Introduction

Word embeddings have been effective in understanding stereotypes and cultural trends, but they haven't been used extensively on corpora prior to the 18th century. Our research uses word embeddings to analyze words related to consumerism within the Early Modern period when the market economy began to dramatically expand in England.

Methodology

1. Data Cleaning

- Scrape data from EEBO TCP Phase I.
- Merge spelling variations with some help from the University of Michigan's Middle English Compendium.

2. Word Embedding Models

- Generate Word2Vec Word Embedding Models. Find cosine similarity between word vectors of interest over time.

\[
\cos(\theta) = \frac{x \cdot y}{||x|| ||y||}
\]

- Validate Models via analogy tests and bootstrapping resampling to evaluate accuracy and consistency.

\[
W = (\frac{\text{Queen} - \text{King}}{\text{Man} - \text{Woman}}) + \cdots
\]

- Create artificial “cultural axes” to track words’ semantic relationships with gender, race, and social class.

Birth of ‘Economic’ Consumption

Consume transforms from a “medical” to an “economic” idea in the beginning of the 17th century. Consumption is seen as increasingly “luxurious”, and its association with “disease” continuously decreases.

Gender Axis Insights

1. “Consume,” “consumption,” and “greed” are mostly neutral overall but become more gynocentric over time, possibly suggesting that women during this period are increasingly being seen as primary consumers in comparison to men.
2. The word “possession” remains consistently androcentric over time, possibly due to dominantly male land-ownership.

Intersectional Axis Insights

1. Jewish-signifying words become more related to masculinity and less to consumption during 1495-1544.
2. There is a sudden spike in the relationship between Jewish-signifying words and words related to the resettlement of Jews in England.

Why did we look at these axes? Antisemitism is often expressed in terms of gender and ideas often related to consumerism.

Next Steps

1. Expand our analysis to EEBO TCP II, a set of 30,000+ more texts being released on Jan 1, 2021.
2. Enhance our word embedding’s accuracy by improving our text-cleaning process, increasing bootstrap resample sizes, and possibly using stop words to reduce clutter.
3. Compare the efficacy of different Word Embedding structures, like GloVe, BERT, and fastText

References & Acknowledgements

Find our code, references, and more visualizations at: https://github.com/albertyusun/For-the-Love-of-Greed

This work was supported by funding from the Data+ program at the Duke Rhodes Information Initiative. Many thanks to Dr. Nikhil Garg at Stanford for help with word embedding models; Paul Bendich and Greg Herschlag at Data+ for research assistance; and Mark Delong for setting up a computer through the Duke Research Computer Clusters for our research.