A Linguistic Analysis of Twitter Tweets for Donald J. Trump

Ming Fong
University of California, Berkeley
Linguistics 55AC

December 12, 2020

1 Introduction

In the past decade, politicians, business leaders, and celebrities have increasingly turned to social media to reach a wider audience. Of the many platforms available, Twitter has become one of the most popular social media outlets for those looking to reach a large follower base. Among Twitter's most influential users is President Donald J. Trump. President Trump's use of Twitter has ranged from attacks against political opponents to updates on actions he has taken as President of the United States. The President's frequent use of Twitter is unprecedented among American leaders and has led to criticism from opponents and praise from supporters. In this paper, we will explore linguistic trends of President Trump's Twitter posts with respect to the November 3, 2020 US Presidential election.

2 Methodology

We will collect Twitter Tweet text and the associated metadata from Tweets sent by from the account of Donald J. Trump (@realDonaldTrump). Using that data, we can perform analyses on the various types of data we collect.

2.1 Data Collection

Our data will be directly sourced from Twitter using the Twitter Developer API. We will read the tweets into a Pandas Dataframe with the columns

$$"id","date","text"$$

Due to limitations with the Twitter Developer API, our dataset is capped at the 3200 most recent tweets. The timeframe of our data ranges from 14 September, 2020 to 11 December, 2020.

2.2 Data Analysis

Before performing any analysis of the tweet text data, we must first clean the text. Any non-alphanumeric characters and elements such as links are removed. Additionally, all retweets are removed since their text were not written by President Trump.

Using the text of the tweets, we can assign a sentiment score to each tweet. We do this using *textblob*, a Python module for text processing.