A Comparative Analysis of Speech Patterns Used by Successful and Unsuccessful Debaters

Tufts University, Medford, MA, 02155
Narimon Farhangi
narimon.farhangi@tufts.edu
Updated 2/11/2017

Abstract

This paper describes an analysis of transcripts of Oxford-style debates. I look for disparities in the speech patterns of winners and loser. I conduct different speech analyses on this text including sentiment analysis and looking at n-grams to analyze the attributes that differentiate the speech of winners from that of losers . By doing this, I find speech patterns that winners employ in order to be more convincing. My analysis revealed speakers on the winning team employ more assertive language.

1 Introduction

The art of persuasion is a crucial skill that is difficult to master due to the complicated nature of human psychology. Aristotle's modes of persuasion, ethos (character), pathos (emotion), and logos (logic), guide the overarching structure of many arguments due its prevalence in academia. Like most prevalent theories, it is a broad suggestion that is difficult to employ and can only influence the organization of an argument. I wanted to answer the question: aside from subjectively having a superior argument, what mannerisms measurably make an argument more or less convincing?

In this paper I will describe my approach to algorithmically determining concrete strategies one can employ to make them more convincing. I analyzed the speech of winners and losers of debates in order to identify how the speech patterns I measured affected the persuasiveness of an argument.

2 Background

In researching the art of persuasion, I found many papers and books that more or less rehash the same basic points. Aristotle's modes of persuasion are almost universally referenced.

However appealing this theory appears, it is predicated on Aristotle's idea about how humans ought to be convinced: not necessarily how they actually are convinced. The majority of relevant literature I came across explained ways of tailoring an argument to its audience and supporting an argument with evidence. They gave broad, subjective advice. The rationale for these suggestions exclusively relied upon conjecture regarding the author's beliefs on what should change an individual's mind: not data on what has been actually demonstrated to. This paper aims to fill that void by identifying evidence-based tactics that convince.

3 Dataset

I chose to analyze speech from Oxford-style debates because their conclusions are not reached by arbitration, but rather by each side's ability to sway the opinions of the audience. I downloaded the transcripts of about 100 debates from NPR's Intelligence Squared program.

Due to formatting issues, I had to ignore all debates after February 2015 as well as some other random ones. Many sentences in certain debates also contain randomly placed characters whose ascii values were greater than 127 that rendered the sentences unintelligible. I ignored these sentences in my analysis. Also, because of the dataset, conclusions I come to are biased toward being effective on NPR's predominantly liberal audience.

4 Speech Analysis

In order to find the efficacy of certain patterns of speech and mannerisms, I aggregated all text spoken by the winners and losers of each debate and analyzed the differences between these two bodies of text. I implemented several different types of speech analysis.

4.1 Semantic Analysis

Semantic analysis is a type of natural language processing that attempts to quantify or categorize different emotional aspects of a body of text. Using CLiPS's libraries, I implemented four different types of semantic analysis and compared the results of how the winners and losers of the debates spoke.

4.1.1 Mood

Mood analysis categorizes a sentence as indicative, imperative, conditional or subjunctive. By analyzing auxiliary verbs and adverbs, it places a sentence into one of the following four categories: indicative, subjunctive, imperative, and conditional. For the two bodies of text, I calculated the count of the number of sentences that fell into each mood category. I then calculated the percentages of the total number of sentences that were each type. I analyze the winning and losing percentages.

4.1.2 Modality

Modality is a measure of the level of certainty of a sentence. This level of certainty is a calculated as a number between -1 and 1, where a larger number represents a higher level of certainty. The example given by CLiPS is: " 'I wish it would stop raining' scores -0.35, whereas 'it will stop raining' scores +0.75." I calculated the percentage of the total sentences that scored each possible modality value, rounding all values to the nearest hundredth. I compare the winning and losing percentages in histogram form using varying bucket sizes.

4.1.3 Subjectivity

CLiPS's subjectivity analysis is a measure of the level of subjectivity of a sentence. This is calculated as a number between 0 and 1, where a larger number represents a more subjective sentence. I calculated the percentage of the total sentences that scored each possible subjectivity value, rounding to the nearest hundredth. I analyzed the subjectivities of the two sides in the exact same manner in which I analyzed modality, as described in the modality section above.

4.1.4 Polarity

Polarity measures the extent to which a sentence expresses a negative, neutral or positive sentiment. The polarity is calculated as a number between -1 and 1, where a larger number represents a more positive sentence. I calculated the percentage of the total sentences that scored each possible polarity value, rounding all calculated polarity values to the nearest hundredth. My analysis of this polarity data is exactly the same as that for modality and subjectivity.

4.2 N-gram Analysis

For the purpose of this project, I focused on n-grams that are a list of 1, 2, or 3 consecutive words. N-grams cannot span across multiple sentences. For each unique n-gram that shows up at least once in the winner or loser text, I calculated the frequency with which that n-gram appeared in both texts. For my entire n-gram analysis, I analyzed lemmatized text.

In order to determine what n-grams were used significantly more by one team than the other, I used an independent two-sample t-test. For each n-gram, I calculated a p-value that measures the significance of the discrepancy in frequency of use of that n-gram by the two teams. I considered an n-gram that occurred at least 50 times in the data set with a p-value greater than or equal to 98% to be statistically significant. By perusing the list of significant n-grams and ignoring those which I believed to not be generally usable, I manually curated the resulting lists of n-grams for the winning and losing teams, creating a list of n-grams correlated with statistical significance with winning and one of n-grams correlated with statistical significance with losing.

4.3 Length Analyses

I calculated the number of times each team used a sentence of each possible length and a word of each possible length. Using these, I calculated the average sentence length and word length for each team. Also I created histograms using this data. I also compared the total number of sentences and words spoken.

4.4 Audience Reactions

The transcripts of the debates I used for this project denote times at which the audience reacted with laughter or applause to statements from the speakers. Using these denotations of applause and laughter, I calculated the number of times each team produced either reaction from the audience. Because one team speaking more than the other could skew a comparison of the number of audience reactions, I controlled for the total length of speech.

4.5 Motion Words

For each debate, I extracted what I felt were the keywords in the debate's motion and designated those as "Motion Words." In each debate, I counted the number of times each team used a motion word in their speech. These numbers were aggregated to come up with a final count of the number of motion words used by the winners and by the losers. I compared the ratio of motion words to total words spoken for each team in order to determine whether or not there was a significant difference in the frequency with which each team used motion words as a ratio of total words spoken.

5 Results & Analysis

I studied each different type of linguistic analysis described above separately, drawing conclusions from both numerical and visual analysis when relevant. If applicable, I utilized histograms to demonstrate a difference (or lack thereof) between the two teams' speaking tendencies in the given category.

5.1 Semantic Analysis

I analyzed Modality, Subjectivity and Polarity in the exact same manner. For each of these three types of semantic analysis, I constructed a histogram that displayed the relative frequencies for each possible value of the category (once rounding to the nearest 100th and once to the nearest 10th). Using both numerical differences and visual differences in the histograms, I determined whether or not one team's general sentiment was different from the other's. Because mood calculates a quaternary value, the comparison was much more simple but was also supplemented by a comparative histogram.

5.1.1 Mood

I compared the percentage of total sentences that fall under each category – indicative, imperative, conditional and subjunctive – between the winning and losing text. The difference in the usages of indicative and conditional sentences were both less than .5%. The winning teams used imperative sentences ~3% more than the losing teams. CLiPS explains that imperative sentences are used to issue a command or warning, both of which are authoritative acts that imply confidence. Although a 3% difference is not large, imperative sentences being more generally convincing is not difficult to believe. The losing team used sentences classified as subjunctive ~5% more than the winning team. CliPS defines a subjunctive sentence as one that expresses a wish or an opinion. As almost the polar opposite of an imperative sentence, use of subjunctive sentences express doubt, or at least lack of certainty.

5.1.2 Modality

Both a histogram of values rounded to the nearest .01 and one of values rounded down to the nearest .1 revealed essentially no difference in the modality with which the winning and losing teams spoke. A mathematical analysis of the differences between usages of different buckets of modality confirmed the conclusions my visual analysis. Regardless of bucket size, there was no significant difference between the two teams.

5.1.3 Subjectivity

As with the modality analysis described above, both a histogram of values rounded to the nearest .01 and one of values rounded down to the nearest .1 revealed very little difference in the subjectivity of the two teams speech. A mathematical analysis also supported this conclusion.

5.1.4 Polarity

Like the Modality & Subjectivity analyses, an analysis of the differences by polarity scores rounded to the nearest .01 or .1 yielded no interesting results. However, the winning team did use sentences with negative polarity scores ~5% more than the losing team. Two interpretations of this I considered are that pessimistic views are seen as more believable and that the winning team does more negating of the losing team's points. The teams spoke positively about the same amount.

5.2 N-gram Analysis

After manually curating the list of significant lemmatized n-grams for both sides, I came up with the following lists.

Lemmatized n-grams strongly correlated with winning:

n-gram	p-value	Winning Count	Losing Count
millennial	1.000	74	10
illegal	1.000	160	66
very	1.000	1596	1232
end up	1.000	107	41
talk	1.000	1165	900
interest	1.000	410	289
trial	1.000	84	39
significant	1.000	70	30
growth	1.000	190	117
policy	1.000	477	350
public	1.000	284	193
be very	1.000	417	302
actually	1.000	891	708
so on	1.000	63	28
that.	1.000	562	427
provide	1.000	187	119
have do	1.000	155	95
benefit	1.000	284	198
talk about	1.000	773	612
information	0.999	205	136
be talk	0.999	386	284
you should	0.999	118	69
fact,	0.999	254	177
impact	0.999	122	73
fact	0.999	776	625
number of	0.999	201	137
should	0.998	1014	836
know	0.998	1942	1670
will be	0.998	315	234
in fact,	0.997	224	159
obama	0.997	226	162
and what	0.996	256	188
children	0.996	227	164
we be talk	0.996	183	128
be talk about	0.995	339	259
simultaneously	0.994	150	103
but	0.994	3463	3078
data	0.993	143	98
address	0.993	149	103
i be sorry	0.992	72	43
we be in	0.991	74	45

Lemmatized n-grams strongly correlated with losing:

n-gram	p-value	Losing count	Winning count
terror	1	94	39
you	1	8252	8111
the american	1	247	164
my	1	1180	1025
government	1	879	748
media	1	155	95
attack	1	224	153
the government	1	298	218
book	1	241	169
part of the	1	157	100
fail	1	158	101
threat	0.9999	154	100
home	0.9999	156	102
deal with	0.9999	183	127
say	0.9998	2516	2401
decision	0.9998	216	158
to look at	0.9997	87	50
it may	0.9997	63	32
can go	0.9997	67	35
look at	0.9997	586	504
part of	0.9996	297	235
investment	0.9995	85	50
probably	0.9995	209	156
community	0.9995	179	130
you have	0.9994	787	704
need	0.9994	827	743
have never	0.9971	73	46
in my	0.9965	154	117
no one	0.9955	104	74
i say,	0.9951	66	42
when i be	0.9947	62	39
i have be	0.9945	81	55
the audience	0.9932	73	49
you be	0.9927	1036	983
do not believe	0.9915	72	49
i have	0.9914	634	585
you be go	0.9912	123	94

5.3 Length Analyses

Not only was the average word length for both sides almost the exact same, a histogram comparing the percentage of all words that are each length depicted almost no difference between the speech of the two sides.

The results of my analysis of sentence lengths almost exactly mirrored that of the word lengths. The team's usages of sentences of all ranges of lengths were basically equal.

The winning team spoke ~6% more than the losing team presumably because the team that speaks more has more opportunities to argue their point and because of this is able to construct a superior argument.

5.4 Audience Reactions

The winning team garnered applause at around the same rate as the losing team given that the audience was caused to applaud ~7% more by the winning team than the losing team but the winning team spoke ~6% more than the losing team. However, the winning team caused the audience to laugh ~31% more than the losing team. There was a clear tendency for the audiences to trust the side that made it them more.

5.5 Motion Words

When comparing the percentage of words spoken by each team that was a motion word, I found that the winning team used motion words ~6% more than the losing team. However, because the winning team also spoke ~6% more, the two teams used motion words at about the same rate.

6 Future Work

While certain aspects of the speech analysis conducted in this experiment yielded interesting results, I was not fully satisfied by the number of concretely utilizable strategies I discovered.

I will use the syntactic parse trees of sentences spoken by the two sides to analyze disparities in the grammatical sentence structures used by the two teams. I would also like to factor the magnitude of victory into my analysis. Instead of binarizing the outcome of a debate, I will find the correlation coefficient of any mannerism with the magnitude of victory. By re-doing all analysis described in this paper using this method, I predict to achieve more interesting results and subsequent conclusions.

7 References

[1] Semantic Analysis Libraries
http://www.clips.ua.ac.be/pages/pattern-en

[2] Statistical Computation Library
https://docs.scipy.org/doc/scipy-0.17.0/reference/generated/scipy.stats.t.html