Understanding Congressional Expertise with Topic-Specific Dictionaries

SoRelle Wyckoff Gaynor, University of Maryland

Prepared for the UMD Government and Politics Department Methods Paper Defense
February 22, 2021

Abstract: Understanding how lawmakers develop and share congressional expertise is an important and ongoing question for congressional scholars. Measuring text complexity provides some insight to speaker expertise, however, existing measures of text complexity fail to accurately measure lawmaker familiarity with policy-specific vocabulary. I propose a new *dictionary-complexity score* to more accurately measure legislature expertise by weaving dictionaries based on word frequency with existing complexity formulas. This new method produces a complexity score that more accurately captures lawmaker familiarity when discussing otherwise complex policy. In addition, the flexibility and accessibility of this approach holds the potential to be applied towards other research questions and showcases a new way to use dictionaries in computational text analysis.

Word Count: 10,155

Introduction

Developing policy expertise is a priority for many members of Congress. Experienced, knowledgeable lawmakers not only benefit constituents, but generate respect among congressional colleagues (Curry 2019, Krehbiel 1991). Beyond district-specific information or prior experience, congressional committees provide the greatest opportunity for rank-and-file members to develop policy expertise (Kiewiet and McCubbins 1991, Fenno 1966). And in an era of congressional centralization, access to committee knowledge provides rank-and-file members with otherwise rare informational and procedural advantages when issues of their jurisdiction come before the chamber (Curry 2019). But despite the prominent role that expertise plays in policymaking, legislative scholars have faced difficulties capturing how this knowledge is developed and measured. Existing work has relied on rich, but often inaccessible observational data (Fenno 1966) or in-depth interviews with congressional staff (Curry 2019). And other measures, such as congressional seniority are blunt reflections of actual expertise.

This paper proposes a new method to analyze congressional expertise in an accessible, replicable, and robust way. Utilizing computational text analysis, I develop dictionaries based on word frequency within a predetermined corpus[1] and apply these dictionaries to established measures of complexity. This method also has the benefit of flexibility—researchers can easily develop dictionaries for individual policy issues, making it possible to compare how expertise differs across congressional committee, chamber, or even individual member of Congress. Accurately capturing congressional expertise is the first step to understanding how expertise is developed and disseminated throughout Congress on a large scale.

[1] Corpus refers to a collection of written or spoken materials that is analyzed. Corpora (plural) can be a collection of paragraphs, documents, or books.

In the U. S. Congress, complexity scores provide one way for researchers to infer the expertise of the speaker, author, or audience. Existing measures of text complexity are widely applied outside of a research setting, often referred to as "readability scores." Several industries, like the New York Times, the U. S. Government, and real estate agencies, use these scores to ensure that public-facing documents fall with a certain reading level.[2] Complexity measures are even embedded into Microsoft Word. These formulas weight the difficulty of texts by the number of letters, syllables, and word familiarity. Yet, these approaches do not accurately capture the realities of a legislative environment and expertise of the speaker. Despite the high frequency of policy-specific or congressional words, because of their otherwise rare usage in the English language, the resulting complexity scores are inaccurate, often weighting words as more or less complex based on their syllabus count, rather than actual familiarity with lawmakers. In short, existing measures of complexity do not accurately take into consideration word familiarity, an element I argue is easily captured by word repetition.

Thus, I introduce a new *dictionary-complexity score* to analyze the complexity of legislative texts. I develop dictionaries based on word frequency, and then apply the dictionaries to existing calculations of text complexity. I find that applying policy-specific dictionaries to existing complexity measures more accurately reflect the complexity of political texts by taking legislator familiarity and word usage into consideration. In addition to improved accuracy, this research presents an additional way to develop and apply dictionaries in computational text analysis. By creating dictionaries based on quantifiable word frequency I bypass resource-heavy approaches that require multiple human coders or pre-measured training texts.

[2] See the U.S. Government's website for criteria and communication requirements: https://plainlanguage.gov.

For an initial demonstration, I apply this method to a case study of congressional documents discussing the Farm Bill from the 109th – 115th Congress (2007-2018). As bicameral, omnibus legislation that is reauthorized every four to five years, the Farm Bill is an ideal case study to test this new approach. One, the policy subject matter is complex to outside observers, yet well-understood by members of Congress—language that under existing methods would be incorrectly scored. Two, agricultural policy is less partisan than other high-profile policy topics, allowing us to observe how policy debates differ by chamber, setting, and speaker. Third, the specialized, yet repetitive, nature of the Farm Bill generates vocabulary that is largely unchanged from year to year. We can imagine a situation in which a policy vocabulary could vary significantly year to year, either given specific legislation or current events that shift the narrative, but this is unlikely to be the case for the Farm Bill. For initial demonstrative purposes, this provides helpful flexibility in dictionary creation, as detailed below.

Theoretically, I expect that complexity will differ across setting and time. Committee documents, intended for members of the committee, and later other legislators, will likely be more technical and complex than public-facing speeches. Floor speeches, however, can either be directed at other members of Congress or their constituents— the complexity of these documents will likely have a wider range, depending on intended audience and topic. Complexity should also differ by hearing type—oversight and informational hearings have different purposes, so we should expect accurate complexity scores to reflect the variation in witnesses, questions asked, and information presented.

I test the accuracy of this dictionary-complexity score with descriptive comparison against existing complexity measures and showcase the potential application of these scores with an in-depth regression analysis. Ultimately, I find that my new measure is more accurate than

existing measures of complexity, and captures notable differences across time, by speaker, and setting. When dictionary-complexity estimates are compared with Dale-Chall (an existing and popular measures of complexity), I find that the addition of a policy-specific dictionary leads to a more accurate and cohesive measure of text complexity. Documents are more complex during years where the Farm Bill is up for reauthorization, but the congressional record from the floor is on average, more complex than congressional hearings, likely due to the larger range of voices on the floor and the opportunity for expert committee members to flex their knowledge to their peers (Curry 2019). Within hearings, House hearings present more complex language, and by hearing topic, oversight hearings are more complex than other congressional settings. And lastly, while there are no substantial differences across rank-and-file members, committee chairs are less complex than their (often expert) witnesses.

This analysis is an important first step to understand how Congress develops expertise through word usage. More importantly, this method holds potential for future analysis of congressional expertise across committee setting, party, and seniority. The remainder of this paper will be organized as follows: First, I discuss current approaches and shortcomings of text analysis processes used to answer questions of complexity and expertise. I will then provide a detailed overview of the development of my data and methods. I compare my results with existing, comparable complexity analysis. Lastly, I will offer considerations for future research and what these findings can tell us about how expertise is measured in Congress.

Current Approaches to Measure Text Complexity

The benefits of analyzing political texts have been well-documented, most notably for the rich detail that words provide. However, the detail that text provides inherently comes with a

heavy analytical lift. The size and scope of text cannot reasonably be coded by humans, and concerns of replication and bias have limited the feasibility of using human-coded text for causal research. And because of the detailed differences between texts, attempts to read and analyze even a manageable sample of text by hand do not provide any substantial insight (Guo et al. 2016).

From classification to scaling, the automation of text analysis has allowed scholars to quickly analyze text and has subsequently ushered in a new era of text as data (Grimmer and Stewart 2013, Wilkerson and Casas 2017). Researchers have answered questions by relying on either human input, such as dictionaries or supervised training models, and computer-generated sorting or unsupervised processes. Political scientists have used these tools to measure the sentiment of documents (Rice and Dorn 2017), measure changes in linguistic sophistication and complexity over time (Cann et al. 2014), and scale texts to establish more ideological positions for members of Congress (Grimmer 2012, Diermeier et al. 2012, Lauderdale and Herzog 2016, Laver et al. 2003, Benoit et al. 2009). The application of neural network analysis to text via "word embeddings" have even made it possible for political scientists to predict the occurrence of certain words (Rheault and Cochran 2020).

In measuring complexity, computational text analysis has provided robust insight into the linguistic choices of political leaders and parties, making it easier for political scientists to systematically analyze variation between speakers. Past work has evaluated the legal ramifications of complex Supreme Court opinions (Owens and Wedeking 2011), how voters respond to party platforms based on their readability (Bischof and Senninger 2018), and how the professionalization of the State of the Union Address has changed over time (Benoit et al. 2019). But successful applications of complexity in political science often focuses on general political

documents or party platforms, rather than specific policy debates (Grimmer and Stewart 2013, Lowe et al. 2011, Regel 2011). For example, State of the Union Addresses are a common focus of computational text analysis, but the subject matter of these speeches covers a wide expanse of policy topics with few, substantial legislative specifics. Applying these complexity measures to specialized topic texts, such as committee hearings, could fail to accurately measure policy-specific vocabulary.

And despite impressive computational gains, challenges remain. For one, there is still an unavoidable reliance on human insight and evaluation (Wilkerson and Casas 2017). Supervised models and dictionary development require human input to establish scales of reference, and even unsupervised methods require human analysis for validation on the tail-end (Benoit et al. 2009, Grimmer and Stewart 2013). Benoit, Grimmer, and Sterling (2019) have developed an impressive method to scale the complexity of political texts, but their approach still requires a heavy human hand. Screeners, human coders, and even powerful technology to perform these analyses on large bodies of text are often out of the realm of possibility for many researchers due to the high cost of resources and time.

Developing dictionaries based on word frequency presents an opportunity to bypass these more complex analytical approaches. The use of applied dictionaries has proven successful for detailed sentiment analysis (Muddiman et al. 2018, Rice and Dorn 2017, Haselmayer and Jenny 2016) but has not been applied to legislative texts, or measures of complexity. Given that measures of complexity rely heavily on word familiarity, word frequency can aptly be used to develop an understanding of familiarity, as well as being replicable and flexible. And while other measures exist that establish a ratio of rare or common words, namely term frequency/ inverse document frequency (TF-IDF), these measures cannot be tailored beyond simple word

frequency. Including topic-specific applied dictionaries to existing measures of complexity allows researchers to more precisely tailor analysis to a given research question.

Flesch Reading Ease score

The most pedestrian approach to measure complexity is the Flesch Reading Ease score (FRE) (Flesch 1948). Flesch (1948) created a scale taking into consideration the document's average sentence length and average number of syllables per words. Later iterations of this formula have updated the scale to equate difficulty with the US grade reading level (Kincaid et al. 1975). The formula is rather simple: subtract a ratio of the document's average sentence length and syllables-per-word from the total from a preordained number.[3] The end result is a number ranking the difficulty of a document on a zero (most complex) to 100 (least complex) scale. Text with longer sentences and multi-syllable words are considered more difficult. The median FRE score for readable text is 60-70, an 8th grade reading level. An FRE score of 90-100 is considered "very easy", or a fifth-grade reading level, while 0-10 is an "extremely difficult to read" document, intended for advanced graduates or a specialized publication. The New York Times FRE score, for example, is considered readable for 10th grade students (Tauberg 2019). This manuscript receives an FRE score of 26.4, a college graduate reading level.

Dale-Chall readability formula

Dale and Chall (1995) updated the FRE by establishing a list of approximately 3,000 words[4] familiar to 80 percent of fourth-grade students and removing them from the numerator of "difficult words," regardless of their length. This more accurately reflects complexity by removing words, such as "accident," "outdoor," or "strawberry," that would otherwise be considered difficult if only measuring by syllable length. The median Dale-Chall score is 6-6.9,

[3] See the full FRE formula in the appendix.
[4] A dictionary of the Dale-Chall common words can be found here: https://github.com/sorellew/data.

readable for the average 8th grader. At the low end, 0-4.9 is below 4th grade reading level, while a score of 10 or higher is readable for a college graduate. The New York Times Dale-Chall reading score is higher than its FRE at 8.1—requiring a high school education to comprehend. However, this paper receives a similar score of 9.3 (college graduate reading level). The difference in these scores is our first indicator that the readability score is susceptible to the list of familiar words included in the formula, which affects the NYT scoring, but not this paper.

This formula has since been updated. The new Dale-Chall reading formula does not score readability as a grade level, but still relies on the list of common words. The lower the readability score, the more complex the document is. This formula also allows for negative scores—if the number of difficult words is high, relative to the number of total words, then the amount subtracted from 64 could be greater than 64, making the score negative. The larger range of this New Dale-Chall reading formula allows for more detailed comparisons between texts.

$$New\ Dale.Chall = 64 - \left[0.95 * \ 100 * \left(\frac{difficult\ words}{words}\right)\right] - 0.69\ x\ (\frac{words}{sentences})$$

But although these formulas are accessible and easy to implement, from a substantive perspective, these methods do not sufficiently capture the substance of the text. Wide-use dictionaries, such as Dale-Chall's list of exempt words, or dictionaries often used to measure document sentiment, do not apply to topic-specific texts; and formulaic approaches alone bluntly weight words regardless of their frequency or setting type. When a word is not included in the dictionary of common words, it is considered difficult, regardless of the frequency of the word or the audience's education level. In policy arenas, words that are commonly used by the audience and speaker are incorrectly measured as difficult and unfamiliar, given their length or otherwise rare usage in the English language.

Development of Topic-Specific Dictionaries to Measure Complexity

In the legislative process, despite the complexities of policy texts, we should expect that participants (members of Congress, interest groups, staff, etc.) will develop a vocabulary that allows them to engage in the policymaking process and follow-along during committee hearings.[5] But without a tailored complexity application, words like "commodity," "international," or "committee," will be ranked as difficult words based on their syllable count, despite being often used and well-known to members of Congress and witnesses that discuss the Farm Bill. This oversight will inaccurately capture the complexity of documents—in many cases, words that are well-known, yet not a part of the common-word dictionary, will skew the complexity score to be higher than it actually is. For congressional researchers, mischaracterizing text complexity can lead to a misunderstanding in congressional expertise, policymaking, and constituent communication. The level of expertise among members, committees, and congressional sessions plays an important role on the quality of policy and even constituent communication. It's important for scholars to know how (and if) information is being absorbed and used within the chamber, and how members of Congress then communicate that information with each other and their constituents.

My solution is to merge existing, accepted measures of complexity with dictionaries of words that are familiar to policymakers and policy experts, but otherwise would be considered complex to a more general audience. I further avoid time-consuming human input by creating policy-specific dictionaries using a quantifiable measure: word frequency.[6] This approach has

[5] While use of policy-specific words does not necessarily translate to policy expertise, it does indicate adaption of the policy vocabulary—in short, it indicates *listening*. Future work should apply this measure of complexity to observe how members of congress adapt policy language.

[6] While an advantage of dictionary development is research flexibility, common approaches are often time-consuming and difficult to replicate. Dictionaries are created by hand, using *a prior* knowledge, reading the

several benefits. First, it generates a list of words that reflect familiarity, thus making complexity scores for policy texts more accurate. Second, the use of frequency to establish dictionaries limits human preprocessing and is easily applicable to any policy area. Dictionaries also provide researchers with flexibility to establishing individual dictionaries depending on the research question, corpus, or frequency threshold.

This method builds off of the new Dale-Chall formula by adding an additional dictionary to the list of familiar words. The addition of the topic-specific dictionary from the numerator of difficult words will more accurately reflect the familiar words ratio that Dale-Chall relies on. The formula is otherwise unchanged, ensuring that tested measures of complexity, such as sentence length, are still captured.

$$64 - \left[0.95\ x\ 100\ x\ \left(\frac{difficult\ words - \boldsymbol{topic\ dictionary}}{words}\right)\right] - 0.69\ x\ \left(\frac{words}{sentences}\right)$$

Building the Topic Dictionary

After combining documents into a corpus and performing the standard text pre-processing (Grimmer and Stewart 2013), I develop a dictionary based on word frequency. Because I am interested in familiar words that speakers will know despite their otherwise complex standing in the English language, I rely on word frequency to establish the dictionary. This paper focuses on selecting familiar words to more accurately capture document complexity, but other research questions could also be answered in this format. Considering ideology for example, researchers could set a frequency threshold for individual parties, committees, speakers, establishing individual dictionaries based on their research question.

documents, or gradually while working with the data. Using word frequency to develop dictionaries is a straightforward, replicable, and underused option for researchers.

Researchers have flexibility in designing this dictionary, first with the selection of documents that make up the corpus, and then the threshold of frequency by document or corpus. Detailed further below, this manuscript develops corpora by congressional session—each corpus contains all of the documents for the congressional session. Document type, speaker, and setting are additional information, used as variables. Then, using this corpus, I establish a threshold of frequency. For this initial analysis, I generate a general topic-dictionary, regardless of speaker or document attribute, but researchers can develop a frequency threshold reflective of their research question. Researchers should also consider whether to base frequency on how many individual documents in which the word occurs, or how often the word occurs throughout the corpus as a whole. For this analysis, I establish a frequency threshold based on how often the word occurs in each individual document within the corpus. Researchers should give thought to what documents make up their corpus and later, their frequency threshold.[7]

Applying the Topic Dictionary to Complexity Measures

After developing the dictionary of frequent words, I apply it to the new-Dale Chall formula topic dictionary formula, so that the formula includes both my dictionary of frequent words as well as the existing Dale-Chall list of common words and calculations based on word and sentence length. Ultimately, this removes common policy words from being included in the "difficult" word measures, more accurately capturing the complexity of the documents. Fortunately for those using R, the "quanteda" package, a common package used for computational text analysis, has several readability measures built in, including FRE and the two Dale-Chall formulas. Building off of the base new Dale-Chall formula, I created a new function

[7] For example, corpora could also be divided by document type, speaker, or document type and speaker—but because I compare complexity across document type and setting, I use congressional-session corpora to establish a session-specific dictionary.

("dictionary") by including an option to insert an additional dictionary to be subtracted from the count of familiar words, in addition to the existing list of familiar words.[8]

My approach of developing a dictionary based on word frequency avoids resource-heaving human coding efforts, as well as necessary *a priori* information. Furthermore, this function presents an accessible and easily modifiable approach to more accurate complexity measures. Researchers should spend considerable time thinking through the development of the dictionary, given their research question. After a thoughtful dictionary is developed, it is relatively easy to add their dictionary to my new "dictionary" function.

.

Application of the Dictionary-Complexity Measure to Congressional Data

Within Congress, information sharing occurs via congressional leaders, fellow rank-and-file members, committees, and informal groups like congressional caucuses (Curry 2015, Curry 2019, Miler 2011, Hammond 2001, Fong 2020). Committee participation in particular grants members access to witnesses and committee documents, as well as committee staff that help guide them towards helpful resources (Curry 2019). We should expect text complexity to differ by legislator, both in experience and intention; setting, from floor speeches to committee documents; and committee hearing type. And while this application focuses on a single policy arena, congressional hearings can differ in complexity and audience depending on the topic at hand, and debates that are highly ideological are likely easier to communicate than complex issues. Lastly, complexity can differ by chamber and congressional session.

To showcase the depth of information complexity can provide us, I apply this method to documents from the 109th – 115th congressional Farm Bill discussions. The benefit of using the

[8] This new function is available for download at https://github.com/sorellew/data.

Farm Bill as a case study is multifaceted: First, the Farm Bill encompasses several policy areas that, while complicated to non-experts, are fundamental issues for members of Congress. As omnibus legislation, the Farm Bill contains titles that range from Supplemental Nutrition Assistance Program (SNAP), international trade agreements, and crop-specific regulations. These policy debates are well-understood to the staff and members who debate the topics, but the policy-specific language is not common outside of Congress, or even outside of the committee room. Second, the legislation is reauthorized every four to five years, rather than built from scratch. This makes both the legislation and vocabulary easily comparable. Thus, when it comes to dictionary development, we can test both a general, multi-year dictionary of common words alongside individual-year dictionaries.[9] Third, the Farm Bill offers a partisan-neutral policy, and vocabulary differences will more likely be found between witnesses and members, rather between the two parties. Thus, we can more accurately analyze how text complexity differs by setting, congressional session, and speaker—rather than by party.

Given these attributes, I have several expectations for how complexity will differ by session, setting, and speaker. As the arena of fact-finding for members of Congress, I expect committee documents to be more technical and complex than floor speeches on average—but because committee members will utilize floor speeches to showcase their expertise (Curry 2019), there should be a wider range of text complexity for floor documents. I also expect that expertise is more developed in years when the Farm Bill is up for reauthorization[10], when more members are invested in the policymaking process. Because congressional hearings and floor speeches are

[9] As noted above, the development of the dictionary is a step that researchers should emphasize when using this method. For researchers interested in how word usage changes overtime, individual dictionaries allow for comparison over congressional session; if the interest is in the word complexity regardless of session, a dictionary could be developed by a corpus of pooled documents.

[10] The Farm Bill is up for reauthorization every five years but is often delayed due to gridlock—so more accurately, the bill is reauthorized every four to six years. Between the 109th – 115th Congress, the Farm Bill was passed by Congress in 2008, 2014, and 2018. In total, there have been 18 farm bills from 1933 – 2018.

the public element of congressional negotiations, these documents are more frequent in years the Farm Bill is being considered—and the more time that members spend on an issue, the more familiar they will become with it. Also, the impending vote should motivate lawmakers to be more informed about a policy issue. Thus, document complexity should be higher during the 110th, 113th, and 115th sessions of our dataset. I also expect that submitted documents will be more complex than spoken text. In addition, I expect that there will be differences in complexity between speakers, given party control or leadership positions. I test this last consideration below with documents from the 109th Congress, which have been separated into individual speakers.

Overall, I expect that my new, policy-specific measure will more accurately capture document complexity and speaker expertise than traditional measures. Given the many potential differences between congressional texts in one policy area alone, the need for a precise complexity measure is clear. For policy debates, particularly technical ones such as trade policy, agriculture commodities, or nutrition, current measures are insufficient to capture nuanced differences in policy texts. Understanding precise complexity differences between congressional documents will paint a clearer picture of how members of Congress change their speech from setting to setting and provide us with an understanding of what type of documents are more likely to be absorbed and shared throughout the chamber.

Data

I apply this new complexity measure to congressional documents that discuss the Farm Bill reauthorization from the 109th – 115th Congress (2005—2018). I retrieved the documents from the Government Publishing Office (GPO), the official record-keeper for the U.S. Congress. I include every congressional document that mentions "Farm Bill."[11] This retrieval process

[11] Researchers should be considerate of collection methods in regard to selection bias. Given the wide scope of the Farm Bill, this approach is appropriate for this research. But in other policy arenas, a general term such as

yields over 3,300 documents, mainly committee hearing transcripts and supplemental documents, and transcripts from the congressional record (speeches delivered on the House or Senate floor). This analysis utilizes committee hearing transcripts and documents, and the congressional record (n= 3,267). The number of documents per session is correlated with the Farm Bill's reauthorizing years, but as Figure 1 demonstrates, there are still hundreds of documents even during "off years".

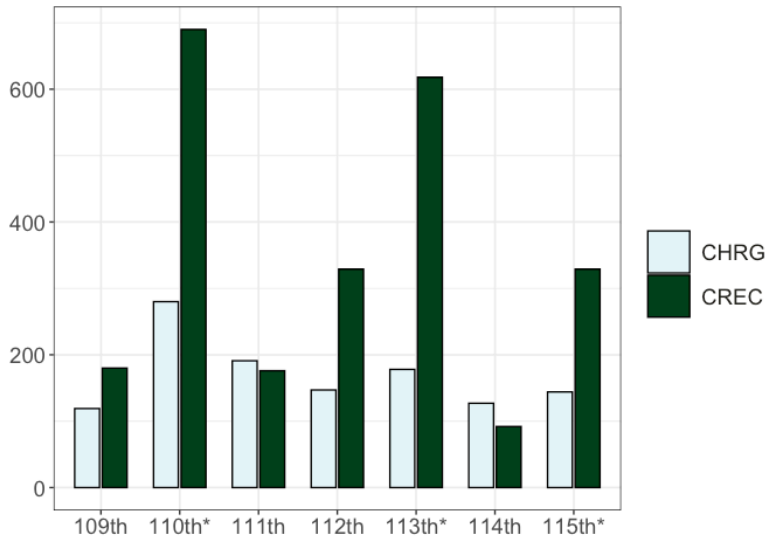


Figure 1: Document distribution, 109th – 115th Congress

Note: Data collected from GPO. Asterisk indicates year that a Farm Bill was passed by Congress.

The text length varies by setting. Committee documents are significantly longer, and between the two chambers, the Senate congressional record is, on average, longer than the House's record. This is largely due to the time limit on House floor speeches, compared to the free rein Senators have on the floor.[12] These documents also contain variety in speaker type. Committee documents contain language from witnesses and committee members alike. The

"immigration", or "Coronavirus" is likely more appropriate. In the case of the Farm Bill, this search pulled in any document that was Farm Bill adjacent—which I found appropriate for my question of complexity. How the Farm Bill is discussed for regional or personal purposes is relevant to establish how familiar the topic is with members of Congress.

[12] In the House, members are largely limited to one-minutes or five-minutes of speaking on the House floor. There are exceptions special order speeches, which are delivered when the House is in pro-forma session and thus with a limited audience. Comparatively, the Senate does not have a time limit for speeches by individual Senators.

majority of the committee documents in my dataset are from the House and Senate Agriculture committee but given the wide jurisdiction of the Farm Bill reauthorization, other committees such as the Financial Services, Ways and Means, and Judiciary committees (among others) are also included. Committee documents can contain spoken text as well as text submitted for the record (but not spoken). The committee record is similar to additional evidence, and can be submitted by members, witnesses, and external organizations. Members will often ask organizations or individuals to submit documents for the record, and some questions to witnesses will be submitted in this way. The congressional record texts, however, only feature members of Congress. But like the committee record, can also contain spoken and submitted text. Members will submit additional documents, such as news clippings or bill text to the record, rather than reading it out loud in the Chamber.

Dictionary Development

Before developing the dictionaries, I cleaned the data by removing numbers, punctuation, and peripheral information (such as GPO report numbers and website links). I then combined the individual texts into corpora. This approach breaks down the sentences into individual words, forming a "bag of words". However, unlike common approaches of text pre-processing, I do not stem the words.[13] Since I am forming dictionaries based on most commonly spoken and written words, I believe it's important to fully capture the actual word. Stemming the words in this dataset introduces unnecessary, and arguably detrimental, ambiguity.

After establishing the corpora, I then develop the dictionaries based on word frequencies. Researchers should emphasize this stage of the analysis, both in which documents make up the corpus, and how the frequency threshold is determined. The establishment of the corpora is an

[13] Stemming words reduces all words to their base word. For example, "running", "runner", and "runs" would all be converted to "run".

essential first step that warrants critical thinking. Given the data, I can either create seven individual corpora (one for each congressional session), or one, larger corpus that contains all seven congressional sessions pooled.[14] Each of these approaches have their own benefit. On the one hand, individual corpora for each congressional session will create a dictionary that reflects the popular vocabulary of that session and will allow us to see how word choices change over time. On the other hand, a larger, pooled corpus will provide one dictionary that we can then compare across congressional sessions. As an initial exploration, Figure 2 lists the most frequent 25 words from the corpora and the pooled corpus. Figure 1A in the Appendix displays the Word Clouds for each corpus.

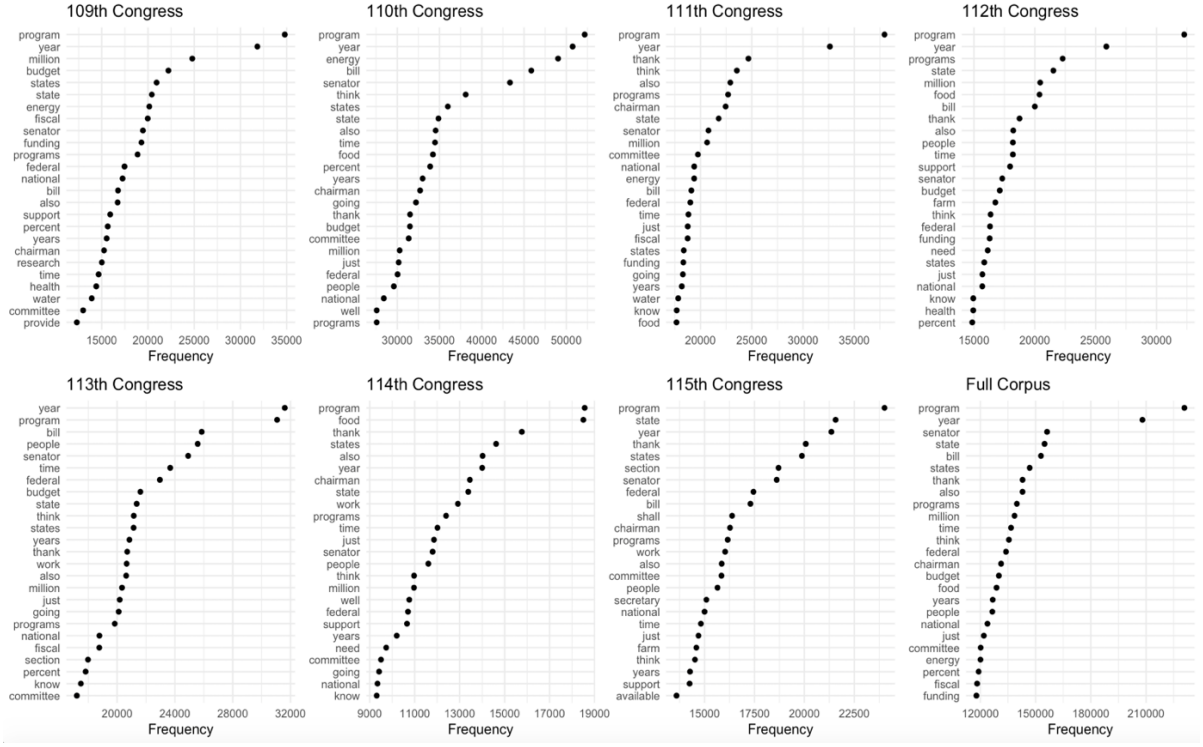


Figure 2: Top 25 Words, Per Corpora, 109th – 115th Congress

[14] Additionally, I could create individual corpora for each document *type*—one for congressional hearings, and another for the congressional record, for example. However, given that I'm interested in how text complexity compares across setting, it's unnecessary for this research question.

As Figure 2 illustrates, there is a great deal of similarity across the corpora. "Program" is noticeably the top word every year except for the 113th Congress (in which it comes in a close second). Because I'm interested in how word complexity changes overtime, I use individual corpora for each congressional session to develop my dictionaries. Given that there are different members of Congress represented in each congressional session's corpus, this approach ensures we are fully capturing the frequent vocabulary of the session. This approach also has the benefit of reduced computational power—a large corpus can be unwieldy. However, one could theoretically use the pooled corpus dictionary, depending on the research question.

After creating these corpora, I then establish a frequency threshold to build the dictionary. Here, again, there is a methodological choice to be made. Researchers can establish frequency by word occurrence throughout the entire corpus (*word frequency measure)*, or by the percentage a word throughout the individual documents (*document frequency measure).* In other words, researchers should consider what they would like their baseline for word frequency to be—total words in the corpus, or total words per documents.

$$Word\ Frequency\ Measure = \frac{word}{total\ words\ in\ corpus}$$

$$Document\ Frequency\ Measure = \frac{\frac{word}{document}}{total\ documents}$$

Because the denominator varies widely by measure, the results of the two frequency measurements have (understandably) varied results. The *word frequency measure* relies on how many words are in the corpus. Larger corpora, both in document size as well as document length, will lead to a larger dictionary. The *document frequency measure* will depend on the number of documents in the corpus, regardless of how large the document is. But because out *document frequency measure* establishes a threshold based on how often the word occurs across all

documents, meeting that threshold is easier in corpora with fewer documents. For example, if the word "committee" occurs in in five documents out of 10, it will have a higher frequency score than if those same five documents are part of a corpus of 100.

Table 1 below details how dictionaries differ in size based on the measurement used. Both dictionaries capture the top quarter of words for each measurement—the most frequent words (top 25 percent) of the entire corpus, compared with the most frequent across individual documents (top 25 percent).

Table 1: Dictionary size, based on frequency measures

Congressional Session	**Total Documents in Corpus**	**Total Individual Words in Corpus**	**Dictionary size, Document Frequency Measure**	**Dictionary size, Word Frequency measure**
109th	299	60,084	93	15,024
110th	970	93,444	35	23,361
111th	367	71,339	90	17,834
112th	143	73,577	35	18,394
113th	796	70,499	31	17,624
114th	219	58,741	227	14,686
115th	473	68,029	31	17,007

The vast difference between the two approaches introduces an important empirical decision for researchers. In selecting a method to develop a dictionary, researchers should consider the normative factors that influence their documents. In the case of this dataset, given that every Member of Congress is not in attendance for every floor speech or hearing (even the ones of the committees they're on), basing the dictionary on how frequently words appear per document (Document Frequency Measure) is the more appropriate selection. If we were to use the *word frequency measure,* a word may have a high frequency score simply for appearing several times within a few documents. This distinction is important because words that transfer to the majority of documents indicate greater familiarity across the chamber than perhaps a single word that is said several times in a select few documents. However, this approach does

deny researchers the opportunity to capture words that are undoubtedly well-known, appearing thousands of times in the political landscape.

Building off prior analyses (Muddiman, et al. 2019), I develop a dictionary based on words that appear in 60 percent of documents as the threshold for establishing familiarity. Each congressional session has a different number of documents discussing the Farm Bill (Figure 1). Therefore, the 60 percent threshold will create dictionaries of varying size.[15][16] I also develop an additional dictionary based on common congressional words—such as "committee," "appropriations," "chairman"—that are familiar to members of Congress, regardless of the policy area, but might not fall into the document frequency threshold. This additional dictionary is developed by frequency within the corpus, and like other dictionary developments, is malleable based on corpus and research needs. A full list of the congressional dictionary is included in the appendix. Lastly, in addition to flexibility on the frequency baseline, researchers can also utilize the frequency measure to create dictionaries that reflect their research question. For example, if a researcher is interested in how word usage differs within a single policy but across different committees, individual dictionaries could be created by establishing the most frequent words for each committee.

Results

These measures score the complexity of each individual text in the dataset. Figure 3 provides a boxplot of the dictionary-complexity measure across all documents for each congressional session (the mean point is the average complexity score for all of the documents in

[15] The full dictionaries can be found at https://github.com/sorellew/data.
[16] Checks were performed with dictionaries of words include din 75 percent of documents, but this ultimately excluded too many words, making the dictionaries unusable. Given the approach selected (document-based threshold) is conservative compared to overall word usage, I felt the 60 percent threshold was appropriate.

a given session). Of note, the interpretation is not necessarily intuitive: the higher the measure, the more familiar the words in the document (or the *lower* the complexity). As expected by the similarities in word frequency noted above, the complexity between congressional sessions is very similar. However, some sessions, such as the 110th, 113th, and 115th Congress have noticeable outliers of highly complex documents. Given that these are years the Farm Bill passed Congress, this is likely a reflection of the wider range of discussion of the legislation, both in subject matter and (non-expert) members discussing the legislation. There is also the potential that in years where the legislation is under floor consideration, the term "Farm Bill" is enveloped into larger discussions about ongoing legislative debates. Because I include the entire text (whether it be floor speech, committee hearing, or committee document), it is likely that during these "on" years, there is more extraneous vocabulary not captured by our dictionary-complexity dictionaries.

Overall, though, there is little difference in the mean complexity of documents, including whether or not the Farm Bill was up for reauthorization. Figure 4 further divides each set of documents by committee hearing documents and speech from the congressional record, for each congressional session. As expected, when the data is divided by document type, I find that documents from the congressional record cover a wider range of complexity. However, unexpectedly, the congressional record is on average, more complex—the majority of the outliers denoted in Figure 4 are from the congressional record.

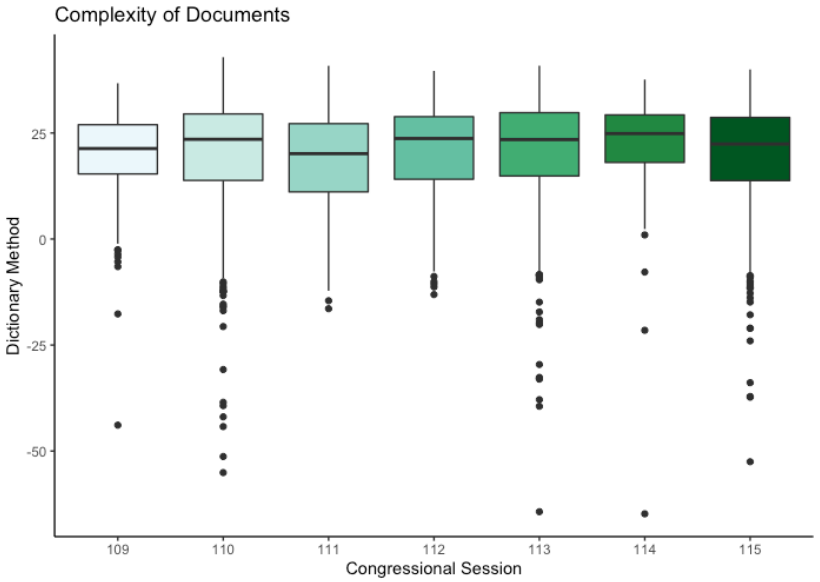


Figure 3: Dictionary-Complexity Measure, by Congressional Session

Note: The higher the score on the Y axis, the more readable (less complex) the text is.

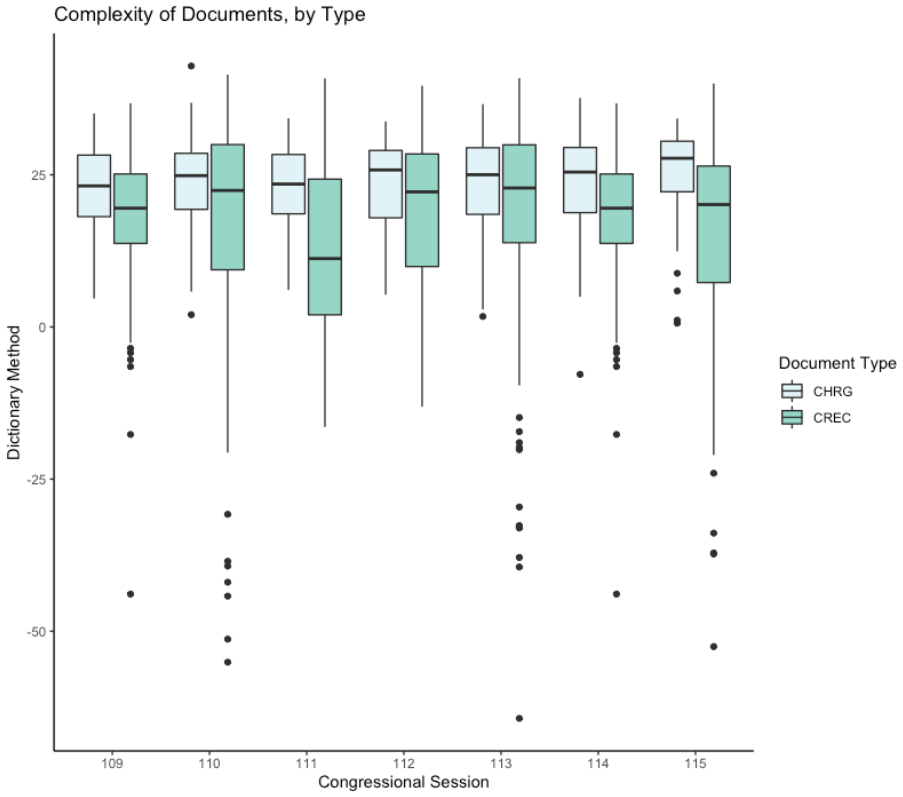


Figure 4: Document Complexity, Dictionary-Complexity measure, 109th – 115th Congress

Note: CHRG refers to Committee Hearing Documents, and CREC refers to Congressional Record documents.

To take a look under the hood, the lowest complexity point in our dataset, with a score of 42.88, is a congressional hearing document of submitted interviews to the Select Committee to Investigate Voting Irregularities, during the 110th Congress. The interviews, centered on personal experiences rather than expertise, included a discussion about the turnout in rural communities and how it impacted representation in the Farm Bill. The most complex data point is a House floor speech in the 113th Congress by Rep. Bill Flores (R-TX) welcoming a delegation of Polish farmers to his hometown, which contained several foreign words. The mean text—with a 20.042 complexity score—is a House floor speech in the 110th Congress by Rep. Virginia Foxx (R-NC) about taxes and job security in her rural community, as well as a discussion on how her local economy was impacted by the most recent Farm Bill negotiations. Each of these documents highlight the relationship between word familiarity and complexity. The least and most complex documents include familiar and foreign language, respectively. The mean document includes policy discussions beyond the Farm Bill, but it's otherwise constituent-centered message contains familiar words likely included within the Dale-Chall dictionary, or with few syllables. Presumably, if a topic-specific dictionary were to be developed for other economic issues discussed, it's complexity score would fall.

Comparing complexity measure performance

Comparing this dictionary-complexity measure to existing and accepted scores of text complexity will also provide us with an idea of its accuracy. The most complex document, based on the FRE score is a Senate floor speech by Sen. John Kennedy (R-LA) from the 115th Congress, discussing the differences between the House and Senate 2018 Farm Bill. Compared to the most complex document by the dictionary-complexity measure, this document is not only

easier to follow, but focuses on the Farm Bill components.[17] The least complex document is a short resolution to designate an agriculture memorial in Louisiana. This document likely received this score due to the short nature of the text and the short sentences within it. The document closest to the average FRE score (46.204) is a Senate floor speech by Sen. Patrick Leahy (D-VT), congratulating Sen. Charles Grassley (R-IA) on his 12,000th Senate vote. Sen. Leahy referenced Sen. Grassley's expertise, as a prior farmer and representative of a large agriculture community and his positive impact on farm bill negotiations.

The most complex and least complex documents in the Dale-Chall measures are the same as the dictionary-complexity measure—the floor speech by Rep. Bill Flores (R-TX) containing several foreign words, and the collection of interviews about voting irregularities, respectively. However, the document closest to the average Dale-Chall complexity score (19.22) is a 2015 Senate Foreign Relations committee hearing led by Sen. Bob Corker (R-TN) on the importance of international food aid, which played a large role in the 2018 Farm Bill.

Overall, the new measure lowers the complexity score across all documents, as expected. By classifying words that are otherwise considered complex as familiar, based on the frequency of their usage, it more accurately captures the complexity of the text. Figures 2A and 3A in the appendix detail how these dictionaries compare across congressional session. Because the new dictionary-complexity measure is most comparable with the Dale-Chall complexity measure, I also plot the difference between the dictionary and Dale-Chall scores, for the 109th-115th Congress.[18] While the range varies by individual text, the dictionary-complexity measure rates texts consistently less complex than the existing Dale-Chall measure. This measure more

[17] Future iterations of this project should implement additional coders for intercoder reliability to evaluate the complexity of the document. The full texts of these documents are available upon request.
[18] Calculated by subtracting the Dale-Chall score from the new complexity measure. Figures found in the Appendix, Figure 3A.

accurately captures the complexity of congressional texts and shows that the use of dictionaries can be applied to other measures.

Applying Complexity Measures to the 109th Congress

To show the full application of this measure, I break down the texts from the 109th Congress (2005-2007) into individual speaker. Because this congressional session is directly prior to the 2008 Farm Bill passage, there were several hearings and floor speeches centered on title development, but the session does not contain the high volume of hearings and floor speeches that surrounds bill passage. This congressional session is also a representative sample of the partisan control within our dataset. The 109th Congress was under fully unified control (Republican Senate, House, and George W. Bush administration), as was the case for the 111th Congress (Democratic House, Senate, and Barack Obama administration) and 115th Congress (Republican House, Senate, and Donald Trump administration). In addition, as the first year of our dataset, we are able to establish complexity patterns that can be compared with later years' analysis. I separate each of the 299 documents in the dataset by individual speaker. The result is 952 individual entries, including members of Congress, committee witnesses, and external submissions. In cases where a speaker appears multiple times in a committee hearing document, each moment of speech is an individual entry. This allows me to measure the difference across statement type (for example, opening statements versus the question and answer period).

As stated earlier, one of the benefits of this approach is the ability to generate a numeric score for text. In this dataset, each entry is given a complexity score, based on the text of the speaker. This allows me to analyze differences by setting or speaker attribute. Figure 5 below compares the complexity of texts between chamber, and Figure 6 further compares complexity

across committee and floor setting, by chamber. When pooled together, the two chambers overlap in their document's complexity scores, with the Senate documents scoring a lower complexity score on average. When I consider complexity by committee and chamber though, there is a slight variation between committee hearings, with House Committees being the source of the most complex language, on average.

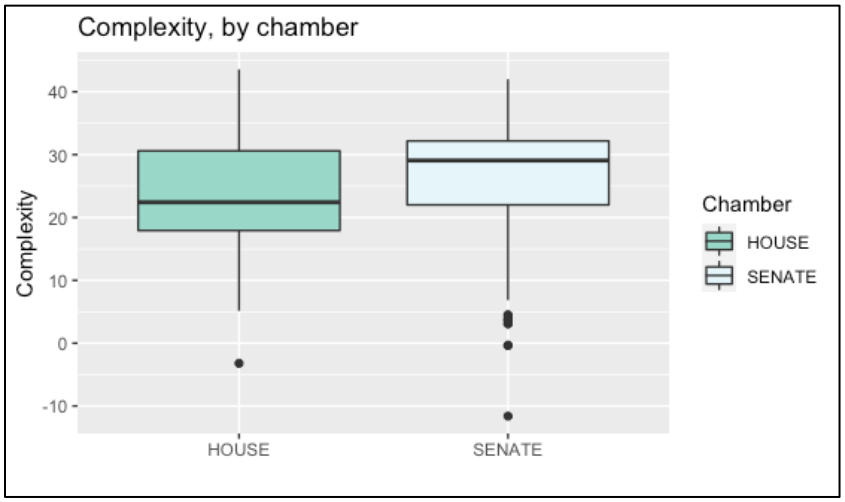


Figure 5: Complexity by Chamber, 109th Congress

Note: Complexity is measured by the new dictionary-complexity approach detailed in this paper.

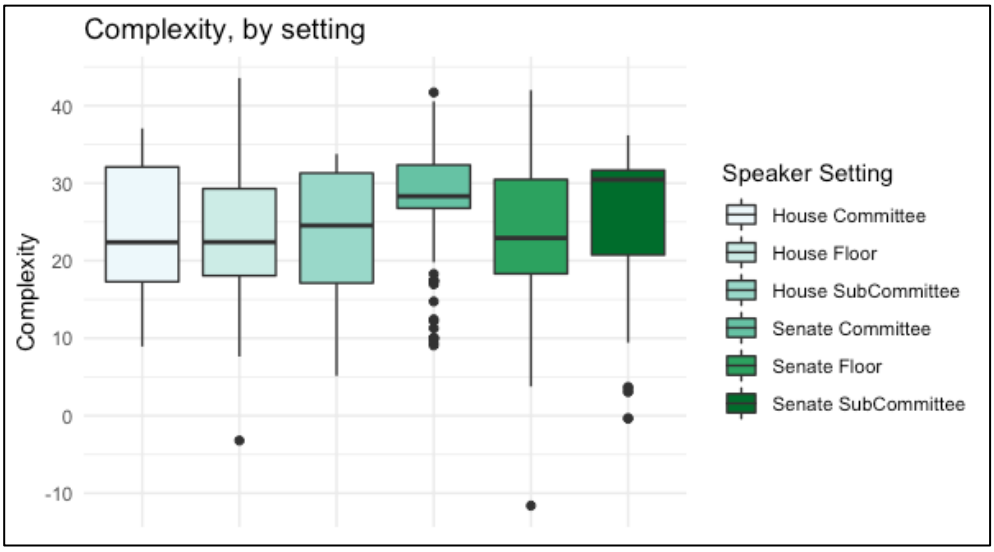


***Figure 6*: Complexity by speaker setting, 109th Congress**

Note: Complexity is measured by the new dictionary-complexity measure detailed in this paper.

We can also divide speaker by party. Figure 7 compares the complexity of speaker across Democratic and Republican members of Congress and administrative officials, and non-partisan witnesses. While the average complexity of the three groups is similar, the nonpartisan witnesses have the most complex speakers in the dataset.

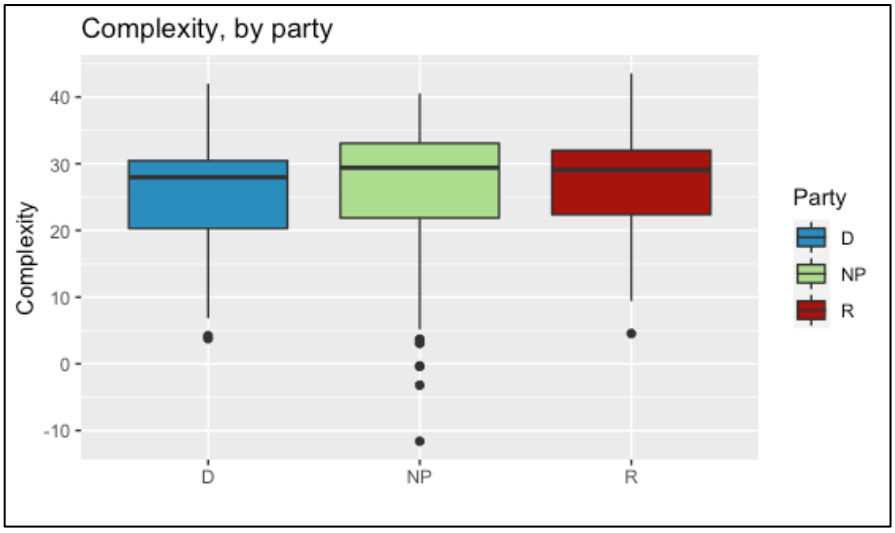


Figure 7: Complexity by speaker party, 109th Congress
Note: Complexity is measured by the new dictionary-complexity measure detailed in this paper. Republican (R) and Democratic (D) speakers contain members of Congress and administrative officials. Non-partisan (NP) speakers are committee witnesses.

These results are somewhat surprising given my expectations on the variability of complexity by chamber, setting, and even political party. However, that there is so much similarity across these various groups could also indicate that word usage throughout Congress is much more permeable that perhaps previously considered. Preliminary evidence indicates that words used and introduced in congressional committees are often used on the House Floor. The similarity among complexity score could be a reflection of this mirrored language. However, future research should consider a more in-depth look at how word usage differs by committee or ideology.

In addition to descriptive comparisons, the complexity score can also be used for more advanced computational inquiry such as regression analysis. By serving as a continuous variable,

we can see which factors lead to increased complexity. While future applications would use this measure as an independent variable to capture congressional expertise, for demonstrative purposes, Table 2 presents two models predicting the likelihood of document complexity using complexity scores as a dependent variable, and document variables as independent variables. The baseline documents for Table 2 are closing statements in committee hearings. Closing statements are generally formalities and summaries of the hearings. As such, other texts are rated as more complex. Spoken text—like opening statements, closing statements, and Q&A are more simplistic than the submitted documents for the congressional record, such as submitted Q&A and prepared statement. However, floor speeches are comparatively complex, indicating that they are used beyond constituent communication. This calls into question earlier hypotheses about the purpose of floor speeches—their higher complexity score indicates that members are speaking about policy, perhaps to other members of Congress, rather than constituents.

Overall, compared to the Dale-Chall estimate, the dictionary-complexity measure produces a lower estimate (and thus lower complexity), on average. For example, while prepared statements from congressional hearings are rated at a Dale-Chall estimate of 24.509 on average, the dictionary-complexity estimate places those same documents at 11.345. Without dictionaries of familiar words, Dale-Chall estimates consider these documents at a professional, grad level. However, by taking into consideration the familiarity of common-used policy words, the new measure appears to more accurately capture the complexity of congressional documents. In addition, the tighter range of the dictionary estimates indicates potential for it to be a more accurate measurement.

Table 2: Statement Type, 109th Congress

Variable	Dictionary estimate	Dale-Chall estimate
Intercept	29.646***	27.941***
	(2.079)	(2.316)
Opening Statement (hearing)	-1.300	-1.224
	(2.122)	(2.364)
Prepared Statement (hearing)	-5.137*	-16.596***
	(2.220)	(2.473)
Q&A (hearing)	-1.714	-2.176
	(5.325)	(2.384)
Submitted Q&A (hearing)	-4.056	-17.873***
	(2.312)	(2.575)
Floor Speech	-5.117***	-5.358*
	(2.133)	(2.376)
Submitted for Record (hearing)	-11.676	-14.145***
	(7.779)	(2.947)
Submitted for Record (floor)	-12.541***	-14.145***
	(2.645)	(2.947)
N	952	
Adjusted R-squared	0.337	
F-statistic		

*Note: Standard errors are reported in parenthesis, with confidence levels reported as follows: *p<0.1; **p<0.05; ***p<0.01. The dependent variable is the dictionary-complexity score.*

I also consider how the speaker or setting impacts the text complexity score. Table 3 considers the role of oversight and nomination hearings, compared to the baseline of policy hearings (N=686). During the 109th Congress, the White House (Bush), Senate, and House were under unified Republican control—but that did not lead lawmakers to ignore their Article One responsibilities. Initial findings in Table 3 indicate that oversight hearings were comparatively complex—rating 9.095 points more complex than a standard committee hearing, on average. Whether this is in part due to shared policy beliefs by the chair and witness and the desire for substantial policy discussion, or "respective oversight" or prior administrations (McDonald and McGrath 2016) requires additional analysis, but it does indicate that oversight hearings, even under unified government, are used legitimate policy discussions. But future research should consider how this changes across congressional session, taking into consideration the role of

minority or majority status. For example, given these initial findings, there is reason to wonder if cross-party oversight hearings lead to more partisan (and less complex) oversight hearings.

Table 3: Text complexity, by hearing type, 109th Congress

Variable	Dictionary estimate	Dale-Chall estimate
Intercept	27.8169***	23.256***
	(0.284)	(0.324)
Oversight Hearing	-9.095***	-11.556***
	(1.742)	(2.576)
Nomination Hearing	-0.952	1.372
	(1.045)	(1.54)
N	686	686
Adjusted R-squared	0.0632	0.030

*Note: Standard errors are reported in parenthesis, with confidence levels reported as follows: *p<0.1; **p<0.05; ***p<0.01. The dependent variable is the dictionary-complexity score.*

Lastly, in Table 4, I consider how the position of the speaker impacts the complexity score in committee hearings. Democratic rank-and-file members are held as the baseline. In this comparison there are significant differences between the dictionary-complexity and Dale-Chall measures. In our dictionary-complexity measure, as expected, witnesses are considered more complex than legislators. This is logical given that they are, after all, the external experts. However, committee chairs are slightly less complex than rank-and-file member. This could be due to the over-representation of their dialogue in the dataset, given that committee chairs are present at nearly every committee hearing. However, when compared with floor discussions, this is unlikely to be the source of the estimate. More likely, this is due to the responsibilities of committee chairs and ranking members. Committee leaders are the main questioners in committee hearings. In several hearings in our dataset, they are the solo questioner. In addition to questioning, they also introduce the witnesses and provide a basic overview of the hearing. House Agriculture Chairman Conway (R-TX) even opens each hearing with a prayer—familiar words with low complexity. So, while their vernacular is slightly over-represented in the

dictionary, their questions and opening statements are also the most basic. However, these findings are not statistically significant at traditionally accepted levels of significance.

The Dale-Chall estimates have greater swings in their levels of complexity, depending on speaker type. The baseline is more complex than the dictionary-complexity measure but witnesses and committee chairs are considered easier to interpret than the scores they receive under the dictionary estimates. In particular, the finding that witnesses—undoubtedly the most complex speakers in our dataset—are more simple than democratic rank-and-file members reflects potential inaccuracies with using unspecified complexity models. While table 4 does raise future normative questions about partisan differences of complexity and speaking habits, it does emphasize the need for policy-specific dictionary models when considering measures of speaker complexity in the U.S. Congress.

Table 4: Impact of Speaker and Setting on Document Complexity

Variable	Dictionary estimate	Dale-Chall estimate
Intercept	27.134***	19.286***
	(0.6065	(0.878)
Committee Chair	0.809	2.522*
	(0.894)	(1.007)
Witness	-0.469	5.524***
	(0.777)	(1.117)
Republican MC	0.699	2.978**
	(7.074)	(1.024)
N	686	686
Adjusted R-squared	0.0696	0.0325

*Note: Standard errors are reported in parenthesis, with confidence levels reported as follows: *p<0.1; **p<0.05; ***p<0.01. The dependent variable is the dictionary complexity score.*

This preliminary analysis indicates that developing dictionaries is a useful method to quantitatively measure political texts. The limited variation in these dictionary-complexity estimates indicate potential for scholars interested in legislative texts and political communication. Furthermore, the continuous output of dictionary-complexity scores makes it

easily applicable to quantitative analysis. For example, future research could interact this dictionary-complexity measure text similarity or the DW-Nominate scores of speakers. In addition, the ease and flexibility of dictionary development presents researchers with an accessible way to evaluate other research questions. Future work could generate ideological dictionaries based on most frequent words by partisans (within or outside of Congress), or a policy-specific sentiment analysis with frequency-based dictionaries. This initial analysis indicates this approach has great potential and ease of use for researchers.

Conclusion

Correctly evaluating the complexity of spoken word and legislative texts of the U.S. Congress holds great importance for scholars. As Congress continues to face a capacity crisis (Drutman et al. 2020), understanding how expertise is developed and disseminated throughout Congress matters deeply to understand the policymaking process at large. Yet existing measures of congressional expertise are time-consuming, or blunt measures like years of seniority. And although text complexity measures do provide some insight to speaker expertise, existing measures rely on measures that fail to take into consideration both the specificities of the policy dialogue and the speaker's familiarity with otherwise complex language. Although some texts have had particular success with supervised scaling, such as party manifestos (Lowe et al. 2011), this application is not universal. And more complex training models used to predict document complexity requires *a priori* information that certain texts, like committee hearings and legislation, cannot leverage (Laver 2014, Laver et al. 2003).

This research has provided a new way to measure congressional expertise by developing a policy-specific dictionary based on word frequency. This approach has several benefits. First, it

offers researchers flexibility in both the development of the text corpus and dictionary. Dictionaries can be formed within specific policies and time periods, or they can be formed from pooled corpora. The malleability of dictionary development by word frequency threshold and speaker attribute also gives scholars flexibility, depending on their research question—whether it be across time, ideology, or committee setting. In addition to its flexible development, it is an accessible method that can be easily applied, circumventing time-consuming or costly methods often required for text analysis. The result is an easy-to-understand complexity score that can be interacted with other variables and research questions. In addition to the ease of use, this approach results in a more accurate complexity score for policy-specific texts. In our Farm Bill sample texts, the dictionary-complexity approach results in a more accurate, and tighter range of scores that more accurately captured the complexity of words spoken, as well as differences between speakers and setting. Future work should apply this method to other policy areas, and across committee setting. Researchers should consider how the development of policy-specific dictionaries could paint a more accurate picture of speaker ideology or sentiment.

Lastly, this research does not consider beyond speculation the role of the audience, focusing instead on what complexity can tell us about the speaker. But there are several interesting lines of research, considering what complexity measures can us about an intended audience. Congressional text is an opportunity for members not only frame policy narratives to each other, but their constituents as well. Members and their staff will strategically consider how best to communicate with their given audience. Thus, language complexity is both a reflection of congressional expertise and a strategic decision (Grimmer 2013). Just as the federal government has readability requirements for its public-facing documents, members also know that they must communicate clearly with constituents, which means keeping text simple and clear. At the same

time, increased complexity can also be used as a tool to mask policy positions or project expertise. Thus, the complexity of text is likely to reflect legislative strategy. Future iterations should consider this theoretical angle, as well as how how direct constituent messaging, such as press releases, tweets, or newsletters, compare with chamber texts.

In an era of gridlock and partisanship, the spoken and written word of politicians becomes more essential in order to fully understand legislative debates. Committee hearings and floor speeches are a window into the expertise, policy positions, and communication strategies of legislators. This approach makes it comparatively easy to consider how expertise manifests itself throughout Congress. This insight is important for scholars of Congress who want to understand how members develop policy expertise and express that knowledge with their peers and constituents.

References

Arnold, R. D. (1990). *The logic of congressional action*. Yale University Press.

Benoit, K., Munger, K., & Spirling, A. (2019). Measuring and explaining political sophistication through textual complexity. *American Journal of Political Science*, *63*(2), 491-508.

Benoit, K., Laver, M., & Mikhaylov, S. (2009). Treating words as data with error: Uncertainty in text statements of policy positions. *American Journal of Political Science*, *53*(2), 495-513.

Bischof, D., & Senninger, R. (2018). Simple politics for the people? Complexity in campaign messages and political knowledge. *European Journal of Political Research*, *57*(2), 473-495.

Cann, D. M., Goelzhauser, G., & Johnson, K. (2014). Analyzing text complexity in political science research. *PS: Political Science & Politics*, *47*(3), 663-666.

Crosson, J. M., Furnas, A. C., Lapira, T., & Burgat, C. (2019). Partisan Competition and the Decline in Legislative Capacity among Congressional Offices. *Legislative Studies Quarterly*.

Curry, J. M. (2019). Knowledge, Expertise, and Committee Power in the Contemporary Congress. *Legislative Studies Quarterly*, *44*(2), 203-237.

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

Diermeier, D., Godbout, J. F., Yu, B., & Kaufmann, S. (2012). Language and ideology in Congress. *British Journal of Political Science*, 31-55.

Esterling, K. M. (2007). Buying expertise: Campaign contributions and attention to policy analysis in congressional committees. *American Political Science Review*, 93-109.

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

Fenno Jr, R. F. (1966). *The power of the purse: Appropriations politics in Congress*. Little, Brown.

Grimmer, J. (2013). *Representational style in Congress: What legislators say and why it matters*. Cambridge University Press.

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, *21*(3), 267-297.

Guo, L., Vargo, C. J., Pan, Z., Ding, W., & Ishwar, P. (2016). Big social data analytics in journalism and mass communication: Comparing dictionary-based text analysis and unsupervised topic modeling. *Journalism & Mass Communication Quarterly*, *93*(2), 332-359.

Hammond, S. W. (2001). *Congressional caucuses in national policymaking*. JHU Press.

Haselmayer, M., & Jenny, M. (2017). Sentiment analysis of political communication: combining a dictionary approach with crowdcoding. *Quality & quantity*, *51*(6), 2623-2646.

Hall, R. L. (1998). *Participation in congress*. Yale University Press.

Kiewiet, D. R., & McCubbins, M. D. (1991). *The logic of delegation*. University of Chicago Press.

Laver, M., Benoit, K., & Garry, J. (2003). Extracting policy positions from political texts using words as data. *American political science review*, *97*(2), 311-331.

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

Lowe, W., Benoit, K., Mikhaylov, S., & Laver, M. (2011). Scaling policy preferences from coded political texts. *Legislative studies quarterly*, *36*(1), 123-155.

MacDonald, J. A., & McGrath, R. J. (2016). Retrospective congressional oversight and the dynamics of legislative influence over the bureaucracy. *Legislative Studies Quarterly*, *41*(4), 899-934.

Maltzman, F. (1998). *Competing principals: Committees, parties, and the organization of Congress*. University of Michigan Press.

McGann, J. G. (2016). *The Fifth Estate: Think tanks, public policy, and governance*. Brookings Institution Press.

Miler, K. C. (2011). The constituency motivations of caucus membership. *American Politics Research*, *39*(5), 885-920.

Muddiman, A., McGregor, S. C., & Stroud, N. J. (2019). (Re) claiming our expertise: Parsing large text corpora with manually validated and organic dictionaries. *Political Communication*, *36*(2), 214-226.

Owens, R. J., & Wedeking, J. P. (2011). Justices and legal clarity: Analyzing the complexity of US Supreme Court opinions. *Law & Society Review*, *45*(4), 1027-1061.

Rheault, L., & Cochrane, C. (2020). Word embeddings for the analysis of ideological placement in parliamentary corpora. *Political Analysis*, *28*(1), 112-133.

Rice, D. R., & Zorn, C. (2013). Corpus-based dictionaries for sentiment analysis of specialized vocabularies. *Political Science Research and Methods*, 1-16.

Rich, A. (2005). *Think tanks, public policy, and the politics of expertise*. Cambridge University Press.

Silge, J., & Robinson, D. (2016). tidytext: Text mining and analysis using tidy data principles in R. *Journal of Open Source Software*, *1*(3), 37.

Tauberg, M. (2019). How Smart is Your News Source? *Towards Data Science: Medium.* https://towardsdatascience.com/how-smart-is-your-news-source-1fe0c550c7d9

Wilkerson, J., & Casas, A. (2017). Large-scale computerized text analysis in political science: Opportunities and challenges. *Annual Review of Political Science*, *20*, 529-544.

Appendix

Flesch Reading Ease score:

$$FRE = 206.835 - 1.015\left(\frac{total\ words}{total\ sentences}\right) - 84.6\left(\frac{total\ syllables}{total\ words}\right)$$

Tf-Idf approach for word frequency, results:

$$idf(term) = \ln\left(\frac{n\ documents}{n\ documents\ containing\ term}\right)$$

Congress dictionary: *"committee", "house", "congress", "hearing", "session", "washington", "dc", "printing", "office", "government", "administration", "mr", "senator", "u.s", "chairman", "graphic", "statement", "bill", "american", "law", "america", "subcommittee".*

Figure 1A: Word Clouds for Congressional Corpora, 109th – 115th Congress

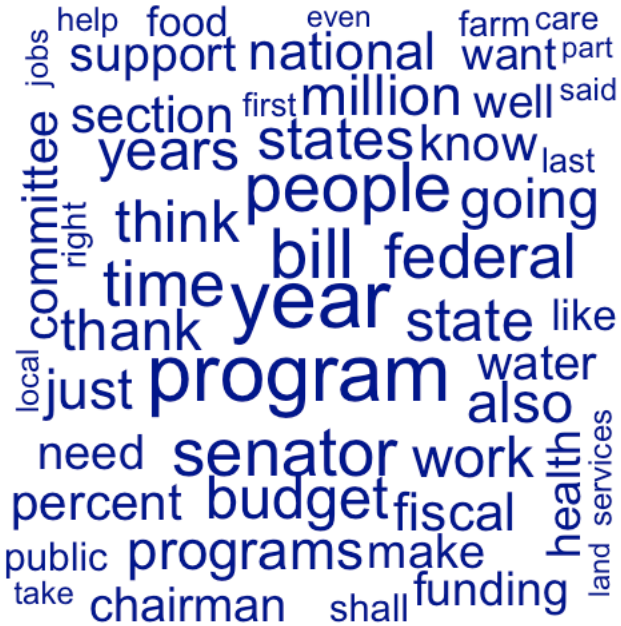
109th Congress

110th Congress

111th Congress

112th Congress

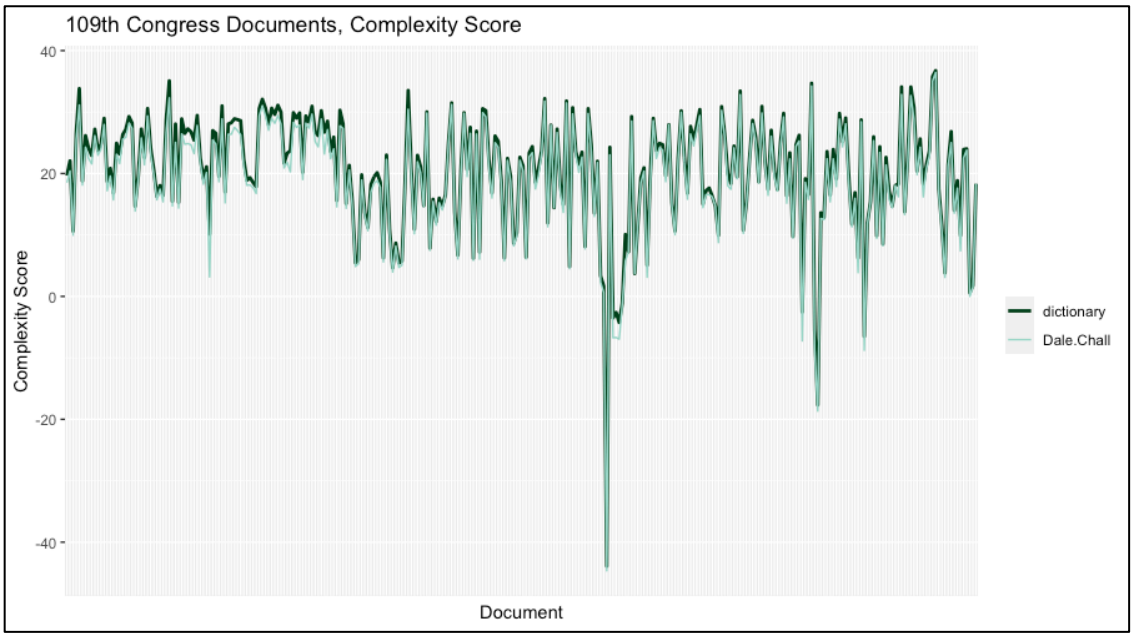
113th Congress

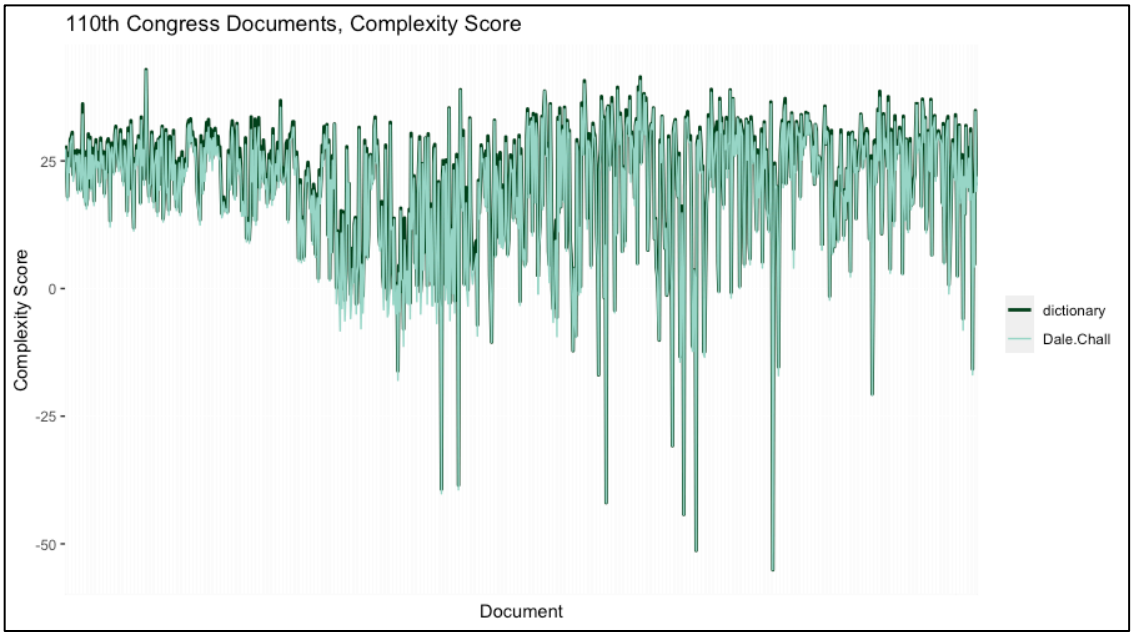


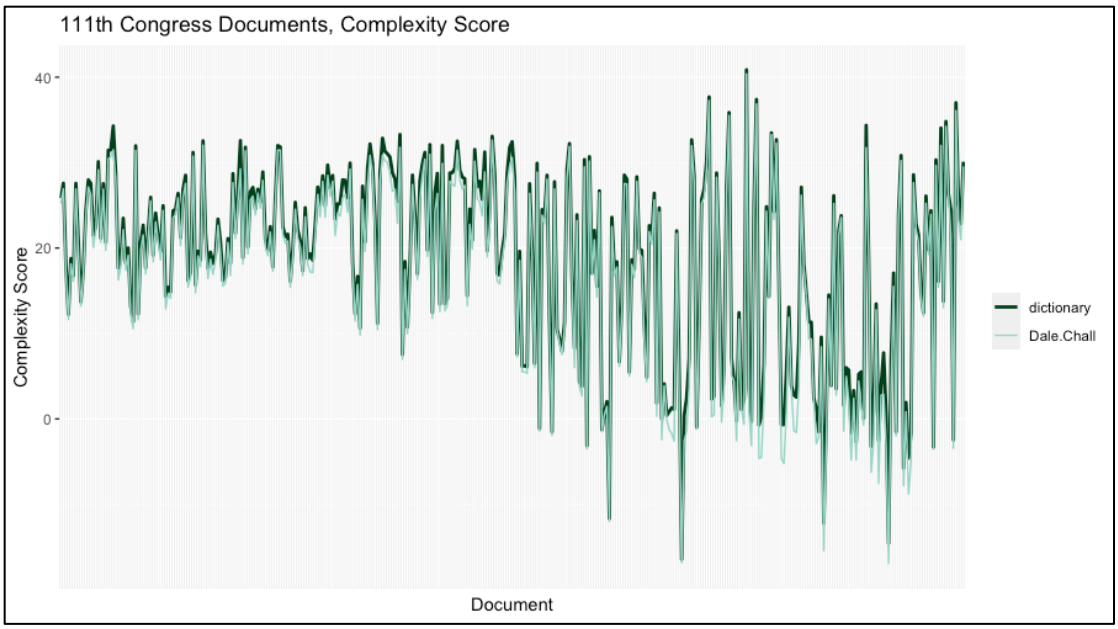
114th Congress

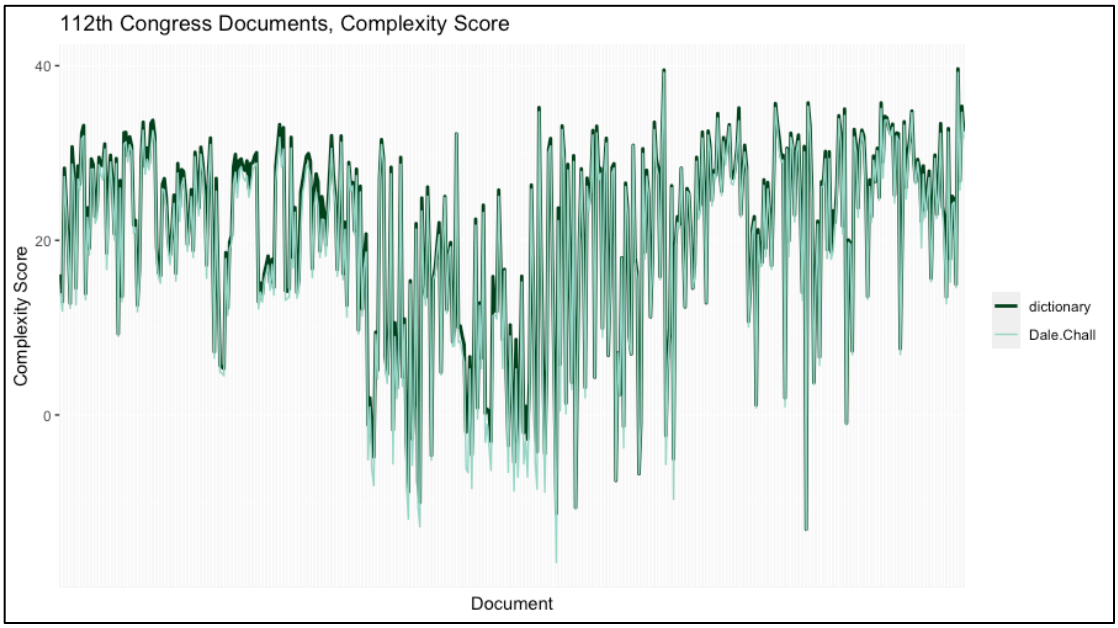
115th Congress

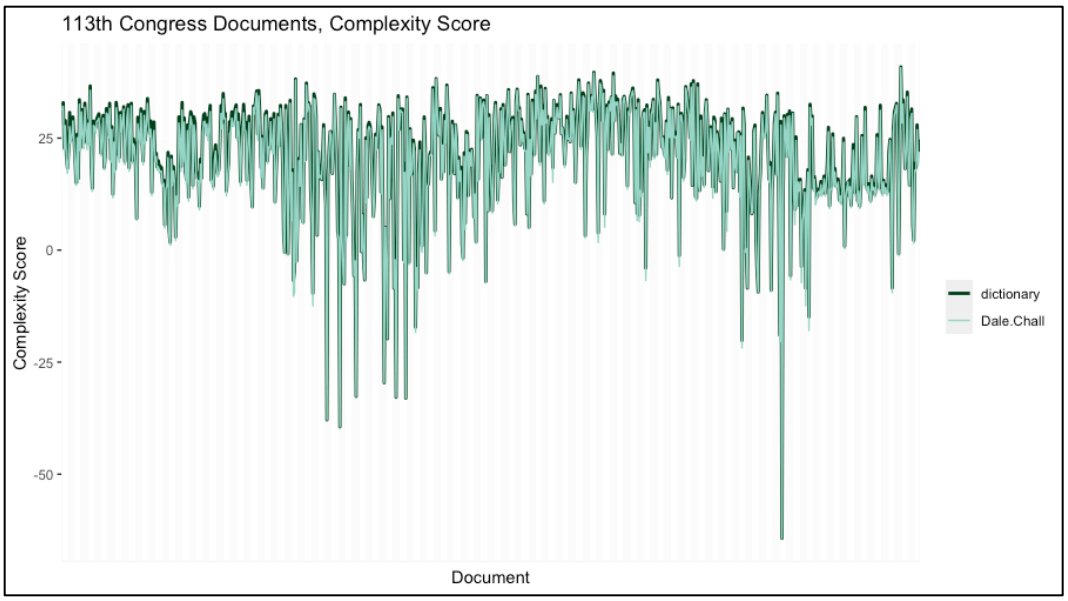
Figure 2A: Complexity score comparison, 109th-115th Congress

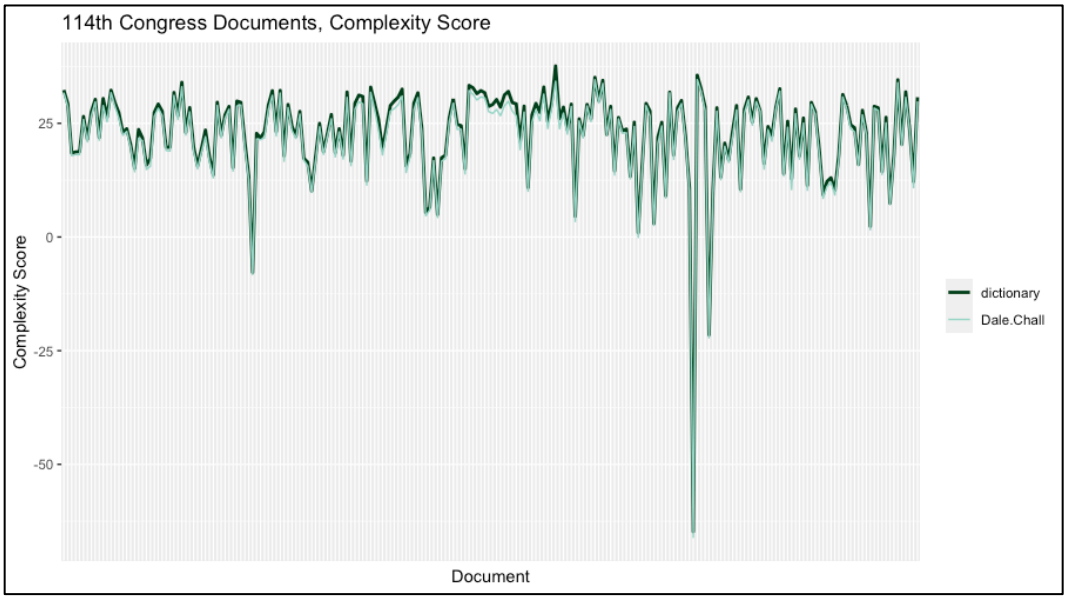





111th Congress Documents, Complexity Score
40
20
0
Complexity Score
Document
dictionary
Dale.Chall


112th Congress Documents, Complexity Score
40
20
0
Complexity Score
Document
dictionary
Dale.Chall


113th Congress Documents, Complexity Score
Complexity Score
25
0
-25
-50
Document
dictionary
Dale.Chall


114th Congress Documents, Complexity Score
Complexity Score
25
0
-25
-50
Document
dictionary
Dale.Chall

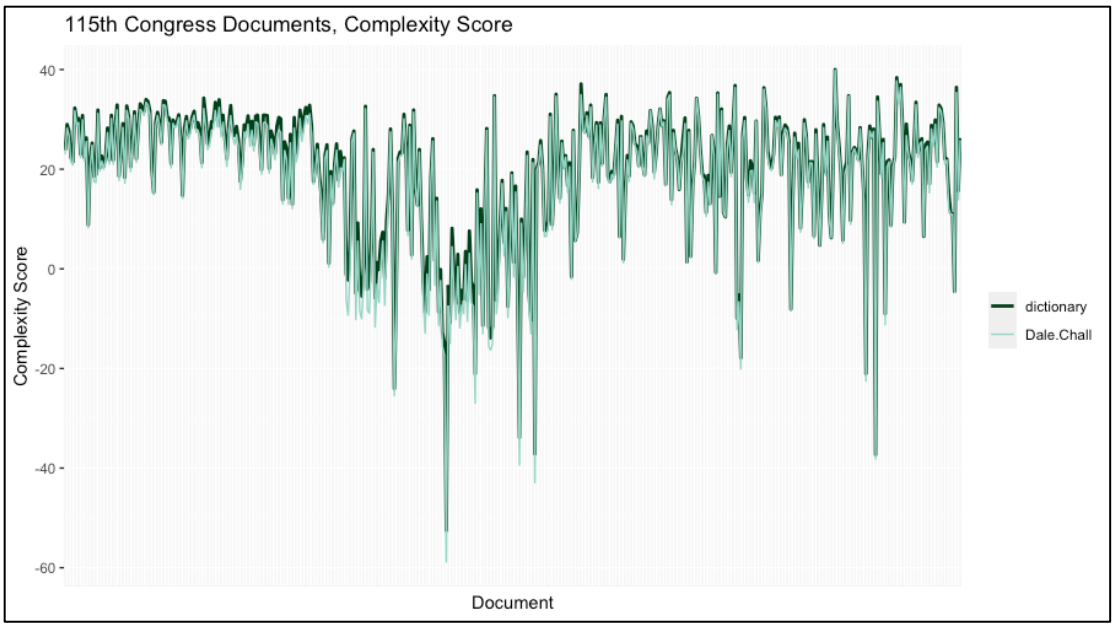
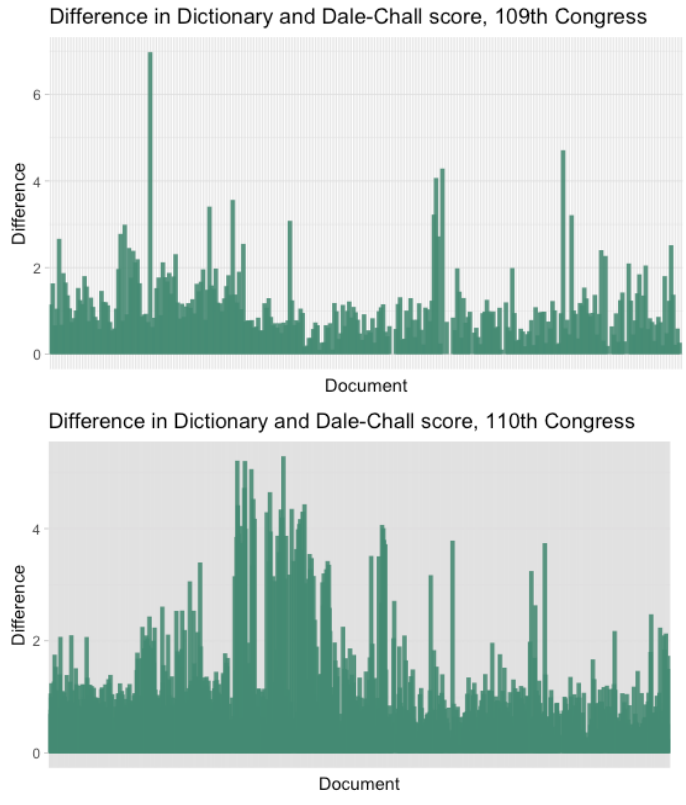
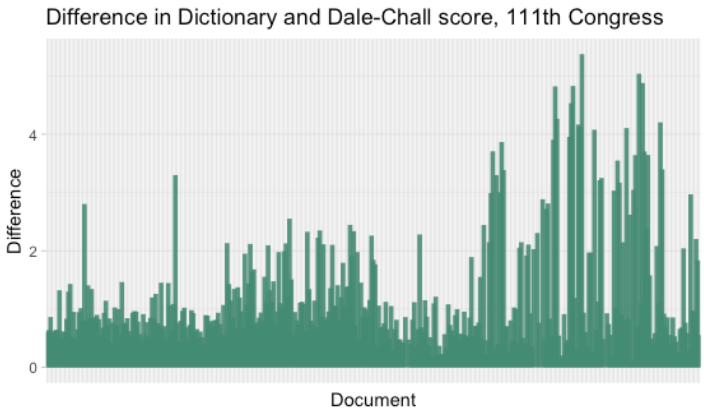
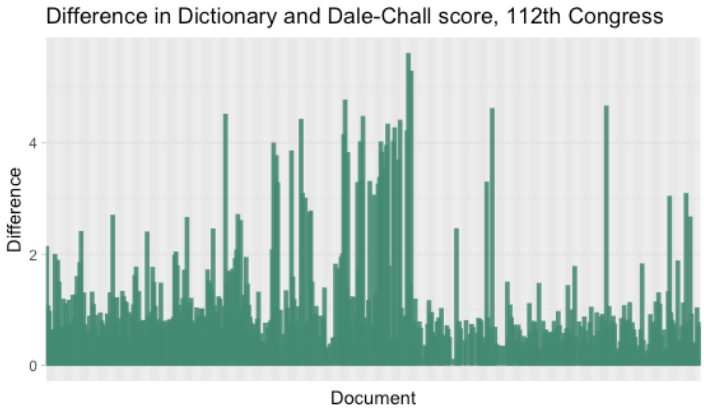


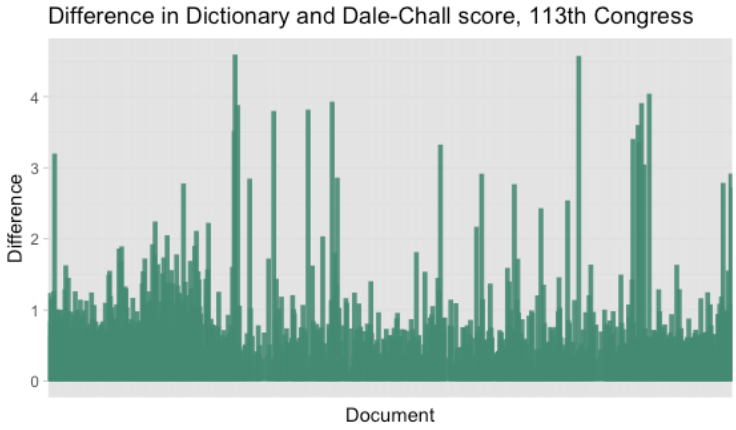
Figure 3A: Difference in Dictionary-complexity and Dale-Chall measurement, 109th – 115th Congress

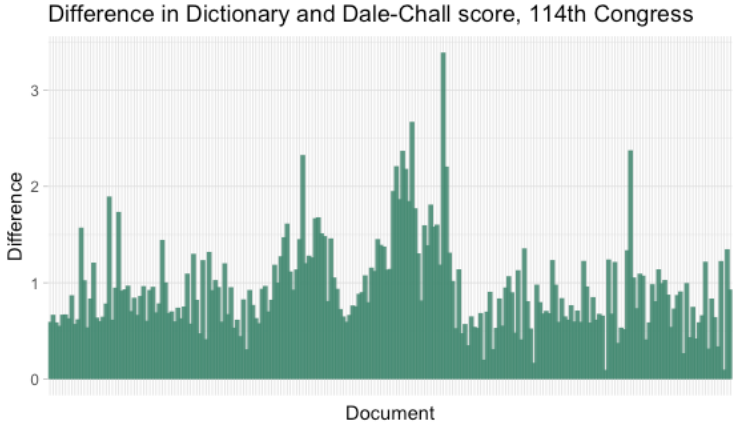
Note: The difference is calculated by subtracting the Dale-Chall measurement from the new, dictionary-complexity measurement.

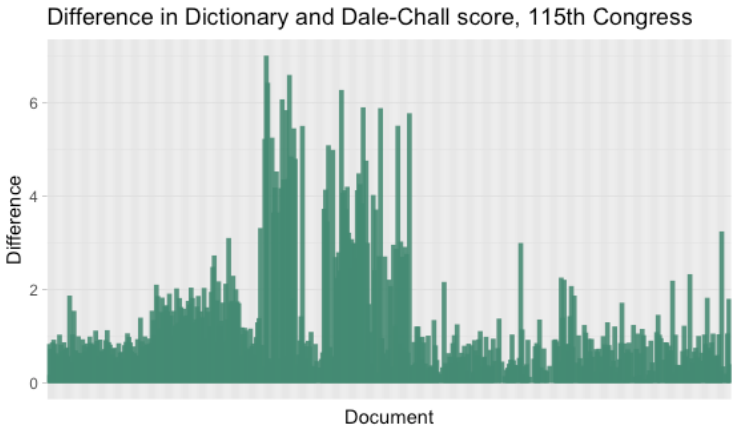



Difference in Dictionary and Dale-Chall score, 111th Congress
Difference
0
2
4
Document


Difference in Dictionary and Dale-Chall score, 112th Congress
Difference
0
2
4
Document


Difference in Dictionary and Dale-Chall score, 113th Congress
Difference
0
1
2
3
4
Document


Difference in Dictionary and Dale-Chall score, 114th Congress
Difference
0
1
2
3
Document


Difference in Dictionary and Dale-Chall score, 115th Congress
Difference
6
4
2
0
Document