

# Analysis of the style and the rhetoric of the 2016 US presidential primaries



Презентацию подготовил  
Краснов Станислав, МКЛ171

# О работе

Автор: Jacques Savoy

Источник: Digital Scholarship in the Humanities, Volume 33,  
Issue 1, 1 April 2018, Pages 143–159,

<https://doi.org/10.1093/llc/fqx007>

Опубликовано: 13 February 2017

# Тема исследования

Анализ стиля речи и риторики кандидатов в президенты США на выборах 2016-ого года.

**Данные:** записи телевизионных дебатов политиков в ходе праймериз.

**Инструменты анализа:** выделение наиболее частотных лемм, коллокаций и тем высказываний в речи каждого из кандидатов, классификация политиков по этим признакам.

# Вопросы исследователей

- Как Трампу и Клинтон удалось победить в дебатах в ходе праймериз?
- Позволяет ли анализ теледебатов в полной мере определить стиль речи и риторику кандидатов?
- Можно ли выделить риторические особенности речи, способствовавшие победе Трампа и Клинтон в праймериз?
- Насколько удалены друг от друга (с точки зрения стиля) манеры речи кандидатов каждой из партий?

# Другие работы в области политического дискурса

- Boller P. F.Jr. 2004. Presidential Campaigns. From George Washington to George W. Bush
- Labbé D., Monière D. 2013. La campagne présidentielle de 2012. Votez pour moi!
- Pauli F., Tuzzi A. 2009. The end of year addresses of the presidents of the Italian Republic (1948–2006): discourse similarities and differences
- Labbé D., Monière D. 2003. Le discours gouvernemental. Canada, Québec, France (1945-2000)
- Sylwester K., Purver M. 2015. Twitter language use to reflects psychological differences between democrats and republicans
- Laver M., Benoit K., Garry J. 2003. Extracting policy positions from political texts using words as data
- Yu B. 2008. Classifying party affiliation from political speech
- Yu B. 2013. Language and gender in congressional speech
- O'Connor B. et al. 2010. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series.
- Hart R. P. 1984. Verbal Style and the Presidency. A Computer-based Analysis
- Tausczik Y. R., Pennebaker J. W. 2010. The psychological meaning of words: LIWC and computerized text analysis methods
- Slatcher R. B., Chung C. K., Pennebaker J. W., Stone L. D. 2007. Winning words: individual differences in linguistic style among U.S. Presidential and Vice Presidential candidates

# Корпус записей

В эксперименте использовались транскрипции с теледебатов, взятые с сайта <http://www.presidency.ucsb.edu>:

- 12 записей для республиканцев (Jeb Bush, Ted Cruz, John Kasich, Rand Paul, Marco Rubio, Donald Trump)
- 9 записей для демократов (Hillary Clinton, Martin O'Malley, Bernie Sanders)

Measurements	Bush	Cruz	Kasich	Paul	Rubio	Trump	Clinton	O'Malley	Sanders
Voc.	2,367	3,373	2,539	1,542	3,203	2,614	3,537	2,019	2,954
Token	23,207	37,459	33,807	11,856	44,415	49,718	61,991	15,824	53,172

# Самые частотные леммы

Toutanova K. et al. 2003. Feature-rich Part-of-speech Tagging with a Cyclid Dependency Network.

10 наиболее частотных лемм английского языка (**The Brown Corpus**):

1. the
2. be
3. of
4. and
5. to
6. a
7. in
8. he
9. have
10. it

The top ten most frequent lemmas according to TV debates

Bush	Cruz	Kasich	Paul	Rubio	Trump	Clinton	O'Malley	Sanders
be	be	the	be	be	be	be	<b>we</b>	be
the	the	be	the	the	<b>I</b>	the	the	the
to	and	<b>we</b>	to	to	the	to	be	<b>I</b>
<b>we</b>	to	to	<b>I</b>	and	<b>we</b>	<b>I</b>	and	to
that	<b>I</b>	and	<b>we</b>	that	and	and	to	and
and	that	<b>I</b>	an	<b>we</b>	to	<b>we</b>	of	of
an	<b>we</b>	an	and	an	have	that	an	that
<b>I</b>	of	have	that	<b>I</b>	an	have	that	<b>we</b>
of	an	in	have	of	<b>you</b>	of	<b>I</b>	an
have	<b>you</b>	<b>you</b>	in	in	<b>it</b>	an	in	in

Note. The personal pronouns are depicted in bold.

# Статистический анализ

Measurements	Bush	Cruz	Kasich	Paul	Rubio	Trump	Clinton	O'Malley	Sanders
MSL	17.3	19.4	18.3	17.2	18.7	<b>13.7</b>	20.5	<b>21.9</b>	19.7
LD (%)	40.7	<b>44.6</b>	38.4	40.3	39.2	<b>36.6</b>	40.4	43.1	43.6
BW (%)	24.4	<b>26.4</b>	20.8	21.9	23.8	<b>18.3</b>	24.1	25.6	<b>26.4</b>
TTR	36.1	37.3	33.9	34.0	33.3	<b>29.7</b>	36.2	<b>37.9</b>	36.1

Mean sentence length (MSL, tokens/sentence): Отцы-основатели (39.6), Обама (18.5).

Lexical density (LD) – коэффициент лексического разнообразия.

$$LD(t) = \text{lexicalword}(t) / n(t) = 1 - \text{functionword}(t) / n(t)$$



# Статистический анализ

Measurements	Bush	Cruz	Kasich	Paul	Rubio	Trump	Clinton	O'Malley	Sanders
MSL	17.3	19.4	18.3	17.2	18.7	13.7	20.5	21.9	19.7
LD (%)	40.7	44.6	38.4	40.3	39.2	36.6	40.4	43.1	43.6
BW (%)	24.4	26.4	20.8	21.9	23.8	18.3	24.1	25.6	26.4
TTR	36.1	37.3	33.9	34.0	33.3	29.7	36.2	37.9	36.1

Frequency of big words (BW) – процент слов, состоящих более, чем из 6 букв.

Type-Token Ratio (TTR)\* – соотношение размера словаря к длине текста.

*\*в данной работе дополнительно усреднялся (Covington and McFall, 2010)*

# Распределение частей речи

Measurements	Bush (%)	Cruz (%)	Kasich (%)	Paul (%)	Rubio (%)	Trump (%)	Clinton (%)	O'Malley (%)	Sanders (%)
Noun	18.8	19.1	17.8	17.8	18.1	15.9	17.7	20.6	19.9
Name	5.4	8.4	4.9	5.6	5.5	4.6	5.2	5.1	6.3
Pronoun	1.9	2.3	2.5	2.1	2.5	1.9	2.5	3.0	2.4
Adjective	6.9	6.5	5.4	6.5	6.1	6.3	6.6	7.6	7.4
Verb	25.7	24.8	27.2	27.8	25.4	30.2	26.7	21.1	23.6
Adverb	6.7	6.1	7.7	7.7	8.1	10.2	7.8	7.0	7.1
Determiner	13.1	11.4	11.6	11.9	12.5	10.7	10.6	11.6	11.5
Preposition	16.0	14.5	16.2	15.2	16.0	13.8	17.2	17.2	15.6
Conjunction	3.7	4.7	5.0	3.9	4.2	4.4	4.3	4.8	4.0
Other	1.8	2.1	1.8	1.6	1.6	2.0	1.4	2.0	2.0
Q-index	78.8	79.2	83.5	81.1	80.7	82.6	80.1	73.4	76.0

Q-index (Kubát and Cech, 2016) – соотношение процента глаголов в тексте к сумме процентов глаголов и прилагательных. Определяет направленность речи (активную, дескриптивную).

# Тематическая концентрация

Measurements	Bush	Cruz	Kasich	Paul	Rubio	Trump	Clinton	O'Malley	Sanders
<i>h-point</i>	52	47.8	66.75	39	76	77.3	84.5	43.5	85
PTC (%)	4.99	6.19	8.49	4.82	08.04	9.89	9.42	4.26	12.12
	need	know	people	think	people	people	think	need	people
	people	Donald	know	say	go	go	people	people	think
	country	say	go	want	president	say	know	country	country
	president	president	want	go	country	know	say	make	say

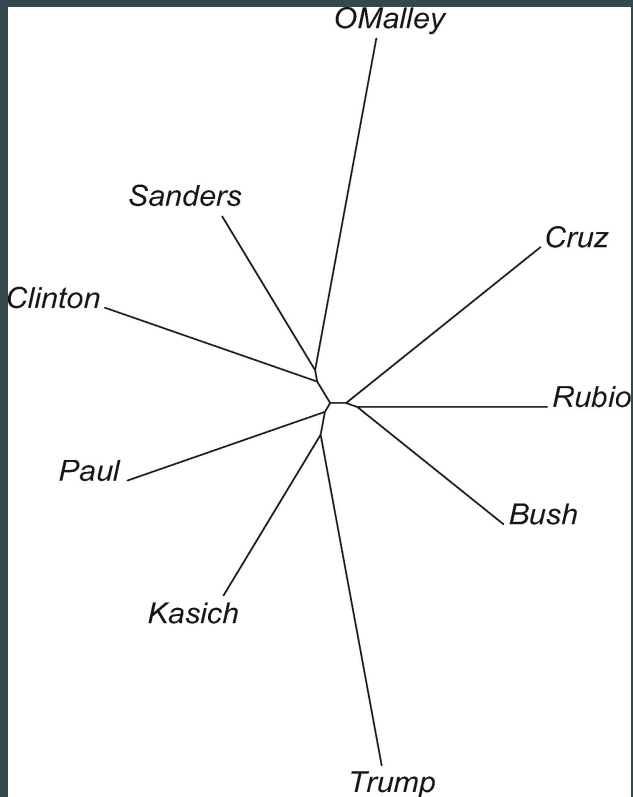
*h-point* – точка в ряду слов, расположенных по убыванию частоты, где порядковый номер слова равен кол-ву его употреблений в тексте.

Слова, расположенные до *h-point*, признаются функциональными, после которых чаще всего идут “ядерные” для темы слова.

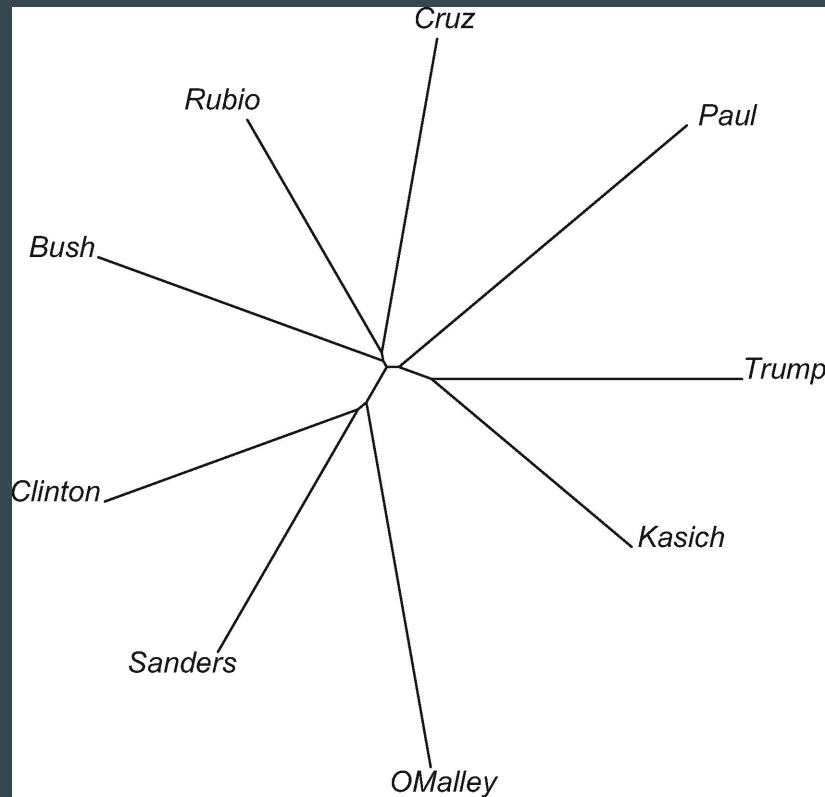
$$n = a \cdot h^2 \text{ or } h = \sqrt{n/a}$$

$$PTC = 1/n_h \sum_{r' < h} f(r')$$

# Стилистические и тематические расстояния



Граф стилистических расстояний



Граф тематических расстояний

# Наиболее специфичные термины

Bush	Cruz	Kasich	Paul	Rubio	Trump	Clinton	O'Malley	Sanders
relates	Donald	Ohio	conservative	be	I	Senator	actually	Secretary
DC	flat	balanced	war	century	very	Sanders	Anderson	Street
signal	amnesty	budget	think	why	tremendous	to	Andrea	major
proven	note	we	Ferguson	someone	nobody	comprehensive	Maryland	Wall
serious	court	formula	Kentucky	issue	going	I	we	campaign
strategy	IRS	surplus	bowling	America	Mexico	income	ISIL	fossil
need	Islamic	secondly	scrutiny	they	not	try	Lester	fuel
caliphate	Washington	hole	borrow	this	Jeb	Republicans	freedom	class
brother	tax	Pentagon	fault	he	excuse	affordable	sort	billionaire
status	Texas	growth	Fed	operating	deal	more	nation	wealth

# Наиболее специфичные предложения

'They *don't* like seeing *bad trade deals*, they *don't* like seeing higher taxes, they *don't* like seeing a loss of their jobs where our jobs *have just been* devastated.' (D. Trump, 10 March 2016)

'I'm spending all of *my* money, I'm not spending, I'm not getting any, I turned down, I turn down so much, I could *have* right now from special interests and donors, I could *have* double and triple what *he's* got.' (D. Trump, 16 September 2015)

'Just excuse me, one second, Rand, ... if you *don't* mind, Rand, you know, you are on last, you *do* have your 1 percent.' (D. Trump, 16 September 2015)

# Наиболее специфичные предложения

‘Look, *I have* the greatest respect *for Senator Sanders* and *for his supports* and I'm going to keep working as *hard* as I can to reach as many people of all ages about what *I will do*, what the experience and the ideas that *I have* that I will bring to the White House and I hope to have their support when I'm the Democratic nominee.’ (H. Clinton, 17 January 2016)

‘*I think* now what I've called for is *counsel for every child* so that no child has to face any kind of process without someone who speaks and advocates for that *child* so that the right decision hopefully can be made.’ (H. Clinton, 11 February 2016)

‘Let's make the *Affordable Care Act* work *for everybody*.’ (H. Clinton, 4 February 2016)



# Результаты

Согласно результатам, речь Дональда Трампа представляется атипичной для дискурса, так как в основном содержит короткие предложения с простыми словами и повторяющимися аргументами. Политик делает акцент на собственной личности, многократно повторяя местоимение “I”, а также уделяет огромное внимание слову “deal”, что отражает направленность его программы на решение экономических и миграционных проблем.



# Результаты

Хилари Клинтон также часто прибегает к личному местоимению “I” и, как и многие демократы, использует в своей речи несколько более сложные конструкции с разнообразным набором лексики, стараясь тем самым звучать более убедительно и рационально.

# Результаты

Неожиданной находкой для исследователей стали результаты анализа речи Теда Круза. Его стиль разительно отличается от сопартийцев, делая видимый уклон в сторону описательной риторики. Отчасти это объясняется тем, что его программа, в отличие от других кандидатов, делает упор на реформы фискальной и здравоохранительной систем.

**Спасибо за внимание!**