

Using Text Analysis to Describe German and English Tweets and News Articles

Tiffany Cheng

April 27, 2018

Introduction

(Include research questions)

Data Collection Methods

(Twitter and scraping news articles)

Location of News Companies

(Names of the five German companies)



Figure 1: Map of Germany

(Names of the six US companies)



Figure 2: Map of the United States

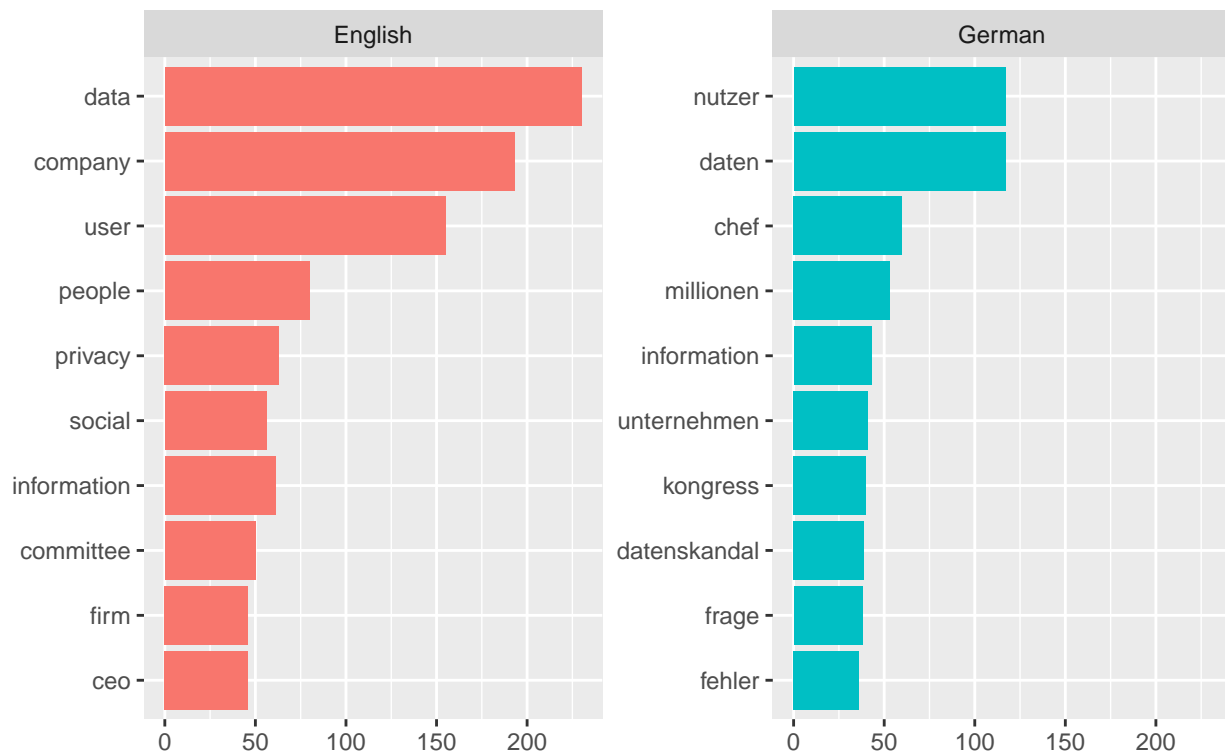
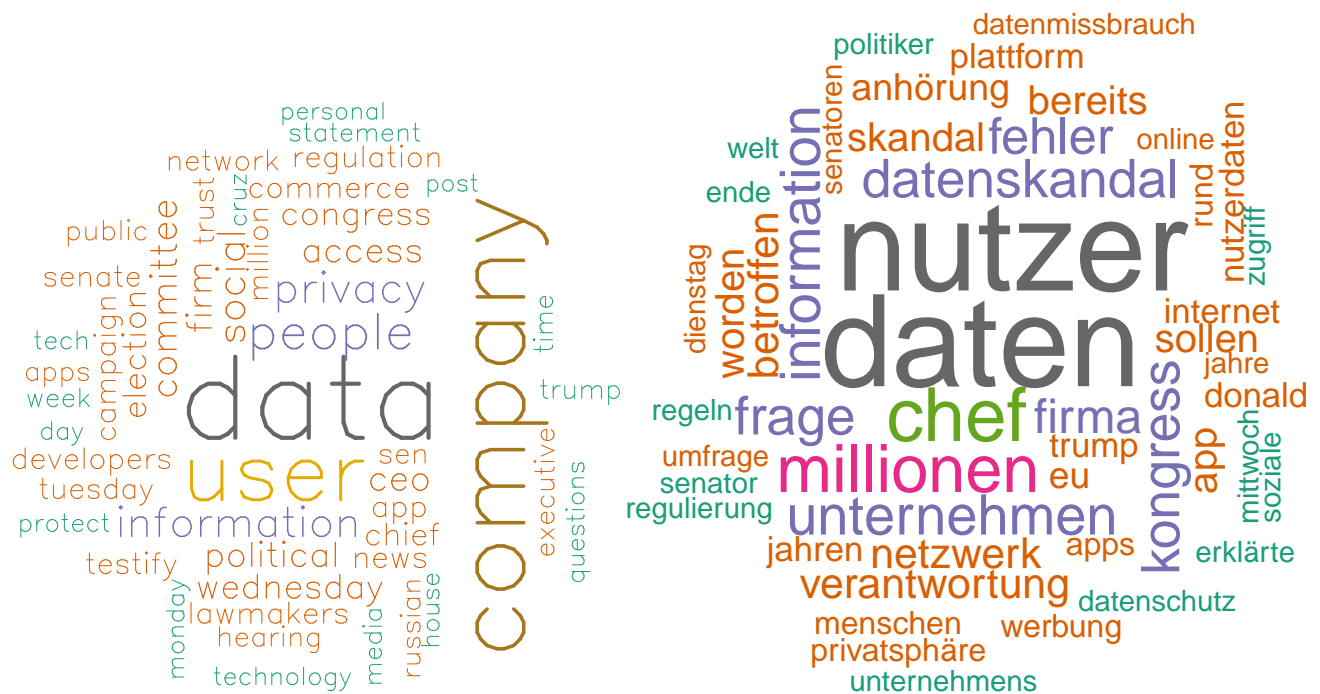


Figure 3: Top 10 Most Commonly Used Words in News Articles

Word Frequency Analysis for News Articles



Word Frequency Analysis for Twitter

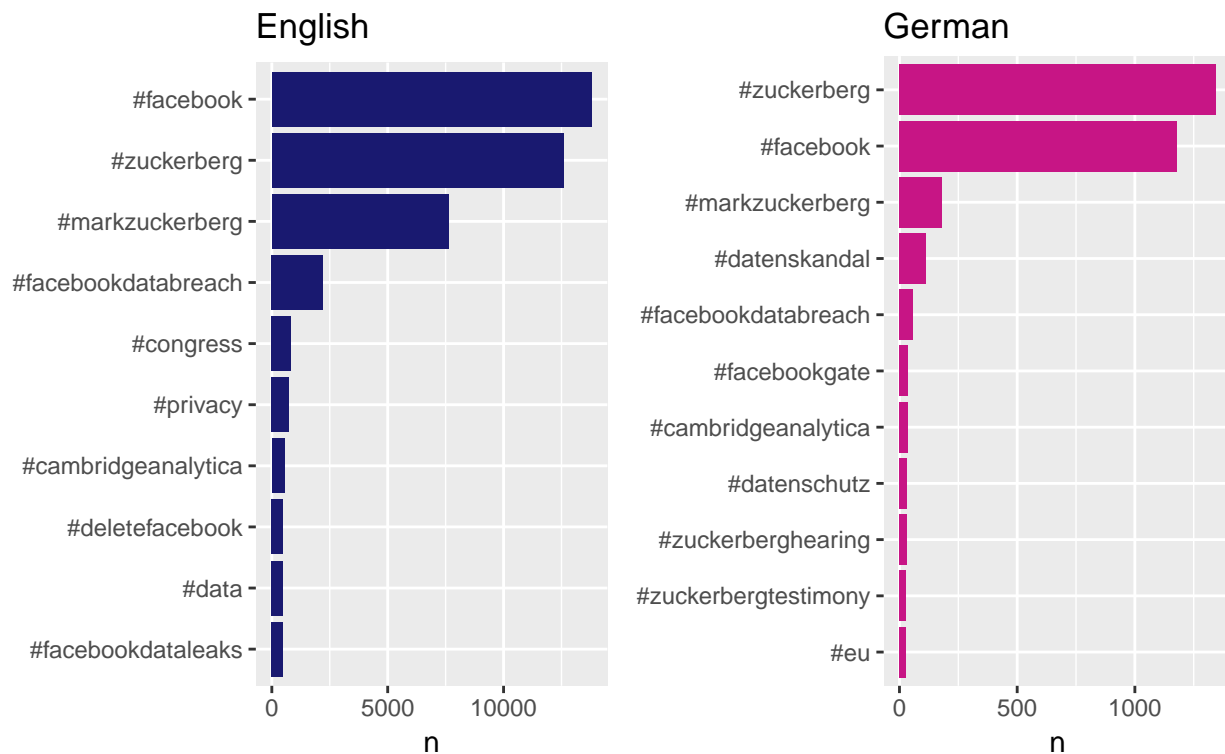


Figure 4: Top 10 Most Commonly Used Hashtags in Tweets

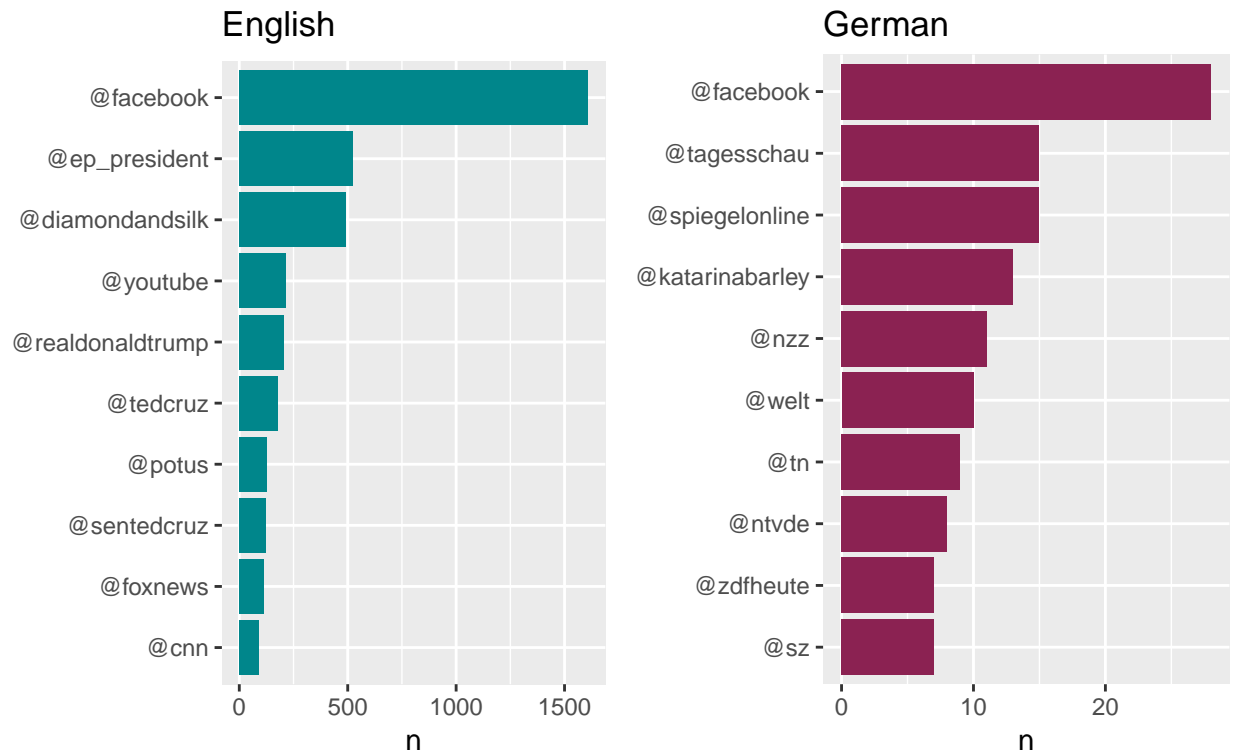


Figure 5: Top 10 Most Commonly Used Mentions in Tweets

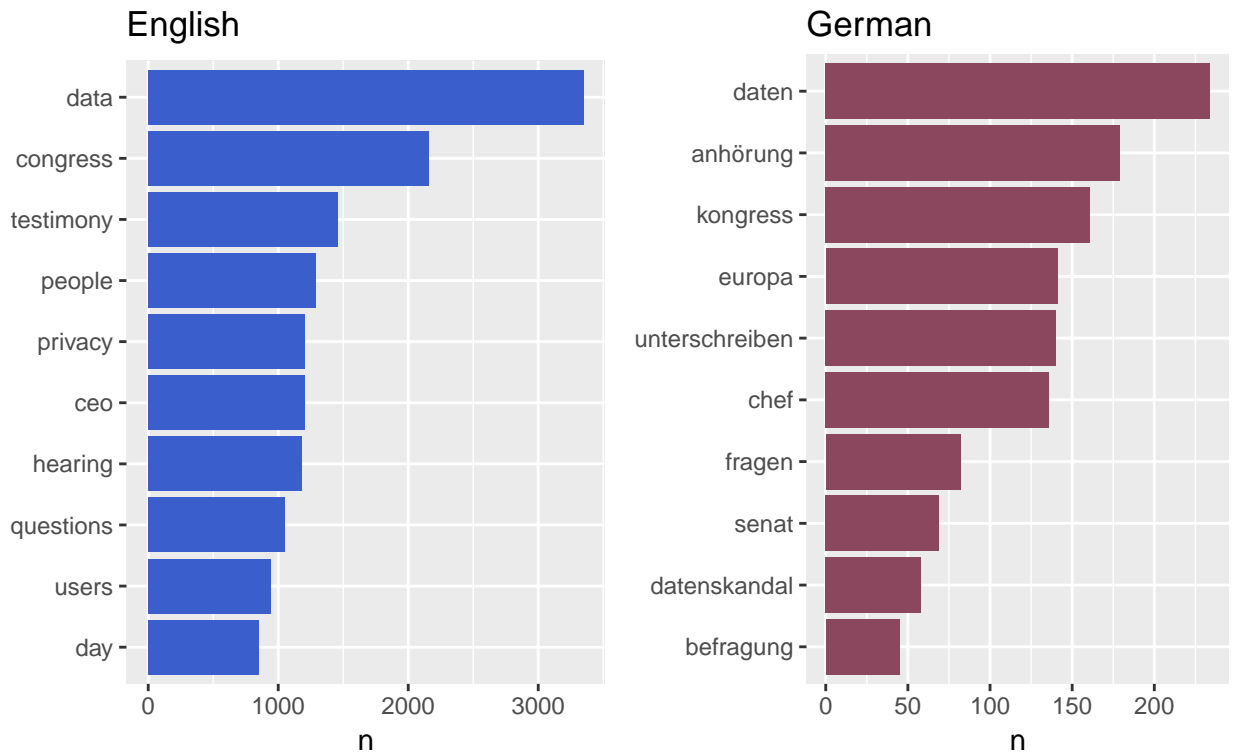


Figure 6: Top 10 Most Commonly Used Words in Tweets

Comparisons Amongst the Same Language

Comparing Word Usage Between Twitter and News Articles (English)

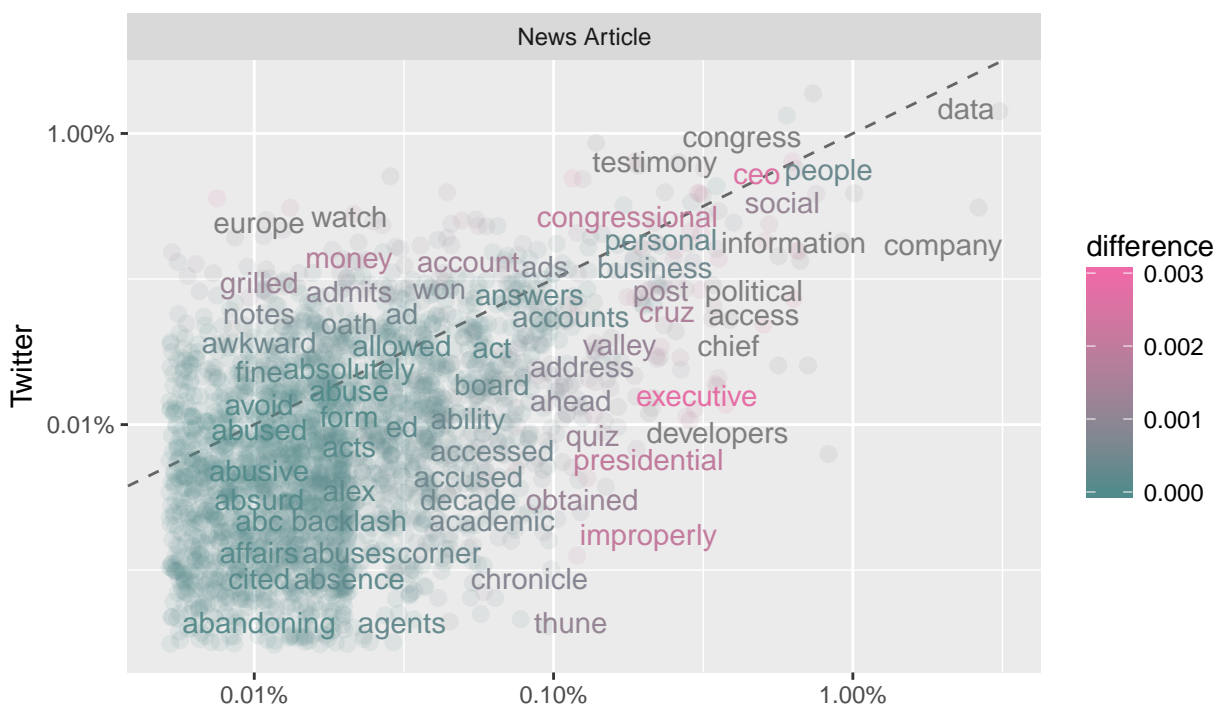


Figure 7: Comparing Word Usage Amongst English News Articles and Tweets

Correlation from the correlation test is 0.6538692.

Correlation from the correlation test is 0.594824.

Sentiment Analysis for News Articles

For English:

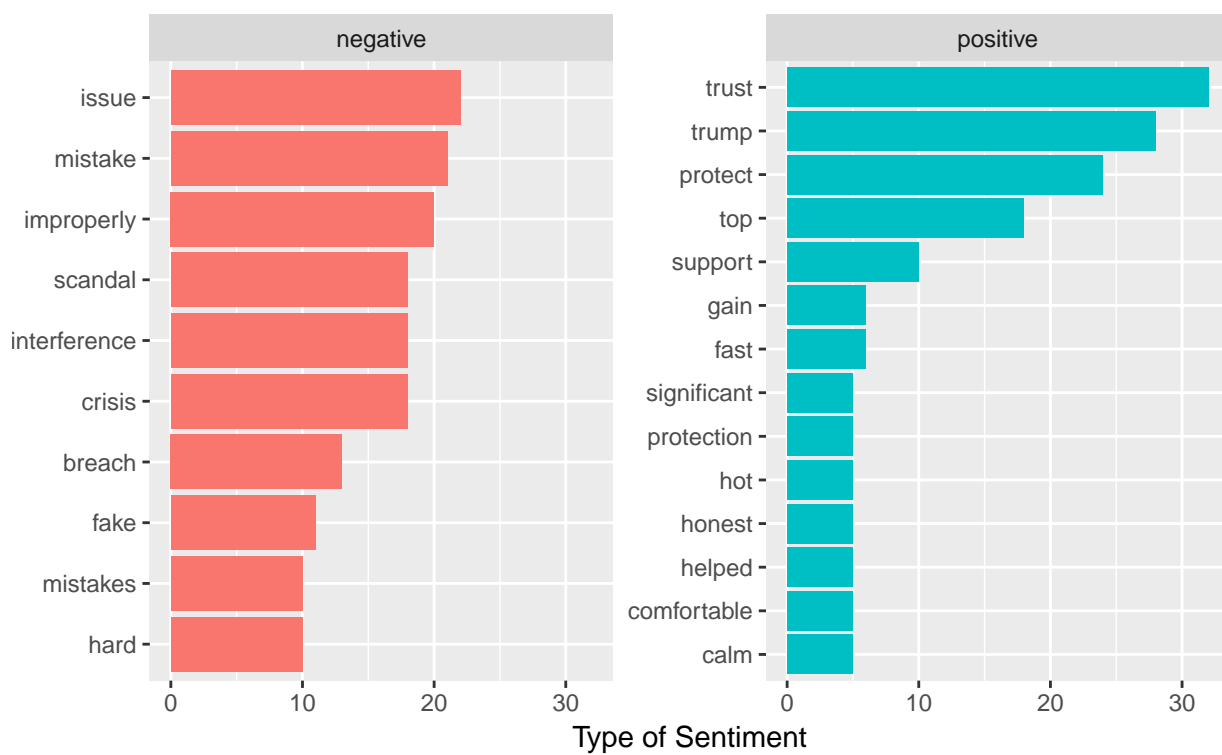
Using **nrc** sentiments lexicon.

A word cloud of terms related to the 2016 US Presidential election. The words are arranged in a circular pattern. The most prominent words, shown in larger fonts, are 'hearing' (orange), 'criticism' (pink), 'violation' (orange), 'deserve' (orange), 'accused' (orange), 'challenge' (blue), 'bias' (green), 'hot' (orange), 'honest' (orange), 'politics' (orange), 'intense' (blue), 'powerful' (blue), 'hate' (blue), 'interference' (pink), 'crisis' (pink), 'breach' (green), 'influence' (green), 'controversy' (green), 'fake' (green), 'misuse' (blue), 'quiz' (green), 'scandal' (pink), 'mistake' (red), 'restriction' (blue), 'government' (green), 'disinformation' (blue), and 'confront' (blue). The words are in various colors including orange, pink, green, blue, and red.



Using **bing** sentiment lexicon results in following table and figure:

negative	positive	sentiment
606	396	-210



negative

interference
improperly
mistakes
scandal
mistake
hard
crisis
issue
issues
controversy
breach
fast
trust
fake
support
significant
trump
top
protect
gain

positive

Using `afinn` sentiment lexicon results in a sentiment score of -262.

For German:

negative	positive	sentiment
247	368	121



negative



positive

Using sentiment assigned to each word results in a sentiment score of -56.43515.

Sentiment Analysis for Twitter Tweets

For English: