**Ben Harwood**

**IST 654 Assignment 1**

**Introduction**

Kentucky is an interesting state (well, commonwealth). As someone who has lived here for almost 30 years, the current author has noticed all kinds of quirks. Perhaps the biggest one is how politically bipolar the people in the state are. Perhaps an explanation is due.

Mitch McConnell is the longest serving KY US Senator in history, winning initial election in 1984 and staying there ever since. He is one of the leading voices (perhaps second only to the current president) of the Republican party in the state, and the nation. In fact, KY has not elected a democrat to the senate since 1992 and has not voted for a democrat in a presidential election since Bill Clinton’s second term. Yet, in the last 100 years there have been only 6 republican governors.

While the reasons behind this could fill volumes (such as in the last election where the incumbent wanted the names of teachers who participated in a walk out in the biggest district in the state to protest a state house bill that would have gutted the state pension system that some of them had been mandatorily paying into for several years, and accused them of putting their students’ lives in danger because some of the kids would be stuck at home with abusive parents), this paper revolves around the inauguration speeches of the past two governors, Matt Bevin and Andy Beshear, in effort to see what issues they seemed most passionate about as they entered office.

**Section 1: Analysis and Models**

This section discusses the data to be used in the analysis, as well as the method of analysis.

**Section 1.1: Data**

In order to analyze the two governors’ speeches, obviously the speeches themselves are needed. Each speech was located on the web (see the references section) and copied verbatim as a text file. Each of the speeches was then, using Python, tokenized (broken into individual words (tokens), and even characters in situations such as a comma at the end of a word indicating a pause where the comma would be separated out). Next, the tokens were reduced to all lowercase letters and non-alphabetic tokens were removed. Finally, both speeches had commonly used words (or stop words) such as “it”, “a”, “is”, etc. removed, as they do not contribute to the overall tone and message of the speeches. Here is a bit of top-level information about each speech.

|  |  |
| --- | --- |
| **Speech A**  Total words (including stop words): 650  Stop word ratio[[1]](#footnote-1): 60.99%  Most frequent word and freq.: Kentucky, 34 | **Speech B**  Total words (including stop words): 279  Stop word ratio: 53.92%  Most frequent word and freq.: Kentucky, 27 |

It can immediately be noticed that while Governor A said more, more of his speech was stop words, so perhaps there was less actual content there; or he was just being verbose. An interesting observation can also be made right away based on each governor’s use of the word “Kentucky”. Here are the first 25 occurrences of the word Kentucky in each speech. The governor speaking will be left out for now, to increase the drama.

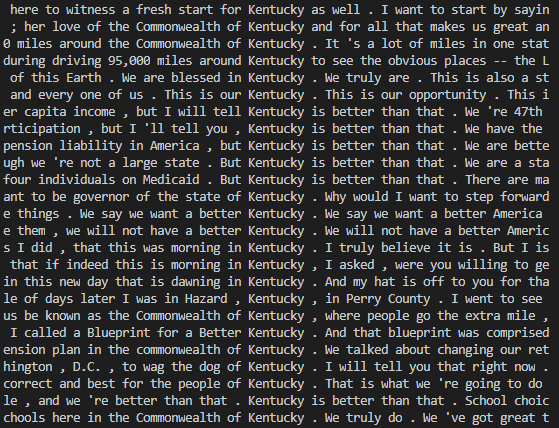


Figure : Governor A uses of the word Kentucky

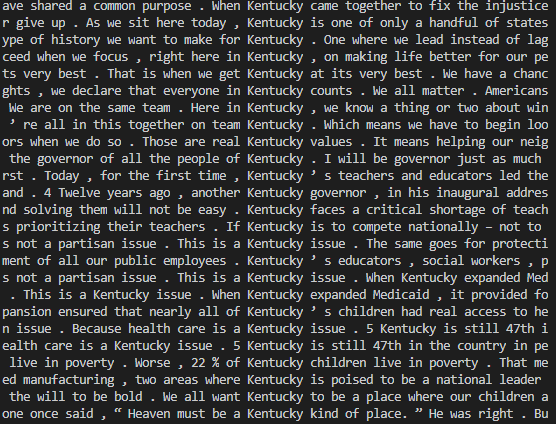
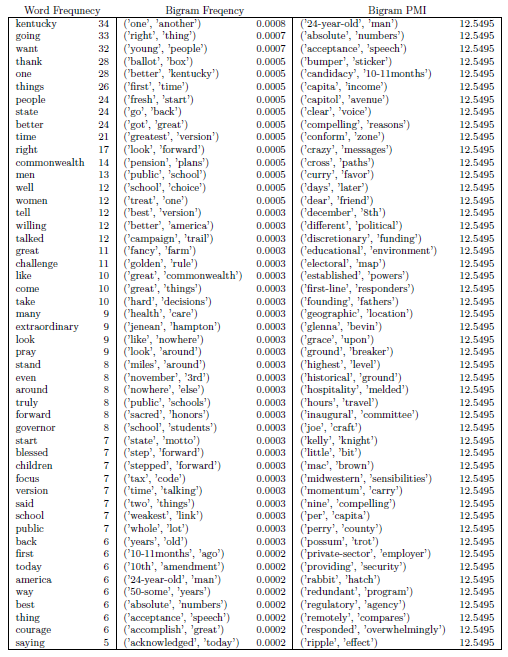


Figure : Governor B uses of the word Kentucky

Notice from Figure 1 that the phrase, “Kentucky is better than that” occurs quite frequently. More on this in Section 2.

**Section 1.2: Analysis**

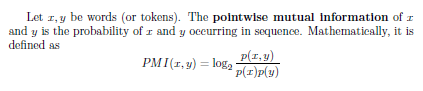
In the previous section, it was noted that both speeches’ most frequent word was “Kentucky”, and this should not come as a surprise. To continue the analysis, each speech will be considered separately, again with the deliverer not revealed. The following table gives the 50 most used words and their frequencies, along with the 50 most common bigrams and 50 bigrams with the highest pointwise mutual information.



Consider that the second most frequent (and substantive) word in this speech is “better” which was used 24 times, which was as frequent as “people” and “state”.[[2]](#footnote-2) This reflects the previous discussion.

Recall that bigrams are just pairs of words that occur together, and their associated frequency is the percentage of the occurrence of the bigram. These can lend additional insight than just the individual word frequencies. For example, the bigram (‘better’, ‘kentucky’) has a frequency of 0.05%. Perhaps more interestingly (and at the risk of giving away which governor’s speech this was), notice that (‘school’, ‘choice’) also has a frequency of 0.05%, as does (‘pension’, ‘plan’). These will come up again in Section 2.

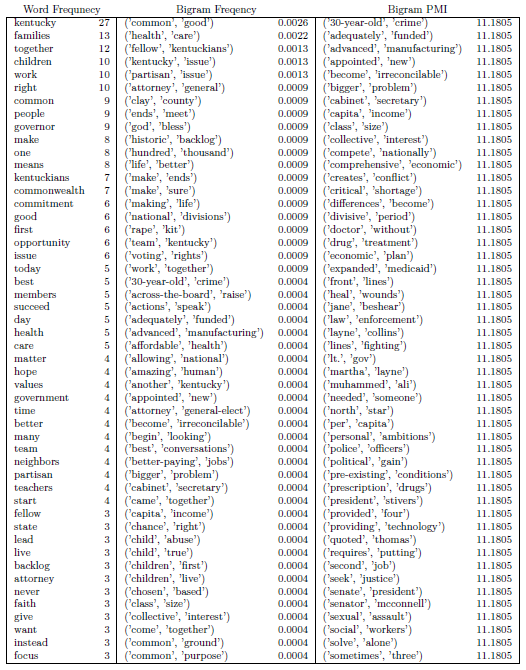
The third part of the table is the 50 bigrams with the highest pointwise mutual information score. Here is a quick refresher:



This measure is similar to frequency, but instead of measuring the percentage that the bigram occurs, PMI measures the probability that the words appear together compared to the probability that the joint appearance was random. Essentially, PMI indicates how much one word appearing is related to the other word appearing. Alas each of the 50 top PMI bigrams in the first speech occurred exactly once, so not much additional insight can be gleaned there.

It should also be noted that trigrams could be considered as well. The table for the second speech is at the top of the following page. In this case, 4 of the top 5 most frequent words are all substantive, with the governor talking about Kentucky, families, children, and work. This is a very distinct difference from the first speech whose first substantive word was sixth on the frequency list, but also note the difference in tone. The first speech’s most frequent words were “better”, “state”, and “people”, while the second speech’s most frequent words were “family”, “children”, and “work”. Ironically, the second collection seems to reflect a speech that is about people more than the first set, and the first set includes the word people! And notice also bigrams in the second speech like (‘ends’, ‘meet’), (‘voting’, ‘rights’), (‘rape’, ‘kit’), (‘child’, ‘abuse’), and (‘children’, ‘first’), which support this.

As with the first speech, each of top PMI bigrams occurred only once.



**Section 2: Results**

It was mentioned in Section 1.1 that the phrase “Kentucky is better than that” was used quite frequently. Indeed, this phrase occurs a total of nine times in the entire speech, and the word “better” precedes Kentucky twice and the shorter phrase “better than that” appears without “Kentucky” five more times. Now, every governor says they are going to make their state better, but when 20% of the state name occurrences are followed by “is better than that” then the motivation is clear. It should be reiterated here that both governors followed governors from the opposite party, so of course each one thought the previous governor did things the wrong way (such is the nature of politics). But compare this to speech B, where the bigrams discussed in section 1.2 convey a message focused on the people themselves, so whereas the first speech appears to be focused on “fixing” things (indeed the governor speaking talks about his Blueprint for a Better Kentucky) the second speech appears focused on helping the people in the state. The difference, from a listener’s perspective, is that one sounds very critical and perhaps even negative while the other is more optimistic sounding as it seems to focus less on the state of the state and more on how the new administration will actually work for the people that live there. Obviously, each approach speaks to the political position of each governor.

It was mentioned briefly at the end of Section 1.2 that trigrams could be looked at as well. Doing so reveals that the most frequent trigram (of any consequence) in speech A is, unsurprisingly, (‘better’, ‘than’, ‘that) with frequency 0.23% and (‘the’, ‘common’, ‘good’) for speech B with frequency 0.21%. This once again solidifies the prior evaluation.

**Conclusion**

In the end, politicians are well-known for saying one thing and doing something completely different. Such is the nature of power. And it is entirely possible that each governor’s delivery of their respective speech was in such a way that regardless of the message they were trying to convey it was misinterpreted by those watching or listening. Again, this is the nature of politics, some of the people you represent would follow you off a bridge and some would rather push you off the bridge.

For those curious, speech A was delivered by republican Matt Bevin on December 8, 2015 and speech B by democrat Andy Beshear on December 10, 2019.

**References**

Beshear, Andy. “Inaugural Address.” *Kentucky Governor Andy Beshear*, Commonwealth of Kentucky, 10 Dec, 2019, <https://governor.ky.gov/attachments/20191210_Inaugural-Address_Gov-Beshear.pdf>.

Bevin, Matt. “Inaugural Address.” *Bevin: ‘Focus on that which unites us’*,” Courier Journal, 8 Dec. 2015, <https://www.courier-journal.com/story/opinion/2015/12/10/focus-which-unites-us-bevin-says-inaugural-speech/77045430/>.

1. How much of the speech was just stop words. [↑](#footnote-ref-1)
2. Again, more in Section 2 [↑](#footnote-ref-2)