

# Анализ стиля и риторики президентских праймериз 2016 года в США

Выполнил:

На основе статьи Жака Савоя (Jacques Savoy)

Максим Лепехин,  
Студент 2 курса программы  
«Русская литература и  
компаративистика»

## Материал для анализа



P.S. Пожалуйста, смейтесь сейчас, дальше может быть не так весело

## Соотношение уникальных слов и общего количества произнесенных слов

Table 1 Some statistics about candidates' speeches and comments

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

Voc. – количество уникальных слов; Token – общее количество слов

## Наиболее частотные лексемы

Table 2 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	we	be
the	the	be	the	the	I	the	the	the
to	and	we	to	to	the	to	be	I
we	to	to	I	and	we	I	and	to
that	I	and	we	that	and	and	to	and
and	that	I	an	we	to	we	of	of
an	we	an	and	an	have	that	an	that
I	of	have	that	I	an	have	that	we
of	an	in	have	of	you	of	I	an
have	you	you	in	in	it	an	in	in

*Note.* The personal pronouns are depicted in bold.

Для их вычисления использован специальный алгоритм - part-of-speech (POS) tagger – который вычленяет из предложений части речи. Например, предложение «**But I also know this problem is not going away**» будет выглядеть как «**But/CC I/PRP also/RB know/VBP this/DT problem/NN is/VBZ not/RB going/VBG away/RB /»**, где NN означает noun, PRP – preposition и т.д.



## Что видно, исходя из этих данных?

- Те из кандидатов, кто раньше был у власти в качестве губернаторов, склонны употреблять «мы», а не «я». Это в первую очередь относится к О'Мейли, а также к Бушу и Кейсику.
- Сенаторы, напротив, употребляют местоимение «я» чаще – см. Круза, Сандерса и Клинтон.
- Те кандидаты, которые дольше оставались в предвыборной гонке, склонны к употреблению «я», а не «мы».

# Замер стилистических особенностей каждого из кандидатов

Table 3 Four global stylistic measurements according to TV debates

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	<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	29.7	36.2	<b>37.9</b>	36.1

*Note.* Extreme values are depicted in bold.

## MSL (средняя длина предложения)

Table 3 Four global stylistic measurements according to TV debates

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	<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

*Note.* Extreme values are depicted in bold.

Самый большой показатель – у О’Мэлли, самый низкий – у Трампа, который использует упрощенный синтаксис, обращается к простому коммуникативному стилю.

## LD (лексическая плотность)

Table 3 Four global stylistic measurements according to TV debates

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	<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

*Note.* Extreme values are depicted in bold.

LD показывает информативность и разнообразие текста. Как мы видим, Трамп имеет самый низкий показатель и по этому параметру – он повторяет одни и те же слова, «кружит» вокруг одних и тех же выражений.



# TTR

Table 3 Four global stylistic measurements according to TV debates

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	<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

*Note.* Extreme values are depicted in bold.

TTR (Type-Token Ratio) показывает соотношение уникальных слов и видов частей речи. Если показатель высокий, значит, кандидат чаще обращается к разным темам или к подходу с разных сторон к одной, используя разнообразные слова. Низкий показатель свидетельствует о тенденции оперировать одними и теми же словами для разговора на разные темы.

## BW (количество длинных слов)

Table 3 Four global stylistic measurements according to TV debates

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	<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

*Note.* Extreme values are depicted in bold.

Длинным считается слово, содержащее шесть и более букв. Предложения с большим количеством длинных слов воспринимаются тяжелее. Трамп и здесь имеет самый низкий показатель.

## Какие части речи чаще всего используют кандидаты?

Table 4 POS distribution according to TV debates

Measurements	Bush (%)	Cruz (%)	Kasich (%)	Paul (%)	Rubio (%)	Trump (%)	Clinton (%)	O'Malley (%)	Sanders (%)
Noun	18.8	19.1	17.8	17.8	18.1	<i>15.9</i>	<i>17.7</i>	<b>20.6</b>	19.9
Name	5.4	<b>8.4</b>	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	<i>1.9</i>	2.5	<b>3.0</b>	2.4
Adjective	6.9	6.5	<i>5.4</i>	6.5	6.1	6.3	6.6	<b>7.6</b>	7.4
Verb	<i>25.7</i>	24.8	<i>27.2</i>	27.8	25.4	<b>30.2</b>	<i>26.7</i>	<i>21.1</i>	23.6
Adverb	<i>6.7</i>	<i>6.1</i>	<i>7.7</i>	<i>7.7</i>	8.1	<b>10.2</b>	7.8	7.0	7.1
Determiner	<b>13.1</b>	11.4	11.6	11.9	12.5	10.7	<i>10.6</i>	11.6	11.5
Preposition	16.0	14.5	16.2	15.2	16.0	<i>13.8</i>	<b>17.2</b>	17.2	15.6
Conjunction	<i>3.7</i>	4.7	<b>5.0</b>	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	<b>83.5</b>	81.1	80.7	82.6	80.1	<i>73.4</i>	76.0

*Note.* The maximum value per grammatical category is shown in bold, and the minimum in italics.

Те, у кого преобладает использование глаголов (в первую очередь это Трамп), склоны к риторике действия, преобладание существительных свидетельствует о склонности к объяснениям. Q-index – отношение количества глаголов к сумме глаголов и прилагательных, показывает склонность к той или иной риторике – «действия» или «объяснения».

## Тематическое распределение

Table 5 *h*-point and most frequently used thematic words per candidate

Measurements	Bush	Cruz	Kasich	Paul	Rubio	Trump	Clinton	O'Malley	Sanders
<i>h</i> -point	52	47.8	66.75	39	76	77.3	84.5	43.5	85
PTC (%)	4.99	6.19	8.49	4.82	8.04	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

*Note.* The maximum value is shown in bold, and the minimum in italics.

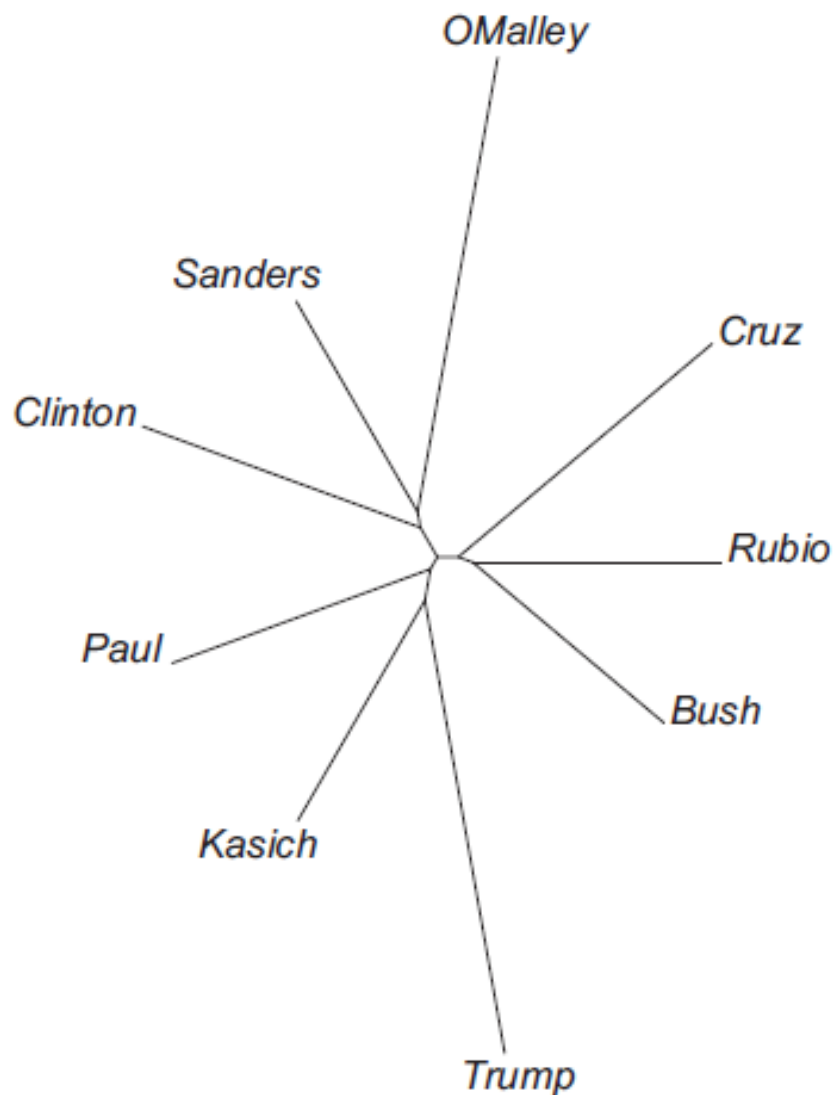
H-point – это параметр, введенный для измерения частотности уникальных, а не служебных слов. Ниже этой точки обычно встречаются как раз служебные слова, если кандидат все время повторяет какие-то понятия и слова, они оказываются ниже этой точки. Показатель РТС (пропорциональная тематическая концентрация) демонстрирует, какой процент не служебных слов встречает ниже *h*-point. Больше всего этот показатель у Трампа и Клинтон – они скорее повторяют свои аргументы, чем заговорят о чем-то совершенно новом.



## Стилистическая дистанция между кандидатами

Длина линии, соединяющей две точки, измеряет стилистическую дистанцию между кандидатами.

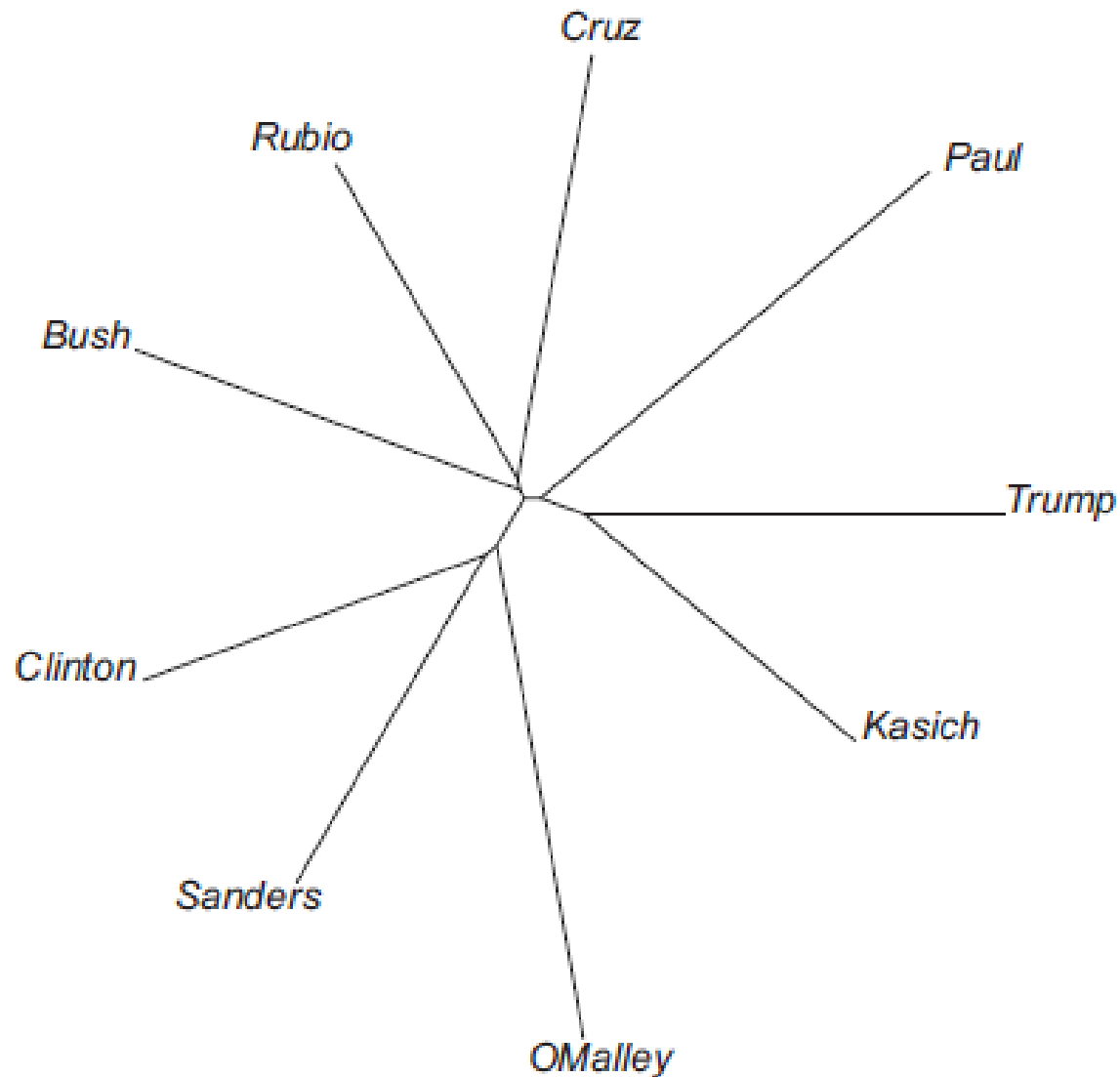
В расчет принимаются только служебные слова.



Можно выделить три кластера – один с кандидатами от демократов (Clinton, O'Malley и Sanders), два других – с кандидатами-республиканцами (Bush, Cruz и Rubio с одной стороны и Kasich, Trump и Paul – с другой).

## Тематическое расстояние между кандидатами

Для расчета тематического расстояния анализируются только не служебные слова.



Кластеризация получилась во многом похожей на предыдущий случай.

## Самые употребляемые понятия

Table 6 The top ten most specific terms per candidate

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

Как видно из таблицы, только у двух кандидатов, Трампа и Клинтон, «Я» – это одно из самых употребляемых слов. Тому, кто хочет остаться в электоральной гонке, стоит себя продвигать.

## Выводы

- Трамп оказался нетипичной для праймериз фигурой: использовал укороченные предложения, ограниченный словарь, повторял одни и те же аргументы простыми очень простыми словами.
- Хиллари Клинтон единственная из демократов отдаёт предпочтение местоимению «Я».
- Среди республиканцев тоже есть кандидат, чья риторика сильно отличается от распространенной в этой партии риторики «действия» - это Тед Круз. В его речах преобладают имена собственные и существительные.



## Выводы о выводах

- С одной стороны, работа проиллюстрировала как интуитивно улавливаемые вещи (популистский стиль Трампа проще синтаксически и тематически), так и не самые очевидные тренды (выигрывает тот, кто напирает на презентацию собственной персоны и многое другое).
- С другой же стороны, как кажется, анализ только «устной риторики» оказывается половинчатым без рассмотрения речей, веб-сайтов и прочих источников, которые автор относит к более риторически выверенным и просчитанным.
- Такое возражение вырастает хотя бы из представления автора о том, что устная речь «выявляет реального человека за созданным ею образом» (reveal more closely the real person behind her/his projected image). Не ясно, как в таком случае соотносятся риторики «устные» и «письменные» - не только в плане результатов её анализа, но и по воздействию на аудиторию.