

Harris McCullers
Leah Herman
MLLC 188
May 7th, 2019

Sexism Semantics: Using Natural Language Processing to Illustrate Sexist

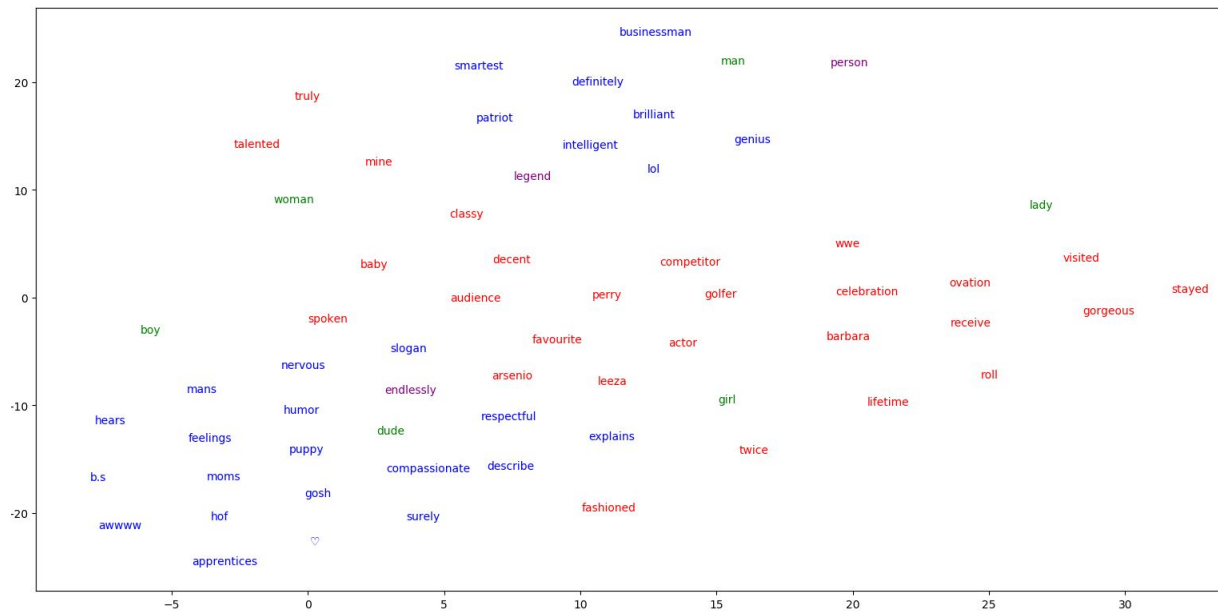
Microaggressions in Semantics.

Rationale Statement

Microaggressions are something that almost everybody has heard about, and many people make a conscious effort in their day to day lives to avoid them. However, one of the most common confusion surrounding microaggressions is simply: what is and isn't one? As a computer scientist, I found myself in a valuable position. Firstly, I work in a field and industry that is struggling with issues of diversity, and secondly I have the computational tools to attempt to better understand microaggressions. It was for these reasons I decided that I should create a machine learning model to help demonstrate microaggressions, specifically against women. In order to do this, I employed an algorithm called Word2Vec using gensim in python. Developed by Google, the algorithm can be used to understand the semantic meaning of language within a training set (otherwise known as a very large text file). I used this algorithm in the core of my code to create a program that could learn from a sample text and then visualize words with similar semantic meanings to male and female specific words. This is done by reducing the vector space of each word from either 50 or 100 down to 2, and plotting those as coordinates on a 2D graph.

Words on these "word clouds" are grouped by similarity, with more semantically similar words appearing closer together. Furthermore, words in blue are more male associated, words in red are more female associated, and purple words share association with both. The first set of

images were trained on a large database I created of nearly all of Trump's tweets. One such word cloud is shown below.



In this graph, semantics hinting towards microaggressions are apparent. Take for example, the proximity of “woman” and “talented”, as opposed to “man” and words like “brilliant, genius, and intelligent”. This seems to suggest an attitude that, where men are smart, women are lucky or gifted. Notice also the proximity between to the word “woman” and “mine”. Lastly, notice how “woman” is closely associated with “baby, while “man” is more closely associated with “businessman”. For the most part, we already know that Trump’s attitude towards women is problematic. However, by using this model we can quantify and easily illustrate exactly what problematic suggestions he makes linguistically. Feel free to look at the rest of the photos for the trump tweet dataset (found in Photos/TrumpTweets) for more.

Within the folder of images, there are some from a few different datasets that I would like to break down. In the folder “Photos/Text8” there are a few images of word clouds based on a nlp (natural language processing) dataset simply called text8, which is a fairly neutral and very large set of text from all sorts of places like literature, social media, and articles. This word cloud is less reflective of any one person's bias and points towards larger societal bias. In the folder “Photos/TrumpSpeeches” There are photos of some word clouds created using a large collection of trump speeches. This word cloud not only includes male and female words but also the names of male and female candidates. There are some apparent microaggressions, but overall this one was less bad than his tweets. I believe that this is because his speeches are proofread and polished while his tweets represent his more natural and colloquial speech. Lastly, in the folder “Photos/novels” there are word clouds from datasets created using classic novels. These, though interesting, are not as relevant due to their age.

Overall, this project has taught me a lot about microaggressions in a more specific way. By reducing words to these vectors, it is easy to quantify these microaggressions and discover where they are hiding and what they say about the sample text. However, data alone is not the only important factor. By understanding *where* these models come from and what text they are sampling, we are able to connect our personal experiences, and those we have learned about in class, to the results. Only by doing this, can we discover not only what the microaggressions are, but where they came from. It is for this reason that the word clouds are left largely untouched, without “irrelevant” words removed; so that each viewer may make their own discoveries and inferences.