

CAMPAIGNING IN POETRY: IS THERE INFORMATION CONVEYED IN THE CANDIDATES' CHOICE OF WORDS?*

Bruce Sacerdote

Dartmouth College and NBER, bruce.sacerdote@dartmouth.edu

Owen Zidar

Dartmouth College, Owen.Zidar@dartmouth.edu

January 27, 2008

Abstract

We construct and analyze a database consisting of the words used in speeches made by the candidates in the 2008 Democratic and Republican Presidential Primary. We present findings in two key areas. First, we estimate which candidates are more negative by counting the number of times they mention their opponents by name. Both Obama and Clinton are among the least negative. John Edwards is 3.7 times more negative than Barack Obama and Mitt Romney is 4.7 times more negative than Obama. We also compare the speeches of the candidates to speeches delivered by famous political figures and orators such as Martin Luther King, Ronald Reagan, and John F. Kennedy. For example, we can ask whether a candidate's words are more like those of MLK versus Ronald Reagan. Using our metric, John McCain and Mitt Romney are the most like Ronald Reagan whereas (within the set of primary candidates) the words of Mike Huckabee and Barack Obama are the closest to those of Martin Luther King. Hillary Clinton is by far the candidate closest in oratory to Bill Clinton.

* We thank Doug Staiger and Dean Lacy for helpful suggestions.

I. Introduction

The process by which voters choose a candidate may be driven by policy differences but can also be driven by appearance (Benjamin and Shapiro [2006]), gut feelings, likeability, or even subconscious reactions and associations. The 2008 campaign has been particularly notable for its vigorous debates, closely contested primaries but also a lack of strong policy differences within each party. In fact Fred Barnes of the *Wall Street Journal* went so far as to dub this "the Seinfeld campaign," implying that it's about nothing, at least as far as substantive policy differences go.

Following methods developed in political science and the humanities, we use textual analysis to unearth additional information about each candidate. In particular we try to quantify which candidates are running the more negative campaigns and we try to identify which candidates have word choices that make them most like historical political figures (and in some cases great orators) like Martin Luther King, Ronald Reagan, John F. Kennedy, Winston Churchill, Lyndon Baines Johnson, and Joseph Stalin.

We take as our data the set of speeches made by the primary candidates during the period January 2006 through January 5, 2008. We identify which words are used and the relative frequency of each word. This allows us to identify the themes emphasized by each candidate and by the campaign as a whole. We can also determine who mentions their opponents or George W. Bush the most often. We compare the words of the candidates to the words of historical political leaders. We rank the candidates as being most like a particular leader such as Ronald Reagan. Several of the Republican candidates wish to be seen as the heir to Reagan while on the Democratic side Barack Obama has been criticized for allegedly admiring Reagan and his policies.¹

Surely there are a large number of ways one could conduct such an analysis and we present here only one example. (A sensible extension of what we have done here would be to look at two and three word phrases rather than single words.) And our results are only one source of insight into a candidate's personality and policies. We are combining both policy and personality since the frequent use of some words like "markets" or "freedom" might tell us something about a candidate's policy views while the frequent use of the pronoun "I" rather than "we" may indicate something about a candidate's personality.

Even with these large caveats in mind, we still believe that there are several advantages to a systematic analysis of the candidates' words. First, textual analysis can process significantly more data at much faster rates than the traditional method of having political scientists read and analyze what candidates are saying. Second, one can make quantitative statements about candidates' word choices and one can make fairly objective comparisons of candidates' speeches to reference speeches. Third, word choices may

¹ After the South Carolina Democratic Primary, some commentators including Bill Bennett have noted a perceived similarity between Obama's and Reagan's speeches. This is a claim we test directly.

reveal something deep or hidden about the candidate's beliefs that go beyond stated positions which potentially have been manipulated to maximize appeal to voters. For example, its easy to simply state that one is a Reagan Democrat, but somewhat harder to demonstrate that likeness of ones ideas to Reagan's over several years of speech making.

A variety of authors have demonstrated the usefulness and reliability of textual analysis. Stock and Trebbi [2003] is an excellent example which uses textual analysis to analyze whether a particular chapter was written by a father or son author. Laver, Benoit and Garry [2003] uses textual analysis to rank policy positions using the words in manifestos from the British Labor and Conservative parties. They arrive at rankings that are extremely similar to the rankings generated by the more traditional method of political scientists reading and scoring the texts on a liberal versus conservative scale. In our own work we have tested how often textual analysis correctly identifies the author of a given speech. We find a great deal of robustness. For example, speeches by Reagan are always correctly identified as being similar to a set of reference speeches by Reagan and very dissimilar to reference speeches by LBJ or Martin Luther King.²

II. Methodology

Our methodology follows closely that of Laver, Benoit and Garry [2003] who developed software for the purposes of counting words within a text file and for ranking different texts for their likeness to a set of reference texts. For each candidate we read in each word from the set of their speeches over a 2 year time period. To determine the degree of negativity of candidate i we simply take the number of mentions of opponents by i summed across all speeches and we divide by the total words used (summed across all speeches). In other words given j speeches by candidate i we calculate the following:

$$(1) \quad \text{Negativity}_i = \frac{\sum_j \text{Mentions of candidate } -i}{\sum_j \text{All Words}}$$

This is essentially the relative frequency with which a candidate mentions her opponents, where by the term *relative* we mean as a fraction of all words used. We include as "opponents" all candidates who are running in either nomination contest. For example, when Mitt Romney uses the words Hillary or Clinton, that is counted as a negative mention. One important objection to our measure is that not every mention of an opponent is done in a negative light. However, some spot checking of our speeches reveals that *most* such mentions are in fact negative. Obviously we are making a huge simplifying assumption by counting all such mentions as negative.³

We also use the data to identify the "themes" brought out by the election and the themes emphasized by each candidate. This calculation is even simpler in that we add up the

² See Mannion and Dixon [1997] and Holmes [1994, 1998] for more on the properties and methods of various estimators within textual analysis.

³ The discussion that occurs during debates does include compliments and points of agreement with other candidates as well as negative remarks, but we do not include transcripts from debates.

total times a word is used in our data and rank the words by frequency of use. In reporting themes, we report only the most commonly used nouns (e.g. change, Iraq, health.). We drop extremely commonly used words that do not appear to convey a theme (e.g. "here").

Finally we compare the words of the candidates to the words used by other orators. In doing this we used a simplified version of the procedure of Laver Benoit and Garry. Our goal is to ask whether a candidate is most similar to one versus the other of two famous political figures, e.g. Martin Luther King (MLK) versus Ronald Reagan (RR). We take a set of reference speeches from both reference speakers, e.g. MLK and RR. For any given word w , we 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.

If the word "freedom" is used twice as often by MLK versus RR, then candidate i 's use of the word freedom would imply that there is 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 (and thinks?) like RR. We define this probability of being like MLK as P_w .

$$(2) \quad P_w = \frac{\left[\frac{\sum_{MLK} \text{use of } w}{\sum_{MLK} \text{all words}} \right]}{\left[\frac{\sum_{RR,MLK} \text{use of } w}{\sum_{RR,MLK} \text{all words}} \right]}$$

We 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 we multiply P_w by the relative frequency with which candidate i uses word w (labeled F_{wi}) and we sum over all words w . By relative frequency F_{wi} we mean the number of times w is used as a fraction of total words spoken.

$$(3) \quad P_i(MLK) = \sum_w F_{wi} P_w$$

This is clearly a very simple measure of the degree to which i 's speeches resemble those of one reference speaker versus another. However, the method does seem to deliver sensible rankings of the probability that a given speaker either has similar policies or at least similar word choices to one of the reference speakers. For example, we always find that any of Reagan's speeches resemble his other speeches much more closely than his speeches resemble those of MLK, JFK, or LBJ. More generally, when we try to rank historical famous politicians as being more or less conservative (for instance by putting them on a Reagan vs LBJ scale) we find very sensible rankings that agree with the existing political science literature.

We can also place a standard error around our estimate of $P_i(MLK)$ by treating each word w , and hence each P_w as separate data point that helps us estimate the mean, $P_i(MLK)$. Currently we are treating each word w as an iid observation; this is likely a gross simplification. However, given the large amount of text we have and the thousands of unique words, we are generating small standard errors, which even if greatly inflated by a factor of 2-3 would still make most of our specific estimates of $P_i(MLK)$ for the different candidates statistically different from one another.

In all of the results below, we rank the presidential candidates as being more like one versus the other of two historical figures. In other words, our reference scale only has two endpoints and with no reference speakers in the middle. We do this for convenience and to avoid having to make additional assumptions: If we attempted to add JFK as a third reference point on an LBJ-Ronald Reagan scale, it would be difficult to justify precisely where JFK's speeches belong.

We are more comfortable with the ranking delivered by our methodology than by the absolute level of the estimated P_i . The scale of the estimates of P_i is somewhat arbitrary unless one insists that the level of our estimate is indeed a sensible probability that the speaker is in fact one reference speaker versus the other. We are thinking of it more of an index with an arbitrary scale. We follow standard practice in the textual analysis literature by inflating our index by the ratio of the variance of the original reference index (in our case .5 since we simply make MLK 1 and RR 0) to the variance of the raw estimated P_i 's.

III. Data

We take as data speeches delivered by the candidates over the period January 1, 2006 to January 14, 2008. We pull the text of these speeches from two places: candidate web sites (as of January 14, 2008) and transcripts of all campaign speeches transcribed by *Congressional Quarterly*. Using *LexisNexis Academic*, we identified and saved candidate speeches by searching for “(unique candidate name) delivers remarks” within the source *Congressional Quarterly* transcriptions.

Although the *Congressional Quarterly* transcripts are not exhaustive, they provide a fairly representative sample of speeches. For instance, the ten Mitt Romney speeches include stump speeches in Michigan on 1/15/08, Iowa on 1/2/08, South Carolina on 12/18/2007, as well as his nationally televised speech on faith in America on 12/6/07.

The number of speeches for each candidate from each source (website versus *Congressional Quarterly*) is as follows:

Candidate	Number of Speeches from Candidate's Website	Number of Speeches from Congressional Quarterly	Total
-----------	--	--	-------

Hillary Clinton	50	47	97
Barack Obama	54	42	96
John Edwards	16	16	32
Fred Thompson	9	5	14
John McCain	37	26	63
Mitt Romney	14	10	24
Rudy Giuliani	n/a	12	12
Mike Huckabee	n/a	8	8
Bill Clinton	n/a	1052	1052
George W. Bush	n/a	1488	1488

Note: Giuliani and Huckabee did not provide transcripts of speeches on their websites. Regarding dates, Bill Clinton's speeches range from January 1, 1996, to January 25, 2008. George W. Bush's transcripts comprise of the ten years prior to January 25, 2008.

The third column of Table 2 shows how many words we have per candidate. For certain candidates such as Obama and Clinton we are including roughly 250,000 words. For others such as Mitt Romney we have about 54,000 words.

Our reference speakers to whom we intend to compare the candidates are Ronald Reagan, Martin Luther King, Lyndon Baines Johnson, John F. Kennedy, Winston Churchill, Joseph Stalin, Bill Clinton, and George W. Bush. For Reagan, MLK, JFK, and Churchill we pull a famous set of speeches by each orator.⁴ For example the set of MLKs speeches includes:

“Birth of a New Nation” (4/7/57), “I Have A Dream” (8/28/63), “Nobel Peace Prize Acceptance Speech” (12/10/64), “A Time to Break the Silence” (4/4/67), and “I See the Promised Land” (4/3/68). For Bill Clinton and George W. Bush, we used transcripts from *Congressional Quarterly*.

IV. Results and Discussion

Our first set of results is our measure of negativity for each candidate. We calculate this measure using the subset of campaign speeches that are available on candidates' websites. Recall that we label negative speech as the frequency with which a candidate mentions an opponent by name (or uses the word opponent). Table 1 reports the frequency of these negative mentions per 10,000 words used. Clinton and Obama are among the least negative candidates; they mention their opponents only 1.3 and 1.8 times respectively per 10,000 words. In contrast, John Edwards is at 6.57 negative mentions per 10,000 words, Edwards is 3.6 times more negative than Obama. Mitt Romney is even more negative still at 8.52 negative mentions per 10,000 words.

⁴ We used the set of speeches from the website www.famousquotes.me.uk/speeches/, which google returns first for a search for “famous speeches.”

In Table 2 we proceed to our comparison of the words of the candidates relative to a reference group of historical political figures. As described above, we take the reference candidates two at a time due to the difficulty of placing three or more reference orators on a single scale.

The first panel of Table 2 ranks the candidates' speeches by their resemblance to Ronald Reagan (designated a 0 on our arbitrary scale) versus Martin Luther King (designated as a 1). Mike Huckabee and Barack Obama deliver speeches which are the most like those of Martin Luther King while John McCain and Mitt Romney are the most like Ronald Reagan. On the surface this seems plausible given Huckabee's background as a Southern preacher and Obama's background as a community organizer and activist. And McCain and Romney have both invoked Reagan's name and policy positions in a positive light and would like to be viewed as heirs to the Reagan legacy and the Reagan coalition of voters.

Presenting this index is potentially useful, but readers may also wonder precisely what word choices make Obama and Huckabee so like MLK and so unlike Reagan. We can answer this question by sorting all the words on the frequency that MLK used a word relative to the frequency that Reagan used the word. (P_w in equation 2 is a transformation of this relative frequency.) We then ask how much the candidates use or fail to use words that are strongly associated with MLK or RR.

In Table 3, we list the top words that make a speaker most like MLK rather than Reagan. Column 1 shows the word's use by Reagan divided by the word's use by MLK. The remaining columns show how often the word is used (per 10,000) words by Obama, Clinton and McCain. Among the top twenty words more frequently used by MLK relative to Reagan are the words black, children, God's, poverty, revolutionary, protest and dream. Obama and Clinton are much more likely to use all of these words than is John McCain. For example, MLK used the word dream 5x more than did Reagan. Obama uses the word "dream" 4.0 times per 10,000 words, whereas McCain uses the word dream .68 times per 10,000 words. Huckabee uses the word dream 2.3 times per 10,000 words. Clinton uses the word protest .12 times per 10,000 words whereas McCain and Romney did not use the word protest at all in our dataset. Obama uses the word poverty 4 times per 10,000 words relative to McCain's .1 times and Clinton's 1.5 times per 10,000 words.

The second panel of Table 2 ranks candidates on their resemblance to LBJ versus Bill Clinton. Perhaps the most striking fact is how strongly Senator Clinton's speeches resemble those of her husband. Interestingly Giuliani and Huckabee also are ranked as being similar to Bill Clinton. Fred Thompson and John McCain are the most like LBJ (and hence the least like Bill Clinton) on this scale.

Some of the key words that are much more frequently used by the Clintons and Huckabee and Giuliani include security, education, million, need, different and think. Bill Clinton used the word security about 5 times more than LBJ. This may seem ironic given LBJ gave us both a major expansion of social security and a major expansion of the Vietnam

War (which allegedly had something to do with National Security). Hillary Clinton uses the word security 8.8 times per 10,000 words while Obama (who falls in the middle of the LBJ-Bill Clinton scale) uses the same word about 6.8 times per 10,000 words. Bill Clinton uses the word "know" 5 times more often than did LBJ. Hillary Clinton uses the word know 57 times per 10,000 words versus 25 times per 10,000 words for John McCain. Bill Clinton uses the word families 3 times more often than LBJ. Hillary Clinton uses the word families about 16.1 per 10,000 words versus 5.8 for Obama and 1.0 for Giuliani. Another popular Bill Clinton word is "back" which Hillary uses about twice as much as the other candidates. The word "prosperity" is used twice as often by Bill Clinton as it is used by LBJ. Hillary mentions prosperity 1.3 times per 10,000 words and Huckabee uses the word prosperity 2.3 times per 10,000 versus .71 times for Obama.

When we instead put the candidates on a George W. Bush – Bill Clinton scale, many of the same patterns emerge as with the LBJ-Bill Clinton scale. Hillary is far and away the candidate who most resembles Bill. McCain and Romney resemble Bush. Huckabee moves away from the Clinton end of the spectrum and towards Bush. Giuliani remains the third most like Bill Clinton. The frequent use of the words "Iraqi" and "disarm" makes McCain resemble Bush while "cold-blooded" is a favorite Bush term that is also widely used by Romney. Hillary tends to use the words Gore, hottest, Americorp, aging, Shumer, and empowerment much more than her rivals and all of these words are frequently used by Bill Clinton.

Finally, and partly for fun we rank the candidates on a Stalin versus JFK index. Words that are very JFK like include believe, threat, civil, responsibility, cooperation, provide, and effort. Words that are very Stalin like are army, enemy, Germans, defeat, enslaved, defend, brothers, and honor. Among the candidates we examine, Huckabee's words are the closest to those of Stalin while Clinton, Obama's and Edward's words are the closest to those of JFK. We should note that this is entirely a *relative* ranking of the candidates. It may be in an absolute sense that Huckabee's word choices are extremely far from those of Stalin. Interestingly, Huckabee frequently uses the words "red," "army" and "october" which were favorite words of Stalin. JFK used the word civil 12.5 times more than Stalin. Obama uses the word "civil" 5.8 times per 10,000 words versus 3.8 for Giuliani, .93 for Romney, and 0 times for Huckabee.

V. Concluding Remarks

Overall we think that there are several advantages to engaging in this textual analysis. Our exercise provides one fairly objective way to measure the words of a candidate. Word use may tell us something about a candidate's policy positions or personality or both. For example, we think it is significant that both Huckabee and Obama resemble Martin Luther King more than their opponents in their frequent use of key words such as children, commitment, poverty, and revolutionary. We are also struck by the degree to which Senator Clinton's word choices strongly resemble those of her husband. We also created an absolute measure of the negativity of the candidates: Clinton and Obama are the least negative while Edwards and Romney are the most negative. Obviously this

exercise makes good use of computing power and is able to handle large amounts of data but it is also only one tool in any effort to understand policy positions of candidates or to predict their future behavior.

References

- Barnes, Fred, "The Seinfeld Campaign," *The Wall Street Journal*, January 3, 2008.
- Benjamin, Daniel J. and Jesse M. Shapiro, "Thin-Slice Forecasts of Gubernatorial Elections," NBER Working Paper No. 12660, November 2006.
- Holmes, David I. 1994. "Authorship Attribution." *Computers and the Humanities*. 28, pp. 87–106.
- Holmes, David I. 1998. "The Evolution of Stylometry in Humanities Scholarship." *Literary and Linguistic Computing*. 13, pp. 111–117.
- Mannion, David and Peter Dixon. 1997. "Authorship Attribution: The Case of Oliver Goldsmith." *The Statistician*. 46, pp. 1–18.
- Michael Laver, Kenneth Benoit, and John Garry (2003) Extracting policy positions from political texts using words as data. *American Political Science Review* 97(2).
- Mosteller, Frederick and David L. Wallace. 1963. "Inference in an Authorship Problem." *Journal of the American Statistical Association*. 58, pp. 275–309.
- Stock, James and Francesco Trebbi, "Who Invented Instrumental Variable Regression?" *Journal of Economic Perspectives* 17 (2003), 177–194.

Table 1
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

A negative mention is counted whenever a candidate mentions an opponent from either party by name or whenever he or she uses the term "opponent." We have also tried including the use of the word "Bush" as a negative mention when used by Democratic candidates. The ranking of the Democrats remains the same. These data are based on 180 speeches gathered from candidate websites. We are working on a broader analysis which uses transcripts gathered by Congressional Quarterly.

Table 2
Comparing the Candidates to Historical Political Figures

We take the set of words used in candidate speeches from candidate websites and from speeches transcribed in *Congressional Quarterly*. We compare to words used in speeches by famous orators as described in the data section. We apply the formula in equation (3) to generate an index for the resemblance of the candidates' speeches to those of the reference speakers. The candidates with the lowest (or most negative) index numbers are the ones that most resemble the reference speaker assigned a 0.

Ronald Reagan (0) Martin Luther King (1)

RR (0) MLK (1)		Standard	Unique	Total
Candidate	Index	Error	Scored	Words
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

Lydon Baines Johnson (0) William Jefferson Clinton (1)

LBJ (0) Bill Clinton (1)		Standard	Unique	Total
Candidate	Index	Error	Scored	Words
Fred Thompson	-0.460	0.111	63,926	3,036
John McCain	-0.063	0.055	289,805	6,762
Mitt Romney	0.124	0.081	128,278	4,601
John Edwards	0.348	0.067	183,923	4,725
Barack Obama	0.857	0.038	589,080	8,151
Rudy Giuliani	0.962	0.081	120,591	3,020
Mike Huckabee	0.993	0.154	34,118	2,009
Hillary Clinton	1.743	0.035	688,837	8,984

Table 2 (continued)

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

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

Table 3
Which Words Indicate that A Candidate Most Resembles Martin Luther King Instead of Ronald Reagan?

Rank	Word	Ratio (Reagan's Use to MLKs)	Use Per 10,000 Words		
			Obama	Clinton	McCain
1	black	0.119	0.236	0.075	0.014
2	get	0.119	2.157	3.435	0.718
3	children	0.136	1.101	1.612	0.472
4	didn't	0.143	0.641	0.763	0.178
5	god's	0.143	0.044	0.012	0.021
6	him	0.143	0.564	0.448	0.438
7	am	0.150	0.898	1.000	1.081
8	justice	0.168	0.338	0.205	0.178
9	refuse	0.178	0.061	0.052	0.048
10	death	0.178	0.155	0.072	0.116
11	came	0.190	0.371	0.569	0.383
12	down	0.194	1.003	0.794	0.486
13	stop	0.204	0.415	0.237	0.219
14	poverty	0.204	0.408	0.153	0.082
15	revolutionary	0.204	0.024	0.012	0.000
16	call	0.204	0.456	0.413	0.239
17	commitment	0.204	0.277	0.286	0.205
18	protest	0.204	0.007	0.012	0.000
19	another	0.204	0.513	0.384	0.937
20	dream	0.219	0.402	0.170	0.068
21	period	0.238	0.030	0.087	0.075
22	bull	0.238	0.000	0.000	0.000
23	aren't	0.238	0.250	0.133	0.109
24	civil	0.238	0.574	0.254	0.219
25	church	0.238	0.176	0.081	0.014
26	tonight	0.238	0.024	0.066	0.055
27	read	0.238	0.135	0.107	0.109
28	able	0.238	0.668	0.650	0.219
29	mind	0.238	0.128	0.098	0.034
30	simply	0.238	0.290	0.124	0.151

Table 3b
Which Words Indicate that A Candidate Most Resembles Ronald Reagan Instead of Martin Luther King?

Rank	word	Ratio RR/MLK	Use Per 10,000 Words							
			Hillary	Obama	Edwards	McCain	Romney	Rudy	Huckabee	Thompson
1	soviet	38.48	0.52	0.34	0.65	1.64	2.47	1.49	0.58	1.87
2	east	25.66	0.49	1.49	0.76	4.45	1.86	5.61	1.17	0.93
3	europa	21.38	0.78	0.61	2.49	1.16	4.02	2.31	0.00	2.80
4	countries	12.83	4.65	4.08	3.90	4.45	3.40	5.61	1.17	4.67
5	she	12.83	14.82	11.14	11.15	4.79	6.03	3.63	18.64	4.98
6	germany	11.40	0.55	0.03	0.32	0.75	0.93	1.82	0.00	0.31
7	since	11.40	3.64	3.51	3.14	4.17	3.40	3.14	0.00	4.05
8	decades	11.40	1.07	2.03	1.30	1.78	1.39	0.33	0.00	0.93
9	arms	11.40	0.14	0.68	0.97	1.44	0.93	0.66	0.58	0.62
10	earth	9.98	0.69	1.55	0.76	2.26	1.39	0.33	2.91	1.25
11	totalitarian	9.98	0.03	0.00	0.00	0.14	0.15	0.00	0.00	0.00
12	tragedy	9.98	0.17	0.57	0.00	0.48	0.62	0.50	0.00	0.62
13	told	9.98	3.44	5.33	1.84	2.19	3.25	3.30	5.83	2.49
14	nuclear	9.98	1.68	3.44	4.44	6.02	10.67	2.64	0.00	1.87
15	courage	8.55	0.69	2.53	5.09	5.82	0.46	0.00	0.58	0.93
16	system	8.55	10.83	8.17	12.45	6.98	4.02	8.91	14.56	4.98
17	open	8.55	1.76	2.46	1.95	2.81	2.16	0.66	1.75	1.87
18	strengthen	8.55	0.78	1.35	4.44	2.87	8.35	0.00	0.00	0.31
19	sign	8.55	1.42	1.32	0.32	0.21	0.93	0.17	0.00	0.31
20	throughout	8.55	1.50	1.35	1.30	3.08	1.55	0.50	1.17	1.87
21	city	7.84	4.10	4.73	2.81	2.33	2.32	35.48	2.91	0.62
22	stands	7.13	0.64	0.44	0.87	0.75	0.46	0.17	1.17	0.62
23	events	7.13	0.52	0.37	0.22	1.30	1.24	0.00	0.00	2.18
24	cultural	7.13	0.12	0.03	0.11	0.00	0.46	0.17	0.00	0.00
25	hard	7.13	9.65	5.13	8.66	6.23	3.71	4.13	2.33	0.31
26	crew	7.13	0.09	0.00	0.00	0.21	0.31	0.00	0.00	0.00
27	economic	6.77	7.57	2.67	5.95	14.64	10.36	3.63	1.17	6.85
28	news	5.70	2.28	2.90	2.92	2.53	2.01	2.97	6.41	3.42
29	toward	5.70	3.12	1.22	1.73	3.22	1.55	2.81	1.17	2.49
30	generation	5.70	3.00	7.36	5.09	2.94	4.95	0.33	5.24	10.58
31	dedication	5.70	0.17	0.07	0.00	0.68	0.31	0.33	0.00	0.93
32	strong	5.70	3.24	4.46	5.41	1.85	14.69	5.94	4.08	5.29
33	gathering	5.70	0.23	0.20	0.22	0.21	0.15	0.00	0.00	0.31
34	prosperity	5.70	1.33	0.71	1.08	5.68	1.55	0.00	2.33	5.91
35	alliance	5.70	0.03	0.10	1.41	0.96	0.77	0.00	0.00	1.25
36	control	5.70	1.10	1.15	1.62	3.56	0.62	1.98	3.50	2.80
37	lived	5.70	1.18	0.64	1.19	0.89	1.55	0.83	3.50	1.25
38	permitted	5.70	0.23	0.00	0.00	0.00	0.00	0.00	0.00	0.00
39	says	5.70	2.40	3.54	2.38	2.12	3.56	2.81	6.99	3.74
40	course	5.70	3.70	3.58	3.25	5.88	13.14	6.44	0.58	8.72
41	repression	5.70	0.03	0.03	0.00	0.00	0.15	0.00	0.00	0.00

Table 3b (continued)

Rank	word	Ratio RR/MLK	Use Per 10,000 Words							
			Hillary	Obama	Edwards	McCain	Romney	Rudy	Huckabee	Thompson
42	party	5.70	0.84	3.85	7.79	4.17	3.71	7.76	1.75	7.16
43	given	5.70	3.18	1.82	2.06	3.42	3.87	2.15	4.08	7.16
44	forward	5.70	3.24	2.19	1.41	2.26	1.55	0.83	1.75	2.18
45	feel	5.70	4.39	2.90	1.30	1.51	2.63	1.82	6.99	3.11
46	much	5.70	15.77	10.67	13.86	15.05	12.06	24.09	12.82	21.79
47	ago	5.34	5.08	5.30	8.66	7.05	5.26	5.12	6.41	4.36
48	future	5.34	7.25	6.21	7.15	9.44	8.81	3.63	3.50	12.76
49	away	4.99	3.32	5.54	5.09	3.15	4.18	4.46	4.08	3.42
50	peaceful	4.99	0.14	0.10	0.54	0.41	0.46	0.66	1.17	0.00
51	strength	4.99	1.33	0.88	3.57	2.81	14.23	1.65	1.17	3.42
52	capacity	4.28	0.81	1.52	0.22	1.23	0.62	0.17	2.33	0.31
53	developing	4.28	0.40	0.64	0.76	1.23	0.62	0.83	0.00	0.00
54	republic	4.28	0.14	0.03	0.22	0.48	0.00	0.83	0.00	0.00
55	behind	4.28	3.35	5.67	3.25	1.51	3.87	0.99	11.07	0.31
56	quest	4.28	0.00	0.24	0.11	0.27	0.00	0.17	0.58	0.00
57	remain	4.28	0.81	0.88	0.32	1.64	2.32	2.15	0.00	1.25
58	took	4.28	3.00	3.00	2.06	2.74	5.88	6.60	2.33	4.05
59	preserve	4.28	0.32	0.17	0.32	1.51	2.16	0.17	1.17	0.31
60	choose	4.28	1.44	1.22	1.62	2.74	1.24	0.50	0.00	0.62

Table 4
What Are the Most Common Words of the 2008 Campaign?

We sum over the set of speeches available on candidate webpages as of January 2008. We omit in this particular table prepositions, pronouns and commonly used words such as "all" "so" and "do" that may not be informative in attempting to divine the themes of the campaign.

word	word_total
care	1073
war	1033
health	880
iraq	798
security	657
energy	632
economy	572
jobs	497
life	496
tax	494
working	448
military	412
economic	408
oil	407
companies	358
pay	354
education	345
congress	325
school	315
insurance	299
troops	297
bush	284
nuclear	272
college	269
trade	257
america	1571
country	1360
american	1333
world	1312
work	1063
government	1059
today	947
americans	886
believe	805
right	788
children	664
well	658
change	631
great	600
families	567
together	523

states	515
national	475
future	475
nation	473
home	464
system	437
washington	425
family	425
support	423
best	420
women	415
job	413
united	412
against	406
money	386
challenges	385
workers	382
power	374
global	373
leadership	341
history	333
political	333
opportunity	331
public	329
lead	328
child	326
state	325
generation	309
america's	299
fight	298
freedom	297
administration	287
federal	287
problems	270
means	270
leaders	269
responsibility	264
politics	262
chance	260
provide	260
century	258
problem	256
course	255
values	254
policy	254
foreign	253
big	252
build	252
