EXAMINING GENDERED LANGUAGE IN POLITICAL ADVERTISEMENTS

Ariana Barzinpour, Ben Gross, Isaac Hilton-VanOsdall, Mia Santomauro

Hypothesis

There is a correlation between the perceived gender of the language used in an ad and the demographics the ad reaches, and we will find a clear distinction in the language used in ads targeted towards males vs females. These language differences will be present in the vocabulary as well as the punctuation used.

Disclaimer: Our group does not believe in a dichotomous understanding of gender, but both the data available to us and the academic literature we found reflected a binary representation of gender and thus our analysis was mostly limited to this description of the world.

Motivation

Studies (such as this one) on the gendered nature of political ads motivate our project. These studies, along with the gender differences in speech patterns observed by Coulmas, warrant an investigation into the relationship between the findings of these studies and the increase in online political advertising.

Data Info and Collection

We collected our data from the facebook Ad Library via the Graph API on March 11th. Each ad had a few critical features: the unique ad identifier, the text of the ad, the link to the ad, and a breakdown by age and gender of the reach of the ad. We also collected information about the region where the ad ran and the page that ran it.

We cleaned and organized this data before storing it in four separate sql files (one for each candidate, as well as one for all candidates).

Figure 1

Challenges

Choosing the appropriate representation of our variables was challenging. For a given ad there were larger spikes within one gender at a certain age group and the opposite spike in gender at another age group. Taking the aggregate reach of each gender may not appropriately represent the nature of the ads' targeting. The formulae we used for gender reach and gender bias may not have been the ideal way to characterize the data. Possibly our biggest challenge comes from our inability to analyze the photos present in ads. Photos occupy a central component of political ads. While there is research on discussing gendered advertisements, that classification was outside the scope of this project.

Methodology

We calculated the following attributes for each ad: *linguistic features* of the ad text, *gender reach*, and *gender bias*.

The *linguistic features* are the number of nouns, adverbs, adjectives, and punctuation marks present in the text. We calculated these using spaCy, an NLP library.

To calculate *gender reach* we first aggregated the reach percentages provided by Facebook across all age groups by summing them. We calculated a final value for gender reach as the percentage of female reach divided by the sum of female and male reach.

To calculate *gender bias* we compared the words in the ad text, using Jaro-Winkler distance, to a bank of words provided by Gaucher et al. 2011. For each ad, we divided the number of female coded words by the number of total coded words in the ad. Ads with no coded words received a score of .5.

After preprocessing our data, we used a multiple regression with the *gender reach* as the dependent variable and the *gender bias*, and *linguistic features* (adverbs, punctuation, nouns, adjectives) as the independent variables. We then ran a t-test to determine if there was a significant variation between the reach of Trump ads and the reach of Biden ads. Following that test, we reran the multiple regression on only candidates as a variable.

Regression Results

	coef	std err	t	P>|t|
const	0.5079	0.013	37.922	0.000
nouns	0.0141	0.001	13.519	0.000
adverbs	0.0049	0.002	3.216	0.001
adjectives	−0.0038	0.002	−2.427	0.015
punctuation	0.0081	0.005	1.586	0.113
gendered lang	−0.1854	0.020	−9.436	0.000

Figure 2

Surprisingly, the gender bias term - though small - had a negative coefficient, which means that we saw essentially an inverse of the trend we expected. More feminine coded words appeared to make the reach of the ad less female. To get a better understanding of the data we created a visualization that showed the gender reach of an ad compared with the gender bias. The scatter plot on the right (Figure 3) demonstrates the inconclusive nature of the data with the mass of gray dots, ads with either equal parts maculine and feminine coded words or no coded words, closer to 100% female reach.

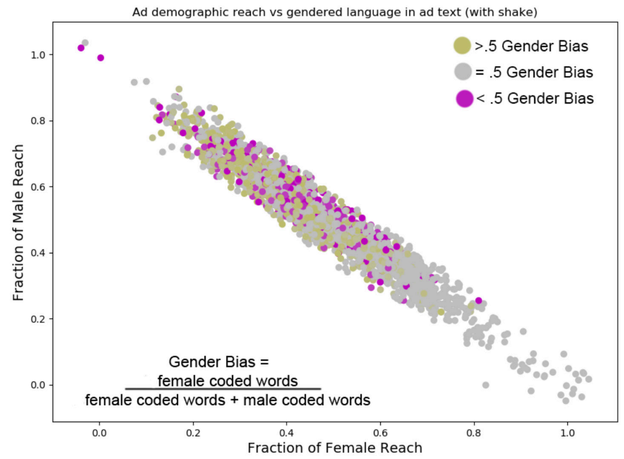

Figure 3

Several data points that had high female reach had no coded words or an equal number of coded words at both ends of the male female reach spectrum. This demonstrated that the correlation between our gender bias term and gender reach was not meaningful, and encouraged us to dig more deeply into the data. Figure 1 shows ads with high gender reach and low gender bias terms.

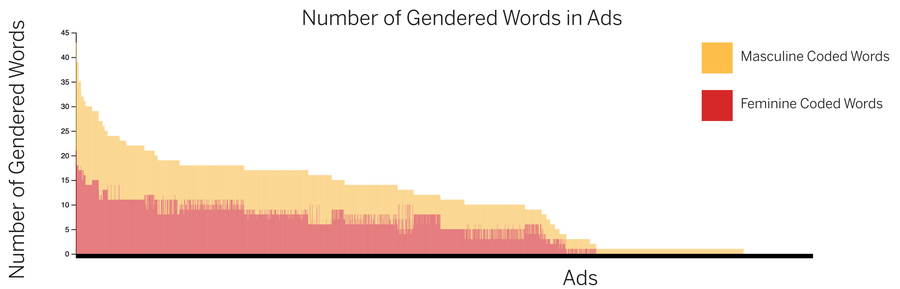

Figure 4

This graph shows that the majority of the ads contained fewer than 20 number of coded words, with a considerable number of ads containing no coded words. This may demonstrate both that the word list we used for determining gender bias was insufficient and that the text of the ad may not be the appropriate lens for understanding the gender targeting of an ad.

Conclusion and Significance

While our results were inconclusive for our broad hypothesis, we still think this research yields interesting results. There are numerous major lawsuits against Facebook for ad targeting. Though Facebook has officially removed the ability of ads to target based explicitly on gender, age, or race, other variables that correlate with those demographics can continue to reinforce the targeting that Facebook was supposed to remove. The ability to identify gendered political ads has significance both for ad targeting at large as well as understanding political messaging. If people have the ability to clearly identify gendered political advertisements, then they can better understand candidates' commitments to gendered issues and hold them accountable. Furthermore, we hope this research demonstrates an increased need in examining digital political advertising as it becomes a growing force in election outcomes.