Annie Williams (maw3as@virginia.edu)
DS 5001
2 May 2021

The *Iliad* and the *Odyssey* are some of the oldest surviving works in Western literature. While these epics are thought to have been written by Homer around 800 BCE, there are still many unanswered questions surrounding who Homer was, and when these works first appeared. Milman Parry, an American Classicist, made the widely accepted claim that these works were originally passed down as a part of oral tradition, and not written literature (Worrall). The *Iliad* and the *Odyssey* have essentially become a centuries long game of telephone, passing through multiple translators and oral performers. Given this background, this project aims to use exploratory text analysis techniques to compare the differences, if any, between two different translations of the *Iliad* and the *Odyssey.*

This project compares the George Chapman (1559-1634) and Samuel Butler (1835-1902) translations of the *Iliad* and the *Odyssey*. Chapman is one of the first authors to translate these works into English, and points out in the introductory letter preceding his translation that he "rejects verbatim translation on the grounds of inelegance" and takes poetic license when necessary ("The Englished Homer"). About two centuries later, Samuel Butler published his translations with a note in the forward that "there would have been less controversy about the proper method of Homeric translation, if critics had recognised that the question is a purely relative one, that of Homer there can be no final translation. The taste and the literary habits of each age demand different qualities in poetry, and therefore a different sort of rendering of Homer."

All of the texts used in this project were downloaded from Project Gutenberg, then tokenized and annotated using the NLTK library. The corpus was then converted into a vector representation and various exploratory text analysis methods and visualizations were applied. The OHCO for this corpus is broken down by document id (each individual translation), book number (similar to chapters, each translation is broken down into individual "books"), stanza number, line number, and token number. The stanza level is particularly interesting because it is completely up to the translator to decide where to split the text into individual chunks.

Principal components analysis was the first unsupervised method used to look for differences between Butler's translation and Chapman's translation. Figures 1 and 2 show the plots of PCA. In Figure one, principal components 0 and 1 show a clear division between Butler and Chapman, while Chapman is much more densely clustered while Butler is more spread. Figure 2 tells a similar story by clustering Chapman's translations tightly together while Butler appears more spread out. A major difference here is that these components place Chapman in the center of the Butler cluster.

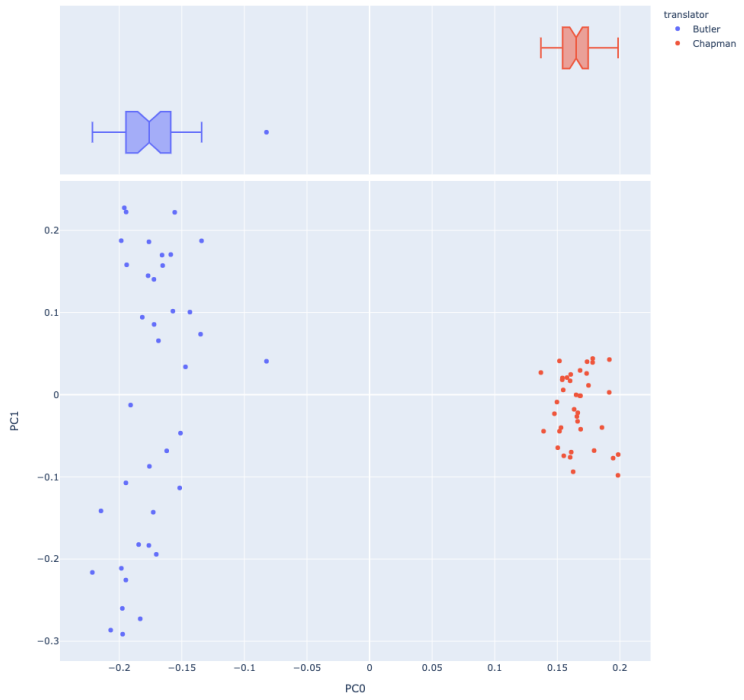

Figure 1: PCA components 0 and 1

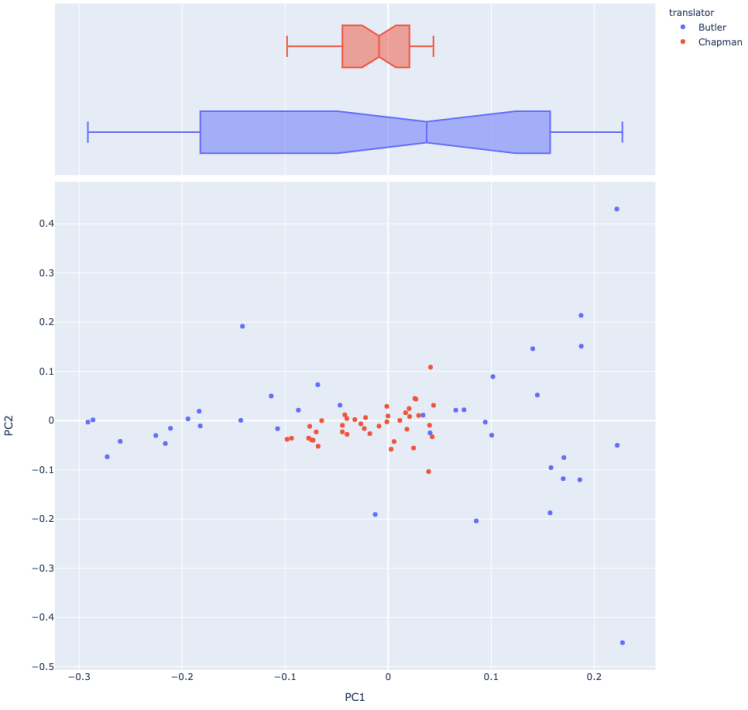

Figure 2: PCA components 2 and 3

Next, latent Dirichlet allocation was used to further understand how these texts cluster and what the underlying structure of the text looks like. After organizing topics by translator, Butler and Chapman fell into two different dominant topics. Butler fell into topic 17 ('son', 'ships', 'spear', 'man', 'men','horses', 'battle', 'hand', 'fight', 'armour'), while Chapman fell into topic 25 ('men', 'son', 'man', 'hand', 'thou', 'thee', 'death', 'king', 'horse', 'mind'). Thematically, these topics feel very similar, and even share some of the same words ('man', 'men', 'son', 'hand', 'horse'). The difference that stands out the most between these two topics is 'death' appearing in topic 25. I was surprised that this was not a more common word across all topics considering that the epics are about war.

The word2vec library was also used to create word embeddings. Two separate models were created, one for Butler and one for Chapman. The bag was changed to stanza in order to get a better sense of differences across chunks of text on a lower level. The word embeddings were also plotted using the sklearn tSNE library, which can be seen in Figures 3 and 4. Outside of the obviously different shapes of the two plots, it is difficult to immediately pick up on any distinct clusters within the plots themselves. In general, the clusters in each tSNE plot seem to be action words that are grouped together, or words that describe a battle. Similar to in the PCA plot, the visualization for Butler is again much more spread out than Chapman's visualization. Perhaps it isn't fair to compare the PCA and tSNE plots, but it would be interesting to look more into this idea of Butler's work being more spread out in the future.

Finally, sentiment analysis was applied to each translation. The sentiment analysis plots can be seen in Figure 5 and Figure 6, where each tuple on the x-axis represents document id and book number. The sentiment analysis plots turned out as I expected they would: very similar but not exactly the same. Across both graphs, the polarity line was quite far below all of the other lines. This makes sense to me given my understanding of the *Iliad* and the *Odyssey* as two dramatic poems about the turmoil of the Trojan War. However, I was surprised that the polarity line on the Butler plot peaks, intersecting the other plots, while the polarity line on the Chapman plot never reaches the other lines. This makes me wonder if the Chapman translation is in some ways more dramatic, or the emotions are heightened further than in the Butler translation.

Applying exploratory text analysis techniques to two different translations of the *Iliad* and the *Odyssey* offered some insights into their similarities and differences. Looking back at this project, there are a few things I would have done differently. More time could have been taken to remove stop words properly from the Chapman text. In retrospect, I feel as though the language is different enough that stopword removal should have been handled differently - I noticed in some of the later analysis that words like "thee" and "thou" were coming through. It might also be interesting in the future to repeat some of this analysis on different OHCO levels in order to further understand the differences between stanza breaks across translations. Additionally, it would be insightful to explore just the differences between the *Iliad* and the *Odyssey* themselves.

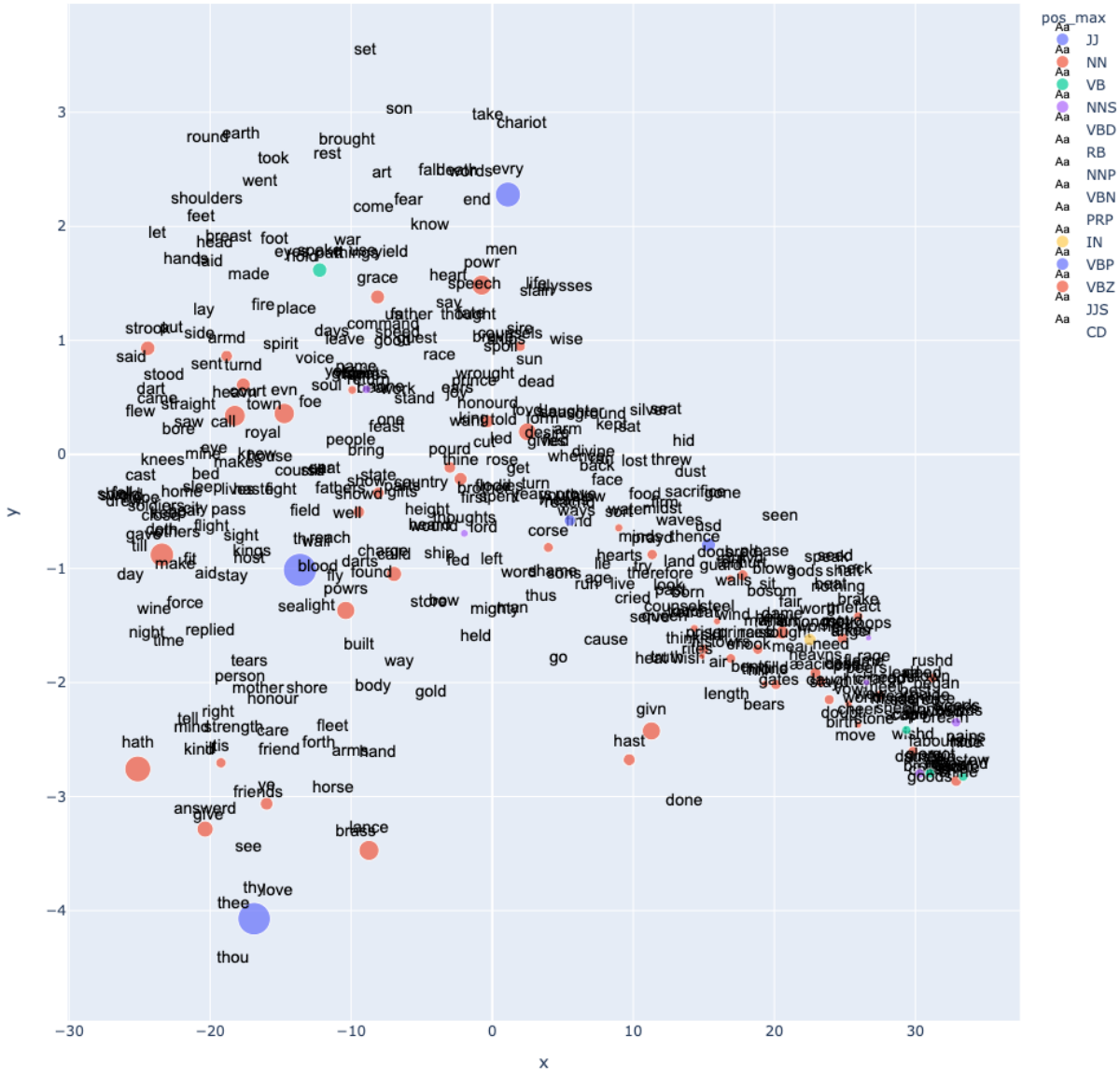

Figure 3: Chapman tSNE

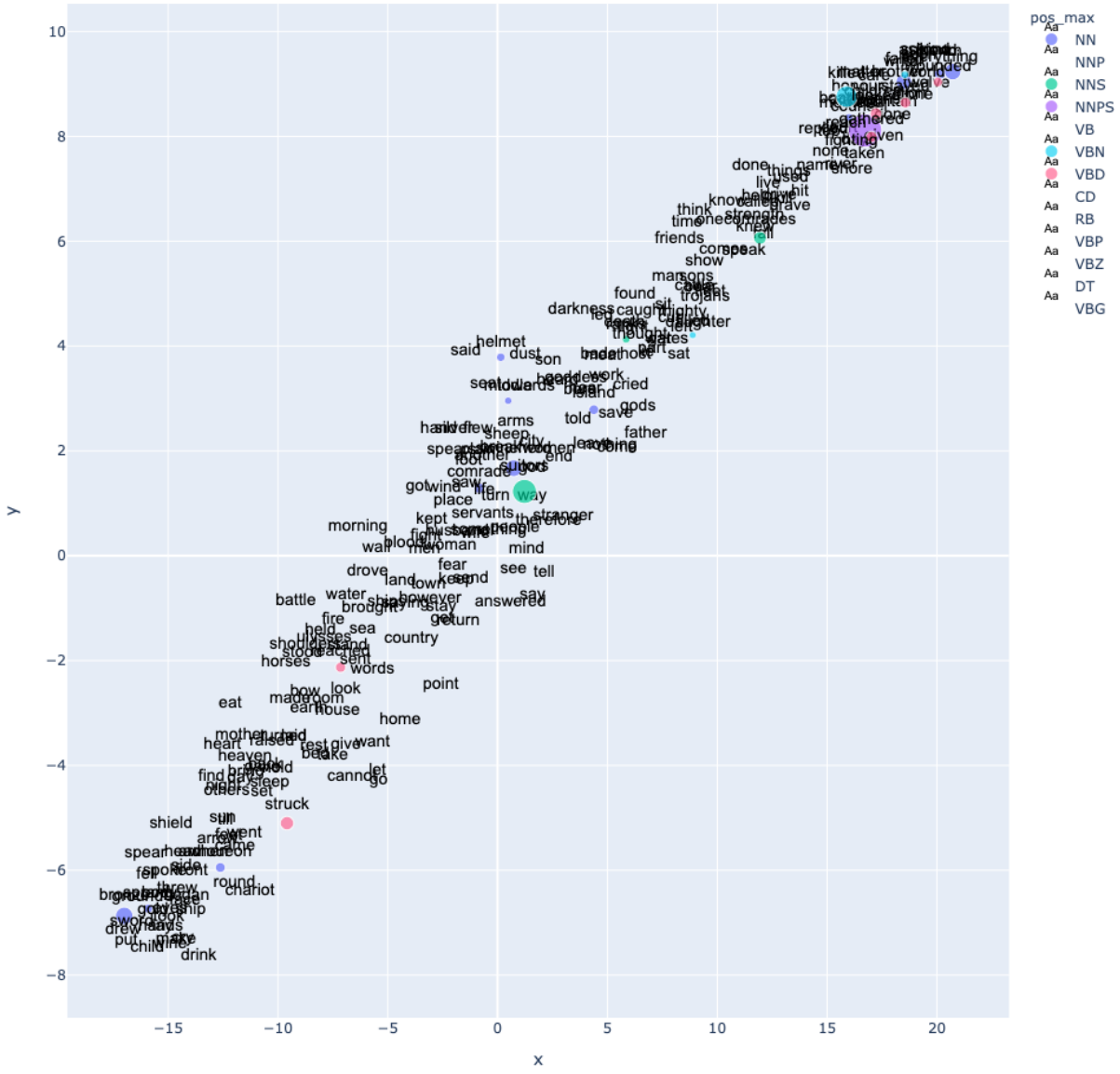

Figure 4: Butler tSNE

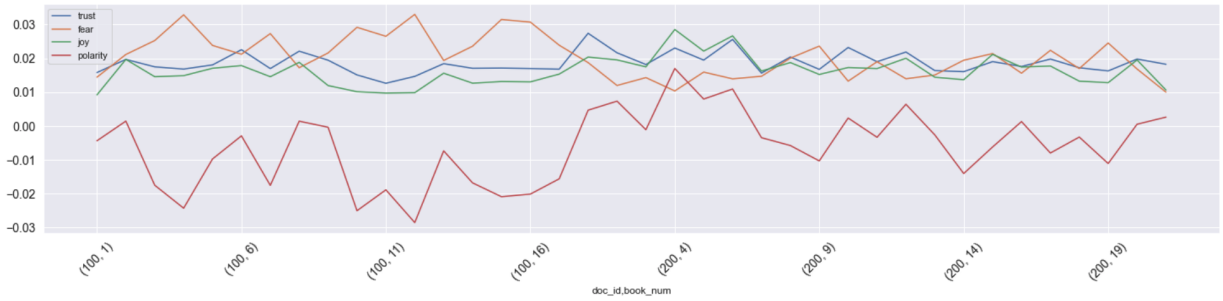

Figure 5: Butler sentiment analysis

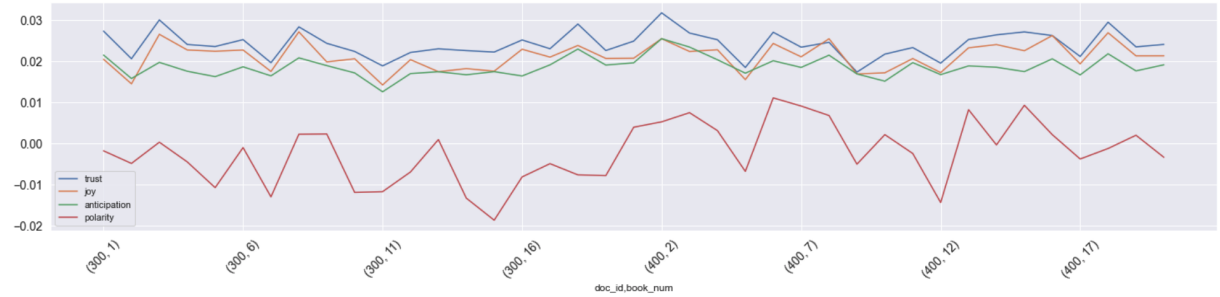

Figure 6: Chapman sentiment analysis

Works Cited

"The Englished Homer." *Omeka RSS*, apps.lib.umich.edu/online-exhibits/exhibits/show/translating-homer--from-papyri/the-english-homer#john-dryden.

"Homer - Modern Inferences." *Encyclopædia Britannica*, Encyclopædia Britannica, Inc., www.britannica.com/biography/Homer-Greek-poet/Modern-inferences#ref21037.

Worrall, Simon. "Author Says a Whole Culture-Not a Single 'Homer'-Wrote 'Iliad,' 'Odyssey'." *Adventure*, National Geographic, 3 May 2021, www.nationalgeographic.com/adventure/article/150104-homer-iliad-odyssey-greece-book-talk-travel-world.