (1) it does not assume that speeches are simply symbolic but rather can be used to persuade or at least influence the final outcome of a bill; and (2) they consider structured floor time rather than unstructured floor time. While these studies posit reasons and conditions under which it is more or less difficult to introduce new information, they do not test these connections; rather they incorporate them as controls in their models or bracket their consideration.

2 Expectations Based on Party Status & Context

These studies leave the primary questions of this chapter unanswered: when is greater policy discussion seen, and when is there a greater emphasis on procedural language and maneuvers rather than policy? While the previous research conducted on the use of the House and Senate floors does not directly answer these questions, it does provide the necessary theoretical groundwork to do so. The five key assumptions that I use to generate expectations for how each party will use each chamber floor are: (1) each party wants to become or maintain the majority in the chamber; (2) the leadership of a party in Congress coordinates its speech and actions; (3) to pursue a policy initiative through legislation, the policy topic must be discussed on the chamber floor; (4) speech on the Congressional floor can be used to speak to the public and the media; and (5) institutional characteristics matter.

What can be learned from these basic assumptions? If the goals of each party differ—regardless of their policy priorities or the behaviors of other institutions—then their behavior within the chamber and specifically on the floor may differ. For those in the minority, the primary goal of the leadership is to gain the majority. To do so, they attempt to highlight why the majority party should not control their chamber. They want to draw attention to the ways in which the current majority party are considering and passing bad policy for the county and incompetent, or unfit to remain in control. To increase the efficiency of these attempts, they coordinate their appeals. As a result, the minority party leadership coordinates and constricts what is said in public.

On the other hand, for the majority party, the primary goal of the leadership of that party is to maintain the majority. To do so, they must balance the collective needs of the party (i.e. maintenance of the party brand) with the individual needs of its members to be reelected (i.e. public connections to work being done and staking out policy positions). In pursuit of the first, party leaders will attempt to pass a legislative agenda that conforms to its brand, or its publicly espoused set of policy positions and specialties, that its members can campaign on in future elections. By doing so, they show voters that they will adhere to campaign promises and deliver what they say they will; in the next round of voting, their party is a safe bet. Not only do party leaders want to pursue a legislative agenda conforming to its brand, but the want to present a unified front while doing so. However, this further implies that both parties will decrease the amount of speech on issues they prioritize and are seen as part of the party brand in pursuit of maintaining a coherent and streamline message.

One way party leadership of each party does so is by attempting to control who speaks, what policy issues they talk about, which arguments they make concerning the issue, and how much is said. This is primarily done to limit the probability that members go off-script and introduce complicating arguments or information into the process. Both of which could confuse the discussion, derail a piece of legislation, or garner unwanted media attention. However, their ability to do so is limited by the degree of power and trust their members put in them: this is higher if there is less internal disagreement, and if a clearly delineated foe exists.

Further, a tension arises: on the one hand the party leadership of each party wants to maintain a streamlined, coherent message; on the other, its members desire enough flexibility to pursue their own goals to increase their likelihood of reelection. This prompts the question: when will leaders relax their control over who speaks and how much discussion takes place?

If the party leadership is confident that their members will toe the party line, then the majority party leadership will relax their control over how much policy discussion takes place. Of all the issues their members could discuss, majority party leadership is most confident that members will stay on brand for issues central to the party. This is because these are the issues that the majority of the party has decided are important to it. Not only are issues central to the party those that the leadership might try to guide members to talk about, but they are also the issues individual members of the party might want to speak about the most. Because individual members of a party need to cultivate a personal vote, they will take advantage of the ability to deliver speeches on policies where they wish to claim credit for the work being done in the chamber in the policy area or publicly take a position on the policy. This implies that the leadership will relax control over how much is said substantively said in a policy area and when that policy area is prioritized by the party.

In order to maximize the effect of these appeals, the each party looks for ways to make their appeals and attacks more relevant and more likely to garner attention. Thus, they look for opportunities where their statements inside of the chamber might gain attention outside of it. As a result, members of either party will be sensitive and responsive to shifts in what the public views as important and what policy areas the media is highlighting. However, each will manifest in different behaviors on the chamber floor due to the differences in source input. First, if the public views a given policy area as important and as a problem, then the minority party can leverage the public's heightened sensitivity to this area to highlight why the majority party is deficient in some way. As a result, members of either party may seek to increase the amount of time spent discussing the policy area itself in order to highlight the ways the legislation the majority party has proposed to address that policy area is deficient, highlight why they current status quo policies in that policy area are deficient, and how the minority party is seeking to make things better. Essentially, the minority party wants to reveal "how the sausage is made," and by doing so decrease the public's approval of the majority party. This serves the added benefit of allowing the minority party to appear responsive to the public's interests and concerns.

Second, media attention to an issue area brings with it the possibility of recognition of more individual members in the media providing public acknowledgment of the status and position of the member. With the increased possibility of being named in association with a policy area, being quoted as a part of a news show or in a news story, or being directly interviewed, each party will clamber to publicly make statements about the policy area (Lovett 2016).

2.1 Hypotheses

From this theoretical discussion, a number of empirically observable implications can be made. Each hypothesis is conditional, and comes from a theory that assumes whether a party is in the majority or minority will change its behavior. Among the studies that examine Congress and party behavior within Congress, this is a common assumption that typically results in the focus on one party in isolation. The hypotheses to be tested are:

> **H1:** As an issue area is more closely identified with a party, the party will discuss it more.
>
> **H2:** As the public considers a policy area more important, each party will talk about policies associated with that policy area more.
>
> **H3:** As media increases its coverage of a policy area, each party will talk about policies associated with that policy area more.

3 Identifying Discussion in the Congressional Record

To measure how the parties their time discussing policy on the chamber floors, I use the text of the Congressional Record from the 104^{th} through 112^{th} congress (1995 through 2012). The Congressional Record is the official account of what is said on the floor of each chamber and the extension of those remarks. This includes commemorative speeches (ex. recognizing the work of a girl scout troop in the member's district), policy speeches, legislative debate, and daily business (ex. prayer, pledge of allegiance, or schedule for the day). Nested within the policy speeches and policy debates is explicit discussion and execution of procedure and construction of the rules for considering policy in legislation. Within the Congressional Record, I focus on policy speeches and legislative debate rather than commemorative speeches or general daily business to capture the division of attention on procedure on the one hand and actually discussing the policy area on the other. Within these speeches, I identify how much time is spent on different possible aspects. I briefly unpack how I do so below from acquisition of the Congressional Record to what members of Congress are talking about to identifying how they speak about those issues.

To acquire the Record, I download it from the Government Printing Office.[5] Then I apply policy topic codes matching those used in the Comparative Agendas Project in three stages to the downloaded speeches, which include all possible speech types. The policy topic codes applied are the major topic codes from the Comparative Agendas Project with an additional code for "no policy topic discussed." The twenty (general) major policy topic codes are: macroeconomics, civil rights, health, agriculture, education, environment, energy, immigration, transportation, law and crime and family issues, social welfare, community development, banking, defense, science and space and technology, foreign trade, international affairs, government operations, and public lands. The three steps I took to apply these codes to the speeches are: (1) I assigned known policy topic codes by matching speeches to a previously coded data set of one-minute speeches given on the House floor (Hughes 2016) *and* by identifying speech content via the speech title (ex. direct mentions of a bill in the title); (2) the speeches whose topic of discussion—if any—is known are then used to train a supervised machine learning algorithm to classify the remaining speeches into the possible policy topics (or lack thereof); and (3) I use the trained algorithm to classify the remaining speeches into policy topics. The overall accuracy of the classification algorithm is 94%. For an extended explanation of the identification of where the data came from and how policy topic labels are assigned to it, see the associated appendix at the end of this paper. After classifying the speeches into policy topics, only those speeches actually on a policy topic are

[5] To download the Record from the GPO, I initially used a parser built by Judd, Drinkard, Carbaugh & Young (2017) and then further cleaned the data.

included in the analysis.

Once the speeches are classified into policy topics, I break them down by paragraph and identify how the paragraphs are used. Speeches are broken down by paragraph to facilitate their identification, because a single speech can contain multiple arguments about a given policy and a mixture of language on policy and procedure. There are 4,261,184 paragraphs in the dataset. This ranges from 66,983 to 672,748 for a single policy topic, with a mean of 213,059 and a standard deviation of 157,363. Paragraphs can be classified as falling into one of three general types, where the largest division is between whether they are geared toward policy discussion or procedure and whether they are or are not substantive. The four general types of speech that emerge are: (1) policy orientated statements within which arguments about the policy itself discussions of procedure, partisanship, or the legislative process are made; (2) procedure I which are procedural statements aimed at shaping considerations or affecting a vote; and (4) procedure II which are procedural statements aimed at facilitating the conversation (ex. yielding the floor), which essentially capture how many times who is speaking changes.

To identify where in the general typology paragraphs fall, I fit a dynamic topic model (DTM) for each policy area. In each dynamic topic model, speech paragraphs are treated as the individual documents and Congresses are the time unit (Greene & Cross 2017). Topic models use the co-occurrence of words in documents in a specific time period to identify latent linguistic structure, commonly called latent topics. Dynamic topic models smooth across those periods to connect identified latent topics over time. They identify k latent topics within the text that are constant over time. The number of overall topics identified varies by policy topic.[6] The four types of speech detailed above emerge from the dynamic topic models as meta categories with multiple latent topics fit within each general type. While in

[6] To identify the number of latent topics, I first fit individual topic models by time period and allow the computer to tell me what number of topics maximizes coherence from a range of topics from 3 to 30. Next, using the best fitting number of topics for each period, I fit a dynamic model across the time periods and allow the computer to tell me what number of topics maximizes coherence from a range of topics from the maximum number of topics in an individual period and 60 topics.

this chapter, only the general typology is used, in the following chapters the specific latent topics identified are discussed in greater detail and used. To identify these general types and labels for each individual latent topic that emerges, I use the twenty most informative terms (i.e. those terms that best distinguish that topic from the others) for each identified latent topic.

For an illustrative example, take four example paragraphs, presented in table 1, from a single speech on the floor by a single speaker. These statements were delivered by J. Exon on January 4, 1995. The overall speech is classified as being about macroeconomics and an example of each speech type is presented. In these statements, the portions that are underlined emphasize what makes this speech about macroeconomics (balancing the budget and the balanced budget amendment), and the portions that are in bold emphasize what makes each speech a member of its overall type. As can be seen, the policy being discussed is a balanced budget amendment and more generally achieving a balanced budget. The first two examples are both policy oriented while the second two are procedurally oriented. Between the paragraphs, the main difference is whether or not the majority of the paragraph directly addresses the policy considered. Between the procedural paragraphs, the main difference is active engagement versus passively passing the torch.

Rather than looking at example paragraphs, I can examine the top twenty words used to classify these latent topics. Continuing the example from above, the latent topics with their associated list of "most informative words" of the paragraphs above are:

- **Policy (1):** people, american, american people, congress, country, people want, speaker, america, believe, deserve, american people want, washington, american families, give, work, people country, people know, working, families, understand
- **Policy (2):** amendment, balanced, balanced budget, budget amendment, balanced budget amendment, vote, constitution, amendments, constitutional, constitutional amendment, budget, votes, amendment constitution, voted, debate, against, offered, budget amendment constitution, balanced budget amendment constitution, motion,
- **Procedure I:** consent, ask, unanimous, unanimous consent, ask unanimous, ask unanimous consent, president ask, president ask unanimous, president ask unanimous con-

Table 1: Paragraphs from "PROPER AND LEGITIMATE ROLE OF GOVERNMENT"

Typology	Paragraph
Policy	Meanwhile back at the ranch we have all kinds of people, well-intentioned people, who are saying, "This has to be off limits. Of course that has to be off limits. We cannot touch this, we cannot touch that." I hope those of us who vote for a constitutional amendment to balance the budget recognize, as we must, that not all of us, maybe not a majority of us, will be here serving in the U.S. Senate and the House of Representatives in the year 2002. Yet we are mandating what people will do then. **We, therefore, in my view, have the responsibility to plow a straight furrow, to tell the people exactly what the situation is, to put the pain and suffering that is going to take place in making these cuts so they are clearly understood**–to recognize that, of all things, we may even have to raise taxes sometime before 2002 to accomplish the ends we are about to vote for. When you mention the tax word around here, though, that is a no-no.
Policy	**I simply say in tackling this proposition this Senator, and I expect two-thirds of the Senate, are strongly in support of and will pass a constitutional amendment to balance the budget. We have the responsibility, not only to vote but we have the responsibility to fully understand what we are tackling and what we are taking on. Therefore, I want to make the point that this S. 9 is a far-reaching measure.** It has to be passed, I believe, to bring some sanity to the Federal Government, to begin to balance income with out-go. Therefore it is a necessity. It is a very, very painful one and the people of the United States who send us here to do their bidding should understand when we do what they want us to do–the vast majority want a constitutional amendment to balanced the budget. I say to the people of the United States of America, it is not going to be easy. I am afraid too many believe if we just eliminate the $1,200 toilet seats and the $500 hammer, and if we cut the salaries of the Members of the House and Senate and their staffs in half, we could do those things and everything would take care of itself. It would be balanced.
Procedure I	Mr. President, **I rise today to lend enthusiastic support** to S. 9, which I think is probably one of the most important, if not far- reaching, measures that have been introduced today, along with very many other important measures.
Procedure II	Mr. President, I **yield the floor**.

Note: Portion that is in bold to emphasize why paragraph fell into the given category. Portion that is underlined emphasizes why the speech is categorized as macroeconomics.

sent, order, consent order, unanimous consent order, ask unanimous consent order, quorum call, ask unanimous consent order quorum, unanimous consent order quorum, consent order quorum, order quorum call, unanimous consent order quorum call, consent order quorum call

- **Procedure II:** yield floor, yield, floor, president yield floor, president yield, chair yield, chair yield floor, thank chair yield, thank chair yield floor, yield back, thank chair, yield floor suggest, floor suggest, yield floor suggest absence, yield floor suggest absence quorum, floor suggest absence, floor suggest absence quorum, president, thank, yield minutes

In the first policy example, the words indicate a latent topic centered on working for the American people, which indicates that the paragraphs are geared toward general policy discussion about a policy. In the second policy example, the words indicate a latent topic centered on discussing the process and politics of it all. In the first procedure latent topic, the top twenty words indicate a topic geared toward actively shaping what is occurring on the floor. Finally, the second procedure latent topic, includes words that explicitly and solely focus on yielding the floor, which is an action taken to move floor business along. Thus it is assigned to the second procedure type. For additional information on the process and validation, see the associated appendix.

4 Explaining Policy Discussion in the Record

4.1 Data

Using the speech data from the Congressional floor, I can then test the hypotheses proposed earlier in this chapter. However, the variables need to be operationalized, and the actual tests still need to be described. For this test, each observation is defined by general policy topic (ex. education, health, etc.), by party (Democratic and Republican),[7] by chamber (House and Senate), and by Congress (104^{th} through 112^{th}).

Using the speech data, the dependent variable is generated. To operationalize how much each party discusses policy by policy area, I use the count of the number of paragraphs classified on policy by type and party status (ex. number of paragraphs of the policy type given by the minority party). As a result, each observation is the number of paragraphs by congress, policy area, and chamber. Table 2 summarizes these values.

[7]Independents are excluded from the analysis.

Table 2: Summary Statistics for the Dependent Variables

	Minimum	Mean	Maximum	Standard Deviation
Policy Paragraphs	140	4,989	29,148	4,190

With the dependent variable explained, I can now turn to the task of operationalizing the concepts predicted to explain variation in strategies. First of these is whether issues are associated with each party. I capture this in two ways. First, I identify whether the public views the party as owning it. Issue ownership is coded as dichotomous, and coding is in line with the definition of issue ownership put forward and measured by Egan (2013). Second, I use the percentage of the number of sentences in each party's platform on each policy topic as defined by the Comparative Agendas project. The raw data for this comes from Christina Wolbrecht's American Party Platforms data set through the Comparative Agendas Project (2016).[8]

Additionally, the possible influence of what is going on outside of the walls of Congress need to be measured. Previously, two outside forces were hypothesized to matter: the media environment and the concerns of the public. The first of these concepts to be operationalized is salience to the public. To measure this, I include the mean proportion of respondents saying a given policy topic "is the most important problem" facing the US from Gallup surveys. This data is drawn from the Comparative Agenda's Project. The modal response is that the issue is not the most important problem (172 of 720 observations), the values range from 0 to 0.53, the mean is 0.05, and the standard deviation is 0.08. In addition to the salience of the issue to the public writ large, members of Congress may be especially sensitive to the salience of the issue topic in the media. To do so, I use Boydstun's New York Time's Index, which is a part of the Comparative Agendas Project. It is a normalized index indicating how much attention the NYT has given to different issue areas based on how many stores

[8]There are two alternative ways to measure this. Results are essentially the same if these are used. Alternatively, the specific legislative priorities expressed at the beginning of each Congress and classify them by CAP topic. This is not done, because the only datasets to have collected this information at this point only contains data for the majority party (Curry & Lee 2017, Curry 2015). However, here I am concerned with the priorities of each.

are written about them each Congress. This variable ranges from 0.71 to 1.52, with a mean of 1.08, and a standard deviation of 0.22.

Finally, I include a number of control variables, which the literature suggests may be important. Hughes (2016) finds that he parties are reactive to each other. This means that if the minority party increases the amount of time they spend directly discussing the policy area, then the majority party feels pressure to at least counter what is being said and vice versa. Alternatively, if the minority party is accusing the majority party of "playing politics" or attempting unfair procedural maneuvers, then the majority party will defend itself and vice versa. To account for what the other party does, I include the proportion of speeches given by the other party for each type. In the models predicting, the number of paragraphs falling into each type, I include the count of paragraphs given by the other party for each type.

Further, I include Five additional control variables. First, I control for the number of bills introduced in that policy area. I do so because the more bills introduced increases the number introductory statements associated with bills in that policy area. Second, I control for the number of speakers involved in the discussion. I do so, because the number of speakers will artificially increase the number of paragraphs. Third, I control for whether or not the set of speeches is given in the Senate or the House. A dummy variable indicating which chamber the set of speeches was given in is included to measure this. I include this, because the traditional orientation towards the House and Senate is that the Senate sees more deliberation than the House. If more deliberation occurs in the Senate, then more policy discussion might as well. Additionally, each has its own rules and procedures that govern how they operate. Fourth, I control for whether or not the party is the Republican party. Once again, this is a binary variable. Each party may adopt different strategies, and the leadership of the parties may have differing abilities to restrict who gives floor speeches and what they say during them. Additionally, while both parties during this time are organized, on average the Republican party has been more internally coherent. As a result, they use every type of speech less than then Democratic party. Fifth, I include a dummy for whether the party is in the majority or in the minority.

4.2 Results

To test the proposed hypotheses, I fit three models. The models predict the logged number of policy paragraphs using a hierarchical linear model (HLM) with random effects for policy topic and Congress. Separate models are fit for the majority party, minority party, and a regression including both. An HLM is used rather than a count model (ex. Poisson regression or negative binomial regression), because the spread and variance of the dependent variable indicate that the underlying data generating process is normal and violates basic assumptions of standard count models. However, because the number of speeches given is bounded by 0, I transform the dependent variable in each model by taking the log of the number of speeches. This ensures the expected number of speeches will range from 0 to positive infinity. Table 3 presents the results of these regressions.

With this, I can evaluate the hypotheses. First, I can evaluate whether if an issue area is closely identified with a party, then that party will discuss it more. Evidence of this would be seen if the "proportion of the party platform" variable and the issue ownership variables are positive. No strong evidence is found for this hypothesis. The signs on the proportion of party platform is positive but not statistically significant for the majority party regression. Both are negative but not statistically significant in the minority party regression. Finally, in the combined regression the signs point in opposite directions. At best, this indicates weak support for the hypothesis that higher prioritization in the party platform increases. At worst, there is no connection. However, a weak connection may have been found but does not reach statistical significance because of slippage between the concept of interest and measure used.

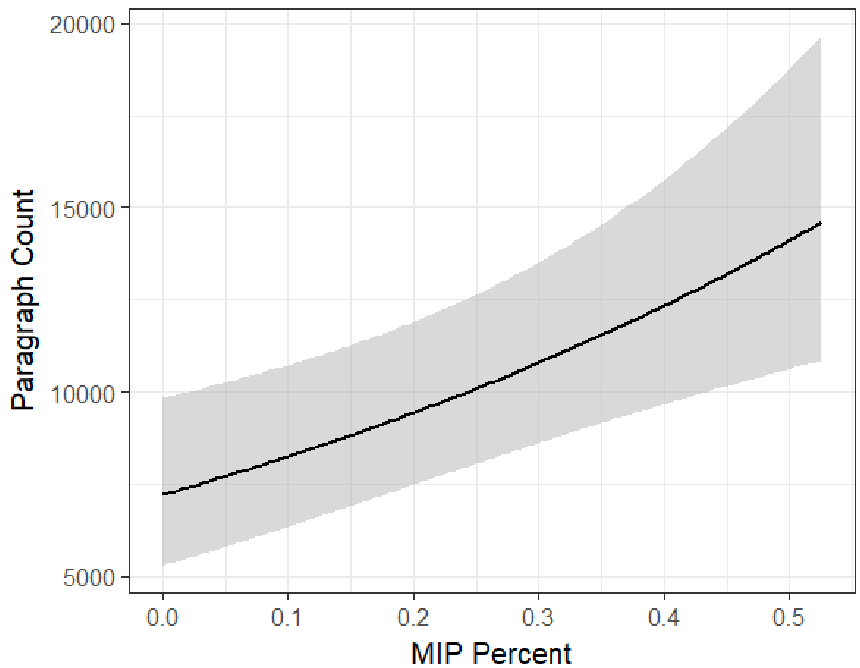
Beyond party issue priorities, I also hypothesized that the broader context will influence what the parties choose to talk about. First, I proposed that as the public considers a

Table 3: Explaining the Number of Policy Paragraphs

	Minority	Majority	Both
(Intercept)	5.97**	5.26**	6.25**
	(0.26)	(0.24)	(0.27)
Party			
Proportion of Party Platform	−0.24	1.14	0.51
	(1.06)	(1.00)	(0.75)
Issue Ownership	−0.08	0.01	−0.04
	(0.25)	(0.22)	(0.26)
Broader Context			
MIP Proportion	2.14**	0.50	1.34**
	(0.57)	(0.50)	(0.40)
NYT Index	0.52**	0.23*	0.38*
	(0.15)	(0.12)	(0.20)
Control Variables			
In the Majority			−0.04
			(0.04)
In the Senate	0.99**	0.83**	0.81**
	(0.11)	(0.09)	(0.07)
Republican Party	−0.42**	−0.10	−0.19**
	(0.07)	(0.06)	(0.04)
Number of Bills Introduced	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)
Number of Speakers	0.01**	0.01**	0.01**
	(0.00)	(0.00)	(0.00)
Policy Paragraphs, Other Party	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)
Procedure I, Other Party	−0.00	−0.00	−0.00
	(0.00)	(0.00)	(0.00)
Procedure II, Other Party	0.06**	0.28**	0.08**
	(0.01)	(0.03)	(0.01)
Congress	*Y*	*Y*	*Y*
Topic	*Y*	*Y*	*Y*
AIC	418.51	334.54	764.96
BIC	545.90	459.43	932.40
Log Likelihood	−176.04	−134.70	−345.50
Deviance explained	0.80	0.82	0.79
R^2	0.78	0.80	0.78
Num. obs.	342	342	684

$^{**}p < 0.05$, $^{*}p < 0.1$

Figure 1: Predicted Number of Policy Paragraphs, Range of MIP Values

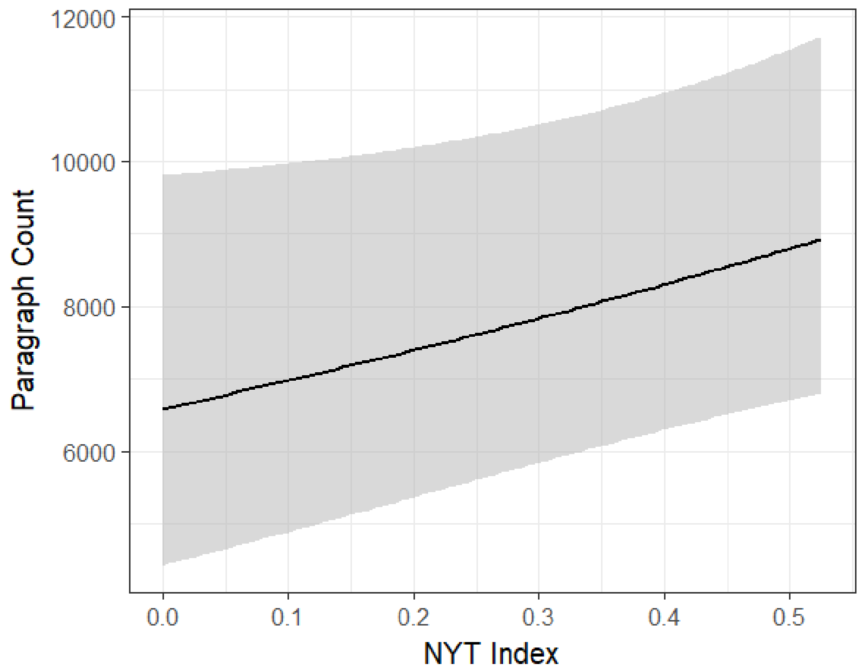


policy area more important, then each party will talk about that policy area more. Evidence of this would be seen if the "MIP Proportion" variable is positive. Once again consistent evidence of this is seen. In all three regressions, the MIP proportion variable, which is the proportion of respondents to a Gallup poll that say the issue are is the most important facing the US today, is positive. Additionally, the variable is statistically significant in two of the three regressions. This indicates that as the proportion of the public sees a policy area as important, then each party will talk about it more.

To put this into context, I can look at the predicted number of policy paragraphs delivered for macroeconomics in the 104th congress. Figure 1 presents this change. The shaded gray area indicates the 95% confidence interval. This shows a change from 7,214 to 14,606 across the range of possible MIP values, which is a change of 7,392 paragraphs. This is equivalent to almost two standard deviation changes for the number of policy paragraphs; a single standard deviation is 4,190.

Not only might the public's perceptions of importance influence whether a party talks about the issue more, but the media's attention to issues may influence each party's

Figure 2: Predicted Number of Policy Paragraphs, Range of the NYT Index



attention to issues within the walls of Congress. Evidence of this would be seen if the "NYT Index" variable is positive. This is seen across all three regressions. The coefficient only reaches statistical significance at the 0.05 level in the minority party regression, but reaches statistical significance at the 0.10 level in the other two regressions. To better contextualize the substantive effect, I once again turn to predicted numbers of paragraphs. These are presented in figure 2. In this figure, I use the predicted values from the combined regression. As can be seen in the figure, the predicted number of paragraphs ranges from 6,585 to 8,928, which is a difference of 2,343 paragraphs. This is approximately half of a standard deviation.

Finally, I can examine whether the control variables adhere to expectations drawn from previous studies. First, one expectation based on common conceptions of the differences between the chambers is that more policy discussion is expected to occur in the Senate than in the House. This found; in each of the regressions, the coefficient for "in the Senate" are positive and statistically significant. This indicates that more policy paragraphs are spoken in the Senate than in the House. Additionally, as the number of speakers increases, the number of paragraphs spoken increases. Finally, if the party in question is the Republican party, then fewer paragraphs are given on policy. However, none of the other control variables adhere expectations provided by previous research or are statistically significant. This indicates the the institutional characteristics of each chamber matters in the degree of policy discussion.

5 Conclusion

So, when is policy discussed more? In pursuit of answering this question, I introduced a new data set using an approach aimed at lowering the costs of large scale text analysis. I showed how a combination of supervised and unsupervised machine learning can be used to identify concepts of interest in the Congressional Record. Further, I showed that context matters. Each party discusses policy more when they can take advantage of increased attention by the media and public on specific issues. Beyond these conditional effects, the parties are shaped by the institution within which they operate: the Senate discusses policy more than the House.

However, in today's Congress, where more steps towards have been taken to curtail public discourse have been taken, this does not produce a rosy outlook. Rather, it feeds into the fears expressed by Senator Durbin in the introduction of this paper. Additionally, it does not give hop that there will be an increase in discussion of the issues anytime soon. However, simply because the leadership and members are wary of public discussion of some policies and pieces of legislation, it does not necessarily mean that anything would change if more discussion occurred.

This examines only one side of the concerns raised by those currently and formerly in Congress and in the media. Senator Durbin, in the tweets quoted at the beginning of this chapter, did not just lament the lack of discussion, but he also implied that if there had been more discussion the outcome would have been different. This prompts another set of questions to be studied: does discussion of policy matter; are the leadership's fears warranted; can discussion alter outcomes? In future research, I explore these connections.

Appendix

A Dataset Coding

Summary statistics are presented in table 4 for percent of the platform devoted to a general policy area, proportion of the population believing a general policy area is the "most important problem" facing the US today according to Gallup polling, and the adjusted number of stories published in the New York Times about the general policy area.

Table 4: Summary Statistics for Continuous Variables in Models

	Minimum	Mean	Maximum	Standard Deviation
Platform Pct.	0.00	0.05	0.25	0.04
Proportion MIP	0.00	0.05	0.53	0.08
Adj. Stories	0.71	1.08	1.52	0.22

Issue ownership coding comes from Patrick Egan's (2013) definitions of which issues are or are not owned by each party over the time period. On average, these are very stable over time. The classification is seen below in table 5.

Table 5: Issue Ownership by Party

	Topics
Neither	Civil Rights, Agriculture, Transportation, Community Development, Banking, Space and Technology, Government Operations, Public Lands
Both	Macroeconomics
Republican	Immigration, Law and Crime, Defense, Foreign Trade
Democratic	Health, Labor and Employment, Education, Environment, Energy, Social Welfare, International Affairs

B CAP Topic Coding of the Congressional Record

Rhetoric and heresthetic are measured by what is recorded in the Congressional Record. The Congressional Record is the official log of speeches given on the floor of the House of Representatives, the floor of the Senate, and extensions of remarks to those recorded speeches submitted at a later date. The electronic version of the Record (.txt files published on line) began in 1995 with the start of the 104th Congress.[9] The electronic version of the Record can either be downloaded speech by speech from congress.gov or in bulk from the Government Printing Office (GPO).[10] For this project, it was downloaded in bulk in batches from the GPO and tagged for its composite parts, such as speaker or text of the speech itself. Then the downloaded and processed files are transformed into flat files by year.[11]

To begin identifying what policy topics — if any — are discussed in each of the speeches, I identify speech segments where the policy topic (including a "no policy" topic code) is known and apply the relevant code. Policy topics are defined using the Comparative Agendas Project (CAP) coding scheme for "major topic" codes, which includes 20 possible general codes. In addition to the 20 policy topic codes, I add a code for "no policy topic." The inclusion of a "no policy" code is important, because not all speeches that lack policy content are excluded at this point.[12] This means there are a total of 21 possible codes that a speech could be given: no policy topic, macroeconomics, civil rights and liberties, health, agriculture, labor and employment, education, environment, energy, immigration, transportation, law and crime, social welfare, community development, banking, defense, science and technology, foreign trade, international affairs, government operations, and public lands. Identifying and tagging speeches that have known codes is done in three stages. These are: (1) merging in CAP codes already assigned to one-minute speeches given on the House Floor (Hughes 2016); (2) merging in CAP codes from the Congressional Bills project using mentions of bill titles and numbers in speech titles; and (3) using a dictionary of known non-policy speech types.

Using the coded documents, I then train a supervised machine learning algorithm

[9]Earlier versions of the record are accessible as PDF files or in hard copy only. Thus, these years are excluded.

[10]For the raw files: `https://www.gpo.gov/fdsys/browse/collection.action?collectionCode=CREC`.

[11]To download the Record from the GPO, I initially used a parser built by Judd, Drinkard, Carbaugh & Young (2017) and then further cleaned the data.

[12]In order to properly train a machine learner to identify CAP policy topics, there must also be a code for speeches that I know include no policy content.

to classify the remaining documents. The coded documents are further divided between a training and testing set. The training set will be constructed with an eye towards combating potentially problematic variation by issue area by drawing a stratified random sample of documents by policy topic. The result is a stratified random sample of 118,021 documents, which amounts to 50% of the speech segments that are labeled.[13] This leaves 118,018 documents in the testing set.

The documents need to become a matrix summarizing their contents: a document feature matrix (DFM) needs to be constructed. The DFM consists of two smaller document feature matrices bound together. In a document feature matrix, the documents are the observations, and the individual terms are the variables or features. The two sources of information are: a DFM of the text of the speeches themselves and a DFM of the titles of the speeches. The speeches and titles are processed in essentially the same manner: lower casing, stemming, punctuation removal, and number removal. Both the unigrams (single words) and bigrams (two words next to each other) are used as features. Additionally, rather than simply the counts of individuals features being used as variables, a tf·idf representation[14] is used. The last step in the process is to limit the features to those that appear in at least three documents. The result is 1,862,893 features being used. The `quanteda` package in `R` is used for pre-processing the data (Benoit & Nulty 2013).

The training set is then used to estimate an algorithm that in turn predicts the labels of the unseen and unknown speeches. The algorithm used is summarized in the gray box below. I use a combination of techniques and models to build the classification algorithm including a support vector machine, LASSO logistic regression, and constructed decision rules. An SVM identifies the maximum-margin hyper-plane(s) that gives the greatest separation between the classes. When performing a classification task that has more than two classes, it does so via pairwise comparisons of each set of classes (Witten, Frank, Hall & Pal 2016,

[13] A variety of sample sizes are experimented with, and 50% is found to produce the most accurate model.

[14] Term frequency · inverse document frequency, tf·idf, is a weight supposed to capture the relative importance of a term. Additionally, the use of term frequency rather than a weighting of the terms is experimented with, however performs uniformly worse.

Friedman, Hastie & Tibshirani 2001). While a number of different kernels can be used in the estimation function to identify the hyperplanes, I choose to use a linear kernel here.[15] A LASSO logistic regression is a special case of an elastic net logistic regression that uses the L1 penalty. The regularization process assists in variable selection by driving some possible coefficients to 0 (Friedman, Hastie & Tibshirani 2001).[16] A summary of the steps of the algorithm is presented here:

1. Construct the training set by drawing a stratified sample of speech-speaker documents.
2. Using the training set, fit an SVM that simultaneously predicts all twenty-one codes.
3. Using the training set, fit twenty-one binary LASSO logistic regressions. The training set actually used to fit the regression should include all instances of the relevant topic code and an equal number of other instances randomly drawn from the broader training set.
4. Repeat step 3 five times resulting in a bootstrap of five classifiers for each topic code. Treat each of the resulting predictions as votes for whether each document is of that topic.
5. From the binary classifier predicting no policy classification, identify whether the document is on a policy topic.
6. If the document is on a policy topic, from the series of binary classifiers, identify which topic received the most votes for the document being of that topic and whether there are multiple topics that have the same number of maximum votes.
7. If only one topic receives the maximum number of votes, then assign that topic.
8. For those where no single maximum exists, use the label applied by the SVM.

The overall accuracy of the model and Cohen's Kappa are examined. Accuracy is calculated by simply dividing the number of documents correctly labeled by the total number of documents. While a score closer to 1 (all documents correctly classified) is always better

[15]The use of radial and polynomial kernels are tested. However, they perform worse.

[16]The `glmnet` package is used to fit the LASSO logistic regressions, while the e1071 package is used to fit the SVM.

than one farther away, ideally the accuracy of the model should match or exceed the degree of agreement one would require of two human coders looking at the same document. Ideally, this will be at least 90%. The overall accuracy of this process is 94.32%, which means that 94.32% of the unseen but labeled speech segments are correctly classified using the algorithm.[17] Additionally, the Cohen's Kappa can be calculated for the performance of the overall algorithm. Convention says that a Cohen's Kappa between 0.811 indicates almost perfect performance (?). Further, intuitively it provides a score for how much better the algorithm performs than simply random chance. Here the Kappa is 0.92. This means that the overall accuracy of the algorithm and Cohen's Kappa exceed the minimum thresholds set out above for when looking at the overall not individual topic evaluations.

C Latent Topic Coding Paragraphs

Once the speeches are classified into policy topics, I identify how speakers are using their time: on policy or procedure. Additionally, speeches are broken down by paragraph to facilitate the identification, because a single speech can contain multiple arguments about the given policy and a mixture of policy and procedure. There are 4,261,184 paragraphs in the dataset. This ranges from 66,983 to 672,748 for a single policy topic, with a mean of 213,059 and a standard deviation of 157,363.

To identify the latent arguments in the Congressional Record, I estimate a separate dynamic topic model (DTM) for each policy topic; this allows for the aspects of language that change over time to be incorporated into the process and holds the general topic of discussion constant. I specifically use the non-negative matrix factorization (NMF) implementation developed and validated by Greene & Cross (2017). I choose to use the NMF implementation rather than the latent Dirichlet allocation (LDA) implementation (?), be-

[17]Additionally, each step of the process also has its own associated accuracy rate. The SVM by itself classified 87.06% of the unseen but labeled documents correctly. The rotating LASSO correctly classified 97.15% of those documents with only one maximum that are unseen but labeled. Of these, the no topic LASSO logistic classifier has an accuracy of 99% and the overall accuracy of the LASSO logistic classifiers for all those possibly on a policy topic is 94.62%.

cause some research has been shown that it is more effective at identifying niche topics that use specialized vocabularies (?). The latent concepts to identify theoretically should use specialized vocabularies, because arguments are specific, the House uses specialized vocabulary when discussing floor procedure and policy, and the Senate uses similar but distinct specialized vocabulary when discussing floor procedure and policy. There are six general steps in the estimation process, which is summarized in the gray box and is described in more detail below. A summary of these steps is presented here:

1. Bin the speech paragraphs by year.
2. Estimate the *word2vec* relationship between the terms.
3. For each year:
    (a) Process the speeches.
    (b) Perform a grid search across the possible number of overtime topics (from 3 to 30 by increments of 1) by fitting a series of topic models and calculating the mean coherence.
    (c) Identify the most coherent number of topics, k.
4. Perform a grid search across the possible number of overtime topics (from 20 to 60 by increments of 1) by smoothing the fit topic models overtime.
5. Identify the most coherent number of over time topics, k.
6. Assign topics based on the final model.

Before discussing the steps of the process in greater depth, how the number of topics k is identified needs to be detailed. The number of topics k that the DTM and each nested topic model should identify is the only parameter that needs to be chosen by the user. As do Greene & Cross (2017), I choose the number of latent topics from a given range that maximizes the average coherence. Average coherence is the mean topic coherence across all estimated topics. Coherence of a single topic is calculated by finding the mean pairwise cosine similarity between corresponding (i and j) term vectors (t) in a *word2vec* space (wv_i and wv_j), seen here (equation one).

$$coh(t) \;=\; \frac{1}{t}\sum_{j=2}^{t}\sum_{i=1}^{j-1} cos(wv_i, wv_j) \tag{1}$$

The *word2vec* tool computes a set of vector representations for all terms in a corpus (**?**). Theoretically, topics whose most important terms to its identification consist of very similar terms should have greater semantic coherence; those with more dissimilar terms should have less semantic coherence.

A separate DTM is fit for each policy topic. As a result, for each of the twenty policy topics these steps are repeated. First, the text must be binned by Congress, and then the *word2vec* associations are calculated using all of the documents at once. Second, a *word2vec* representation is of all of the text is generated. Only terms appearing in at least five paragraphs are used to generate the matrix.

For each year, then, the documents are (pre)processed to create a document-term matrix. In a document-term matrix, the documents make up the observations and the terms make up the variables. The values in the matrix are the representation of the text the computer uses in the estimation of the model. In this case, English stop words are removed from the documents, a normalized tf·idf (term frequency-inverse document frequency) representation is used, and terms appearing in less than 5 documents are removed. Additionally, uni-grams (single words), bi-grams (strings of two contiguous words), tri-grams (strings of three contiguous words), four-grams (strings of four contiguous words), and five-grams (strings of five contiguous words) are used. This is done, because, unlike topics, arguments require multiple words strung together to be meaningful. Using n-grams, assists in more closely approximating this real world semantic structure. Next, a topic model for each year is fit using the same NMF approach. At this stage the number of potential latent topics allowed to emerge ranged from 3 to 30. To choose the number of latent topics, I choose the number of frames that maximizes the coherence measure.

Once the documents in each year are processed and topic models been fit, the identified latent topics need to be aligned and smoothed, which are the last two steps in the process. Using the individually fit topic models, the number of latent topics across the years needs to be identified. The model searches over a range from 20 to 60 possible topics in an increment of 1. To choose, once again maximum coherence is used. Finally, the latent topics are assigned to every document, and the top terms for each identified latent topics are output. Only the primary latent topic is assigned, which means each paragraph receives exactly one label.

References

Austen-Smith, David. 1990. "Information transmission in debate." *American Journal of political science* pp. 124–152.

Cox, Gary W & Mathew D McCubbins. 2007. *Legislative leviathan: Party government in the House*. Cambridge University Press.

Curry, James M. 2015. *Legislating in the Dark: Information and Power in the House of Representatives*. University of Chicago Press.

Curry, James M & Frances E Lee. 2017. Non-Party Government: Bipartisan Lawmaking and Theories of Party Power in Congress. In *Paper delivered at the 2017 Congress and History Conference, Library of Congress, Washington, DC*.

Egan, Patrick J. 2013. *Partisan priorities: How issue ownership drives and distorts American politics*. Cambridge University Press.

Esterling, Kevin M. 2009. *The political economy of expertise: Information and efficiency in American national politics*. University of Michigan Press.

Greene, Derek & James P Cross. 2017. "Exploring the Political Agenda of the European Parliament Using a Dynamic Topic Modeling Approach." *Political Analysis* 25(1):77–94.

Harris, Douglas B. 2005. "Orchestrating Party Talk: A Party-Based View of One-Minute Speeches in the House of Representatives." *Legislative Studies Quarterly* 30(1):127–141.

Hill, Kim Quaile & Patricia A Hurley. 2002. "Symbolic speeches in the US Senate and their representational implications." *Journal of Politics* 64(1):219–231.

Howard, Nicholas O & Jason M Roberts. 2015. "The politics of obstruction: Republican holds in the US senate." *Legislative Studies Quarterly* 40(2):273–294.

Hughes, Tyler. 2016. "Assessing minority party influence on partisan issue attention in the US House of representatives, 1989–2012." *Party Politics* p. 1354068816654323.

Jeong, Gyung-Ho, William R Lowry, Gary J Miller & Itai Sened. 2014. "How Preferences Change Institutions: The 1978 Energy Act." *The Journal of Politics* 76(02):430–445.

Judd, Nicholas, Dan Drinkard, Jeremy Carbaugh & Lindsay Young. 2017. "conressional-record." *A parser for the Congressional Record* .

Koger, Gregory. 2010. *Filibustering: A political history of obstruction in the House and Senate*. University of Chicago Press.

Lovett, John. 2016. "Iindividual Members of Congress and Policy Change: The Use of Issue Salience as Subsidy." *Dissertation* .

Maltzman, Forrest & Lee Sigelman. 1996. "The politics of talk: Unconstrained floor time in the US House of Representatives." *The Journal of Politics* 58(03):819–830.

Mann, Thomas E & Norman J Ornstein. 2006. *The broken branch: How Congress is failing America and how to get it back on track.* Oxford University Press.

Morris, Jonathan S. 2001. "Reexamining the politics of talk: Partisan rhetoric in the 104th House." *Legislative Studies Quarterly* pp. 101–121.

Pearson, Kathryn & Logan Dancey. 2011. "Elevating Womens Voices in Congress Speech Participation in the House of Representatives." *Political Research Quarterly* 64(4):910–923.

Proksch, Sven-Oliver & Jonathan B Slapin. 2014. *The politics of parliamentary debate.* Cambridge University Press.

Riker, William H. 1986. *The art of political manipulation.* Vol. 587 Yale University Press.

Roberts, Jason M. 2010. "The development of special orders and special rules in the US house, 1881–1937." *Legislative Studies Quarterly* 35(3):307–336.

Roberts, Jason M & Steven S Smith. 2003. "Procedural contexts, party strategy, and conditional party voting in the US House of Representatives, 1971–2000." *American Journal of Political Science* 47(2):305–317.

Rohde, David W. 2010. *Parties and leaders in the postreform House.* University of Chicago Press.

Sinclair, Barbara. 2016. *Unorthodox lawmaking: New legislative processes in the US Congress.* CQ Press.

Smith, Steven S. 1989. *Call to order: Floor politics in the House and Senate.* Brookings Inst Press.

Smith, Steven S. 2014. *The Senate Syndrome: The Evolution of Procedural Warfare in the Modern US Senate.* Vol. 12 University of Oklahoma Press.

Theriault, Sean M & David W Rohde. 2011. "The Gingrich Senators and party polarization in the US Senate." *The Journal of Politics* 73(4):1011–1024.

Wolbrecht, Christina. 2016. "American Political Party Platforms." *Comparative Agendas Project* .