

Narrative

Jenna Gibson

12/9/2019

Brief substantive background / goal

In 2017, U.S. President Donald Trump threatened to rain fire and fury upon North Korea. Just a year later, Trump has fully embraced a policy of diplomatic engagement with North Korea's Kim Jong Un, using noticeably more affectionate language to refer to the North Korean leader. At the same time, the State Department appears to have shifted the respect it conferred to Kim in official documents. In their press briefings, State Department officials stopped referring to Kim with no official title, and started almost exclusively calling him "Chairman Kim," his own preferred title. What explains this abrupt change in rhetoric used to confer or withhold respect and status to Kim? And what does that shift say about the bargaining position of major powers vis-a-vis their potential negotiating partner?

To tackle this question, I plan to scrape transcripts of State Department Press Briefings as well as White House Press Briefings and remarks from President Trump. The State Department Spokesperson holds an open press briefing regularly, and posts full transcripts of the briefing on its website. These transcripts are available going back to March 7, 2017, which was the first such briefing held under the Trump Administration. In addition, the White House posts a variety of official material on its site as well, representing official remarks from the President, press briefings, statements from Cabinet-level officials, and more. Thus, by scraping these documents to create a comprehensive data set, I can show a full picture of the Trump Administration's official (non-Twitter) rhetoric when it comes to North Korea policy.

While my data set will represent all the official rhetoric from the Trump Administration on the topics of North Korea, for this project I will focus in particular on the titles that the administration uses when they refer to North Korean leader Kim Jong Un, and how those titles may change over time. The goal is to see if changes in the foreign policy context - the Trump's decision to meet Kim in Singapore in June 2018, for example - shifts the rhetoric the administration uses. I argue that using certain titles - or omitting an official title entirely - indicates the level of respect that the administration wants to confer on Kim. In future analyses, I hope to use this respect level, which is an overt and public signal that can be read by many audiences including Kim and his advisers, as a proxy for how seriously the Trump Administration is taking their negotiations with North Korea.

Collecting data

My data set will consist of text scraped from the State Department and White House websites. In particular, I will collect text from the briefings and statements section of the White House site (filtered by the "Foreign Policy" topic), and the press briefing transcripts from the State Department.

From each of these pages, I first had to scrape the URLs for all documents. I did this by grabbing the URL text from within the html node for each link, and then iterating over the multiple search result pages.

After collecting the full list of URLs, I then wrote a function to scrape text from each page. From every document, I collected the date, title, statement type (White House) or index (State Department), and full text of the document. The statement type label indicated the type of document that each page represented, whether official remarks from the President ("Remarks"), a press briefing ("Statements & Briefings"), or news clips reposted from outside media ("News Clips"). My corpus thus consisted of data from 1,509 White House documents and 176 State Department documents.

Cleaning / pre-processing data

The first step I took in cleaning my data was to fix an issue specific to the State Department website. For many of the documents, the site did not have a date in their date node - presumably the person uploading

the document forgot to include that text. Luckily, the URL for each page included its date. So I was able to subset the URL to isolate the date text and convert it to a month-day-year format.

Next, I wanted to narrow in on documents that specifically dealt with North Korea-related topics. This was quite easy for the State Department, thanks to the site's inclusion of an index on each page that lists all general topics that were covered in that day's press briefing. I filtered out any documents that did not have an index related to North Korea and/or Kim Jong Un. On the White House side, I had to search for mentions of North Korea, the DPRK, Kim, etc. within the text of the document to filter out unrelated documents. In addition, I filtered out the documents labeled as "News Clips", since they were written by other outside actors and thus did not necessarily represent the official rhetoric of U.S. Government officials.

Next, I focused on the text of the documents themselves. I unnested both data sets, which put each paragraph in each document into its own row for easier access. I removed all reporter questions (paragraphs starting with "Q-" for the White House and "QUESTION:" for the State Department), because once again these represented rhetoric from an outside individual, not an official from the U.S. Government. I then threw out any paragraphs that did not specifically mention "Kim", since this project is focused particularly on how Kim himself is addressed within these statements.

Finally, I used the `str_extract_all` function to pull out all mentions of "Kim", including the word before and after the "Kim" match. The additional words allowed me to identify the title that was attached to Kim within the text. But it also helped to confirm that the mention was, indeed, about the correct "Kim" - Kim Jong Un. There were several other "Kim"s that appeared in my matches - U.S. Ambassador to the Philippines Sung Kim, for example, or South Korean First Lady Kim Jung-sook. I removed these unrelated observations and ended up with 240 final observations from the White House and 208 final observations from the State Department for a final total of 448 observations.

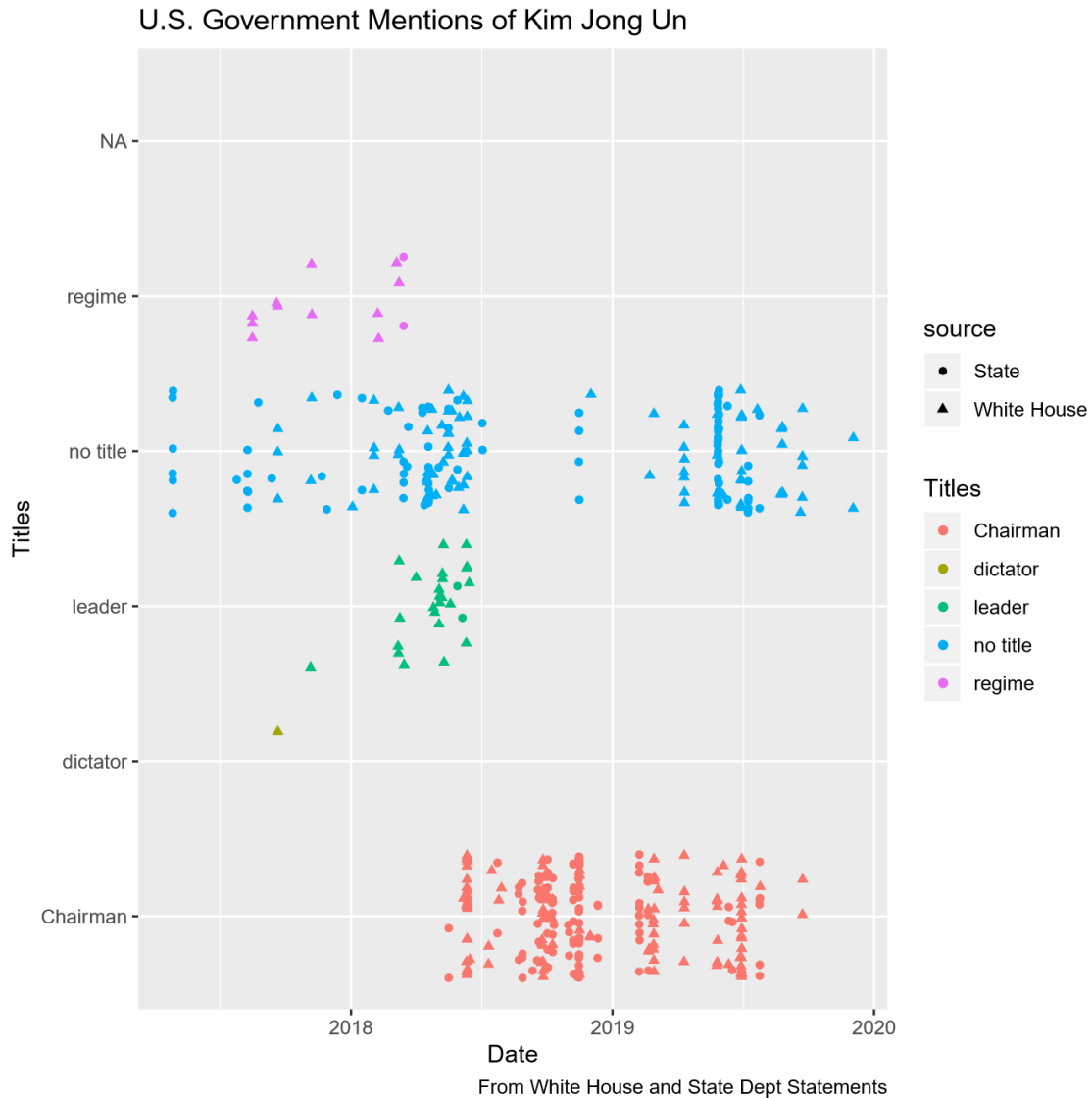
Analysis and visualization

Using the 448 mentions of Kim from the two U.S. Government sources, I could now see what titles were associated with Kim Jong Un in these documents, and how these titles may have shifted over time. I first grouped each mention - which included both "Kim" and whatever word came before and after the name - to see the possible titles that were present in the data. I identified five possible options - "Chairman", "dictator", "leader", "regime", and simply Kim, which I labeled in my data as "no title".

Number of Mentions					
	Chairman	Dictator	Leader	Regime	No Title
WH	117	1	24	11	87
State	107	0	2	2	97
Total	224	1	26	13	184

To visualize how these titles may have changed over time, I created a plot that visualizes each distinct title (by color) coming from both the State Department and White House (by point shape) over time. The resulting graph is below.

While more advanced statistical analysis will be necessary to examine in more detail the potential shift in title (and thus respect) conferred upon Kim Jong Un within these documents, there does appear to be a temporal change captured in the graph. Moreover, the shift appears to take place at a particular time - approximately at the beginning of June, 2018. This corresponds perfectly with President Trump's decision to



become the first sitting U.S. President to meet in person with a North Korean leader. This analysis, therefore, provides preliminary evidence that confirms my hypothesis that increased U.S. signalling of respect could serve as a proxy for desire to seriously negotiate with the North Koreans.

Future work

The most important next step for this research would be to find a statistical model that may be able to analyze how these data have changed over time. While it appears visually that there is a change, does that change actually hold up to statistical analysis? And when exactly were the temporal breaks that are represented in the data?

Only after answering those questions can I then map the data onto changes in negotiating context over time. The change in June, 2018 around the Singapore Summit appears obvious, but there have also been smaller ups and downs in the negotiating process over the last two years. Can the data capture these smaller shifts?

In addition to more detailed analysis, it would be interesting to continue adding to the data set with a broader variety of documents. President Trump is notorious for conducting policy via Twitter - in the future I could include President Trump's tweets about Kim Jong Un to see how the president is referring to the

North Korean leader on a slightly more informal (but still official) platform.

Finally, I could expand this project to attempt to map how the rhetoric of the Trump Administration spread beyond its own official documents. Did the shift from “Kim Jong Un” to “Chairman Kim” within the State Department and White House create a corresponding shift in rhetoric within the mainstream media covering foreign policy? Or vice versa? Scraping media mentions of Kim over the same time period and mapping how the appearance of the “Chairman Kim” title spread could provide great insights into the relationship between the U.S. Government and the mass media.