Exploratory Text Analysis on the Congressional Response to the Withdrawal from Afghanistan

Quantitative Analysis Center
Summer 2022

By: Ethan Brill-Cass
Faculty Sponsor: Logan Dancey, *Government Department*, *Wesleyan University*

Introduction

In August of 2021, the U.S. officially withdrew forces from Afghanistan following nearly two decades of military conflict. The final stages of the withdrawal were marked by hurried efforts to evacuate U.S. citizens and Afghan allies in the face of rapid advances by the Taliban and a suicide bombing outside the airport in Kabul that killed 13 U.S. service members. The evacuation efforts received extensive media coverage, and the Biden administration's handling of the issue faced bipartisan criticism from members of Congress (Sprunt 2021) and negative marks from the public (Van Green and Doherty 2021). Following the withdrawal, members of the Biden administration, including Secretary of State Anthony Blinken and Secretary of Defense Lloyd Austin, testified before House and Senate committees in high-profile hearings that were covered live by numerous media outlets.

The post-withdrawal hearings were certainly not the first time that key congressional committees held hearings on Afghanistan during the 20-year conflict. Indeed, the House and Senate Armed Services and Foreign Relations/Affairs Committees all held hearings on Afghanistan in April and May in advance of the planned U.S. withdrawal. As a result, congressional hearings on Afghanistan in 2021 offer valuable insight on how member behavior and topics of discussion develop on an issue undergoing major changes and a rise in salience.

Topic modeling, described in the methods section, is used in this instance to identify patterns in questioning and discover prevalent topics of conversation within select relevant committee hearings. From these topic models, in conjunction with word count differences, we expect to see changes in the topics discussed over time as well as the emergence of topics seeking to capitalize on the issues' newly high salience by targeting the opposing party. We also expect the parties to differ greatly in the topics they engage with.

Methods

The data we collected came from the 117th Congress from the Armed Services and Foreign Affairs/Relations committees in the House and Senate. We use eight hearings' transcript, with one pre-withdrawal and one post-withdrawal hearing apiece for each of the four committees. Senate hearing transcripts were downloaded from the Senate committee websites, while House hearing transcripts come from ProQuest Congressional.

This data was then modeled using latent Dirichlet allocation (LDA), a popular topic modeling algorithm that utilizes a "bag of words" approach in which order and grammar are ignored (Grimmer, Roberts, and Stewart 2022). Topic modeling produces a set of topics which is meant to represent the latent structure of a text document and its multitude of abstract concepts. LDA is a mixed-membership model which perceives documents as a mixture of distinct topics and topics as mixtures of words. It simultaneously estimates which words are associated with a topic and the amount which each topic is contributes to a document (Silge and Robinson 2017).

A Senator or Representatives full set of questions for a single hearing acted as the documents. The topic probabilities (gamma values) produced inform of us of both the topics that comprise a document and the documents that most prominently feature a topic.
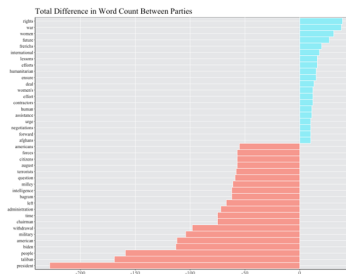
Count differences were calculated by taking the total count of a word across a group (Democrats) and subtracting the total count of that same word across the corresponding group (Republicans) for a given time. These differences illuminate topics and words that differentiate the rhetoric of Democrats and Republicans. Further, these differences highlight the issues that each party in general shifted their focus to as the material situation of the United States' presence in Afghanistan rapidly changed.
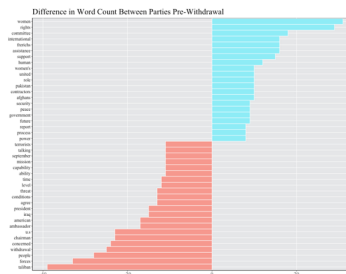
Topic Model – 20 Topics

Conditions Based Withdrawal	Women and Girls	Woodward Book	Specific Accountability	Doha Agreement	Bagram Air Base
president	ambassador	american	president	administration	president
withdrawal	u.s	trump	american	government	time
war	rights	forces	family	question	troops
troops	government	biden	statement	understand	withdrawal
august	withdrawal	administration	question	country	administration
conditions	girls	agreement	frerichs	withdrawal	bagram
agreement	women	women	biden	siv	government
administration	country	president	deserve	president	quote
government	human	book	intelligence	department	biden
u.s	administration	woodward	administration	trump	question
time	peace	bob	wrong	security	mckenzie
biden	alqaeda	terrorists	marine	doha	july
forces	khalilzad	understand	true	information	base
trump	question	time	time	alqaeda	left
security	ensure	committee	accountability	agreement	department

Table 1: 15 words with the highest β value for the six most discernable topics. Some words common words were removed during pre-processing.
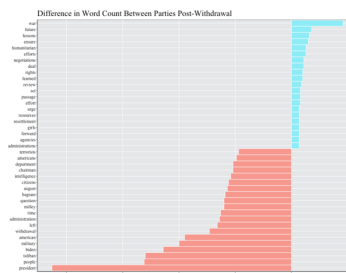
Count Differences

Graph 1: The twenty words with the highest absolute value of difference between parties across all hearings.

Graph 2: The twenty words with the highest absolute value of difference between parties across pre-withdrawal hearings hearings.

Graph 3: The twenty words with the highest absolute value of difference between parties across post-withdrawal hearings.

Results

A twenty topic LDA model produced a range of topics, some of which are more distinct and recognizable than others. Topics 6, 9, 10, 14, 16, and 17 stood out, representing major topics of debate in both Congress and the media.

The Doha Agreement and Bagram Air Base (topics 16 & 17) represent broad areas of debate with which both parties engaged. The Doha Agreement was present in 9 Democratic and 8 Republican documents while Bagram Air Base was found in 6 documents from each party.

Democrats were more likely to discuss the viability of a conditions based withdrawal (topic 6), appearing in 13 Democratic and 6 Republican documents. A similar but heightened disparity can be seen on the issue of women (topic 9) which was present in 26 Democratic documents and only 9 Republican documents.

Topic 14, labeled specific accountability, appears to represent of mix of two issues. The first being the capture of American civil engineer Mark Frerichs and the second being the suicide bombing at Kabul's Airport which resulted in the death of approximately 170 Afghans and 13 U.S. service members. This topic was more common among Republicans, appearing in 8 Republican documents and just 2 Democratic documents.

Topic 10 seemed to represent discussion of General Milley's interview for a Bob Woodward's book which was critical to Trump. While we expected from a representation of top terms that this topic would be much more prevalent among Republicans, it contributed to 4 Democratic documents and 5 Republican documents. Interestingly, while the interview was referenced in the Republican documents, none of the four Democratic documents containing topic 10 referenced this book or interview.

Additionally, the count differences also suggest that pre-withdrawal the topics that set Democrats apart were women, resources/aid, and human rights while Republicans were set apart by discussion of threats, conditions, and capability. While Democrats maintained some focus on women post-withdrawal, they uniquely pivoted to lessons learned. Republicans likewise pivoted to two new subjects, the Biden administration and the intelligence community.

Discussion

This analysis looks specifically at differences in topic probabilities and differentiating words across parties. Given the nature of the topic and frequency at which women are discussed, exploration of differences across sex and military status may provide insight on how speaker-level characteristics impact which topics a committee member engages with.

This analysis was also limited by the relatively narrow scope of the hearings used. Each document already belonged to one topic, the United States' role in Afghanistan, so LDA was in essence searching for sub-topics leading to considerable overlap between topics at every attempted number of topics. Topic 10, General Milley's interview for Bob Woodward's book, illustrates how topic overlap hinders analysis. It may be of use to expand the dataset to better understand how a more robust sample may or may not improve subjective fit of an LDA model on a narrow topic.

References

Grimmer, J., Margaret E. Roberts, & Brandon M. Stewart. 2022. "Text as data: A new framework for Machine Learning and the Social Sciences." Princeton University Press.

Silge, Julia, & David Robinson. 2017. "Text mining with R: A tidy approach." O'Reilly. https://www.tidytextmining.com

Sprunt, Barbara. 2021. "There's a Bipartisan Backlash to How Biden Handled the Withdrawal from Afghanistan." Npr.org, Aug. 17: https://www.npr.org/2021/08/16/1028081817/congressional-reaction-to-bidens-afghanistan-withdrawal-has-been-scathing

Van Green, Ted and Carroll Doherty. 2021. "Majority of U.S. Public Favors Afghanistan Troop Withdrawal; Biden Criticized for His Handling of Situation." Pew Research Center, Aug. 31: . https://www.pewresearch.org/fact-tank/2021/08/31/majority-of-u-s-public-favors-afghanistan-troop-withdrawal-biden-criticized-for-his-handling-of-situation/