

Investigating Political Language Use and Sub-Nationalism Evinced through Online Comments of Democratic and Republican Debates

Tam Dang, Saina Parsi

June 2, 2016

Contents

1	Abstract	1
2	Introduction	2
3	Methods	3
4	Interpretation	5
5	Conclusion	7

1 Abstract

stuff here

2 Introduction

Feelings of nationalism and identity have often risen in place of ideologies and mindsets. In Benedict Anderson's *Imagined Communities*, Vietnam's invasion and occupation of Cambodia was an example of how nationalism had out-weighted the ideology of Marxism. Both countries at the time were Marxist regimes, and yet one had invaded the other. One can surmise that people have a predisposition to identify more with people similar to themselves as opposed to people who only *think* similarly. On a global scale, nationalism divides people by factors including ethnicity, citizenship, language use, and national identity. Similarly, ideologies such as Marxism, capitalism, and democracy can also divide people on a global scale. However, the former has the potential to supersede the latter.

This paper seeks to explore the idea of **sub-nationalism**, how people can choose to identify with groups nested within a nation. Sub-nationalism will be defined in this paper as the idea that there are ideologies that people identify so strongly with that it manifests into a form of *nested nationalism*. This phenomena

acts as a hybrid between nationalism and governmental ideology. This nested form of nationalism has criteria that consists of ideals and opinions rather than ethnic or national identity, yet still has strength comparable to the outer-scoped nationalism.

In other words, ways of thinking, even those that are only marginally different, can ultimately produce different groups with which to identify. Within these groups, even if their ideas are similar, or even the same on several levels (i.e. good intentions for the country such as improving the economy, providing jobs, health care), they will still continue to establish boundaries and regard each other as separate entities.

The Democratic and Republican parties of the United States are such examples of sub-national entities. Through collecting and analyzing the language use between and within these groups from an anonymous and political context, this paper will explore sub-nationalism by recognizing the trends that arise as these groups agree or disagree with each other, support or attack each other, and deduce the environments in which these possibilities occur.

3 Methods

Video Selection

During election season, presidential candidates debate over current issues within their respective parties. Towards the end of each election year, only the strongest candidates remain, along with the strongest supporters of these candidates. While selecting the videos to extract data from, YouTube videos of the latest debates within the Democratic and Republican parties from the years 2008, 2012¹, and 2016 were chosen. Choosing the latest debate from each year increased

the likelihood that the videos were published during the most heated and critical moments of the election year, times in which the Democratic and Republican candidates were soon to be chosen. These would also be times in which people are readily available to argue vehemently for their candidate. As a consequence, the differences between Democratic and Republican supporters would be more pronounced and distinguishable.

Data Extraction

The YouTube comments from each video were first sorted from the default option "Top Comments" to "Newest Comments". This allowed the comments to be sorted chronologically. Only comments that were both relevant in the political discussion and were within a year of the video's publish date were selected, where relevance involves whether or not the comment was constructive or critical towards a candidate or a user. To systematically categorize the data, comments were sorted by intention, positivity versus negativity

which will be referred to as *connotation*, and whether they clearly identified as democrat or republican. Intention and connotation are always packaged as a pair whereas the political position is a stand alone dimension.

Comments can include all three aspects, or they could omit either the political stance or the intention and connotation. Intention is further divided into two directions, the comment can either be directed toward a candidate or towards a non-candidate (typically another YouTube user or other political figures). These direc-

¹In 2012, president Obama was selected as the Democratic candidate as he ran for a second term. The video selected for 2012 therefore was a debate between Obama and Romney instead.

tions are denoted as (c) and (u), respectively. These labels will always be followed by a connotation. For example, a negative comment directed toward another user that was meant to either criticize or insult would be denoted as (u-), where as a positive comment directed toward a candidate that is meant to compliment or agree with the candidate would be denoted as (c+).

For the final dimension, comments that clearly demonstrate an

affinity toward either the democratic or republican party will be denoted with (d) or (r) respectively, in conjunction with their intention and connotation. For example, a comment from someone who identifies as democratic and whose intention was to compliment the candidate they support would be denoted as (d)(c+). For further reference, a table is provided within the appendix each combination explained.

Data Processing

Each comment, along with its three dimensions, were placed in a text file for each video on its own line. Totalling in three files per political party, several Python scripts were written to process the data.

The first script calculated the raw percentages of comments that contained at least one category. In other words, for each video, the percentages of occurrences of each category, (d), (r), (c+), (c-), (u+), and (u-) were calculated.

The second script calculated percentages within the scope of *comment threads*. A comment thread will be defined as having a *top comment* that contains a political stance or intention and connotation followed by comments that are *replying* to that top comment. For example, within each video, if a (r)(c-) comment ap-

pears that has one or more replies to it, the percentage of each category for all the replies (similarly to the first script) are calculated and *mapped* to that top comment. After all the top comments have been processed, the mappings from the example above would be combined with the mappings of all the other (r)(c-) comments to their replies. This process is continued for all top comments for the video for all the possible combinations of political stances and intentions with connotations.

This process of mapping from the top comment categories to the reply categories essentially reveals the environments in which people respond to comments, how often they respond in these environments, and how they respond. In other words, these mappings reveal the sub-

nationalist trends this paper seeks to explore. The interpretation of the data will be referencing visual rep-

resentations of these mappings provided in the appendix.

4 Interpretation

Connotation trends across Comment Threads in Democratic Videos

(d)(c+) top comments (comments that were by democratic-identifying users supporting their candidate) made up 28% of the total top comments for the Democratic 2008 video. 50% of the replies to these types of top comments in this video were (d)(c+), 25% were (u+), and 25% were (d)(u-). This means democratic users supporting their candidate were primarily met with those trying to form solidarity with them by either by replying with another compliment to the democratic candidate, or by agreeing with or supporting the users. For (d)(c-) top comments (democratic users that are against a particular candidate), 100% of replies were (u+), meaning that every reply was in support of what the top comment user had said. In general, there is an overwhelming amount of positive connotation for either intention within replies to top comments for the Democratic video from 2008.

However, in the Democratic 2012 video, there is a dramatic change

in connotation. In 2008, the connotations within replies were by a strong majority, primarily positive. But within 2012, 0% of replies to a (d)(c+) top comment were either (d)(c+), or (u+). About 35% percent of replies to (d)(c+) were now (c-) and 40% (u-). In other words, instead of mostly approval and support, democratic users supporting a candidate are now met with mostly negative comments that are either against that candidate or against themselves. This sharp increase in hostility is further evinced by the fact that for the total sum of replies across all the categories of top comments, (u-) increased from 0% to 46%, and (c-) had increased from 4% to 31%. To summarize, from the democratic video comments from 2008 to 2012, there is a decreased amount of support and a considerably increased amount of disagreement and attacks.

A possible cause for the increase in hostility could be due to the two sub-nationalist groups forming their distinctions. In the Democratic 2008 video, 0% of either top comments and replies contained Republican affiliation. In Democrat 2012, this in-

creased to 24% and 6%, respectively. Republican users now have a presence in the Democratic videos, and based on the sharp increase in negative connotation, their presence is positively correlated with hostility. It is possible that the presence of the other sub-nationalist group increases the tension of the online environment. An environment that was mostly democratic before is now permeated by republican users. Of the top comments that were Republican affiliated in the Democratic 2012 video, the replies were 40% (u-). This means nearly half of the replies to comments with Republican affiliation were met with either insult or disagreement.

Although the majority of negatively connotative comments were done by those without a clear political stance, over 65% of (u-) comments were directed toward either a (d)(c+), (d)(c-), or (r)(c+). Considering the backlash that users with a political stance are experiencing, and the general increase in hostility across all categories of top comments, it can be said that the permeation of republican users in the 2012 Democratic video appears to be threatening democratic ideals while at the same time allows democratic users an opportunity to threaten republican ideals.

This is causing more people to retort or insult to either defend their sub-nationalist ideals, or to disparage

the other sub-nationalist group. Ultimately this is revealing the strength and prevalence of sub-nationalism within this online, political environment, how obvious the distinctions are when these groups are in the presence of each other, and how readily people identify with or retaliate against either the democratic or republican groups.

Equalization of Comment Connotation in the Democratic Videos

The increase in hostility does not continue in the 2016 democratic video, but instead the amount of positive and negative connotation equalize and form a fairly even distribution. Out of all the total replies across all the top comment categories, 15% were (c+), 22% (c-), 22% (u+), and 26% (u-). This is a stark contrast to the previous videos, which were dominantly either positive or negative. This equalization is further pronounced when examining the proportions of (u-) and (u+) replies to top comments with a political affiliation.

40% of the (u-) replies (replies that disagree with or are attacking another user) for the entire 2016 Democratic video were directed toward the republican users, meaning that those users are being met largely with disagreement and derision. On the other hand, half of (u+) replies

(replies that agree with or support another user) were also directed toward users with a republican affiliation. This reveals that republican users are also being met largely with agreement and possibly solidarity. For democratic users, they are being met with 16% of both the (u-) and (u+) replies.

With a 4:5 ratio for republican users and a 1:1 ratio for democratic users in terms of pure support versus backlash, an equalization is clearly evident. The presence of sub-nationalism is now strong enough for the groups to both attack and support at equal rates. The previous videos had shown a dominant positive or negative trend in their language connotations. It can be argued that for the prior videos within the comment sections, the sense of sub-

nationalism was not strong enough, and the threat against either group not severe enough, to elicit the need to both attack the other and defend themselves.

To summarize, from 2008 to 2016, within the comment sections of Democratic debates posted on YouTube, there has been a rise of sub-nationalism shown through the language use of people either identifying with or attacking two clearly defined sub-nationalist groups, YouTube users that identify with the democratic party and users that identify with the republican party.

Trends across the Republican Videos

Trends across the Democratic and Republican Interaction

5 Conclusion