December 2021

Perspective

EXPERT INSIGHTS ON A TIMELY POLICY ISSUE

MAREK N. POSARD, HILARY REININGER, CHRISTIAN JOHNSON

# An Exploratory Analysis of Trends in Text Data from Congressional Oversight Hearings

In general, the U.S. Congress exercises oversight over federal agencies (including the U.S. Department of Defense [DoD]) through various committees (Halchin and Kaiser, 2012). We assume that trends in what is said in these committees could signal the emergence of salient policy issues for policymakers. For example, if members of the House and Senate Armed Services Committees (HASC and SASC, respectively) talk more and more about diversity in the military over several years, then it might suggest that diversity-related issues are becoming more salient. This trend could be a signal for policymakers at various levels within the Pentagon to prepare for questions from Congress about these issues. Leaders of the military branches might want to prepare plans for increasing the diversity of their recruitment pipelines. The Joint Advertising Market Research and Studies program might want to plan for ways to increase the diversity of recruits through joint marketing communications. The Military Health System might want to prepare for concerns surrounding disparities in health outcomes by ascribed features of personnel. If policymakers across DoD can forecast what policy issues are a priority for Congress, it could give them more lead time to prepare thoughtful answers to oversight inquiries, create plans for addressing potential deficiencies in their programs, and improve their overall relationships with legislators and, ultimately, the public they serve.

To this end, we developed a workflow that draws on various tools for acquiring and organizing large volumes of data from HASC and SASC. In this Perspective, we describe a proof of concept for how to acquire and begin analyzing text data for policy analysis. We sketch out a workflow that uses a set of independent resources to analyze publicly available text data from congressional hearings. One could use this workflow to develop a more sophisticated toolkit for analyzing congressional text data.[1] This project is exploratory research on a limited budget to determine the feasibility of using a limited set of analytics on congressional text data for policy analysis.

Research has identified at least two broad ways in which the U.S. Congress fulfills its policy oversight duties (McCubbins and Schwartz, 1984; and Balla and Deering, 2013). The first is "police-patrol" oversight, whereby Congress samples various policy activities within the Executive Branch and looks for potential areas of concern. The other technique is "fire-alarm" oversight, whereby Congress responds to concerns by the public or interest groups on specific issues. Studies find that Congress spends most of its time in hearings conducting police patrols. In 2004, for example, about 89 percent and 95 percent of hearings were categorized as police patrols in HASC and SASC, respectively (Balla and Deering, 2013, p. 34).

The U.S. government documents much of what is said in these congressional hearings. For example, the U.S. Government Publishing Office (GPO) published transcripts for most unclassified hearings for each of the subcommittees and committees within the U.S. House of Representatives and the U.S. Senate. If we assume that the majority of congressional oversight involves routine police patrols, it follows that it would be helpful for policymakers to understand both where these patrols have been and where they are currently located to forecast where they might be headed in the future. Specifically, we hypothesize that the volume of text data from congressional oversight hearings could help policymakers better understand the interests of Congress.

This short Perspective is a proof of concept of a workflow for how to acquire and analyze some of these text data. It has four parts, including this introduction. In the next section, we describe the acquisition of two sources of data: unclassified transcripts from HASC and SASC and answers to advance policy questions (APQs) that are submitted to these committees by political nominees. In the third section, we describe some of the trends we identified in these data. Finally, we conclude with a discussion of next steps for this line of work.

### Abbreviations

| | |
|---|---|
| APQ | advance policy question |
| DoD | U.S. Department of Defense |
| GPO | U.S. Government Publishing Office |
| HASC | House Armed Services Committee |
| NLP | natural language processing |
| OUSD(P&R) | Office of the Under Secretary of Defense for Personnel and Readiness |
| SASC | Senate Armed Services Committee |

## Acquisition of Text Data

We identified two sources of text data for this proof of concept. In the first category are transcripts from unclassified hearings for HASC and SASC. Most of the transcripts are publicly available in the GPO's online portal, which hosts text documents of these hearings and other forms of digital content (GPO, undated-a). The text from these hearing transcripts is organized in a similar manner: Each transcript has the title of the hearing in capital letters, dates of the hearing, number for the congressional session, names of attendees, and details about who said what during these meetings.

The second source of text data is answers to APQs to which presidential nominees responded during their confirmation hearings. Specifically, we analyzed the answers that the former Under Secretary of Defense for Personnel and Readiness Matthew Donovan gave to APQs during his confirmation hearing on March 10, 2020. These APQs are organized by key domain (e.g., recruiting and retention, reserve components, military compensation), specific question asked, and specific response given. In the next section, we describe some of the patterns we identified from these text data.

## Analysis of Text Data

To begin our analysis, we created a script that queried all publicly available hearing transcripts for HASC and SASC from 2005 to 2018 in the GPO portal (GPO, undated-a).[2] We selected this date range because there were complete transcripts available when we queried these data during summer 2019. Our programmer created a means to automatically tag some common features from these hearing transcripts; specifically, whether the hearing was from HASC or SASC and the congressional session in which the hearing was held. We then analyzed these text data using RAND-Lex, the RAND Corporation's in-house text analysis software.[3]

The volume of text data from congressional oversight hearings could help policymakers better understand the interests of Congress.

### Congressional Hearing Transcripts

We organized the HASC and SASC hearing transcripts by congressional year, from the 109th to the 115th Congress (2005–2007 to 2017–2018). We then uploaded these transcripts to RAND-Lex and tried several different analytical approaches.[4] We tried topic modeling, which is a statistical modeling technique that groups words from the text and allows a researcher to identify topics from those groups. However, topic modeling requires a minimum number of documents of a minimum length. Although our documents were long enough, we did not have a large enough number of documents for either HASC or SASC. Each document could have been split into several documents to make topic modeling possible, but we lacked the resources to split our text data into discrete documents and were therefore unable to use topic modeling to find meaningful results.[5]

For each keyness test (which we describe further in the next paragraph), a corpus of one session of either HASC or SASC hearings was compared with a corpus of all of the other hearings in either HASC or SASC. For example, the hearings from HASC's 111th congressional session (2009–2010) were compared with hearings from the rest of the House's congressional sessions in our sample.

The results of keyness testing are words that were either overpresent or underpresent along with two statistical measures of interest to this study: the log-likelihood and percentage difference. Log-likelihood is a measure of statistical significance. Measures with scores that are higher than 11 are statistically significant (99.9th percentile; $p < 0.001$), while measures in the hundreds or thousands are even more significant and very meaningful (Cox et al., 2018; Hardy, 2007). The percentage difference measures effect size. It compares how prevalent a word is in one corpus with how prevalent it is in a baseline corpus. In this study, *percentage difference* is the effect size of the over- or underpresence of words (i.e., the prevalence of words) in one congressional session compared with a baseline of all of the other sessions in the House or Senate. If a word has a large effect size, there is a large difference between how present or absent that word is in that session compared with all of the other sessions.

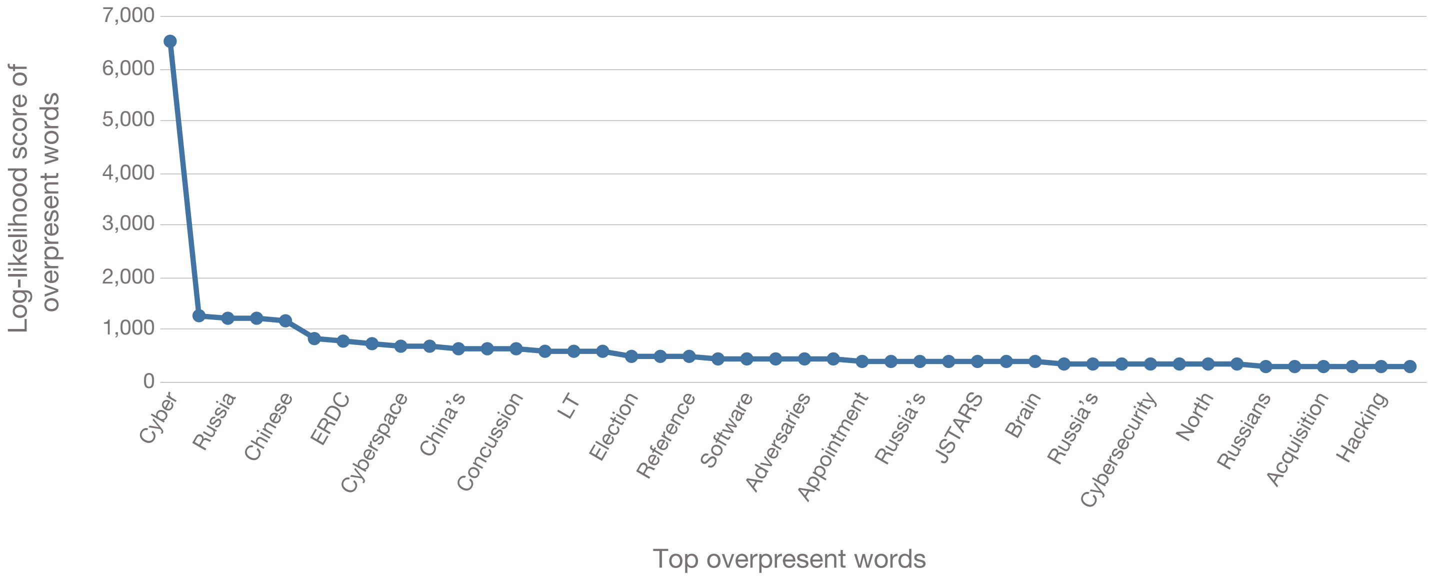
We focused on trends related to the presence of two keywords: *cyber* and *health*.

When a word has a high log-likelihood and a high absolute value percentage difference, it is both significant and large in its effect. For the keyness results of each session, we sorted the words by significance (log-likelihood) and then focused on the top 100 ranked words. We chose the number 100 arbitrarily as a large but manageable number of words that we could analyze and search for patterns. Figure 1 displays an example of some of the top overpresent words for SASC during the 115th Congress.[6] The $x$-axis shows the top overpresent words for the 115th Congress for SASC compared with all other years of SASC. The $y$-axis shows the log-likelihood score, a measure of statistical significance.

Next, we compared the overpresence of keywords for all of the HASCs and SASCs for each year and selected keywords that (1) were relevant to manpower and personnel policies and (2) were overpresent in HASC or SASC for at least two years. Specifically, we focused on trends for two keywords to illustrate this proof of concept: *cyber* and *health*. We selected these words because they were prevalent across multiple Congresses, and we determined that they were relevant as a use case.

Figure 2 displays three trends related to mentions of *cyber*. Because this was an overpresent word in SASC from

FIGURE 1
Top Overpresent Words for the Senate Armed Services Committee, 115th Congress




SOURCE: Authors' analysis of HASC and SASC reports, 115th Congress (January 3, 2017, to January 3, 2019), from GPO, undated-a.

NOTES: This exploratory analysis identified words that were grammatically different in the text data (e.g., Russia versus Russia's). We did not combine these words into a single grouping in this figure because of resource constraints. ERDC = U.S. Army Engineer Research and Development Center. JSTARS = Joint Surveillance and Target Attack Radar System. LT = Mostly lieutenant (including lieutenant and lieutenant general), but this abbreviation might capture other specific acronyms.

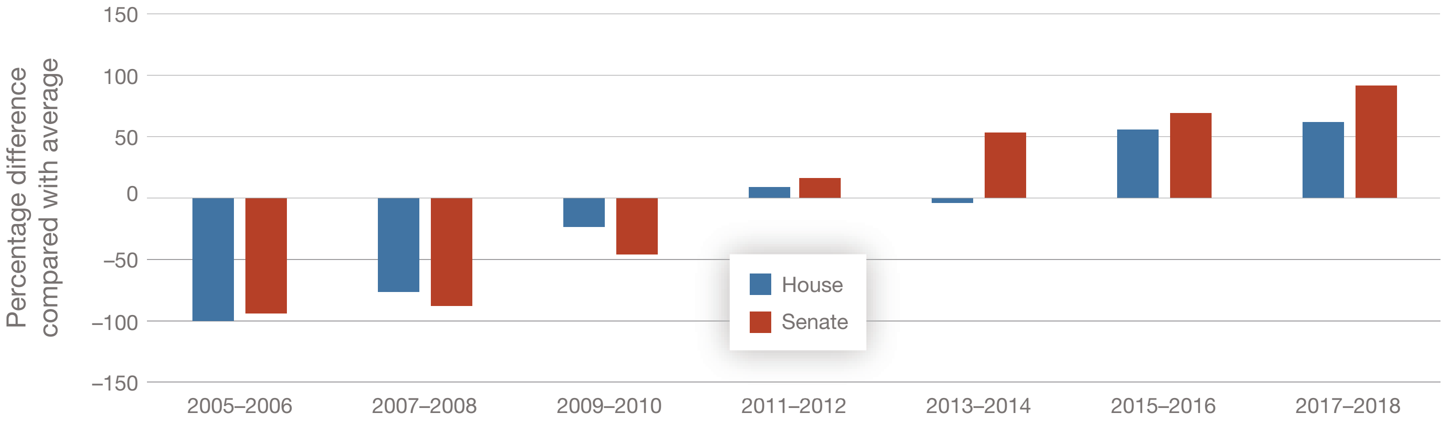
the 115th Congress, it is a useful case for exploring these data further. The blue columns in the figure represent the effect size (percentage difference) of the over- or underpresence of the word *cyber* for a given year in HASC compared with all other years of this committee. Similarly, the red bars represent the effect size of the presence or absence of *cyber* for a given year in SASC compared with all other years for this committee. This figure indicates that, compared with all other congressional years, *cyber* was conspicuously underpresent or perhaps absent from HASC and SASC hearings in 2005 and 2006. However, *cyber* became more present than average for both houses starting in 2011 and continued to rise on average after that year until it was extremely overpresent compared with other years.

Figure 2 highlights two trends: First, the under- or overpresence of mentions of *cyber* in HASC and SASC align over time, and, second, mentions of *cyber* start to become overpresent relative to all mentions within each committee during 2011 and 2012.

Next, we arbitrarily selected a general team—*health*—because military health is a key area of concern for military manpower and personnel policies. Figure 3 shows that, for

FIGURE 2

Trends in Mentions of *Cyber* in the House and Senate Armed Services Committees, 2005 to 2018




SOURCE: Authors' analysis of HASC and SASC reports, 2005 to 2018, from GPO, undated-a.

HASC and SASC, mentions of *health* are overpresent from 2007 to 2010 and begin to dip in later years.

Like in Figure 2, the red and blue bars in Figure 3 represent the average effect size (percentage difference) for the under- or overpresence of the word *health* from 2005 to 2018. This figure suggests that HASC became interested in health in the course of one congressional cycle, from 2005 to 2007. From then on, HASC and SASC sustained similar levels of interest in the topic. *Health* eventually decreased in prominence for both houses in comparison with all other years, as indicated by the large, negative effect size in 2017 and 2018.

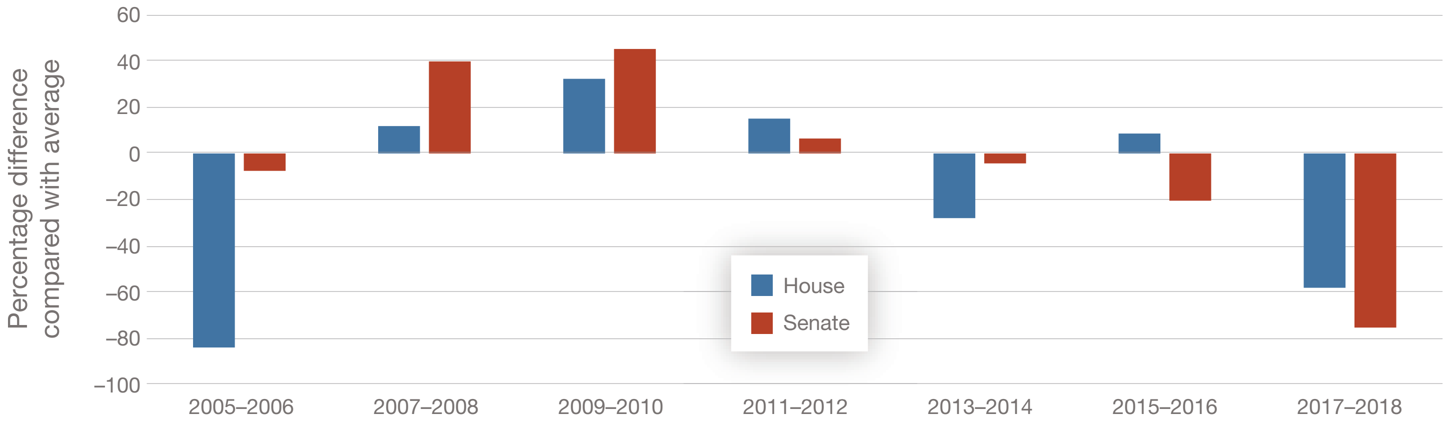
## General Trends in Answers to Advance Policy Questions

We also analyzed the APQs that nominees answer before their confirmation hearings in SASC. As a use case, we examined the APQs that former Under Secretary Donovan answered for his SASC hearing in March 2020 (Presidential nomination 1603, 2020; SASC, 2020).

To conduct this analysis, a RAND programmer began by writing a script that identified, tagged, and queried parts of text within the published document containing the APQs that former Under Secretary Donovan addressed. This script parsed out each domain area of interest listed in the document, separated the questions within each domain, and then separated the answer given to each question by the nominee. The script then queried questions and answers focused on three keywords of interest: *analyze*,

FIGURE 3

Trends in Mentions of *Health* in the House and Senate Armed Services Committees, 2005 to 2018




SOURCE: Authors' analysis of HASC and SASC reports, 2005 to 2018, from GPO, undated–a.

*study*, and *examine*. This script also could be used to identify other keywords of interest. We assumed that if a nominee wanted further analysis of an issue, it was likely to be a salient topic of concern for them.

To illustrate, the following is one example from this APQ document (SASC, 2020, p. 3):

> ***Major Challenges and Priorities***
>
> **If confirmed, what would be your vision for the [Office of the Under Secretary of Defense for Personnel and Readiness (OUSD[P&R])] of today? For the OUSD(P&R) of the future?**
>
> If confirmed, the vision I would set would align to the Secretary's and the imperatives of the National Defense Strategy. Specifically, we need to be able to comprehensively answer the Secretary's question, "are we ready?" by taking advantage of digital modernization and state-of-the-art data management concepts and technologies. This would provide data-driven ***analyses*** to better and more quickly inform Secretary-level decisions required to prepare for and dominate in any future peer conflict. [emphasis added]

The script identified this text as relevant because the answer to this question includes the word *analyses*. The script then parsed the domain name, question text, and nominee's response. We manually summarized the key issue of interest in this answer from the nominee's response. Table 1 shows how we parsed the text in this APQ.

We then counted the number of times this nominee mentioned the keywords of interest by each domain within the APQs. Figure 4 displays these counts, which suggest that overall readiness, recruiting and retention, joint officer

TABLE 1

Domain, Question, and Nominee Response to Advance Policy Question Example

| Domain | Question | Nominee's Response | Key Issue |
|---|---|---|---|
| Major challenges and priorities | 11. If confirmed, what innovative ideas would you consider providing to the Secretary of Defense regarding the organization and operations of the OUSD(P&R)? | If confirmed, I would conduct a comprehensive review of all [personnel and readiness] functions to ensure their continued viability in supporting the National Defense Strategy imperatives, and their effectiveness in enabling and supporting the Military Departments in their Title 10 organize, train, and equip responsibilities. The results of this review would inform my recommendations to the Secretary. | Comprehensive review of all [personnel and readiness] functions |

SOURCE: Authors' analysis of SASC APQs from Under Secretary of Defense for Personnel and Readiness nominee Matthew Donovan (SASC, 2020, p. 3).

management, and the reserve components are salient topics of interest for both members of SASC and this nominee. These findings show how a simple algorithm could quickly parse out key themes from 70 pages of text in this APQ document.

## Next Steps and Conclusion

To summarize, the purpose of this exploratory analysis was to develop a proof of concept for how to acquire and begin analyzing text data for policy analysis. We assume that the majority of congressional hearings are characterized as routine police patrols whereby Congress samples various issues and looks for areas of concern. Research suggests that a small percentage of hearings are fire alarms in which Congress responds to concerns from the public or specific interest groups (Balla and Deering, 2013).

Using this assumption, we analyzed two sources of text data from congressional oversight hearings. The first source involved publicly available hearing transcripts from HASC and SASC from the 109th to the 115th Congresses. The second source drew from answers to the APQs that political nominees submit to SASC. For both types of text data, we discussed how we acquired and analyzed some basic descriptive trends.

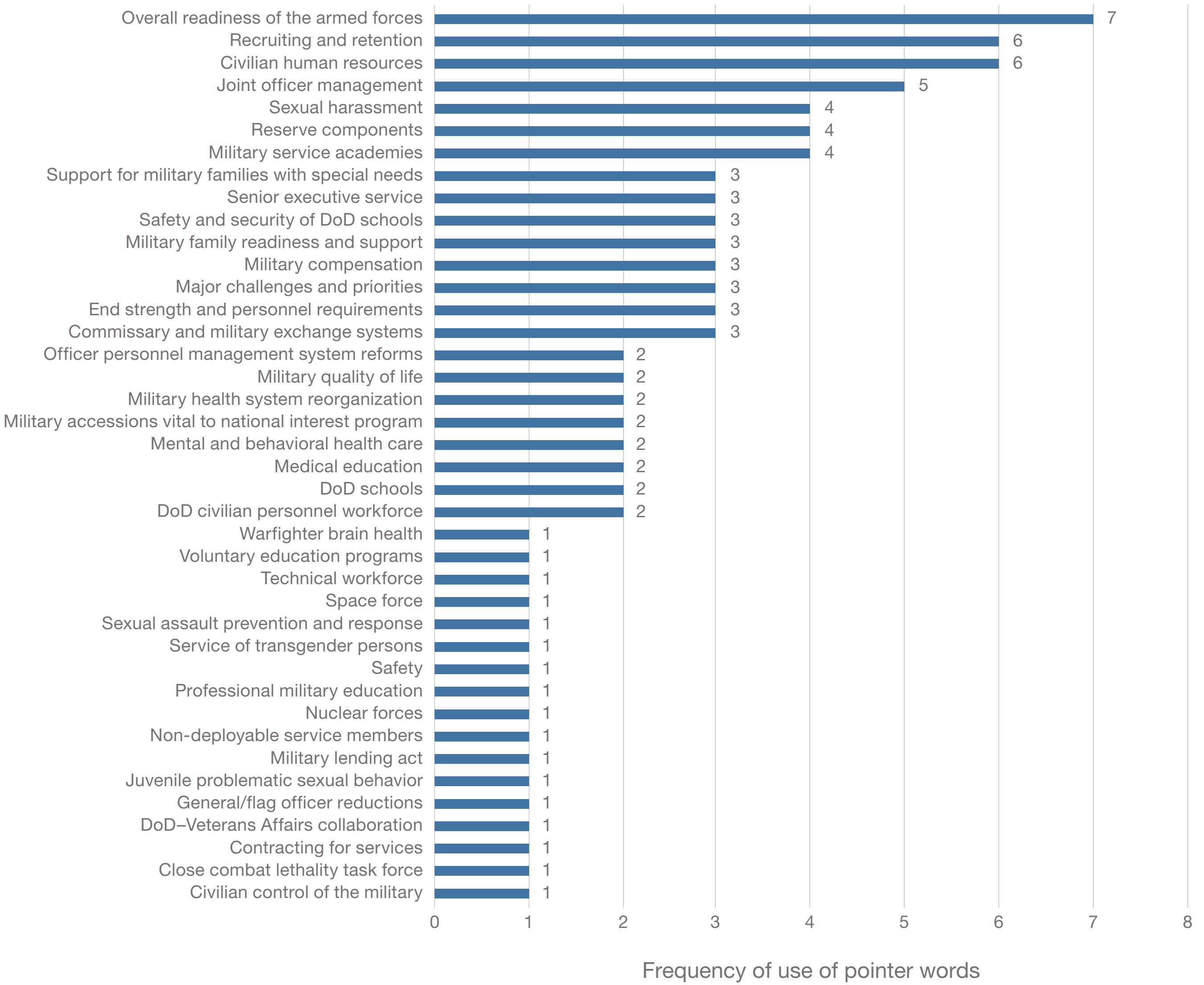
This proof of concept suggests that there are some meaningful patterns in these text data. For the analysis of congressional transcripts, we found increases in mentions of topics related to cyber and health issues within HASC and SASC. The patterns in one chamber appeared to follow those in the other. Furthermore, we explained how a simple script could automatically parse relevant text from APQs and organize that text into a usable format.

### Next Steps

With the large volume of text generated by lawmakers on a daily basis, it is interesting to consider ways in which machine learning and artificial intelligence can augment analysis of congressional data. The field of natural language processing (NLP) has recently come into its own, and NLP methods are capable of understanding rich text-based data in ways that more-traditional techniques could

FIGURE 4

## Frequency of Mentions of *Analyze*, *Study*, or *Examine* in Advance Policy Question Example




SOURCE: Authors' analysis of SASC APQs from Under Secretary of Defense for Personnel and Readiness nominee Matthew Donovan (SASC, 2020).

not. In this section, we describe a handful of potential ways that NLP could be used in analyzing congressional texts.

Topic modeling—identifying what a document is discussing—is a well-known aspect of NLP. Traditional techniques, such as latent Dirichlet allocation, allow one to analyze a corpus of documents, each of which might pertain to one or more topics, and determine both the topic of each document and how pervasive the topic is within each document. This type of analysis lends an extra dimension to conversation modeling: Topics can be viewed as stand-alone issues if they appear heavily only in certain documents (e.g., chemical weapons) or as ubiquitous across different conversations (e.g., China). Networks can then be created between different topics depending on whether the topics appear in conversation together (e.g., the 5G topic might be mentioned frequently with the cybersecurity topic). Adding time data would open up a new dimension; for example, knowing that *China* and *nuclear weapons* were not mentioned together often but that overlap between the two topics has become more common, the savvy reader of congressional data perhaps can infer something about future arms control treaties.

Another use case for NLP is text summarization (i.e., automatically identifying the key sentences from a long document to produce a condensed version). Text-summarization algorithms typically are trained on databases of news articles but could be fine-tuned on transcripts of committee hearings to produce even more-accurate summaries. Summarized versions of a document can be more than 95 percent shorter than the original, which can save readers considerable time. Some text-summarization algorithms simply lift passages from the original document, meaning that the summarized version is guaranteed to be readable, while others synthesize summaries from scratch, which could capture more meaning.

One potential drawback of text summarization is technical. Algorithms based on deep learning, generally speaking, can ingest sequences of only about 500 words at a time. Long documents, therefore, must be dealt with in a piecemeal fashion, which is not necessarily a flaw (stacked summarization algorithms can easily produce summaries of any length desired), but if the most-important exchanges happen at one point during testimony, they might be washed out by text from earlier and later in the transcript. The other potential drawback is common to many deep learning systems: a lack of interpretability in terms of why certain decisions were made.

More-speculative techniques drawn from NLP could be useful when trying to understand how Congress might act in the future. Sentiment analysis—determining whether a piece of text is broadly positive or negative—is another well-established NLP task for which software libraries have been built. Identifying which key members of Congress take a positive view of different DoD programs could be done in an automated way, and software could be written to track statements of interest over time and analyze them for changes in sentiment.

Question-answering algorithms are another potential method for mining interesting information from lengthy documents. In this method, a user who is interested in only a particular part of a document (but who is not sure where their topic appears) can query the algorithm and either be answered in natural language or be pointed to the passage in the text where the answer is likely to appear. The use case is simple to imagine: For example, a congressional hearing about chemical weapons takes place, and the user is interested in whether a particular member of Congress is in favor of or opposed to arms control measures. Performing a text search for *arms control* might not produce useful results because the phrase may have been mentioned by many participants. Instead, the user can write the query "Is Jane Doe in favor of arms control?" and be presented with a clear-cut answer.

## Conclusion

This Perspective describes a proof of concept for how to acquire and analyze text data from congressional hearings. It represents a basic overview of how to acquire, organize, and describe some key trends. All of the text data that we queried were in the public record and freely available from the GPO. Using these results, we recommend the following next steps:

- First, **using text data from public records is a relatively inexpensive way to construct roadmaps to understand the interests of congressional oversight committees**. If we assume that most of what Congress does is akin to a routine police patrol, it follows that policymakers would want to understand where these patrols are located. By describing the interests of Congress, policymakers could anticipate the emergence of salient issues.
- Second, **there is a growing volume of text data, and there are far more-sophisticated methods to analyze these data streams**. This proof of concept shows that these data exist, that they can be acquired relatively quickly and inexpensively, and that there appear to be useful trends once these data are organized. There are many different sources of data that could be acquired, organized, and linked with other sources of text data. For example, one could couple what members of Congress say in one subcommittee transcript with what they say in a related committee transcript a few months later. And one could link these data to what the members' offices tweet online and to interview transcripts from television appearances. One could use more-sophisticated techniques to begin to identify patterns in such speech across different means of communication. The goal is to help policymakers understand what policy areas Congress is most interested in to ultimately improve communication between the Executive and Legislative Branches.

By describing the interests of Congress, policymakers could anticipate the emergence of salient issues.

## Notes

1 One could use this workflow to query and analyze data in a useful manner. See, for example, the Comparative Agendas Project, undated. Furthermore, with more resources, one could develop machine-learning tools to begin forecasting trends using the workflow described in this Perspective.

2 For an example of what a text document looks like, see GPO, undated-b.

3 For details on the use of RAND-Lex, see Kavanagh et al., 2019; and Marcellino et al., 2020.

4 Many different people (e.g., witnesses, representatives, senators) make statements during these hearings. We did not parse who said what, which made it difficult to attribute issue salience to members of Congress versus nonmembers.

5 The focus of our search was on military manpower and personnel policies. Topic modeling produced words that, in general, were not relevant to these types of policies (e.g., *committee, don't, tiff*) or were too broad for meaningful interpretation (e.g., *force, army, security*).

6 We removed words referring to specific people at the committee hearing (e.g., senator, congresswoman).

## References

Balla, Steven J., and Christopher J. Deering, "Police Patrols and Fire Alarms: An Empirical Examination of the Legislative Preference for Oversight," *Congress and the Presidency*, Vol. 40, No. 1, 2013, pp. 27–40.

Comparative Agendas Project, homepage, undated. As of April 6, 2021: https://www.comparativeagendas.net/

Cox, Kate, William Marcellino, Jacopo Bellasio, Antonia Ward, Katerina Galai, Sofia Meranto, and Giacomo Persi Paoli, *Social Media in Africa: A Double-Edged Sword for Security and Development: Technical Annex*, New York: United Nations Development Programme, November 4, 2018.

GPO—*See* U.S. Government Publishing Office.

Halchin, L. Elaine, and Frederick M. Kaiser, *Congressional Oversight*, Washington, D.C.: Congressional Research Service, October 17, 2012. As of August 24, 2020: https://fas.org/sgp/crs/misc/97-936.pdf

Hardy, Donald E., *The Body in Flannery O'Connor's Fiction: Computational Technique and Linguistic Voice*, Columbia, S.C.: University of South Carolina Press, 2007.

Kavanagh, Jennifer, William Marcellino, Jonathan S. Blake, Shawn Smith, Steven Davenport, and Mahlet G. Tebeka, *News in a Digital Age: Comparing the Presentation of News Information over Time and Across Media Platforms*, Santa Monica, Calif.: RAND Corporation, RR-2960-RC, 2019. As of October 26, 2020: https://www.rand.org/pubs/research_reports/RR2960.html

Marcellino, William, Krystyna Marcinek, Stephanie Pezard, and Miriam Matthews, *Detecting Malign or Subversive Information Efforts over Social Media: Scalable Analytics for Early Warning*, Santa Monica, Calif.: RAND Corporation, RR-4192-EUCOM, 2020. As of October 20, 2020: https://www.rand.org/pubs/research_reports/RR4192.html

McCubbins, Mathew D., and Thomas Schwartz, "Congressional Oversight Overlooked: Police Patrols Versus Fire Alarms," *American Journal of Political Science*, Vol. 28, No. 1, 1984, pp. 165–179.

Presidential nomination 1603, "Matthew P. Donovan—Department of Defense," 116th Congress, 2019–2020, Washington, D.C., March 23, 2020. As of November 2, 2020: https://www.congress.gov/nomination/116th-congress/1603

SASC—*See* U.S. Senate Armed Services Committee.

U.S. Government Publishing Office, govinfo homepage, undated-a. As of April 6, 2021: https://www.govinfo.gov/

U.S. Government Publishing Office, "Congressional Hearings," webpage, undated-b. As of April 6, 2021: https://www.govinfo.gov/app/collection/chrg/116/senate/Committee%20on%20Armed%20Services

U.S. Senate Armed Services Committee, *Senate Armed Services Committee Advance Policy Questions for the Honorable Matthew Donovan Nominee to Be Under Secretary of Defense for Personnel and Readiness*, Washington, D.C., March 10, 2020. As of April 6, 2021: https://www.armed-services.senate.gov/imo/media/doc/Donovan_APQs_03-10-20.pdf

## About This Perspective

The U.S. Congress exercises oversight over the U.S. Department of Defense through its various committees. The patterns of what is said in these committees could signal emerging topics of interest for policymakers. This exploratory research examines the feasibility of acquiring, organizing, and analyzing text data from Congressional oversight committees. The authors discuss a workflow for a set of basic tools for analyzing transcripts from Congressional hearings and from answers to advance policy questions by nominees for political appointments. The results from a basic descriptive analysis suggest that useful patterns exist in these data. Using text data from public records is a relatively inexpensive way to analyze trends in policy analysis. Furthermore, more-sophisticated methods exist that could link and analyze different streams of text data (e.g., Congressional transcripts, media interviews, Twitter data). These text data could provide policymakers with a roadmap to anticipate the concerns of members of Congress.

The authors thank Michael Ryan of the RAND Corporation and Keller Scholl of the Pardee RAND Graduate School for their research programming help. They are grateful to Gordon Lee, whose dedicated work improved the prose of this Perspective. Finally, the authors thank William Marcellino and Jasmin Leveille of RAND and Steven J. Balla of George Washington University for their thoughtful reviews.

This Perspective was completed in February 2021 and underwent security review with the sponsor and the Defense Office of Prepublication and Security Review before public release. This Perspective was sponsored by the Office of the Secretary of Defense and conducted within the Forces and Resources Policy Center of the RAND National Security Research Division (NSRD), which operates the National Defense Research Institute (NDRI), a federally funded research and development center sponsored by the Office of the Secretary of Defense, the Joint Staff, the Unified Combatant Commands, the Navy, the Marine Corps, the defense agencies, and the defense intelligence enterprise.

For more information on the RAND Forces and Resources Policy Center, see www.rand.org/nsrd/frp or contact the director (contact information is provided on the webpage).

## About the Authors

**Marek N. Posard** is a military sociologist at the RAND Corporation and a professor of policy analysis at the Pardee RAND Graduate School, where he conducts research on manpower and personnel issues, personnel vetting, and election security–related topics. He holds a Ph.D. in sociology.

**Hilary Reininger** is an assistant policy researcher at the RAND Corporation, where she conducts research on contemporary information operations, Truth Decay, and countering foreign disinformation efforts. She holds an M.S. in conflict analysis and resolution.

**Christian Johnson** is an associate physical scientist at the RAND Corporation, where he builds and analyzes machine learning and artificial intelligence systems and applies them to public policy problems. He holds a Ph.D. in physics.



For more information on this publication, visit www.rand.org/t/PEA779-1.



PE-A779-1