Advanced Quantitative Techniques (Class 4)

Gregory M. Eirich
QMSS

Agenda ~ Intro to Text Analysis

1. The promise of text as data, or -- Why text analysis?

2. Getting started: Where to get texts; formatting texts; organizing texts; capturing meta-data; units of analysis; arranging text through stemming, stop-words, and other pre-processing

3. The main method for today: Counting, combined with “bag of words”

4. Sentiment analysis via automatic dictionary-based methods

1. Why text analysis?

Why text analysis?

- Most of the world is text: we communicate, negotiate, advertise, complain, use the Internet, etc. in text

- Text is both easy to see and hard to see at the same time (i.e., we have a “feel” for what we are hearing or reading, but it is hard to put our finger on any larger pattern of word usage)

Why text analysis?

- The ability to rigorously meld the qualitative (raw text) with the quantitative (patterns) that is so exciting about this emerging field

- There are a tremendous number of new questions you can answer with textual analysis

2. Getting started

Where to get text

Formatting text

- There are packages in R are “tm”, “quanteda” and “tidytext”
- There are packages in Python: Natural Language Toolkit (NLTK), TextBlob, spaCy

Other considerations

- How do I organize my texts?
- How do I encode meta-data?
- What is the appropriate unit of analysis?

How should I arrange my text?

- Should I stem my words?
- Should I keep or discard stop (or function) words?
- Should I maintain capitalization?
- What about punctuation?
- What about special characters and emojis, like :) and :P?

PA

Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It

Matthew J. Denny[1] and Arthur Spirling[2]

[1] 203 Pond Lab, Pennsylvania State University, University Park, PA 16802, USA. Email: mdenny@psu.edu
[2] Office 405, 19 West 4th St., New York University, New York, NY 10012, USA. Email: arthur.spirling@nyu.edu

Abstract

Despite the popularity of unsupervised techniques for political science text-as-data research, the importance and implications of preprocessing decisions in this domain have received scant systematic attention. Yet, as we show, such decisions have profound effects on the results of real models for real data. We argue that substantive theory is typically too vague to be of use for feature selection, and that the supervised literature is not necessarily a helpful source of advice. To aid researchers working in unsupervised settings, we introduce a statistical procedure and software that examines the sensitivity of findings under alternate preprocessing regimes. This approach complements a researcher's substantive understanding of a problem by providing a characterization of the variability changes in preprocessing choices may induce when analyzing a particular dataset. In making scholars aware of the degree to which their results are likely to be sensitive to their preprocessing decisions, it aids replication efforts.

Keywords: statistical analysis of texts, unsupervised learning, descriptive statistics

1 Introduction

Every quantitative study that uses text as data requires decisions about how words are to be converted into numbers. These decisions, known collectively as 'preprocessing', aim to make the inputs to a given analysis less complex in a way that does not adversely affect the interpretability or

These decisions can matter a lot

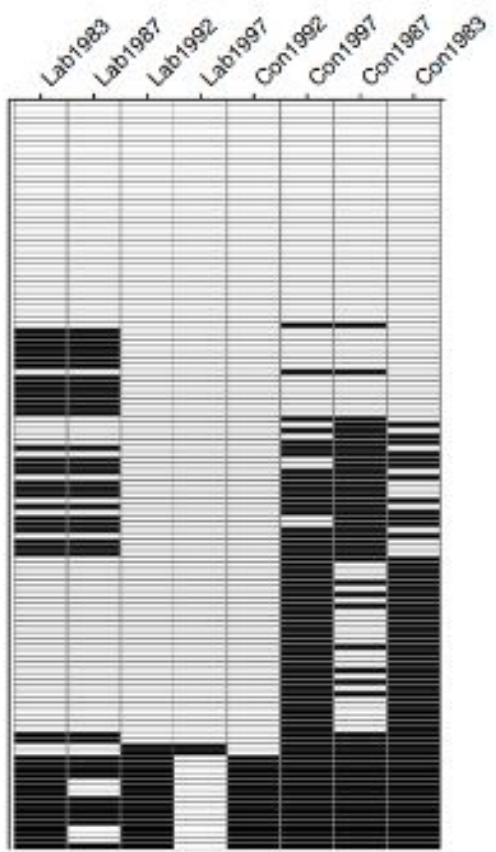


Figure 1: Wordfish results for the 128 different preprocessing possibilities. Each row of the plot represents a different specification. A white bar implies that the manifesto for that year is in the correct place as regards our priors. A black bar implies it was misplaced

Many preprocessing choices make mistaken identity more likely

3. The main method of the day … Counting

A very simple example: “key words”

- Shin, Taekjin, and Jihae You. "Pay for talk: How the use of shareholder-value language affects CEO compensation." *Academy of Management Proceedings*. Vol. 2013. No. 1. Academy of Management, 2013.

Their question

- Can CEOs benefit from signalling like they care about their shareholders, even if they do not have policies or governance institutions that are especially shareholder-friendly?

Their answer

- Can CEOs benefit from signalling like they care about their shareholders, even if they do not have policies or governance institutions that are especially shareholder-friendly?

- Yes, they can … by mentioning the word “shareholder” more often in their annual reports

Their findings

- Top managers' use of shareholder language leads the Board of Directors to evaluate those managers more favorably and to grant a higher level of compensation.

- Analyze the letters to shareholders from 160 U.S. firms

- They found that the use of shareholder-value language is significantly related to a higher level of CEO compensation, regardless of the actual implementation of shareholder-value strategies or any changes in firm performance. Adding 1 more "shareholder" word per 1,000 words is worth half a million in additional compensation.

Their findings

- They also found that the effect of shareholder-value language on CEO compensation is stronger when the board is less independent, firm performance is poorer, and shareholder activism is stronger.

- They also perform fixed effects and see within the same firm, when the use of shareholder words goes up over time, so does compensation

Many other “key word” examples

- Steven Levitt on good and bad words for real estate ads (in Chicago real estate ads, words like *charming*, *spacious*, *great neighborhood* and *fantastic*, or even having exclamation points(!), are associated with lower sale values, net of other factors)

- Dan Jurafsky found the same effect on words for food on menus (in *The Language of Food: A Linguist Reads the Menu*), but the opposite on “fancy” adjectives on items, like “aged” balsamic

Another “key word” example

- Picking a handful of key democratic words

ctions

The Washington Post

Monkey Cage | Analysis

Unlike all previous U.S. presidents, Trump almost never mentions democratic ideals

By David Beaver and Jason Stanley February 7

Most

Recent “key word” example

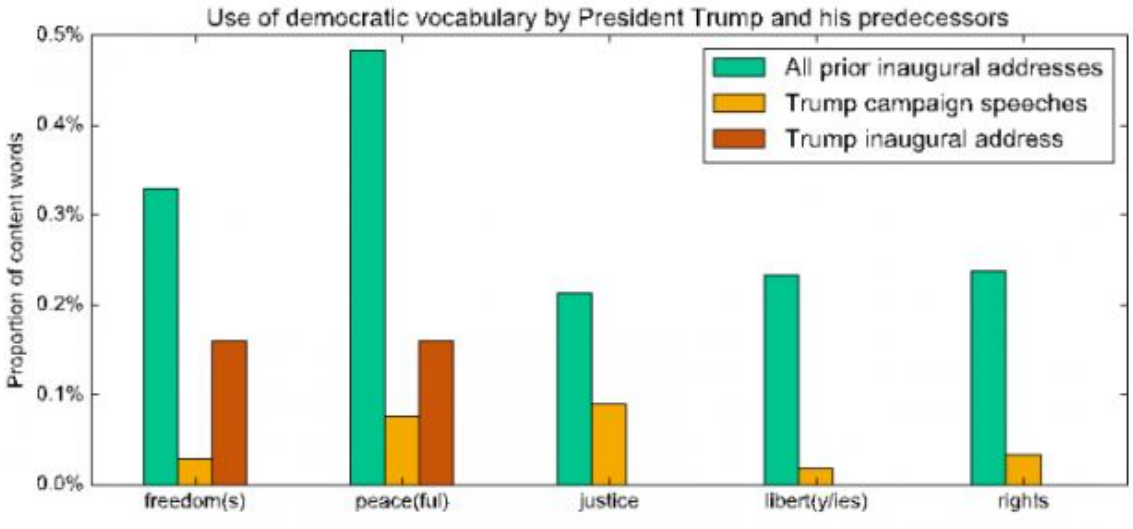


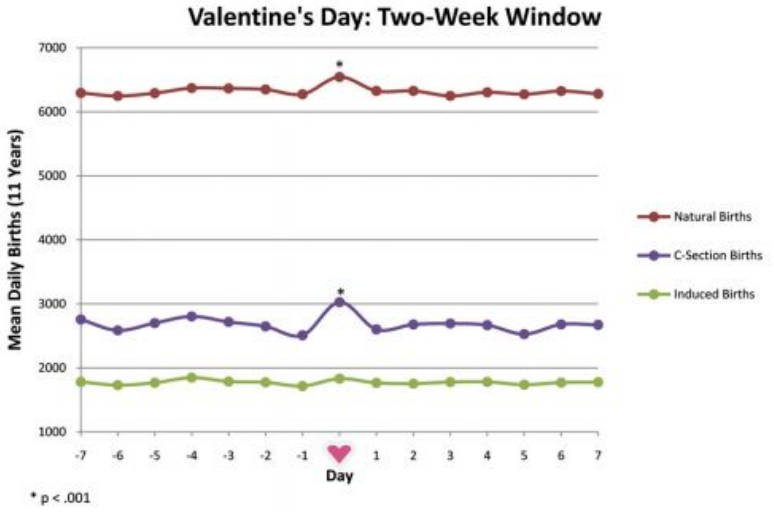
Figure: Christopher Brown

- BTW, “in previous inaugural speeches, ‘liberty’ is the 27th most common content word used in previous inaugural speeches. ‘Trump’ appears to play a similar rhetorical role for the president; in the 62 of his speeches we analyzed, it was the 10th most common content word.”

Any problems with “key words”?

Let me highlight one issue...

- Look at these results: Levy, Becca R., Pil H. Chung, and Martin D. Slade. "Influence of Valentine's Day and Halloween on birth timing." Social Science & Medicine 73.8 (2011): 1246-1248.



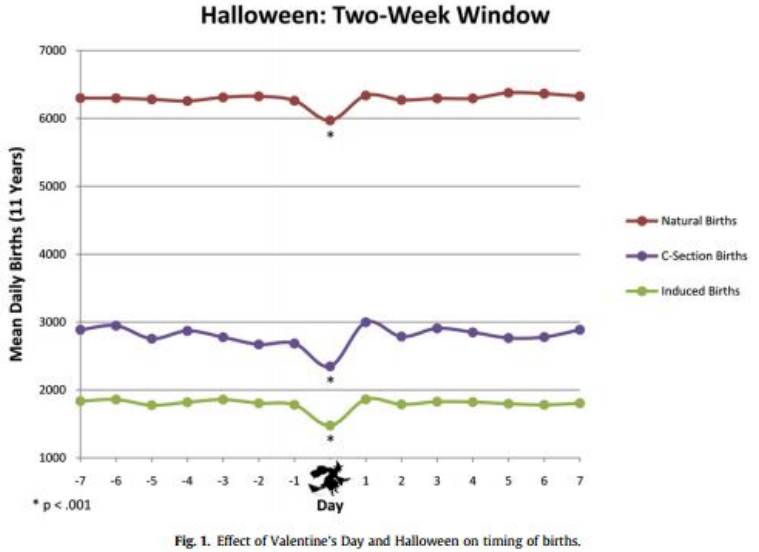
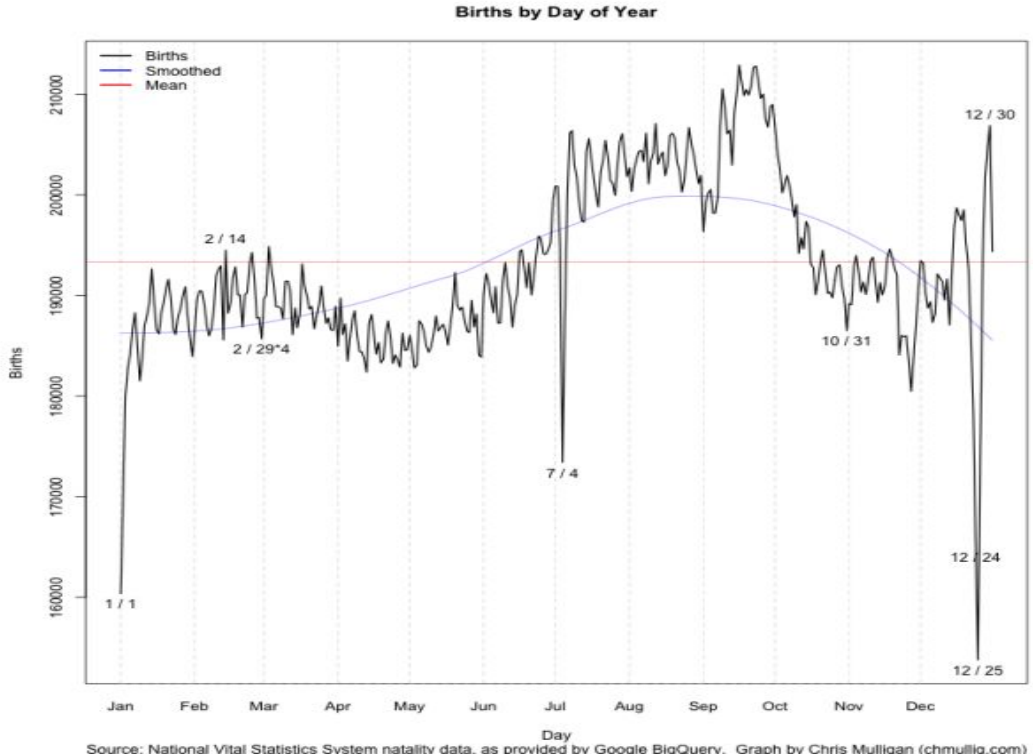


Fig. 1. Effect of Valentine's Day and Halloween on timing of births.

Compare all the days!

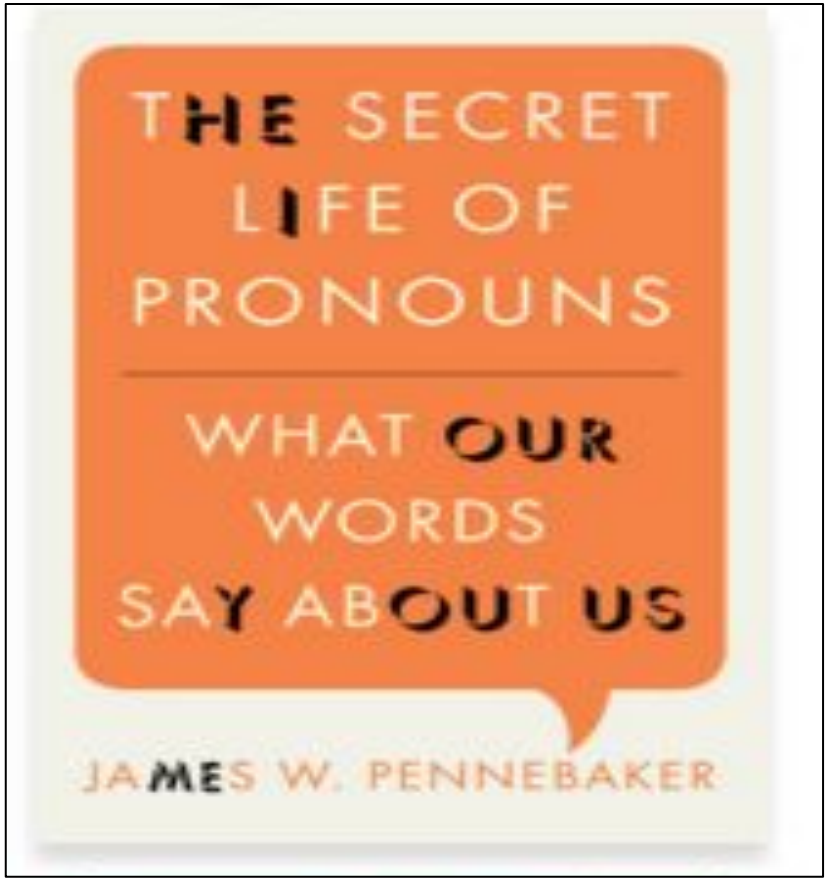
- Does love or fear explain July 4th? Etc.



A less simple example: Counting all the words

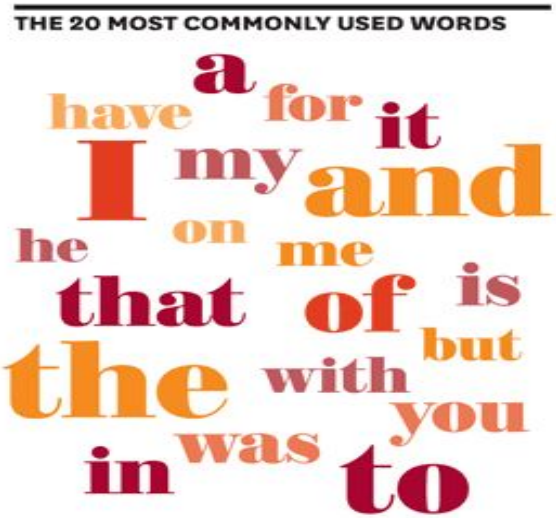
Remember this?

Text analysis-- One example



The question

- Are there clear correlates to increased usage of function words?



Function words

- Out of 100,000 words in the average English speaker's vocabulary, function words account for only about 500, or 0.5%.

- 55% of what we speak, hear, and read in typical speech, however, is made up of these function words.

A set of questions

In daily conversations, e-mails, informal talks, blogs, and even most formal writing, who uses the following parts of speech more, men or women?

For each of the questions, circle the correct answer:

1. First-person singular (e.g., *I*, *me*, *my*):
   - **a.** women use more
   - **b.** men use more
   - **c.** no difference between women and men

2. First-person plural (e.g., *we*, *us*, *our*)
   - **a.** women use more
   - **b.** men use more
   - **c.** no difference between women and men

3. Articles (*a*, *an*, *the*)
   - **a.** women use more
   - **b.** men use more
   - **c.** no difference between women and men

4. Positive emotion words (e.g., *love*, *fun*, *good*)
   - **a.** women use more
   - **b.** men use more
   - **c.** no difference between women and men

5. Cognitive words (e.g., *think*, *reason*, *believe*)
   - **a.** women use more
   - **b.** men use more
   - **c.** no difference between women and men

6. Social words (e.g., *they*, *friend*, *parent*)
   - **a.** women use more
   - **b.** men use more
   - **c.** no difference between women and men

The answers

In daily conversations, e-mails, informal talks, blogs, and even most formal writing, who uses the following parts of speech more, men or women?

For each of the questions, circle the correct answer:

1. First-person singular (e.g., *I*, *me*, *my*):
   - **a.** women use more
   - **b.** men use more
   - **c.** no difference between women and men

2. First-person plural (e.g., *we*, *us*, *our*)
   - **a.** women use more
   - **b.** men use more
   - **c.** no difference between women and men

3. Articles (*a*, *an*, *the*)
   - **a.** women use more
   - **b.** men use more
   - **c.** no difference between women and men

4. Positive emotion words (e.g., *love*, *fun*, *good*)
   - **a.** women use more
   - **b.** men use more
   - **c.** no difference between women and men

5. Cognitive words (e.g., *think*, *reason*, *believe*)
   - **a.** women use more
   - **b.** men use more
   - **c.** no difference between women and men

6. Social words (e.g., *they*, *friend*, *parent*)
   - **a.** women use more
   - **b.** men use more
   - **c.** no difference between women and men

Males and females in dialogue

Who writes dialogue for male characters that most closely mirrors prototypical male speech, and vice versa?

Based on 110 scripts (from 70+ playwrights) --

- Their females talk like prototypical females and males talk like prototypical males: Spike Lee (*Do the Right Thing*), Thornton Wilder (*Our Town*) and Joan Tewkesbury (*Nashville*)

- Their females AND males talk like prototypical females: Nora Ephron (*Sleepless in Seattle*), Callie Khouri (*Thelma and Louise*) and Woody Allen (*Hannah and her Sisters*)

Males and females in dialogue

Who writes dialogue for male characters that most closely mirrors prototypical male speech, and vice versa?

Based on 110 scripts (from 70+ playwrights) --

- Their females AND males talk like prototypical males: Quentin Tarantino (*Pulp Fiction*), Cameron Crowe (*Almost Famous*), and Shakespeare (*Romeo and Juliet*).

Another example: Do political magazines write about female and male public figures differently?

(with my co-author, Feijia Chen, QMSS, Class of 2015)

Different words for males and females

- It is well established that individuals tend to evaluate males and females in different ways, even when they are engaged in the same behavior (Heilman & Chen 2005)

- This difference is more pronounced -- and unfavorable towards women -- if individuals hold strong gender normative expectations (Alexander & Andersen 1993).

The media, male and female

- The media is often thought to generate and perpetuate gendered views of male work and female work (Wood 1994)

- We want to focus on one segment of the media-- i.e., political magazines with explicit ideological commitments

Our goal

- Large-scale text analysis was applied to two political magazines, the *National Review* and the *New Republic*, with different ideological commitments, conservative and neo-liberal, respectively

- What role do their divergent explicit political ideologies play in how, and to what extent, females and males are written about?

Our first hypothesis

- The neo-liberal magazine, the *New Republic*, will write about males and females in essentially the same way, since it has gender equality as an ideological commitment

- The conservative magazine, the *National Review*, may be more comfortable highlighting differences between male and female politicians (due to their more traditional gender values), thereby using different words to describe public figures and issues

An opposing hypothesis

- American politics is structured in such a way that no matter the ideological orientation of the magazine, the types of words associated with female *vs*. male in politics will be equivalent across the different magazines

- This could be due: (**1**) either to the distribution of issues championed by females *vs*. males (so-called “female issues”), or (**2**) because there are strong constraints on journalists in terms of how they can report news and necessitates a standardized script around gender

The inspiration for this paper

Caren's *New York Times* analysis

- Neal Caren asks if males and females come up in different contexts in the *New York Times?*
- He compared the words in sentences that discuss females with the words in sentences that discuss males from the Feb. 27- March 6, 2013 issues
- Specifically, for every sentence in each article, he sought to identify a single subject with a gender (based on a dictionary of words associated with one gender or another, like "her," "she," and "chairwoman," or "he," "his," and "pope"), and he collected all the words in each such sentence

This is obviously a simplistic starting place regarding gender

- This method only focuses on a binary for gender, which should be deepened further in future work

What did he find?

Male	Female	Word
72	02	prime
70	02	baseball
92	03	official
61	02	capital
61	02	governor
75	04	fans
120	07	minister
51	03	sequester
118	07	league
58	04	failed
57	04	cardinals
54	04	finance
78	06	reporters
50	04	winning
73	06	finally
116	10	players
56	05	acknowledged
67	06	address
66	06	attack
108	10	opposition
54	05	rest
53	05	camp
52	05	costs
91	09	goal
50	05	crowd
118	12	bank
57	06	referring
66	07	sports

Male	Female	Word
0	29	pregnant
0	17	husband's
1	16	suffrage
2	25	breast
4	16	gender
6	22	pregnancy
10	21	dresses
13	23	birth
13	22	memoir
25	37	baby
17	25	disease
14	20	interviewed
12	17	abortion
24	34	dress
23	32	married
12	16	activist
25	33	author
14	18	drama
30	36	hair
18	21	rape
24	27	dog
19	21	novel
99	108	children

BTW--

- He could classify 25.9% of all the *Times* sentences as having a gender
- There were 19,681 sentences about males and 6,242 sentences about females
- That means that there are **3.2** sentences about males for each sentence about females

What did he find?

He writes:

> “My quick interepretation: If your knowledge of men's and women's roles in society came just from reading last week's *New York Times*, you would think that men play sports and run the government. Women do feminine and domestic things. To be honest, I was a little shocked at how stereotypical the words used in the women subject sentences were.”

BTW--

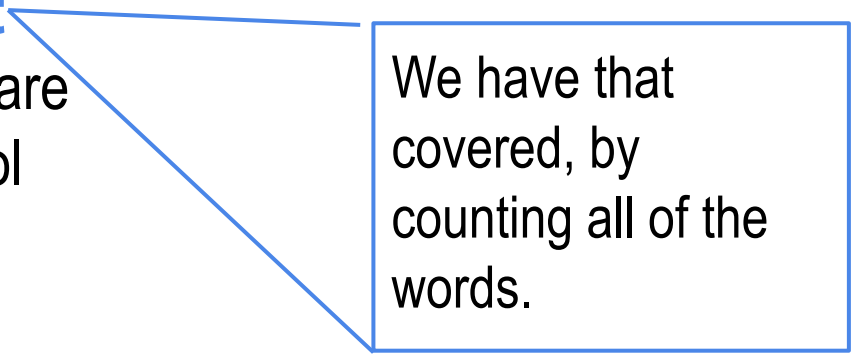
- Caren and Phil Cohen also looked the gender of *Times* reporters and what words show up in their respective headlines --

This got me thinking ...

- What do quantitative social scientists do?
  1. Count
  2. Compare
  3. Control

This got me thinking ...

- What do quantitative social scientists do?
  1. **Count**
  2. Compare
  3. Control



This got me thinking ...

- What do quantitative social scientists do?
  1. Count
  2. **Compare**
  3. Control

We already have a comparison of male *vs*. female, **but** what about one publication *vs*. another publication too?

This got me thinking ...

- What do quantitative social scientists do?
  1. Count
  2. Compare
  3. **Control**

The *Times* covers too many topics, **but** what if we only focused on one domain. By focusing on politics and public life alone, we seek to hold the subject matter constant, while allowing other factors to vary.

The other inspiration for this paper

Political speech similarity

- Can we figure out which speakers are more similar to each other than they are to others?

- Yes. Bruce Sacerdote and Owen Zidar. "Campaigning in Poetry: Candidates Choice of Words in the 2008 Election." Dartmouth Working Paper, 2008.

Political speeches

Can we say which Presidential candidates in 2008 were most like Ronald Reagan (RR) or Martin Luther King (MLK) in their word choice?

Step 1: For any given word *w*, they calculate the relative frequency of its use by the reference authors (RR and MLK). For candidate *i*, each time they use the word *w*, it tells us something about the probability that *i* is more like RR versus MLK.

Political speeches

- For instance, if the word "people" is used twice as often by MLK versus RR, then candidate *i*'s use of the word people would imply that there is a 2/3 chance that *i*'s speech was delivered by someone who speaks like MLK and a 1/3 chance that the speech was delivered by someone who speaks like RR.

Political speeches

- They define this probability of being like MLK as P_w:

$$P_w = \frac{\left[\sum_{MLK} use\ of\ w \Big/ \sum_{MLK} all\ words\right]}{\left[\sum_{RR,MLK} use\ of\ w \Big/ \sum_{RR,MLK} all\ words\right]}$$

Political speeches

- Step 2: They then average over all words *w* to get an estimate of the total probability that a speech was delivered by a candidate who resembles one of the reference orators versus the other, e.g. MLK versus RR. To do this they multiply P_w by the relative frequency with which candidate *i* uses word *w* (labeled F_{wi}) and they sum over all words *w*. By relative frequency F_{wi} they mean the number of times *w* is used as a fraction of total words spoken.

$$P_i(MLK) = \sum_w F_{wi} P_w$$

Results

Ronald Reagan (0) Martin Luther King (1)

RR (0) MLK (1)			Unique	Total
Candidate	Index	Standard Error	Scored Words	Words Scored
John McCain	-0.926	0.054	265,745	2,316
Mitt Romney	0.172	0.079	118,769	1,907
Fred Thompson	0.310	0.110	60,261	1,501
John Edwards	0.606	0.067	171,427	1,923
Rudy Giuliani	0.611	0.079	112,686	1,442
Hillary Clinton	0.751	0.035	635,177	2,506
Barack Obama	1.051	0.037	545,123	2,547
Mike Huckabee	1.466	0.146	31,666	1,061

Results

George W. Bush (0) William Jefferson Clinton (1)

GWB-WJC			Unique	Total
		Standard	Scored	Words
Candidate	Index	Error	Words	Scored
John McCain	-0.8133	0.0392	7,187	291,173
Mitt Romney	-0.2770	0.0564	4,798	128,800
Mike Huckabee	0.5876	0.1047	2,054	34,221
Fred Thompson	0.6509	0.0731	3,118	64,074
John Edwards	0.7887	0.0437	4,876	184,266
Rudy Giuliani	0.8073	0.0546	3,099	120,794
Barack Obama	0.8391	0.0251	8,643	590,639
Hillary Clinton	1.4210	0.0229	9,550	690,330

Results

Stalin (0) John F. Kennedy (1)

Stalin (0) JFK (1)			Unique	Total
		Standard	Scored	Words
Candidate	Index	Error	Words	Scored
Mike Huckabee	-0.587	0.231	31,561	1,108
Fred Thompson	0.007	0.168	60,316	1,655
Rudy Giuliani	0.236	0.122	112,519	1,589
Mitt Romney	0.318	0.120	118,997	2,124
John McCain	0.577	0.081	269,096	2,752
John Edwards	0.973	0.101	172,399	2,139
Barack Obama	1.081	0.057	546,643	2,865
Hillary Clinton	1.691	0.053	640,346	2,949

BTW--

Mentions of Opponents in Campaign Speeches During 1/1/2006 to 1/14/2008

Candidate	Negative Mentions Per 10,000 Words	Number of Speeches Analyzed
Hillary Clinton	1.376	50
Barack Obama	1.803	54
John Edwards	6.568	16
Fred Thompson	2.283	9
John McCain	3.018	37
Mitt Romney	8.517	14

How this applies in our case

- Instead of asking the question: Given that these words are used this often by Obama, for instance, how close is his speech to being like MLK *vs*. RR?

- We ask: Given that these words are used this often by the *New Republic* to talk **about females**, how close is their writing on women to being like the *New Republic*'s writing **for males** *vs*. the *National Review*'s writing **for females**?

Data and methods

- We collected all the articles from the *National Review* and the *New Republic* from Dec. 2012 to March 2013, representing half a million of words.
- Applied Caren's methods

Python code to collect our data

```
from __future__ import division

import glob
import nltk
from string import punctuation

tokenizer = nltk.data.load('tokenizers/punkt/english.pickle')

male_words=set(['guy','spokesman','chairman',"men's",'men','him',"he's",'his','boy','boyfriend','boyfriends','boys','br
other','brothers','dad','dads','dude','father','fathers','fiance','gentleman','gentlemen','god','grandfather','grandpa','gr
andson','groom','he','himself','husband','husbands','king','male','man','mr','nephew','nephews','priest','prince','son','
sons','uncle','uncles','waiter','widower','widowers'])
female_words=set(['heroine','spokeswoman','chairwoman',"women's",'actress','women',"she's",'her','aunt','aunts','b
ride','daughter','daughters','female','fiancee','girl','girlfriend','girlfriends','girls','goddess','granddaughter','grandma','
grandmother','herself','ladies','lady','lady','mom','moms','mother','mothers','mrs','ms','niece','nieces','priestess','prin
cess','queens','she','sister','sisters','waitress','widow','widows','wife','wives','woman'])
print len(male_words)
def gender_the_sentence(sentence_words):
    mw_length=len(male_words.intersection(sentence_words))
    fw_length=len(female_words.intersection(sentence_words))

    if mw_length>0 and fw_length==0:
```

Data and methods

- We ended up with tens of thousands of words that are associated with “male” sentences in each one of these magazines in these two years, and likewise for “female” sentences
- We then removed all proper names and all “stop” or “functional” words (the, a, with, from, is, *etc.*)
- In the end, we had a classic “bag of words” dataset

What that looks like in R

```
male = scan(file.choose(), what="char") ## get your “male” text file  ##

male <-tolower(male) ## make everything lower case ##

install.packages("tm")
library(tm)

removePunctuation(male) ## remove punctuation ##

install.packages("SnowballC") ## so as to stem the words ##
library(SnowballC)
stemDocument(male)

male.list<-strsplit(male, "\\W+", perl=TRUE) ## make a list of words ##
male.vector<-unlist(male.list)
male.freq.list<-table(male.vector)

male.sort.freq.list<-sort(male.freq.list, decreasing=TRUE) ## have the list go from
most frequent to the least frequent ##
```

What that looks like in R

```
male.sort.table<-paste(names(male.sort.freq.list), male.sort.freq.list, sep="\t")

cat("Word\tFREQ", male.sort.table, file=choose.files(), sep="\n") ## send this file to
some Folder and give it a name, and then later, you can open it with Excel ##

## now do the same thing on the other text file ###

female = scan(file.choose(), what="char") ## get your “female” text file  ##

(etc.)
```

What that looks like in R

```
### and to get the merged dataset all together, do this ####

exceloffrequencies1 = read.csv(file.choose()) ## choose Bag #1 Excel of frequencies,
with one variable called mfreq and the other called word ##

exceloffrequencies2 = read.csv(file.choose()) ## choose Bag #2 Excel of frequencies,
with one variable called ffreq and the other called word ##

merge<-merge(exceloffrequencies1,exceloffrequencies2, by=c("word"), all=T)


#### now you have your whole dataset of frequencies by type of male or not ####
```

… I will show you other ways too

Note well

- We did not “stem” the words for this paper, because our particular literature did not do so, but in general, stemming makes a lot of sense

What we found

Result #1

- Males are talked about more often (i.e., more sentences) than females, and even more so at the *National Review*

	2012-2013		
	Male	**Female**	**Ratio**
New Republic	3,000	617	4.9
National Review	3,030	462	6.6

A common finding

American Sociological Review
2015, Vol. 80(5) 960–984
© American Sociological Association 2015
DOI: 10.1177/0003122415596999
http://asr.sagepub.com

A Paper Ceiling: Explaining the Persistent Underrepresentation of Women in Printed News

Eran Shor,[a] Arnout van de Rijt,[b] Alex Miltsov,[a] Vivek Kulkarni,[b] and Steven Skiena[b]

Abstract

In the early twenty-first century, women continue to receive substantially less media coverage than men, despite women's much increased participation in public life. Media scholars argue that actors in news organizations skew news coverage in favor of men and male-related topics. However, no previous study has systematically examined whether such media bias exists beyond gender ratio imbalances in coverage that merely mirror societal-level structural and occupational gender inequalities. Using novel longitudinal data, we empirically isolate media-level factors and examine their effects on women's coverage rates in hundreds of newspapers. We find that societal-level inequalities are the dominant determinants of continued gender differences in coverage. The media focuses nearly exclusively on the highest strata of occupational and social hierarchies, in which women's representation has remained poor. We also find that women receive greater exposure in newspaper sections led by female editors, as well as in newspapers whose editorial boards have higher female representation. However, these differences appear to be mostly correlational, as women's coverage rates do not noticeably improve when male editors are replaced by female editors in a given newspaper.

sociological science

A Large-Scale Test of Gender Bias in the Media

Eran Shor,[a] Arnout van de Rijt,[b] Babak Fotouhi[c]

a) McGill University; b) Utrecht University; c) Harvard University

Abstract: A large body of studies demonstrates that women continue to receive less media coverage than men do. Some attribute this difference to gender bias in media reporting—a systematic inclination toward male subjects. We propose that in order to establish the presence of media bias, one has to demonstrate that the news coverage of men is disproportional even after accounting for occupational inequalities and differences in public interest. We examine the coverage of more than 20,000 successful women and men from various social and occupational domains in more than 2,000 news sources as well as web searches for these individuals as a behavioral measure of interest. We find that when compared with similar-aged men from the same occupational strata, women enjoy greater public interest yet receive less media coverage.

Keywords: gender; bias; media coverage; computational analysis

MEDIA attention is crucial for individuals in various social and occupational domains and can have substantial consequences for their success, and even

Public Interest vs. Media Attention too

Result #2a

- How much does a sentence about females from the *New Republic* look like a sentence about males from the *New Republic* (=1), *vs*. a sentence about a females from the *National Review* (=0)?

Result #2b

- How much does a sentence about females from the *National Review* look like a sentence about males from the *National Review* (=1), *vs*. a sentence about a female from the *New Republic* (=0)?

Result #2, conclusion

- The evidence suggests that overall, male sentences and female sentences at each magazine are very similar to each other (though not exactly the same), implying that an overriding editorial voice is at work for both genders

- Our hypothesis that political magazines express different attitudes towards males and females depending on their gender ideology, therefore, does not find support

Alternate statistical measures

- We could have used alternate measures to compare one bag of words to another, such as:

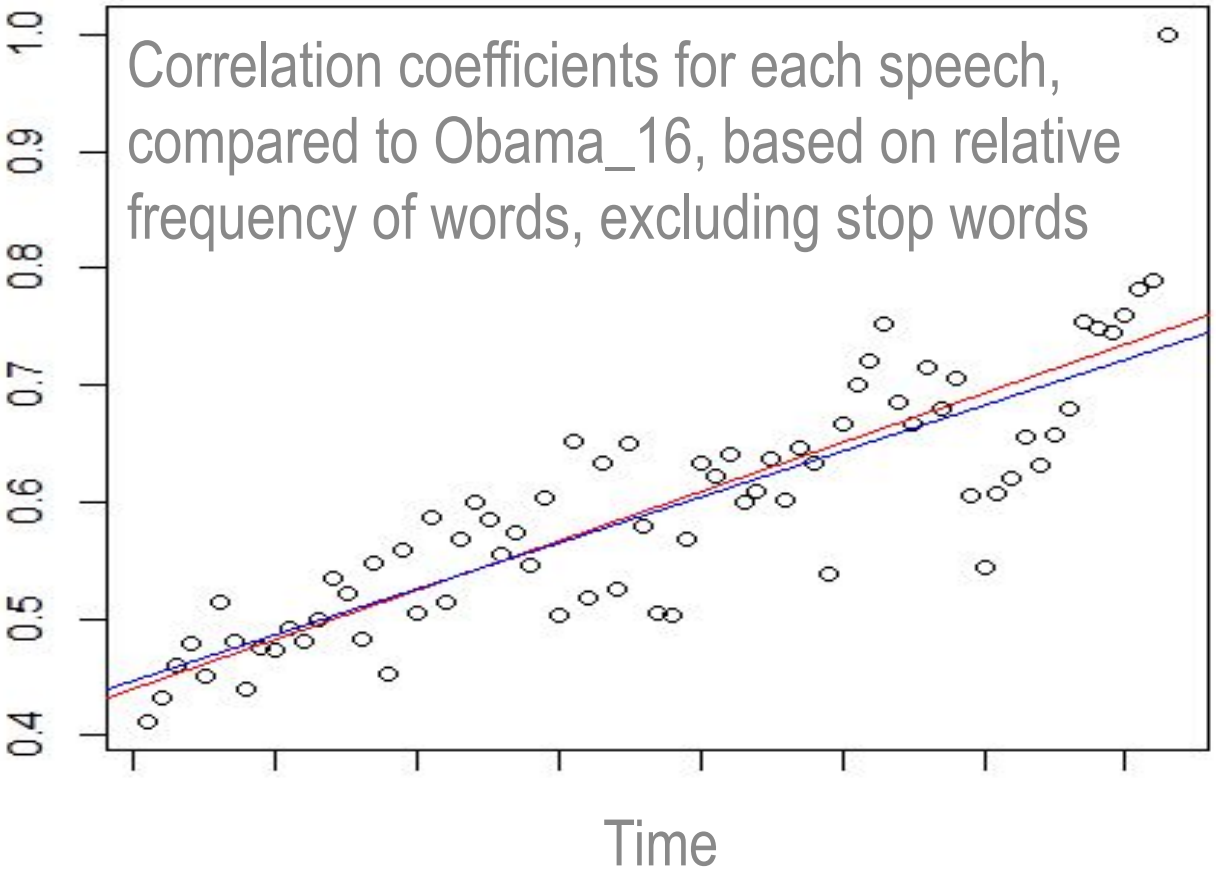
1. Pearson correlations of word frequencies (or of ranks)
2. Chi-square tests for differences in expected frequencies

- Sometimes there are additional corrections done to re-weight relative frequencies by the sheer commonness of the words too

Let's see a correlation …
Remember this question?

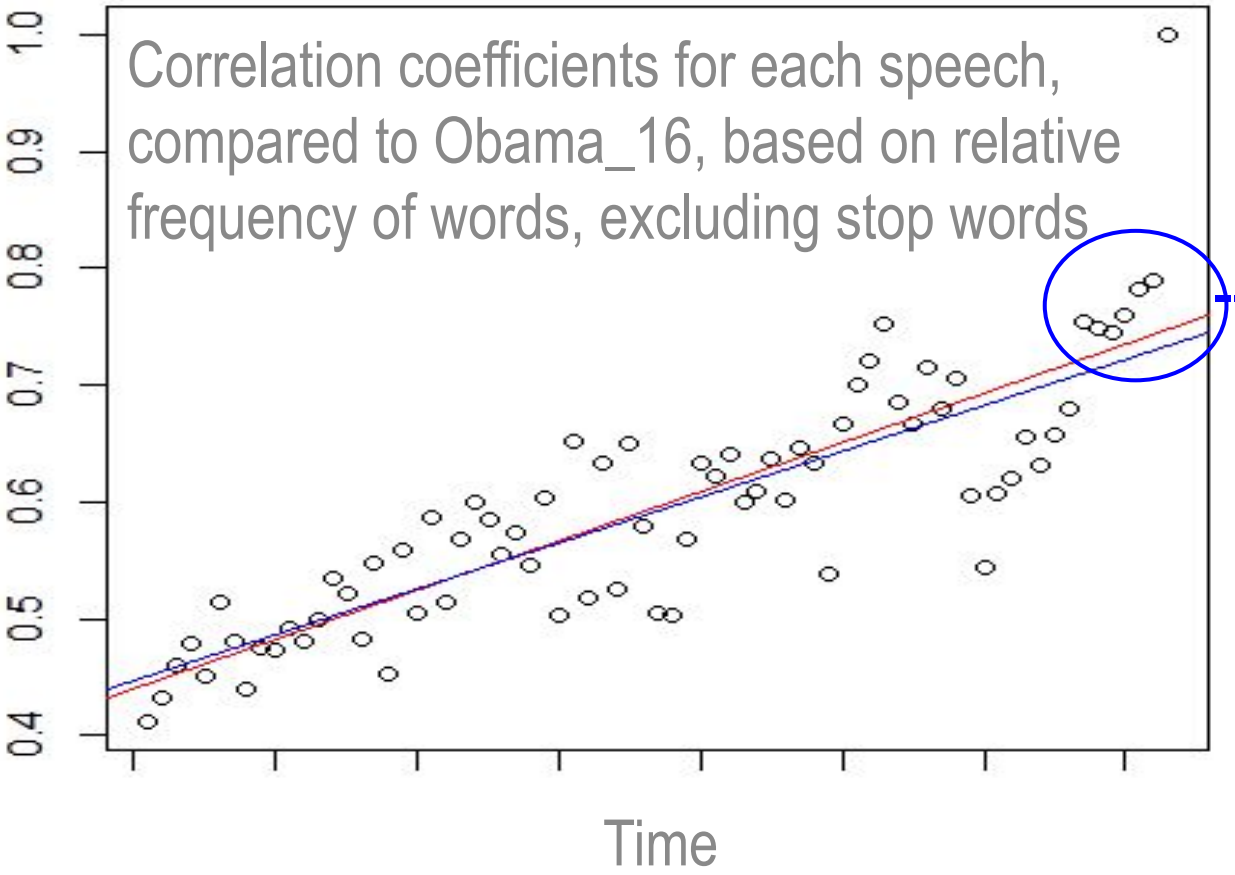
- Can we figure out from word choice which President delivered which States of the Union?

Obama's 2016 State of the Union



Who does Obama in 2016 most speak like?

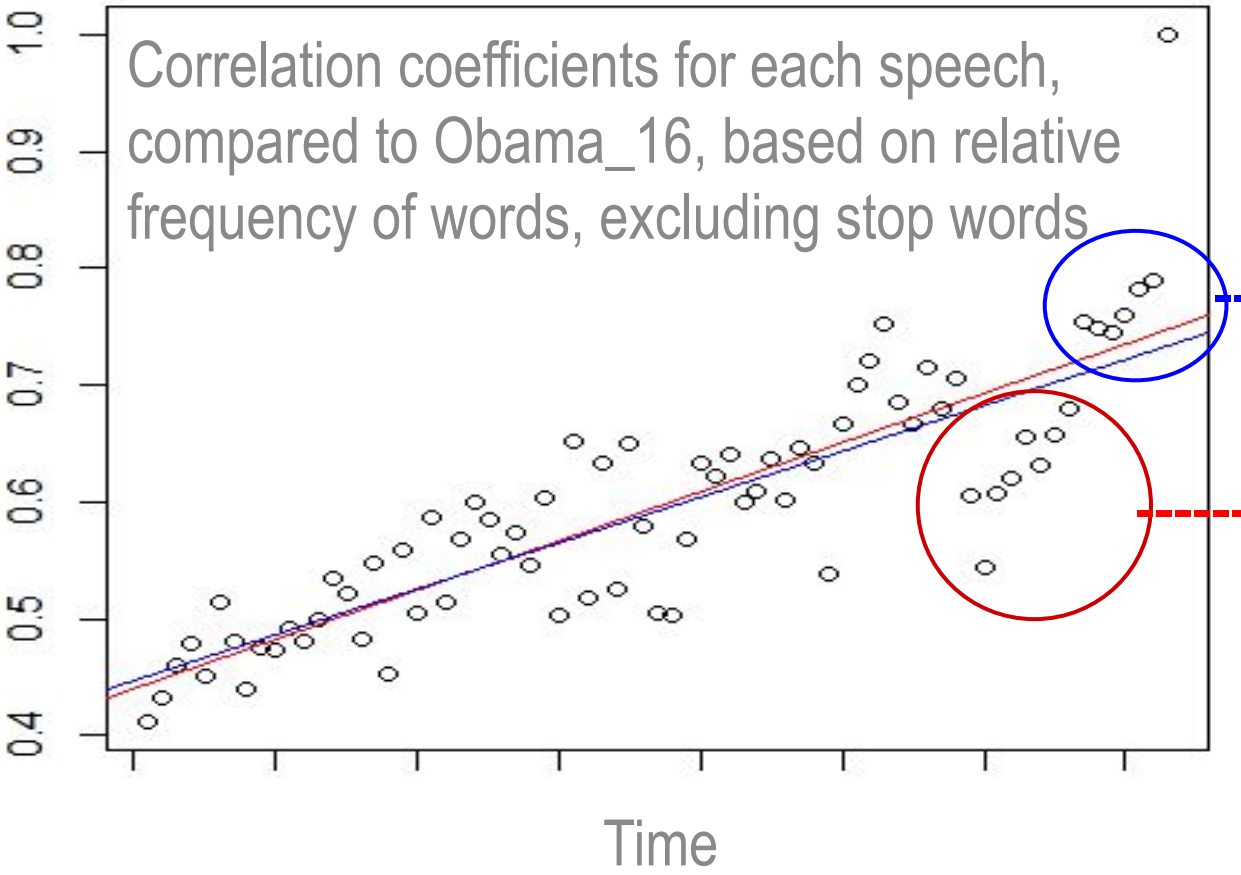
Obama's 2016 State of the Union



Who does Obama in 2016 most speak like?

----------Obama

Obama's 2016 State of the Union

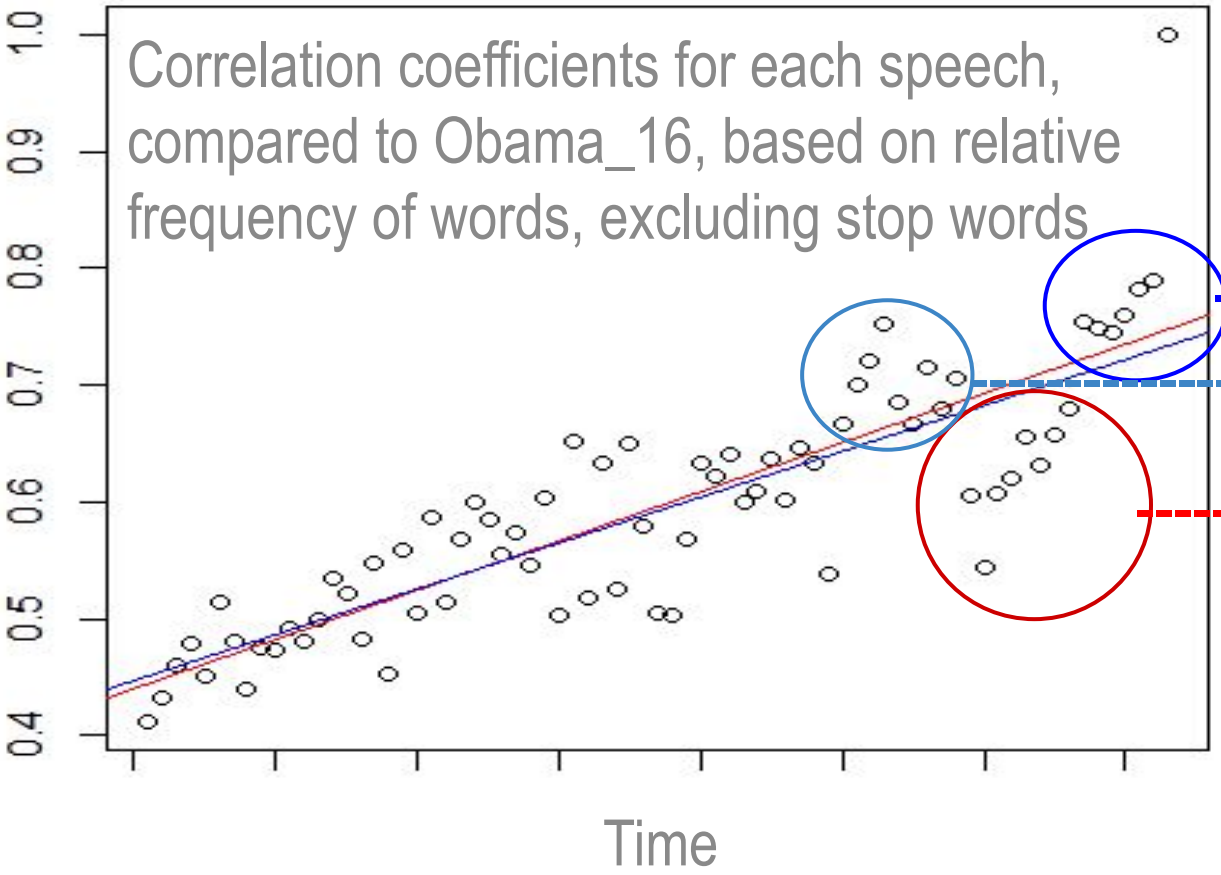


Who does Obama in 2016 most speak like?

----------Obama

----------W. Bush

Obama's 2016 State of the Union



Who does Obama in 2016 most speak like?

Obama

Clinton

W. Bush

On comparison issues, see for instance:

- Rayson, Paul, Damon Berridge, and Brian Francis. "Extending the Cochran rule for the comparison of word frequencies between corpora." *7th International Conference on Statistical analysis of textual data (JADT 2004)*. 2004.
- Baron, Alistair, Paul Rayson, and Dawn Archer. "Word frequency and key word statistics in corpus linguistics." *Anglistik* 20.1 (2009): 41-67.

Another interesting comparison

SHORT ARTICLE

Have State Policy Agendas Become More Nationalized?

Daniel M. Butler, Washington University in St. Louis
Joseph L. Sutherland, Emory University

Previous work has shown that US voters are focused on national news and national issues and that US elections have become more nationalized. We explore whether state policy agendas have become more nationalized over time. We measure the state agenda by analyzing governors' State of the State addresses from 1960 to 2016. Our analysis shows that state agendas have become more similar to each other over time and that state agendas are more similar to the national agenda (as laid out in the State of the Union address). The nationalization of US politics is not only affecting voters and elections; it is also seen in the nationalization of the policy agenda.

Governors have incentives to please their constituents by putting forth an agenda focused on the issues they care about. Because Americans are consuming more national political news and less state and local political news (Hopkins 2018; Martin and McCrain 2019), governors have incentives to focus on these national issues. Yet, the ability to define political priorities shapes political conflict and is a source of political power (Schattschneider 1960). Governors may use that power proactively rather than simply responding to voters' increased focus on national issues (e.g., Butler and Hassell 2018).

$p < .01$). Likewise, the SOTS addresses have become more similar to the presidential SOTU address. In 1960, one of those topics would be covered in both the SOTS and SOTU; in 2016, four to five of those topics would be covered in both the SOTS and SOTU (a nearly fourfold increase in topical cosine similarity over the period, $p < .01$). These patterns hold for both Republicans and Democrats and for all regions in the country. The nationalization of US politics is also observed in the policy agenda.

BUILDING THE SOTS ADDRESSES CORPUS

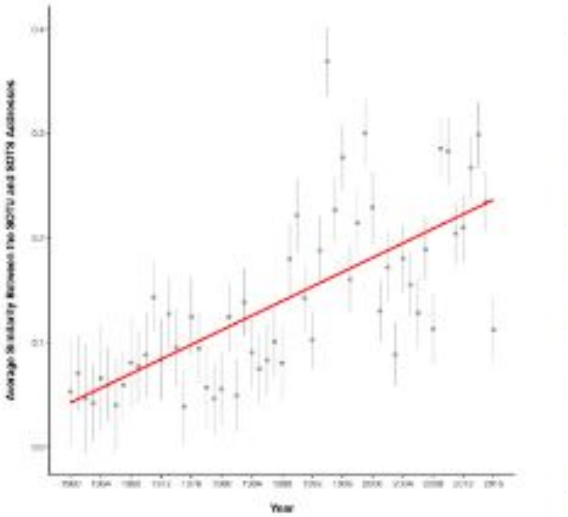


Figure 4. Similarity of SOTS and the SOTU addresses has increased over time. Coefficient estimates from a regression of similarities between SOTS addresses, given by governors, and SOTU addresses, given by the president, on a set of dummy variables for year. Each point is the coefficient estimate for that year. Lines intersecting the points are 95% confidence intervals for the coefficient estimates. A thick line of best fit overlays the points.

I will have a lot more to say about how to compare texts next week

Result #3

- Despite the results of our quantitative methodology, our qualitative sense still supports the idea that males and females are talked about differently, independent of magazine ideology
- There were still qualitative differences in how some words were disproportionately applied to males *vs*. females, and in ways consistent with previous investigations and theories

For the *National Review*

Predominantly “male” words

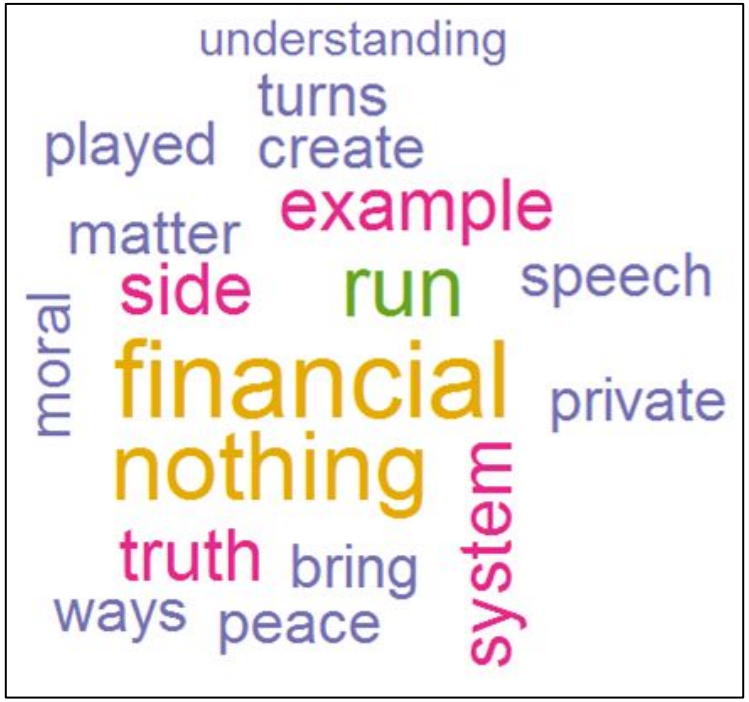
Predominantly “female” words

What that looks like in R

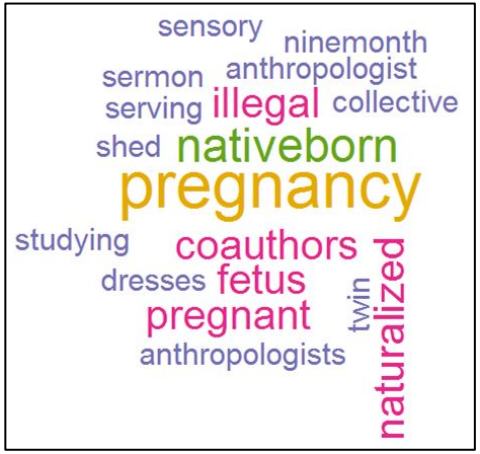
```
male = scan(file.choose(), what="char") ## get your “male” text file  ##

male <-tolower(male) ## make everything lower case ##

install.packages("tm")
library(tm)

removePunctuation(male) ## remove punctuation ##

install.packages("wordcloud")
library(wordcloud)
wordcloud(male, min.freq=2) # draw a word cloud of male words #
```

For the *New Republic*

Predominantly “male” words



Predominantly “female” words



Result #3

- Predominantly female words concerned motherhood and passivity, like “pregnancy, ” “fetus,” “nine-month” and “dresses (in the *New Republic*) and like “baby,” “husband’s,” “cute” and “beaten” (in the *National Review*).
- Predominantly male words concerned leadership and activity, like “run,” “bring,” create” and “moral” (in the *New Republic*) and like “big,” “effort,” “winning,” and “taking” (in the *National Review*).

Result #3

- These are the most heavily gender skewed words in each magazine, and in fact, most of the male words appeared dozens of times, without a single instance of those words applying to women.
- This still suggests that there may be a relatively narrow script about how some small subset of topics are discussed concerning men and women that transcend magazine ideology

Result #3

- This is what the previous literature primed us to expect a whole lot of, but we seemed to only get a bit, though the bit is damning
- Through the content analysis of newspaper coverage, Niven (2005) finds that women in Congress receive less coverage and less issue coverage
- Women in Congress are less likely to be portrayed as successful in their legislative efforts and less likely to be portrayed as competent

Result #3

- Niven (2005) also finds that females are more likely to be portrayed as advocating for a specific group, more likely to be portrayed as partisan, and more likely to be portrayed as ineffective. They often are only reported on concerning “female issues.”
- Females are also more likely to have their personal life details – their age, appearance, family life, personality – discussed in the media than their male counterparts, even though male legislators are more likely to post that information to their websites

In conclusion

- From our word counts, we get the overall impression of similar depictions of males and females within each magazine
- Yet, there is a clear sense that males and females are not being talked about quite the same way or on the same issues

Future steps and considerations

- We also have earlier years for *New Republic* and *National Review*, to compare the past -- and we also have a "non-partisan" magazine (*National Journal*) as another control
- Counts vs. n-grams vs. context vs. sentiment
- Are we capturing topics or evaluation?
- More, much more to still work on ...

Another example (also from Caren)

His question

- What words are most associated with the abstracts for high status sociological journals? What words are most associated with the abstracts from low status sociological journals?

Neal Caren

← A Sociology Citation Network

The Most Cited Articles in Sociology by Journal →

"Our findings show" What words to use in an abstract

Posted on May 23, 2012 by Neal Caren

One of the fun and simple things to do with large text databases that are already categorized into groups is to see what words are used more frequently in certain kinds of texts than in others. For example, Google Book's Ngram Viewer lets you track how different words were used in different years across the twentieth century. Or, the Sunlight Foundation's Capitol Words project lets you search through the Congressional Record for specific terms and graph word frequencies by date and party.

Sociologists are well aware that language use and status are highly correlated, so I thought it would be fun to investigate how the words used in high status journals differs

Search

Neal Caren in an assistant professor of Sociology at the University of North Carolina, Chapel Hill.
Email
Twitter

Another example (also from Caren)

- From Web of Science, he collected all the information available on articles appearing in 37 general interest sociology journals over the last five years, including the abstract.
- Bag of Words #1 = Abstracts from the 4 highest status journals: *American Journal of Sociology, American Sociological Review, Social Forces* and *Social Problems*.
- Bag of Words #2 = Abstracts from 3 three lower status, general interest journals.

Then he compared relative frequencies

Table 1. Abstract words sorted by the likelihood of appearing in a high status journal abstract compared to a low status journal abstract. Words must have appeared in at least 10% of high status abstracts and have a ratio of 1.2 of higher.

Word	Ratio	Proportion high	Proportion low
outcomes	2.21	0.10	0.05
inequality	2.15	0.12	0.05
through	1.89	0.22	0.11
population	1.79	0.10	0.06
model	1.77	0.11	0.06
analyses	1.73	0.16	0.09
models	1.70	0.12	0.07
find	1.65	0.28	0.17
theories	1.61	0.10	0.06
effect	1.60	0.15	0.09
us	1.60	0.10	0.07
first	1.60	0.13	0.08
characteristics	1.56	0.11	0.07
longitudinal	1.53	0.12	0.08

- "As a measure of what is valued in high status sociology, I think the list has a great deal of face validity. We value developing and testing theory through the scientific process, and words like, outcomes, new, model, find, process, effect and findings reflect these values. High status work is also a collective process-"we" or the "authors" are doing the finding."

Then he compared relative frequencies

Table 2. Abstract words sorted by likelihood of appearing in a low status journal abstract compared to a high status journal abstract. Words must have appeared in at least 10% of low status abstracts and have a ratio of 1.2 of higher.

Word	Ratio	Type 1	Type 2
sociological	1.96	0.12	0.06
race	1.58	0.12	0.08
based	1.55	0.11	0.07
been	1.52	0.17	0.11
well	1.49	0.13	0.09
i	1.37	0.21	0.16
cultural	1.36	0.15	0.11
two	1.36	0.22	0.16
examines	1.36	0.19	0.14
relationship	1.28	0.14	0.11
research	1.26	0.36	0.28
theoretical	1.24	0.10	0.08
gender	1.24	0.15	0.12
has	1.24	0.30	0.24
as	1.23	0.59	0.48
some	1.22	0.14	0.11

- “Interestingly, the key word to avoid is “sociological”. Don’t tell people your work is sociological; show them. “Has” and “been” both show up, so the passive voice is also to be avoided.”

Which abstract sounds the most high status?

This article investigates the effect of family life course transitions on labor allocation strategies in rural Chinese households. We highlight three types of economic activity that involve reallocation of household labor oriented toward a more diversified, nonfarm rural economy: involvement in wage employment, household entrepreneurship, and/or multiple activities that span economic sectors. With the use of data from the China Health and Nutrition Survey (CHNS 1997, 2000, and 2004), our longitudinal analyses of rural household economic activity point to the significance of household demography, life course transitions, and local economic structures as factors facilitating household labor reallocation. First, as expected, a relatively youthful household structure is conducive to innovative economic behavior Second, household entrances and exits are significant, but their impacts are not equal. Life events such as births, deaths, marriage, or leaving home for school or employment affect household economy in distinctive ways. Finally, the reallocations of household labor undertaken by households are shaped by local economic structures: in particular the extent of village-level entrepreneurial activity, off-farm employment, and out-migration.

The winner:

Chen, Feinian, and Kim Korinek. "Family life course transitions and rural household economy during China's market reform." *Demography* 47.4 (2010): 963-987.

- As Caren notes, "I was expecting an article from one of the journals used to train the algorithm to win the prize. That an article from a high status journal that wasn't included in developing the model won means overfitting likely isn't a problem."

How to write your abstract:

- Use big blue words
- Avoid big orange ones

Or as Caren says:

- “More practically, feel free to use this list as a how to when writing your own abstract. For example, take the paper you are working on now and delete the line, “My research examines the relationship…” and replace it with “Our findings show…” Feel free to add me as co-author, at least of the abstract.”

A former student's exercise

- This is just another bag of words analysis
- I have posted the R code for this exercise on Courseworks

I did some quick text mining of today's inaugural address and compared it to Obama's address from 2009. some observations:

1. Obama used 1,176 unique words; Trump used 745 unique words.

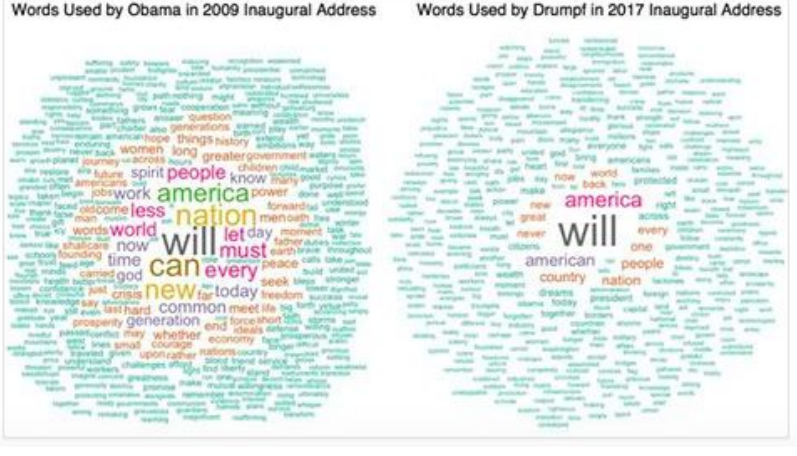
2. Obama used those unique words with more diverse frequency. For example, Obama's 12 most frequently used words represent almost 10% of his speech, while almost 10% of Trump's speech consisted of only 3 words: "american" (12 times), "america" (17 times), and "will" (40 times).

3. Obama said "women" 4 times; Trump said "women" 2 times.

4. Words Obama used that Trump did not even once: "generation" and "generations" (8 times combined), "peace" (4 times), "economy" (3 times), "father" (3 times), "founding" (3 times), "freedom" (3 times), "history" (3 times), "hope" (3 times), "ideals" (3 times), "ambitions" (2 times), "cooperation" (2 times)

5. Words Trump used that Obama did not even once: "triumphs" (2 times), "victories" (2 times), "winning" (2 times), "carnage" (1 time)

6. yes I cherry picked the above words. #missyouobama



Lots of text analysis going on ...

Unique Trump words:

sprawl, ignored, windswept, overseas, tombstones, rusted-out, trapped, neighborhoods, landscape, flush, carnage, unrealized, robbed, stolen, likes, listening, hardships, transferring, politicians, reaped, stops, subsidized, disagreements, bedrock, Islamic, reinforce, solidarity, unstoppable, brown, mysteries, arrives, politicians, hire, infrastructure, trillions, depletion, allowing, disrepair, redistributed, tunnels, stealing, ravages, issuing, blood

The Fix

Trump's inaugural address was demonstrably bleak

By Philip Bump January 20

On Jan. 20, 2017, President Trump took the oath of office, pledging in his inaugural address to

Mos

1

2

3

4

4. Sentiment analysis via automatic dictionaries

An example

- Ryan C. Black, Sarah A. Treul, Timothy R. Johnson, and Jerry Goldman. Emotions, oral arguments, and Supreme Court decision making. *The Journal of Politics*, 73(2):572–581, April 2011.

- Can measuring the emotional content of Supreme Court justices' questions and comments made during oral arguments allow us to predict the decisions they subsequently make?

Classifying Justices' words

- The *Dictionary of Affect in Language* was used to gauge the emotional content of the justices' words during oral arguments.

- Emotion in language can be described adequately and efficiently in terms of a two-dimensional space defined by the pleasantness and activation of words. She measures each dimension on a 3-point scale; words are unpleasant, neutral, or pleasant as well as passive, neutral, or active.

Classifying Justices' words

- ~800 (10% most unpleasant) unpleasant words: chaos, failed, hostile, nightmare, and phony.

- ~800 (10% most pleasant) pleasant words: award, confidence, favorable, quality, and respect.

Their results

TABLE 1 Logistic Regression Models Predicting Case and Vote Outcomes

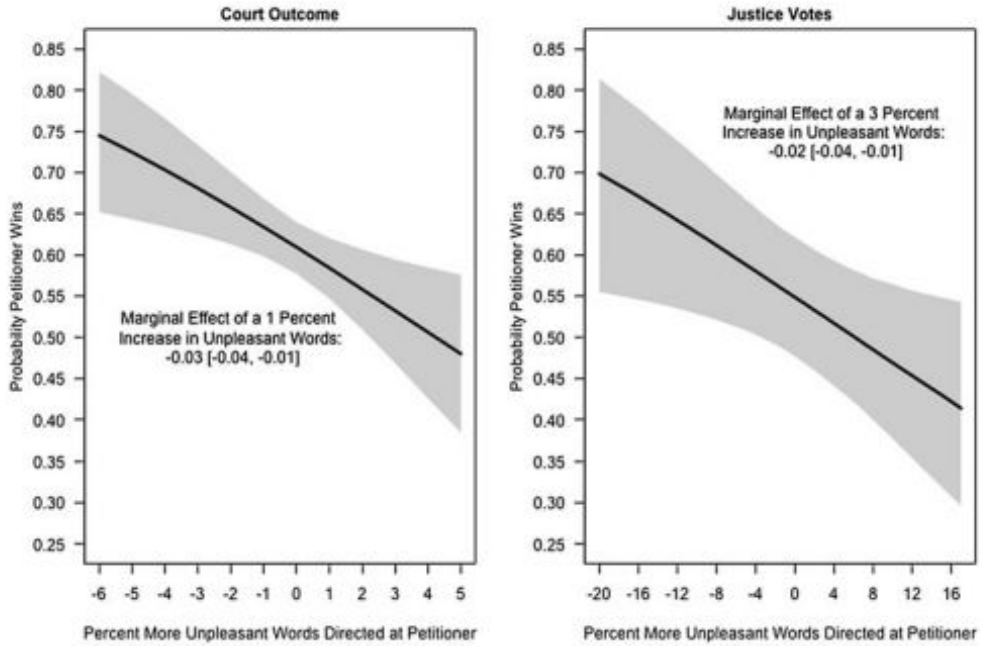
	Court Outcome (1979-2008)	Justice Votes (2004-2008)
Percent More Unpleasant Words Directed at Petitioner	−0.105* (0.037)	−0.032* (0.013)
Percent More Pleasant Words Directed at Petitioner	0.081 (0.042)	−0.011 (0.014)
Number More Questions Directed at Petitioner	−0.021* (0.002)	−0.061* (0.007)
Political Ideology	−0.334 (0.203)	0.276* (0.039)
Lower Court Decision Was Conservative	0.310 (0.201)	1.018* (0.243)
Political Ideology x Lower Court Conservative	−0.321 (0.300)	−0.531* (0.052)
Solicitor General as Amicus Supporting Petitioner	0.737* (0.124)	1.079* (0.268)
Solicitor General as Amicus Supporting Respondent	−0.891* (0.149)	−0.663* (0.255)
Number of Amicus Briefs Supporting Petitioner	0.080* (0.019)	0.054* (0.023)
Number of Amicus Briefs Supporting Respondent	−0.084* (0.019)	−0.038* (0.016)
Petitioner's Level of Resources	0.091* (0.019)	0.078 (0.043)
Respondent's Level of Resources	−0.067* (0.018)	−0.113* (0.043)
Constant	0.483* (0.241)	0.240 (0.478)
Observations	2996	3042
Log Likelihood	−1765.667	−1673.836
Psuedo R2	0.107	0.161
Percent Correctly Predicted	68.7	72.7
Proportional Reduction in Error	15.7	24.8

*denotes $p < 0.05$ (two-tailed test).

Note: Robust standard errors are reported in parentheses next to maximum-likelihood parameter estimates. Standard errors for the Justice Votes model are clustered on each unique case (N = 347). Political Ideology (and its part in the interaction term) for the Court Outcome model is the median justice's ideal point estimate. Political Ideology (and its part in the interaction term) for the Justice Vote model is the voting justice's ideal point estimate.

Graphing the effect

FIGURE 1 Predicted Effect of Unpleasant Language on Court and Justice Vote Outcomes



Note: The left panel represents the Court outcome dependent variable, where a 1 percent increases is slightly less than one standard deviation (1 S.D. = 1.11 percent). The right panel represents the justice vote dependent variable, where a 3 percent increase is the same as one standard deviation. Both marginal effects are calculated using the sample mean (i.e., 0 percent) as the baseline value. All other variables were held at their sample means or modes as appropriate. Shaded regions represent the 95 percent confidence interval obtained through stochastic simulations.

Using Justices' words

- Supreme Court decisions are surprisingly predictable, even just using ~1,600 (rather severe) words as clues

What is going on with sentiment analysis?

- Counting up words and scoring them

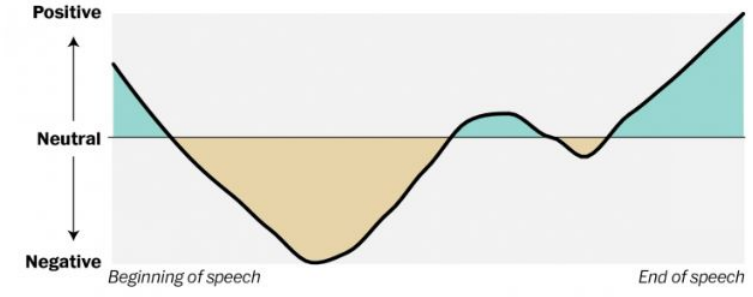
Another analysis of Trump

How Trump's inaugural address compares to his predecessors, charted

By **Christopher Ingraham** January 20

Trump's inaugural address

Sentiment structure of Donald Trump's inaugural address



WAPO.ST/**WONKBLOG** Source: NPR address transcript

Note: Analyzed using the Syuzhet R package by Matthew Jockers of the University of Nebraska-Lincoln

How to do sentiment analysis

```
import pandas as pd
import numpy as np
import glob
import nltk
from textblob import TextBlob
import matplotlib.pyplot as plt


test = ['I am a very happy person.', 'I am a very sad person',
        'I've always understood happiness to be appreciation. There is no greater happiness than \
        appreciation for what one has- both physically and in the way of relationships and \
        ideologies. The unhappy seek that which they do not have and can not fully appreciate the \
        things around them. I don't expect much from life. I don't need a high paying job, a big \
        house or fancy cars. I simply wish to be able to live my life appreciating everything around
me.',
        'I love life and never sad wedge abhor']
```

How to do sentiment analysis in R

2. Import some text to sentiment analyze

```
for d in test:
    d = TextBlob(d)
    print(d.sentiment)
```

```
Sentiment(polarity=1.0, subjectivity=1.0)
Sentiment(polarity=-0.65, subjectivity=1.0)
Sentiment(polarity=0.14057851239669422, subjectivity=0.38772727272727275)
Sentiment(polarity=0.375, subjectivity=0.8)
```

Lots of issues here

- How to weight the words
- How to take into account context at all (more on this in 2 weeks)

Sentiment analysis

- We could use any dictionary for any types of words, not just happy and sad ones, to count instances

When counting gets quite complicated ...

One more example: LIWC

Wordwatchers

Tracking the language of public figures

States of the Union: Truman to Obama

by James W. Pennebaker

Most years since George Washington, the President of the United States has addressed the joint sessions of Congress along with leaders in the military, judiciary, and other parts of government in a public speech. The purpose of the address is to summarize the accomplishments and problems of the nation and to lay out plans and expectations for the coming years. Although the tone of the State of the Union addresses change from year to year, the occasion is generally a mixture of a sober analysis and political undertones.

The address is typically written, at least in part, by the president with help from experts, speechwriters, and aides. Nevertheless, it generally reflects the leader's intentions, values, emotional and thinking styles, and personality. Unlike the inaugural address, which is delivered to the nation once every four years, States of the Union (SOU) talks are delivered annually to the country's governing body. The SOU, then, is a more business-like and detail oriented communication intended to direct Congress to move in specified directions.

Language and Personality

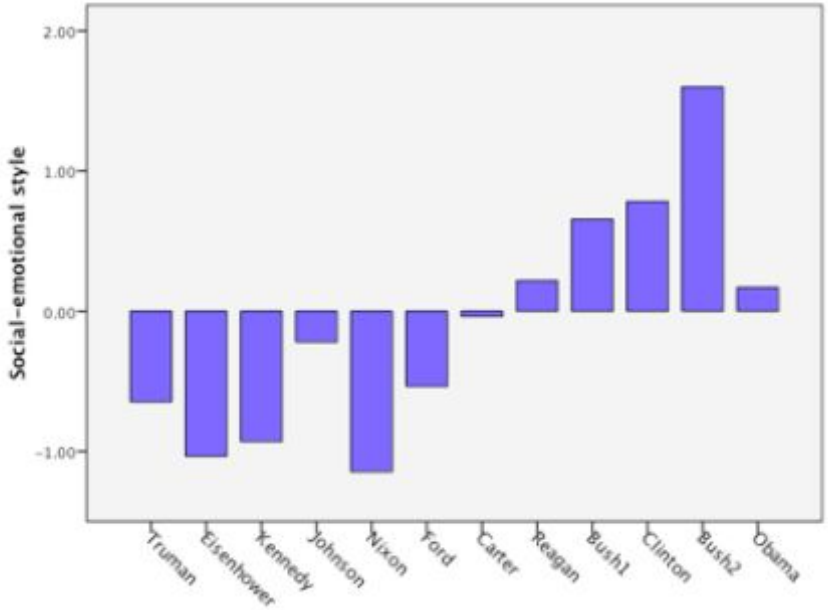
This site explores how we can learn about the candidates' personalities, motives, emotions, and inner selves through their everyday words.

Blogroll

- How do we do it?
- LIWC: Computerized text analysis program
- Pennebaker website
- Questionnaires and psychological tests — very cool indeed
- References
- Relative Semantic Analysis: Here's how we

Me, myself and I

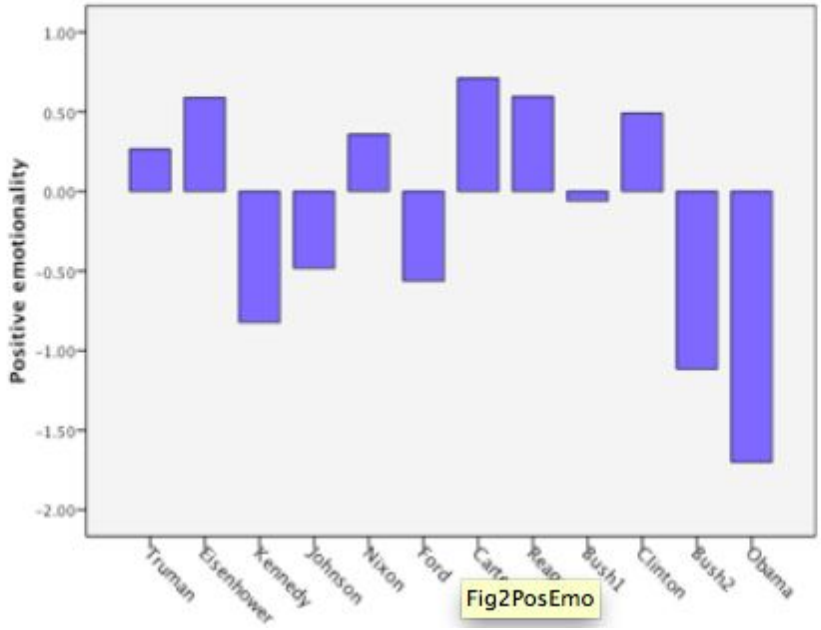
Figure 1: Social-emotional style. Higher numbers reflect use of more personal pronouns, references to other people, and emotional words.



- Using the LWIC
- Obama's style is "cooler" than most recent presidents

The down-beat

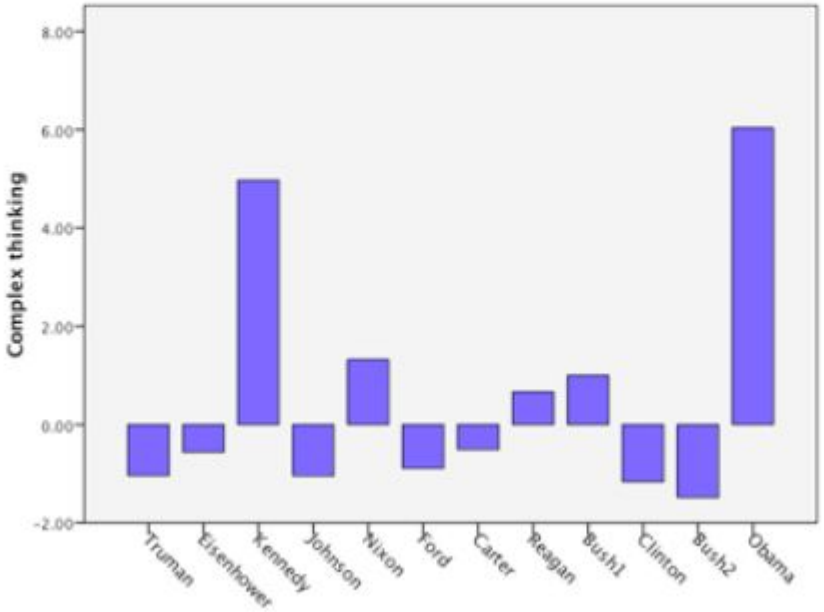
Figure 2. Positive emotionality. The higher the number, the more the person uses positive emotion words relative to negative emotion words.



- Obama's emotional tone is very negative

Complexity of thought

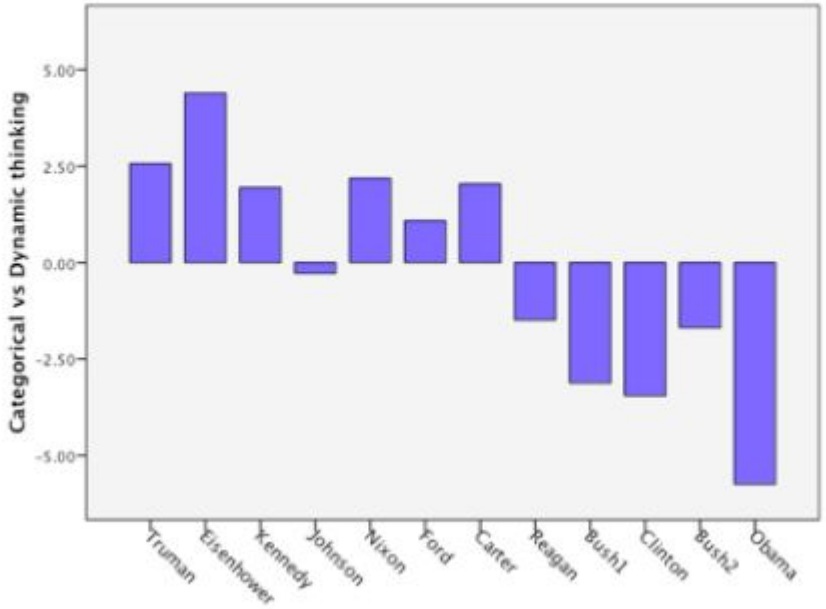
Figure 3. Complexity of thinking. The higher the number, the more complex and nuanced the language in the presentation of arguments.



- Obama was a law professor, you know

Dynamic, not categorical, thinking

Figure 4. Categorical versus dynamic thinking. Higher scores reflect categorical thinking whereas lower (or more negative) scores indicate dynamic thinking.



- Dynamic thinkers refer to history, evolution -- not pre-established categories

In summary ...

- “Barack Obama thinks and relates to people differently from most of his predecessors. His thinking style is both highly complex and, at the same time, dynamic. Socially and emotionally, he is surprisingly cool and distant. The word “cool” is not ill-advised. In his SOU addresses, as well as his press conferences, he is detached. His use of both positive emotion and negative emotion words is much lower than recent presidents. Although his personal pronouns in his SOUs are slightly above average, they are actually quite low when talking informally in interviews or press conferences. His is the language of the confident leader as opposed to the close buddy.”

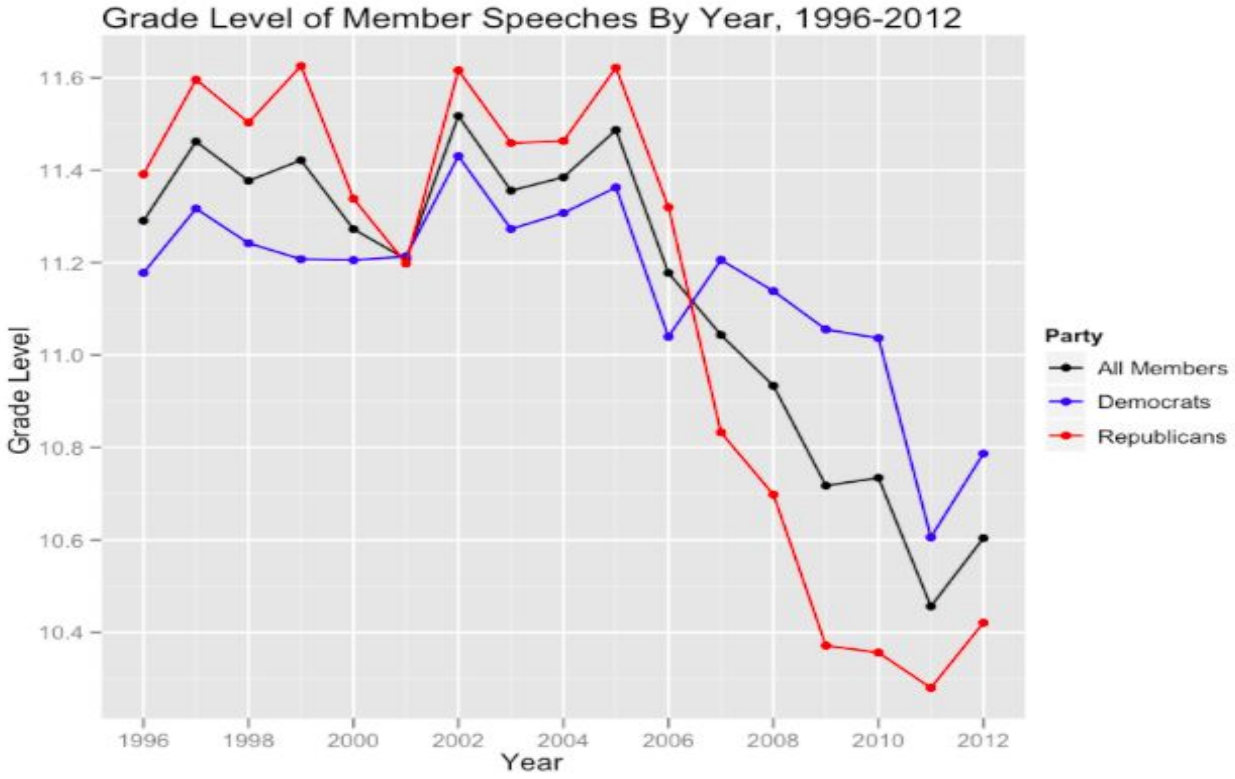
Another project, this time involving sentiment ...

Political speeches

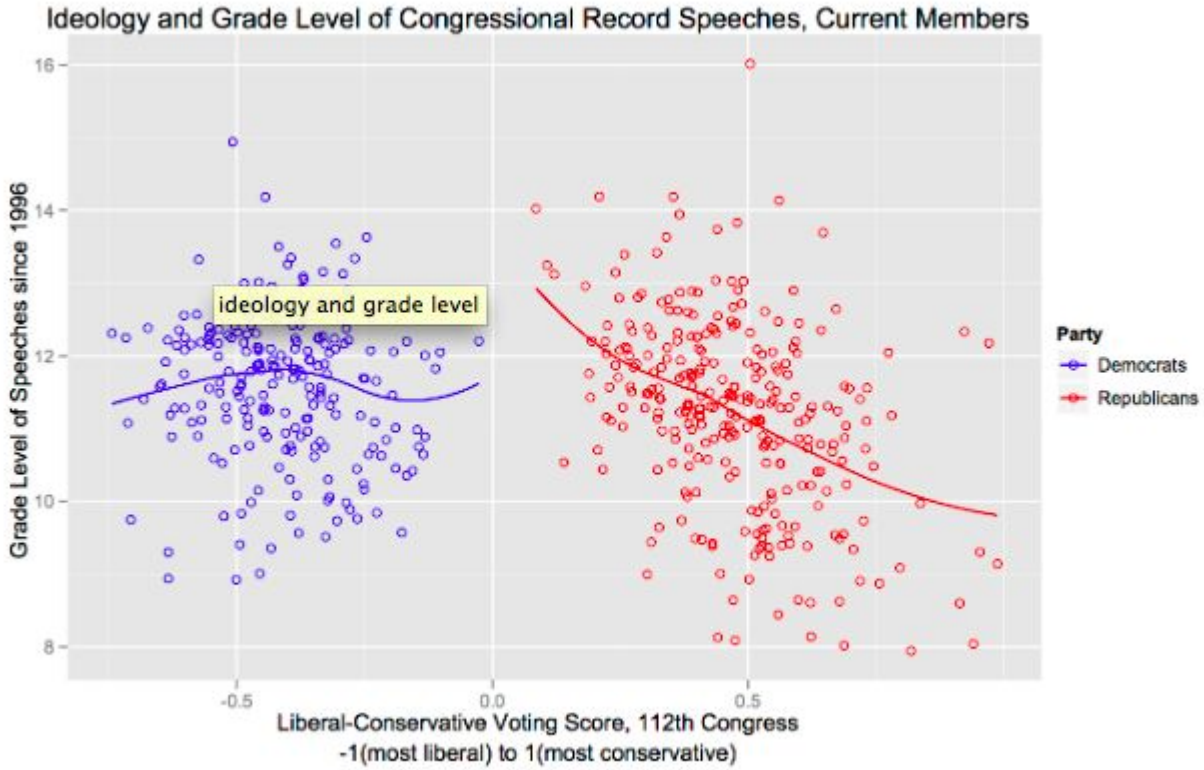
- How have the complexity and sophistication of Congressional speeches changed over time?

From the Sunlight Foundation, Capitol Words -- “The changing complexity of congressional speech” by Lee Drutman MAY 21, 2012, 12:01 A.M.

The answer ...


Grade Level of Member Speeches By Year, 1996-2012
Grade Level
11.6
11.4
11.2
11.0
10.8
10.6
10.4
1996
1998
2000
2002
2004
2006
2008
2010
2012
Year
Party
All Members
Democrats
Republicans

BTW --

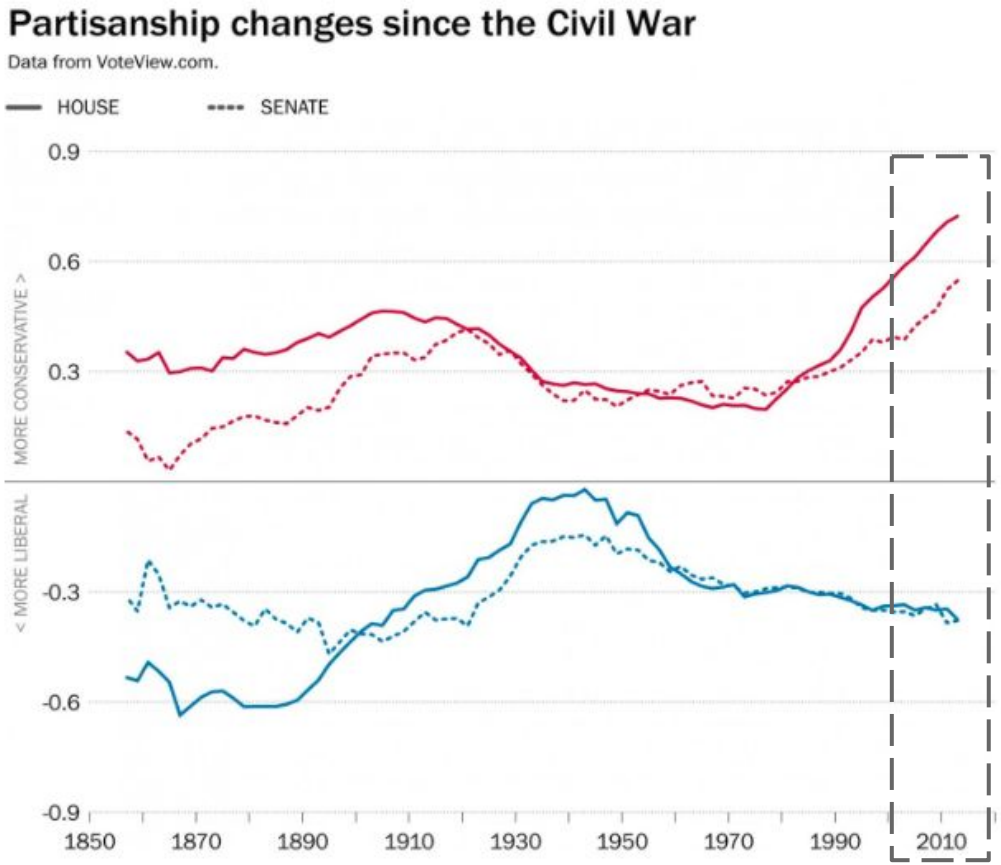

Ideology and Grade Level of Congressional Record Speeches, Current Members
Grade Level of Speeches since 1996
16
14
12
10
8
ideology and grade level
-0.5
0.0
0.5
Liberal-Conservative Voting Score, 112th Congress
-1(most liberal) to 1(most conservative)
Party
Democrats
Republicans

Sentiment analysis of Congress

- Has the aggregate mood of speeches in Congress changed in some clear way across time, or is it tied to certain external events (recessions, wars, etc.)?

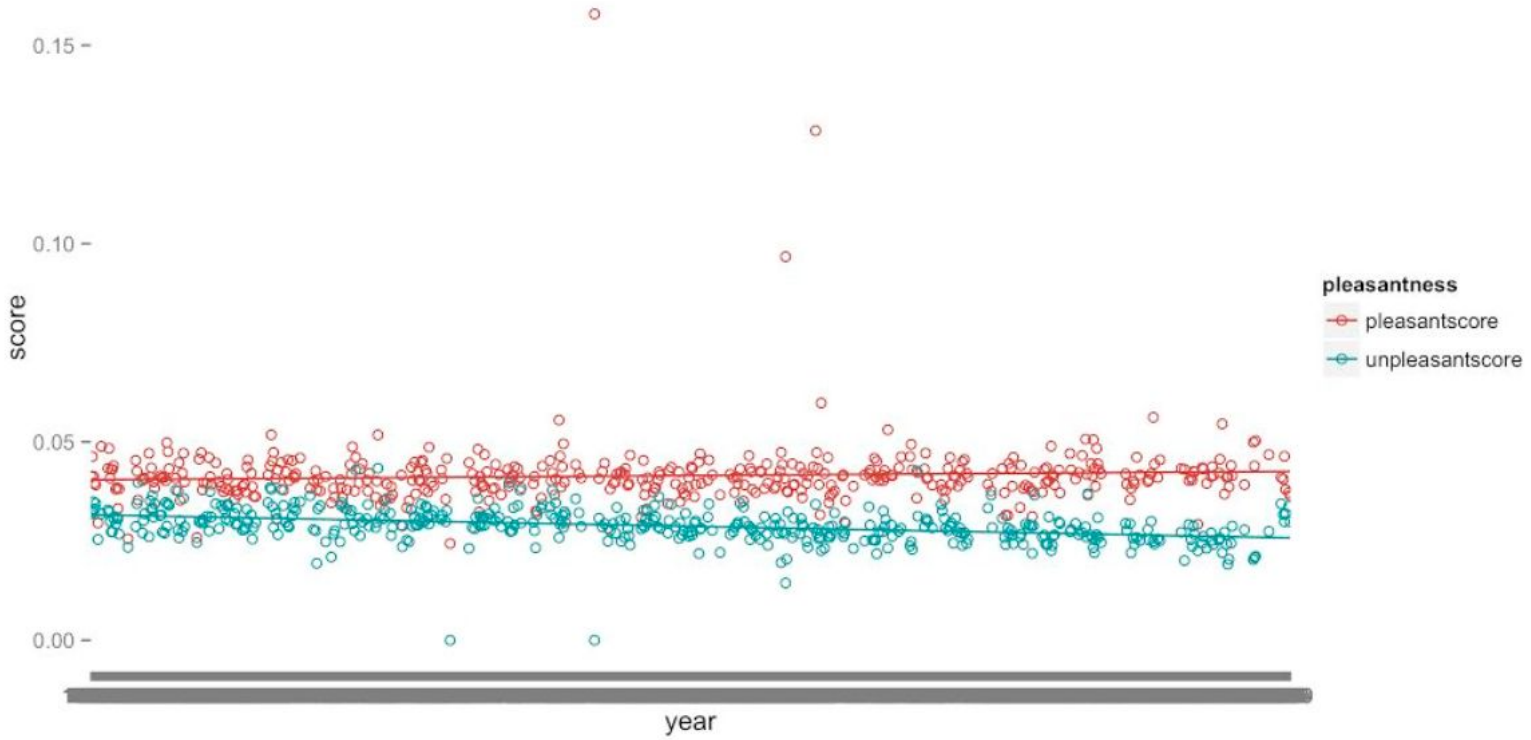
- For individual Congress-people, do they show constant sentiment or does their mood trend upward or downward?

Increased partisanship



- Republicans have gotten more conservative and Democrats more liberal
- From "The unprecedented partisanship of Congress, explained" by Philip Bump, Washington Post, Jan. 13, 2016

Increased rancor? Not that I can see

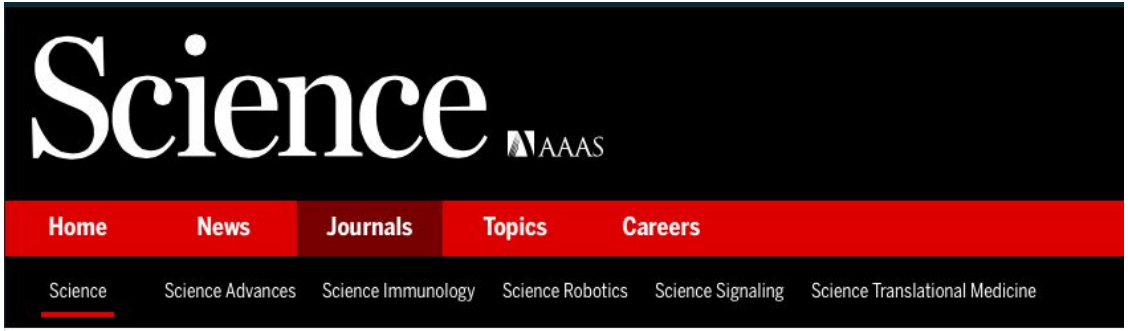


- Virtually no movement in sentiment in Congress over this period

How is this possible?

- What else should we be considering?

Another take on this same issue



SHARE

1

0

REPORT

Conservatives report, but liberals display, greater happiness

Sean P. Wojcik[1,*], Arpine Hovasapian[1], Jesse Graham[2], Matt Motyl[3], Peter H. Ditto[1,*]

[1]Department of Psychology and Social Behavior, University of California, Irvine, CA 92697, USA.

[2]Department of Psychology, University of Southern California, CA 90089, USA.

[3]University of Illinois, Chicago, IL 60607, USA.

↵*Corresponding author. E-mail: swojcik@uci.edu (S.P.W.); phditto@uci.edu (P.H.D.)

Science 13 Mar 2015:
Vol. 347, Issue 6227, pp. 1243-1246
DOI: 10.1126/science.1260817

Article | Figures & Data | Info & Metrics | eLetters | PDF

Emotion and Reason in Political Language*

Gloria Gennaro† Elliott Ash‡

February 2021

Abstract

We use computational linguistics techniques to study the use of emotion and reason in political discourse. Our new measure of emotionality in language combines lexicons for affective and cognitive processes, as well as word embeddings, to construct a dimension in language space between emotion and reason. After validating the method against human annotations, we apply it to scale 6 million speeches in the *U.S. Congressional Record* for the years 1858 through 2014. Intuitively, emotionality spikes during time of war and is highest for patriotism-related topics. In the time series, emotionality was relatively low and stable in the earlier years but increased significantly starting in the late 1970s. Comparing Members of Congress to their colleagues, we find that emotionality is higher for Democrats, for women, for ethnic/religious minorities, and for those with relatively extreme policy preferences (either left-wing or right-wing) as measured by roll call votes.

Key Words: Political Rhetoric, Word Embeddings, Emotions, U.S. Congress

Word Count: 10 331

This just came out

This is “emotionality”

Figure 8: House Member Emotionality by Party and by Party Majority

Time series of emotionality in the House of Representatives for Democrats (blue) and Republicans (red), 1900-2014. Blue and red areas indicate Democratic majorities in the House of Representatives.

Sentiment analysis in R

How to do sentiment analysis in R

1. Start with the sentR package

```
install.packages('devtools')
require('devtools')
install_github('mananshah99/sentR')
require('sentR')
```

How to do sentiment analysis in R

2. Import some text to sentiment analyze

```
# Words to test sentiment
test <- c('I am a very happy person.', 'I am a very sad person',
'I’ve always understood happiness to be appreciation. There is no greater happiness than appreciation for what one has- both physically and in the way of relationships and ideologies. The unhappy seek that which they do not have and can not fully appreciate the things around them. I don’t expect much from life. I don’t need a high paying job, a big house or fancy cars. I simply wish to be able to live my life appreciating everything around me.
', 'I love life and never sad wedge abhor')
```

How to do sentiment analysis in R

3. Import our dictionary list of positive and negative words

```
# Create small vectors for happy and sad words
positive = c('happy', 'love')
negative = c('abhor', 'sad')

## in terms of importing a whole list of words, this almost worked, but not quite:
## neg <- scan(file.choose(), what="") ## it is just a text file of a long list of words ###
## new = as.list(neg)
## negative = new
```

I will show you better ways later ...

4. Run sentR

```
out <- classify.aggregate(test, positive, negative)

> out
  score
1     1
2    -1
3     0
4    -1

text
1
I am a very happy person.
2
I am a very sad person
3 I’ve always understood happiness to be appreciation. There is no greater happiness than
apprec[...omitted...]I simply wish to be able to live my life appreciating everything around
me. \n
4
I love life and never sad wedge abhor
```