Text Analysis of Secretary of State Hillary Clinton's Emails:
Hard Power vs. Soft Power

By
Imgesu Cetin
Talor Gruenwald
Nicole (Niki) Kalmus

UCSD POLI 175
Professor Kirk Bansak
June 6, 2020

Summary

Through creation of novel dictionaries used to describe hard and soft power and sentiment analysis, we determined that the content of Hillary Clinton's emails tends to be slightly more related to soft power than hard power. Our results show that Clinton's use of hard power was centered in the Middle East, while her use of soft power centered around development and agreements. This is important because she is widely regarded as a war hawk—this analysis may help to recharacterize her foreign policy stance.

Methodology and Findings

In this analysis, we use sentiment analysis and topic modeling to analyze text data from Hillary Clinton's emails, released by the State Department in 2015. The data include approximately 8000 emails that were sent and received on Secretary Clinton's private server. The data has been subsetted to include only the top 3000 frequent unigrams, which we found to be the cleanest and strongest predictors of sentiments and topics in the data. It has been cleaned to exclude common stop words,[1] alphanumeric names such as "X30," and single letter unigrams such "t" or "u".

Sentiment analysis is used in this study to indicate the distinction between hard and soft power in Secretary Clinton's emails. According to Joseph Nye in 1990, soft power, or "non-coercive power," can be combined with hard power, or "military power" to solidify U.S. power.[2] In 2015, Foreign Affairs claimed that when he was president, Barack Obama emphasized soft power.[3] Thanks to the release of Secretary Clinton's emails, we are able to evaluate whether this claim has validity. In addition, through the use of topic modelling, we can assess the types of hard and soft power Clinton exerted during her tenure.

Sentiment Analysis

In order to analyze if Secretary Clinton leaned more towards hard or soft power, we first created our own dictionary of words that fit into hard and soft power categories using a combination of pre-categorized words from the General Inquirer's Harvard IV-4 and Lasswell value dictionary categories.[4] For example, one category from Harvard IV-4 consists of words defined as "military" words which we added to the hard power dictionary, and another, "Econ@", consists of words used to describe economic power. See Appendix for specific dictionaries employed. We then weighted each word by multiplying

---

[1] Stop words retrieved from Princeton University:
https://algs4.cs.princeton.edu/35applications/stopwords.txt
[2] Li, 2018
[3] Klare, 2015
[4] Harvard University General Inquirer descriptions: http://www.wjh.harvard.edu/~inquirer/homecat.htm.

it by positive one if the word appeared in the soft power dictionary and negative one if the word appeared in the hard power dictionary. All other words received a value of 0. This allowed us to weight each email by its "hard" and "soft" words and classify the email as a hard or soft power leaning email.

As shown in the histogram below, Hillary Clinton's emails tend to be characterized more by soft power than hard power. This is evidenced by an average global sentiment greater than 0. Less than 0 would imply more emails demonstrating hard power. This is consistent with Hillary Clinton's role as Secretary of State, which distinctly attempts to resolve conflicts in international relations through means other than military action.

Figure 1: Histogram of Hillary Clinton's Emails Classified as Hard or Soft Power
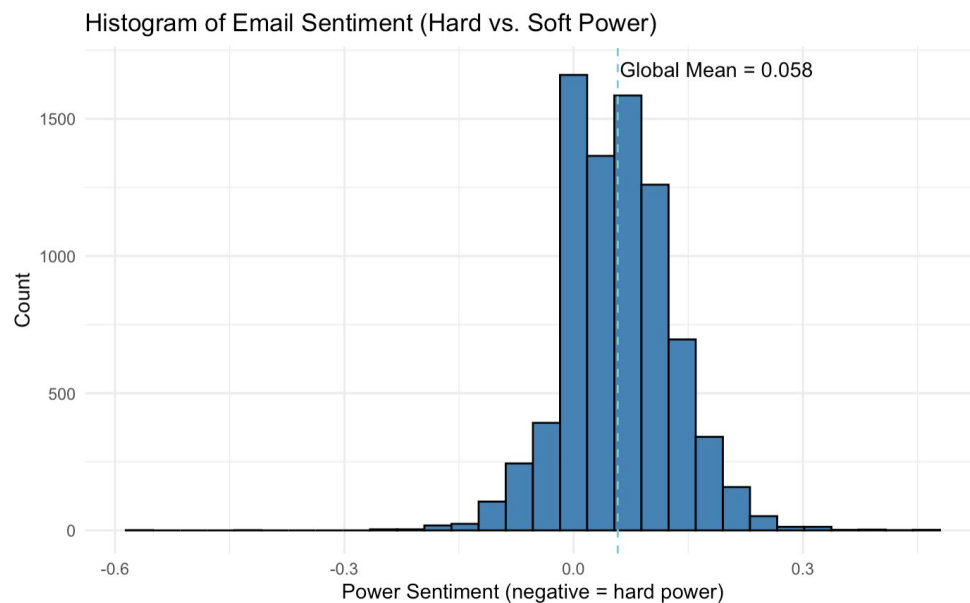
Below are word clouds of most frequent words in emails characterized as describing hard power (left), and soft power (blue scale). They demonstrate that the unsupervised machine learning performed by our sentiment analysis model is performing well, at least by human detection standards. The hard power words are militaristic and forceful, while the soft power words are more "diplomatic" in the broad sense of the term.

Figure 2: Top Word Cloud of Clinton's in emails classified as hard power (left) and soft power (right)

time
war march press
general make deal
foreign issue
military

security
agreement
team
white office
press support
human bill administration
great peace deal
make minister back
economic group
health good
secretary
call sensitive

Structural Topic Modeling

To further analyze Hillary Clinton's use of and hard power as Secretary of State, we used topic modeling as another unsupervised classification method in order to group alike topics in Clinton's emails. We ran our topic modeling using Latent Dirichlet Allocation, which is the most common algorithm for topic modeling. To find our optimal k, that is the number of groups, we ran multiple LDA models with different k values. When we ran our model with k=3 we observed that there was no clear difference between topics. We then went on fitting our LDA model with k= 5 groups, which gave us more meaningful results.

To visualize our results, which can be found below, we found that, as expected, words classified as soft power were all more placid in nature. Words classified as hard power were related to the Middle East, war, security, and in a more negative tone. Figures 3 and 4 display the results of LDA with the top ten words for each defined topic on the y-axis and the per-topic-per-word probability on the x-axis. We can see that emails that are classified into hard power seem to be dealing with Middle Eastern affairs. The emails that were classified into soft power are more about American domestic affairs. Words such as 'obama', 'american', and 'government' show that soft power-classified emails related more to American affairs.

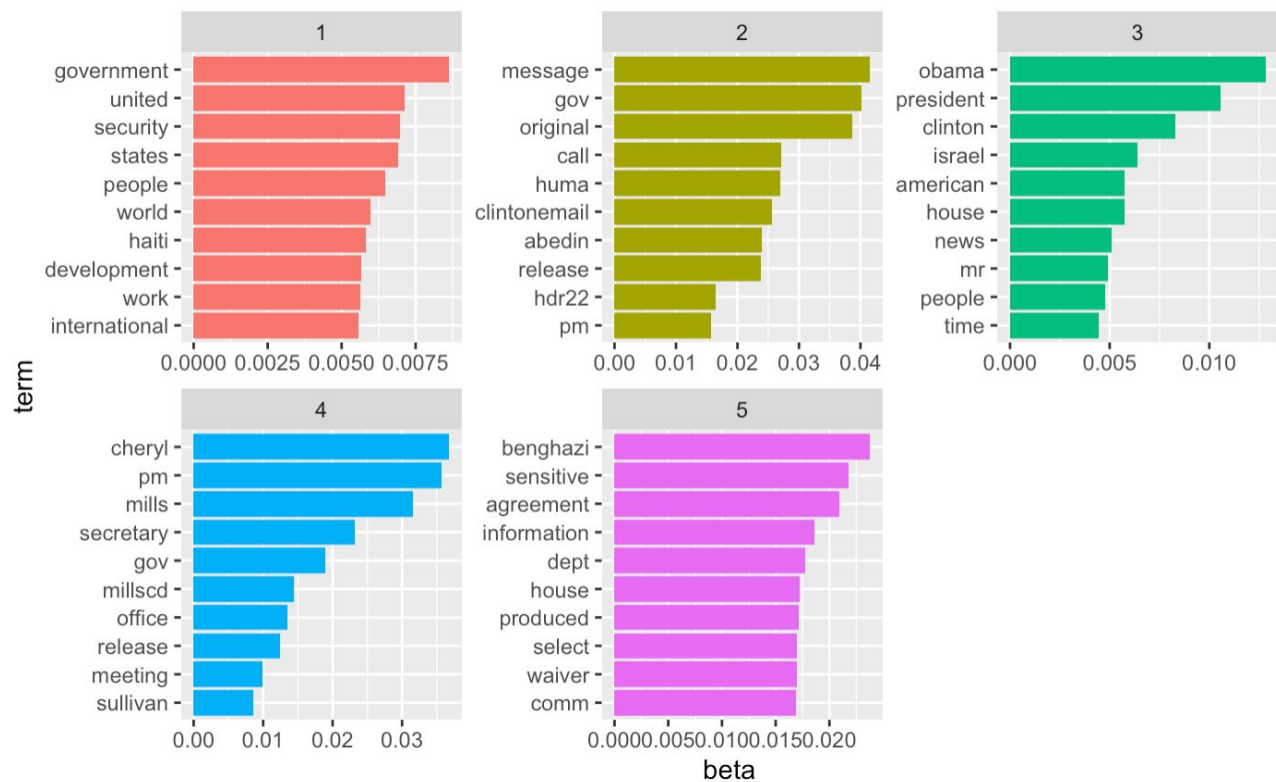**Figure 3: Soft Power Topics with Top Words (5 Topics)**

Figure 4: Hard Power Topics with Top Words (5 Topics)

Discussion of Results

We conclude that the content of Hillary Clinton's emails tend to be slightly more related to soft power than hard power, though she was widely regarded as a war hawk. This analysis may help to recharacterize her foreign policy stance while simultaneously confirming Foreign Affairs assessment of President Obama's foreign policy. Our analysis demonstrates Hillary Clinton's use of hard power was focused in the Middle East, while soft power emails dealt with topics such as development and meetings.

The process of topic modeling pre-classified emails was hindered by the data. LDA with 3 topics yielded results that were less conclusive, as several words seemed to appear in every email, making it more difficult to identify the subject of each topic. LDA with 5 topics results in more insight; however, many of the topics' most frequent words include the names of Clinton's staffers and commonly released email words such as "message" and "release." Further analysis might be more systematic in removing these common words and seeing what topics are uncovered as a result.

Bibliography

Harvard University General Inquirer Descriptions of Inquirer Categories and Use of Inquirer Dictionaries http://www.wjh.harvard.edu/~inquirer/homecat.htm.

Klare, Michael T. (2015). Hard Power, Soft Power, and Energy Power. www.https://www.foreignaffairs.com

Li, Eric. (2018). The Rise and Fall of Soft Power. Foreign Policy. www.foreignpolicy.com

Sedgewick, Robert and Kevin Wayne. (2020). Algorithms, 4th Edition. Princeton University. https://algs4.cs.princeton.edu/35applications/stopwords.txt

Appendix

Table A: Dictionaries used in hard and soft power classification

Hard power	Soft power
Milit - Harvard IV-4	Econ@ - Harvard IV-4
Ngtv - Harvard IV-4	Pstv - Harvard IV-4

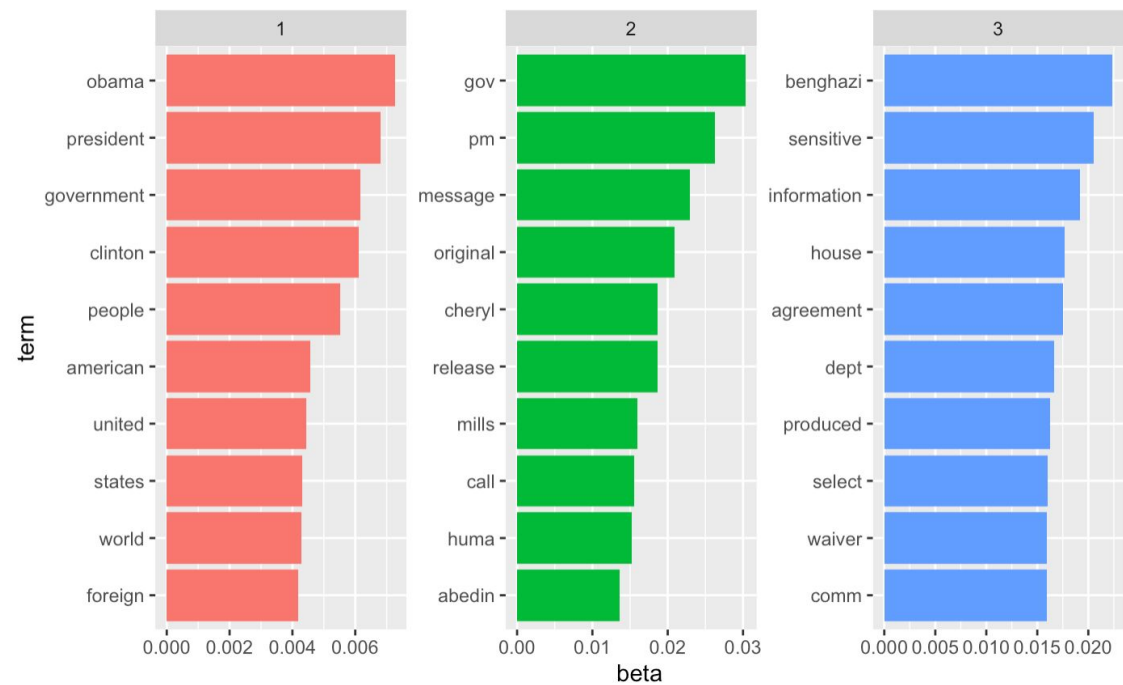| Hostile - Harvard IV-4	PowAuth - Lasswell value
PowCon - Lasswell value	PowCoop - Lasswell value

Table B: Soft Power Topics with Top Words (3 Topics)

Table C: Hard Power Topics with Top Words (3 Topics)