

Does the Presidential Cycle Influence the Speeches of the President?

A Research Paper for Methods of Corpus Linguistics



Abstract: This paper studies whether there are differences in keywords between speeches given during elections compared to speeches given during tenure with a comparative keyword analysis. The results indicate that foreign politics are an important topic for (future) presidents during all times. During election and re-election the election itself seems to be the most important topic. Further, candidates use more informal language, perhaps to come across as more relatable. During the first 3 years of presidency presidents talk about a variety of different subjects and they have a positive, future-oriented tone.

Eline De Witte (r0653377)

Methods of Corpus Linguistics [F0TU1a]

Master in Digital Humanities

Year: 2021-2022

Background information

The speeches of American presidents are well documented and provide rich information on how the president communicates with the American citizens. It is known that the speeches of the president have an important impact of what Americans find important (Cohen, 1995). When a president addresses certain problems the public pays more attention to them, and the problems that he does not talk about receive less attention. The president influences the public agenda in that manner.

The themes in speeches during elections certainly matters. The candidate who manages to speak about themes that move most citizens can win the election (Liu & Lei, 2018). Moreover, voters tend to vote for the presidential candidate that talks most about policy, instead of character (Benoit, 2003). However, politicians are often accused of saying one thing but doing another. Do politicians really continue to care about the same topics once they are elected? There seems to be little research that compares the topics. There are a few differences found between speeches on different moments in the presidential cycle.

Hager and Sullivan (1994) found that the amount of speeches per year depends on several factors, such as the presidential cycle. During the re-election year, e.g. the 4th year of the presidency, presidents produce more speeches than in the other years of their presidency. It has been assumed that presidents produce less speeches over time because their influence decreases over time (Light, 1999). However, this is not reflected in recent research, as Eshbaugh-Soha (2010) saw no significant decline in the average number of speeches per year.

Another difference that is found based on the timing in the presidential cycle, is the complexity of the speech. Presidents give speeches lower in complexity during election campaigns, both in the first- and re-election (Tetlock, 1981). The speeches of one particular president keep the same complexity during the first three years of the presidency.

Hypothesis

I am interested in knowing whether these politicians present themselves differently during their election campaign than when they are elected. I will conduct a comparative keyword analysis and hope to get insight in the different topics that successful presidential candidates and presidents talk about.

Method

Corpus

I will use the corpus that contains speeches from several American presidents, from before and during their presidency. I made the decision to only include the presidents after world war 2. A keyword analysis is namely sensitive for word use and this changes over time. Lumping speeches given in the 19th century together with speeches in the 21st century thus will not make much sense. The corpus contained speeches from the following presidents that were in office after 1945: Reagan, Obama, Kennedy, Nixon, Ford, Carter, Bush, Clinton, Lyndon B. Johnson, Eisenhower and Truman. The speeches of these 11 presidents will be used.

Data analysis plan

The main analysis will be a comparative keyword analysis in RStudio with the use of functions from the `mclm`-package and the `tidyverse`-package. Before that main analysis the data will be divided into three groups: before the elections, the first three years of tenure and the fourth year. I put the first three years together in one group because there seems to be no evidence in the literature that these years would be significantly different from each other. Moreover, only the speeches from presidents after 1945 will be used. The pre-election group consists of the 2 years prior to the presidency. This selection and division is done with a program written in Python that takes all the data from the presidential speech corpus and the result is 3 txt-files that contain all the speeches of 1 group. Then in RStudio 151 common words are removed from the analysis. This is done because 'the' and 'and' are almost always in the top of most frequently used words and do not inform us on what subjects the politicians talk about. Next, the comparative keyword analysis is done. A significance measure, G-test, and an effect size measure, pointwise mutual information, are used.

Effect Size Measure: Pointwise Mutual Information

Pointwise mutual information (PMI) measures the frequency of a word in the target corpus and in another corpus and compares this ratio to the frequency with which other words occur in the target corpus and in another corpus (Stubbs, 1995; Evert, 2008).

In the following analyses a cut-off value of $PMI > 0,3$ will be used. This means that any word has to have a PMI higher than 0,3 to end up in the top keyword list. This is done to ensure that all words in the list are actually more typical of the target group than another group.

Significance Measure: G-test

Considering effect size measures tend to favour low frequency words a second measure is added: the G-test. The G-test is based on chi-squared but is more robust when dealing with small frequencies. Chi-squared is a statistical test that checks whether there is enough evidence against the null-hypothesis (Evert, 2008). The null hypothesis states that there is no difference in how frequent this word is in one text compared to other. Only when the evidence against the null hypothesis is strong enough it can be assumed that the null hypothesis is false and there is in fact a real difference.

In the keyword analysis chi-squared will be used to order the words, in descending order, after the words have past the requirement of a score above 0,3 on pointwise mutual information.

Results

Data Cleaning in Python

The program in Python showed the amount of speeches per category. The pre-election-group contained 15 speeches and consisted of 693 970 characters. This equals an average of 46 264,67 characters per speech. The first three years, or the tenure-group, contained 211 speeches, with 83 speeches given in the first year, 60 in the second and 68 in the third. The group had a total of 3 908 386 characters, which translated to 18 523,16 characters on average. The re-election group had 48 speeches that were composed of 1 134 113 characters. This equals an average of 23 627,35 characters per speech.

Data Cleaning in RStudio

I added more types to the 'stop list'. This is the list of things to remove out of the frequency-lists of the groups because they occur often but are not very informative on the topics of the text. I namely noticed some reoccurring types in the frequency-lists of all three groups. The types 'n' and '--' are in all 3 frequency lists, but it is part of the annotation instead of the actual text said by the politicians.

Analyses in RStudio

	Keyword	PMI	G_signed	dir
1	Mr	1.68	1047.91	1
2	Uh	3,01	792.86	1
3	Nixon	2,78	584.03	1
4	Ford	2,80	563.46	1
5	Kennedy	2,33	554.32	1
6	Carter	2.59	521.10	1
7	Moderator	2.97	362.39	1
8	Senator	1.88	353.44	1
9	Vice	2.18	348.68	1
10	Question	1.59	261.17	1
11	Governor	1.71	255.73	1
12	President	0.85	247.26	1
13	't	0.97	241.24	1
14	Administration	1.57	237.17	1
15	Think	1.02	222.98	1
16	Dukakie	2.96	196.42	1
17	Mcgee	2.99	159.66	1
18	Islands	2.83	152.51	1
19	Formosa	2.90	141.99	1
20	Shadel	2.99	134.44	1
21	Candidate	2.31	134.58	1
22	Don	1.08	122.58	1
23	Candidates	2.31	111.28	1
24	Howe	2.99	109.92	1
25	Quemoy	2.99	109.92	1
26	Matsu	2.98	101.65	1
27	Eight	2.16	98.97	1
28	Kraft	2.98	97.51	1
29	Isn	1.90	96.20	1
30	Prestige	2.59	95.89	1
31	Believe	0.88	95.73	1
32	've	0.73	83.63	1
33	Indicated	2.13	81.80	1
34	Going	0.85	80.60	1
35	Comment	2.14	78.50	1
36	Position	1.42	77.37	1
37	Bush	1.53	76.63	1
38	Communists	1.79	74.04	1
39	Georgia	2.46	69.11	1
40	Mccain	2.76	66.74	1
41	Campaign	1.42	65.86	1
42	Frankel	2.97	64.50	1
43	Ought	1.37	63.66	1
44	Republican	1.30	63.30	1
45	Castro	2.24	63.66	1
46	Minutes	1.83	62.97	1
47	Voted	1.81	62.84	1
48	Cent	2.80	62.58	1
49	Depletion	2.97	60.39	1
50	Maynard	2.97	60.39	1

Table 1. Top 50 keywords for the pre-election group

Table 1 shows the top 50 keywords for the pre-election group. In the column 'keyword' the word is displayed. The column 'PMI' shows the PMI-score of the word and all these scores are bigger than 0,3, as this was a requirement. The column 'G_signed' displays the score for the G-test for a particular word. The last column 'dir' stands for the direction. A score of 1 signifies that the word occurs more then would be expected, a score of -1 means a word occurs less than expected.

I divided the 50 top keywords into categories based on their meaning. Some words clearly belonged to a category, f.e. 'governor', while some required a surface colocation analysis to be classified, such as 'Formosa'. These are the 6 categories for the 50 top keywords:

- (1) Reference to the political system: senator, vice, governor, president, administration, candidate, candidates, communists, campaign, republican, voted, cent, position
- (2) Situation of campaign: moderator, question, Shadel, Howe, Kraft, comment, Frankel, Maynard, minutes, mr, eight
- (3) Names: Nixon, Ford, Kennedy, Carter, Dukakis, McGee, Bush, McCain
- (4) Foreign politics: islands, Formosa, Quemoy, Matsu, Georgia, Castro
- (5) Informal language: 't, don, isn, 'v
- (6) Other: uh, think, prestige, believe, indicated, going, ought, depletion

The most important topic in these speeches seems to be politics and the presidential campaign specifically. For example, 'vice' refers to the vice president that the presidential candidate choses and 'cent' is a keyword because it is part of the expression 'per cent' used to refer to the percentage of voters are currently supporting the candidates. Further, the situation that these speeches are given in frequently occur in the text. It becomes clear that these texts are not all speeches and that some of them are debates, where journalists and other candidates speak as well. This causes some names of journalists to be in the keyword list, such as 'Howe' and 'Kraft'. These names are partly annotation, but are also part of what the candidates responded. Other words that refer to debates in this group are 'comment', 'question' and 'minutes'. Next, a lot of names from politicians are in the top 50 keywords, both from successful presidential candidates and their opponents during election. They are again partly part of the annotation and partly what is said by the candidates.

The 3 previous categories do not give an insight yet on the message of the candidates. However, the fourth group contains names of foreign places and people, showing that presidential candidates talk about foreign politics.

Lastly, I added the category informal language even though it is not a content based category. I found it striking that in this group there are abbreviated forms of verbs, since this is indicative of informal language.

	Keyword	PMI	G_signed	dir
1	Viet-nam	0.56	235.54	1
2	Q	0.42	131.54	1
3	Cooperation	0.38	61.04	1
4	Space	0.36	54.61	1
5	Iraq	0.39	51.32	1
6	University	0.44	49.72	1
7	Nthis	0.36	49.05	1
8	Peaceful	0.33	46.54	1
9	Seek	0.30	46.32	1
10	Mankind	0.41	42.08	1
11	Lasting	0.47	38.52	1
12	Alabama	0.54	38.33	1
13	Hemisphere	0.42	36.55	1
14	Negro	0.49	34.78	1
15	Steps	0.35	33.23	1
16	Coverage	0.44	33.15	1
17	Investment	0.31	32.97	1
18	Nso	0.35	32.02	1
19	West	0.32	31.75	1
20	Western	0.35	31.30	1
21	1961	0.50	31.04	1
22	Palestinian	0.54	29.37	1
23	Stability	0.37	29.09	1
24	Peoples	0.33	29.03	1
25	Financial	0.36	28.65	1
26	April	0.49	28.29	1
27	Request	0.42	28.14	1
28	Concern	0.33	28.11	1
29	Danger	0.35	26.93	1
30	Affirmative	0.47	26.77	1
31	Solution	0.38	26.70	1
32	Palestinians	0.53	26.41	1
33	Recovery	0.53	25.96	1
34	London	0.55	25.95	1
35	Global	0.36	25.73	1
36	Berlin	0.32	25.72	1
37	De	0.53	25.67	1
38	Graduates	0.53	25.67	1
39	Ensure	0.36	25.52	1
40	Region	0.31	25.50	1
41	Institutions	0.35	24.50	1
42	Attacks	0.43	24.31	1
43	Dangers	0.42	24.17	1
44	Corps	0.40	24.10	1
45	1963	0.50	23.93	1
46	Sought	0.39	23.05	1
47	Qaeda	0.45	23.04	1
48	Germany	0.38	22.76	1
49	Role	0.32	22.12	1
50	Afghan	0.53	22.00	1

Table 2. Top 50 keywords for the tenure-group

The top 50 keywords depicted in table 2 are divided into categories. Again surface colocation analysis is used to place some words into a category.

- (1) Places: Vietnam, Iraq, Palestinian, Palestinians, London, global, Berlin, Germany, Afghan, Alabama, west, western, region
- (2) War: danger, attacks, dangers, qaeda, recovery, stability, solution, peaceful, lasting
- (3) Time: 1961, April, 1963
- (4) Annotation: q, nthis, nso
- (5) Others: space, university, seek, mankind, hemisphere, negro, steps, coverage, peoples, request, concern, affirmative, graduates, ensure, institutions, sought, role, de (part of expressions: arc de triumph and de gaulle), cooperation, corps, finance, investment

Thirteen words refer to places around the globe. Most of them are places outside the U.S.A. and 5 of them refer to warzones. War seems to be a frequent subject in the speeches of a president as the second category is named war. There are both negative words in the category that relate to danger and words with positive associations, such as 'peaceful'. The word 'lasting' is in this category because it often occurred in the composition 'lasting peace'. Presidents use these 2 words often when talking about warzones and American interventions.

The third category time is a bit surprising, but it implies that presidents frequently refer to certain times. Surface colocation analysis reveals that presidents refer to their own timing: they promise to achieve certain goals by a certain time and report which goals they have achieved in a certain amount of time. Unfortunately these times are also part of the annotation.

There are a few other examples of annotations that got into the keyword list. After every sentence there is a sign, '\n', to instruct the computer to print the following text on the next line. In R however 'n' is seen as part of the words and so the word 'nthis' is taken into account for the analysis, instead of 'this'.

Lastly, the category with words that could not be divided into categories is quite big. Presidents seem to talk about a lot different subjects. A similar tone can be noted in many of these words: they are positive and future oriented. The presidents talk about young people, want to ensure certain things, invest... The word 'steps' for example occurs often because presidents refer to the progress they are making. The presidents talk about their goals, such as going to 'space' and health care 'coverage' for all American citizens.

	Keyword	PMI	G_signed	dir
1	Clinton	2.24	363.85	1
2	Dole	2.23	357.66	1
3	Reagan	2.06	317.28	1
4	Vietnam	1.41	257.46	1
5	Lehrer	1.79	202.68	1
6	Perot	2.31	198.67	1
7	Mondale	2.12	174.56	1
8	Applause	1.11	156.92	1
9	're	0.71	108.45	1
10	Governor	0.99	106.56	1
11	've	0.60	95.10	1
12	't	0.51	93.98	1
13	Newman	2.24	91.73	1
14	President	0.41	82.12	1
15	Minute	1.75	81.39	1
16	Get	0.70	73.06	1
17	Bridge	1.74	70.42	1
18	Smith	1.54	68.46	1
19	Four	0.99	62.01	1
20	America	0.44	61.58	1
21	Hanoi	1.67	56.27	1
22	Got	0.84	54.18	1
23	1966	2.25	53.63	1
24	Republican	0.94	50.11	1
25	Party	0.83	48.14	1
26	Mr	0.38	46.92	1
27	'll	0.80	46.79	1
28	Bush	1.02	46.66	1
29	Children	0.63	45.73	1
30	Rebuttal	1.97	44.36	1
31	Followup	2.27	42.49	1
32	Kalb	2.27	42.49	1
33	Cut	0.73	41.19	1
34	Medicare	1.24	40.15	1
35	Ellis	2.27	39.34	1
36	Senator	0.66	39.24	1
37	80 th	2.22	38.14	1
38	Schools	0.91	37.70	1
39	Drugs	1.16	36.31	1
40	--to	2.26	36.20	1
41	Hilliard	2.26	36.20	1
42	Kondracke	2.26	36.20	1
43	Opponent	1.40	36.08	1
44	Didn	0.88	36.05	1
45	Family	0.74	35.57	1
46	Thank	0.66	35.34	1
47	Know	0.39	34.95	1
48	Victims	1.32	33.87	1
49	Right	0.48	33.77	1
50	jobs	0.56	33.63	1

Table 3. Top 50 keywords for the re-election group

The top 50 keywords for the re-election group, shown in table 3, are divided into 7 categories. Once again I used surface colocation analysis to place some words into a category. The categories are shown here, starting with the biggest category:

- (1) Situation of campaign: Lehrer, Newman, minute, Smith, rebuttal, followup, Kalb, Ellis, Hilliard, Kondracke, opponent, thank, four, mr
- (2) Names: Clinton, Dole, Reagan, Perot, Mondale, Bush
- (3) References to the political system: governor, president, America, republican, party, senator
- (4) Abbreviated forms: 're, 've, 't, 'll, didn
- (5) Annotation: applause, --t
- (6) Foreign politics: Vietnam, Hanoi
- (7) Others: get, bridge, got, 1966, children, cut, medicare, 80th, schools, drugs, family, know, victims, right, jobs

The first category contains words that refer to the presidential campaign and the political debates for the elections. The category consist of the names of journalists, and words such as 'four' and 'minute'. These refer to the time a candidate gets to answer a question. Unfortunately this shows that the text does not only contain what is said by the targeted presidential candidates, but it also includes what is said by journalists and opponents.

Further, talking about presidential candidates and the political system are typical for this group. For example 'party' is in this list because it refers to a political party. Half of the 50 keywords are in one of first 3 categories which all clearly are said in context of the presidential campaigns. The elections are clearly the most talked about topic in this group.

Another remarkable category consists of abbreviated forms. These forms are considered to be more informal. This might suggest that during the re-election presidents use more informal speech then they did before.

There are 2 types in the top keywords of this group that are part of the annotation, namely 'applause' and '--t'. 'Applause' is not said by the president but the reaction of the audience was included in the text. This does suggest that there is more applause during electoral speeches and debates then during the speeches given in the first 3 years of tenure.

The category foreign politics has 2 words, which are 'Vietnam' and 'Hanoi', the capital of Vietnam. It is the first category in this list that shines light on the matters that the presidential candidates talk about. In the rest category however there are also several words that refer to the core tenets of the candidates, such as 'medicare', 'jobs' and 'drugs'.

COMPARISON BETWEEN GROUPS OF KEYWORDS

A first remark is that the election and the re-election group yield very similar top keywords. Firstly, they have almost exactly the same categories. The only difference is that the election group has no separate annotation category. Moreover the groups have more than half of their keywords refer to presidential candidates, the democratic system, the election and the electoral debates. It is on one hand unfortunate that these analyses do not give insight into the core tenets of the presidential candidates but on the other hand these results are also indicative of what is important at the moment: winning the election.

A second similarity is that both during the elections and the re-elections more abbreviated forms of verbs get used, indicative of more informal language use. There are several possible explanations for this. One explanation is that the presidential candidates cannot prepare what they are going to say exactly, as they get questions from journalists and react to their opponents. They thus do not start from a written text. Presidents give more speeches that are prepared beforehand and that might result in more formal language use. However some of the speeches in the tenure group are press conferences. Then the president answers questions from journalists and he too has no prepared text. Another explanation is that presidential candidates might use less complex language during elections as a way too be understandable and relatable too voters. It has been shown before that presidents give less complex speeches during elections than during tenure (Tetlock, 1981).

A last surprising resemblance between the election and the re-election group is a few of their keywords. Some words do actually appear in both lists, such as 'Bush', 'governor', 'president' and 'republican'. This was surprising at first but it does make sense when thought through. The target group gets compared with the 2 other groups at the same time. For example the election group got compared with the tenure and re-election group. When these 2 groups get joined it is possible that the ratio of a target word compared to other words in the tenure and re-election group is much lower then in the election group because the word might not occur often in the tenure group. It lowers the combined ratio of the tenure and re-election group. There are also many other words that are remarkable similar in the election and re-election 2 keyword lists, such as 'minute' and 'minutes'. This is another proof that the speeches during these years resemble one another.

The tenure group yields very different results compared to the election and the re-election groups. Firstly, most of the keywords refer to the policies and core tenets of the presidents. Clearly presidents talk about foreign politics a lot, but there are many different subjects as well, such as finance, young graduates, jobs and space. The category 'others' is much bigger in the group indicating that presidents talk about the larger variety of things.

A second notable difference between the tenure top keywords and the other 2 groups is the connotation of the words. In the election and re-election group most words seem

to be neutral, f.e. 'followup' and 'president'. The category foreign politics is common among the 3 groups and has quite negative connotations. Tenure however has the category 'war' that is obviously holds words with negative meanings. Further tenure has many keywords that refer to progress and positive development. Some words that the presidents use in the context of 'war' involve improvements, such as 'stability' and 'solution'. Many words in the others-category also refer to the future, such as 'young graduates', 'steps'.

There are similarities between all three groups as well. Unfortunately, all three keyword lists have some annotation in there. This is noise in the results and it does not provide additional information on the speeches. In future analyses it would be a good idea to filter these out. As contents regards the 3 groups have the category foreign politics in common. Clearly, American presidents and top politicians often talk about the subject. The men often refer to places outside the U.S.A. where there has been war in the last 70 years. Vietnam is most often mentioned, occurring in every list in some form, i.e. the country's name or the name of a city.

Discussion

Results

The results show that American presidents and top politicians devote a lot of time in their speeches to foreign politics and more specifically to wars. The speeches given during elections and re-elections were very similar in which top key words they yielded, while the tenure had different keywords. These results indicate that speeches given during the 4th year of presidency, during re-elections, should not be automatically lumped together with speeches given during the first 3 years of presidency. After all, the keywords indicate that these speeches are less about the policies for this tenure and more about the presidential campaign for the next tenure. The keywords of the election and re-election are very different from the keywords of tenure. This suggests that successful presidential candidates do not talk about the same topics they will talk about during their presidency. Furthermore, they use more informal language. During tenure the keywords overall have more positive connotations and are more future-oriented.

Limitations

First of all, keyword analysis is an exploratory technique. Further research on the topic done with other techniques can provide more insight. These results can however, serve as an inspiration for other researchers and shine a first light on important differences. A second limitation is that these top keywords contained a lot of noise. The texts did not just contain what was said by a president or successful presidential candidate and it was not filtered out before the analyses. The texts also contained questions by journalists, answers from opponents in the presidential campaigns and annotation. This definitely influenced the results. It would thus be fruitful to conduct these analyses again, with texts where this noise has been removed.

References

- Benoit, W. L. (2003). Topic of presidential campaign discourse and election outcome. *Western Journal of Communication*, 67(), 97-112.
- Cohen, J. E. (1995). Presidential rhetoric and the public agenda. *American Journal of Political Science*, 87-107.
- Eshbaugh-Soha, M. (2010). The politics of presidential speeches. *In Congress & the Presidency*, 37, 1-21.
- Evert, S. (2008). Corpora and collocations. *Corpus linguistics. An international handbook*, 2, 1212-1248.
- Hager, G. L., & Sullivan, T. (1994). President-centered and presidency-centered explanations of presidential public activity. *American Journal of Political Science*, 1079-1103.
- Lewis, M. L., & Frank, M. C. (2016). The length of words reflects their conceptual complexity. *Cognition*, 153, 182-195.
- Light, P. (1999). *The president's agenda: Domestic policy choice from Kennedy to Clinton*. JHU Press.
- Liu, D., & Lei, L. (2018). The appeal to political sentiment: An analysis of Donald Trump's and Hillary Clinton's speech themes and discourse strategies in the 2016 US presidential election. *Discourse, context & media*, 25, 143-152.
- Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of language*, 2, 23-55.
- Tetlock, P. E. (1981). Pre-to postelection shifts in presidential rhetoric: Impression management or cognitive adjustment. *Journal of Personality and Social Psychology*, 41, 207.