

Lost in Translation: Using Sentiment Analysis to Analyze Translations of Homer's *Odyssey*

Erin Shaheen

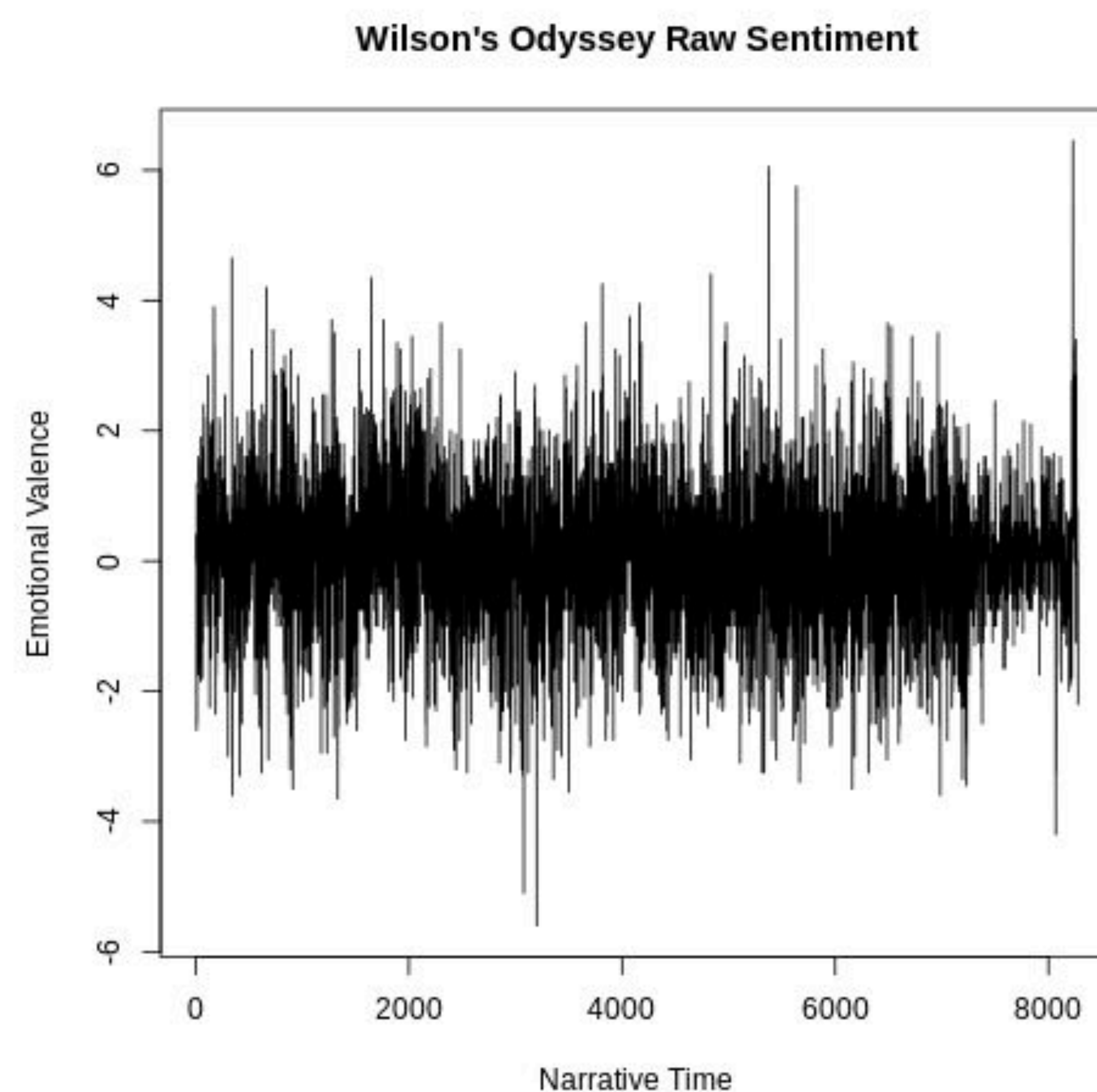
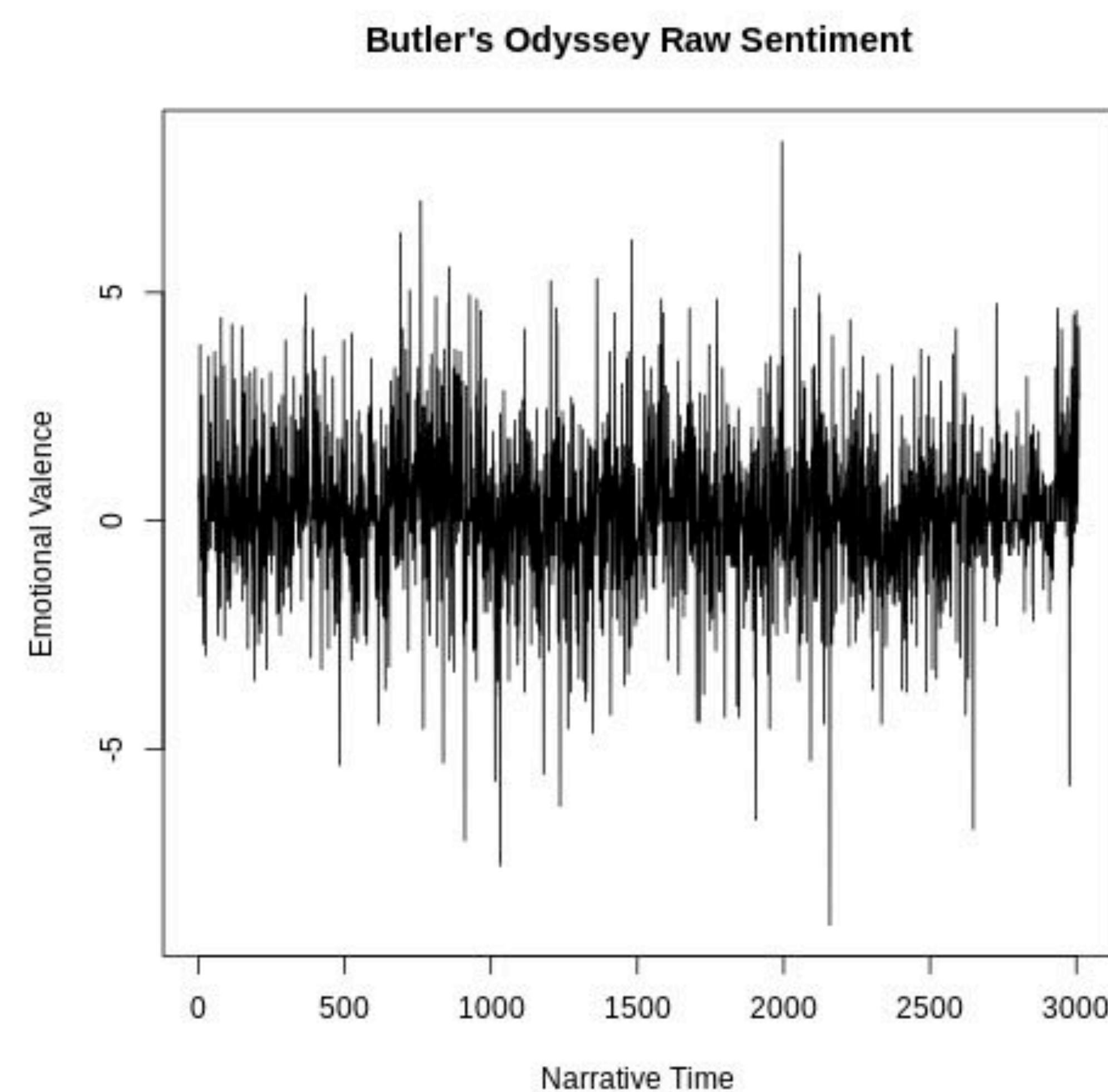
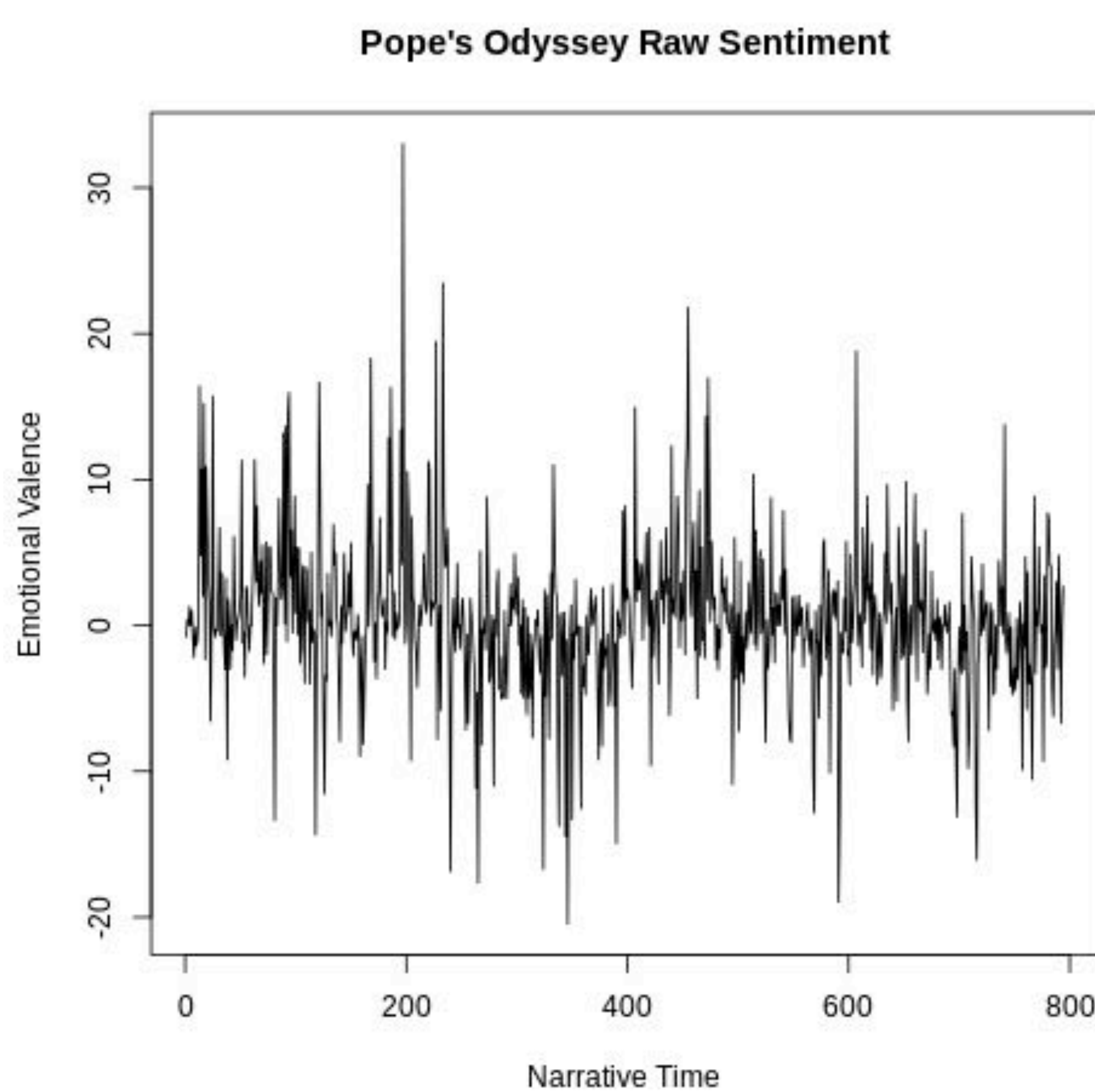
Introduction

For my final project in AI for the Humanities, I knew that I wanted to examine the creative role of translators in classic literature. I ran three different translations of Homer's *The Odyssey* through Syuzhet.R, a sentiment analysis tool that has grown increasingly popular in the Digital Humanities field. Syuzhet tracks the emotional arc (also called "emotional valence") of a text by giving each word a different score. For example, on a scale of -1 to 1, "terrible" might score -1 and "amazing" might score 1. A more neutral word like "okay" would score closer to 0.

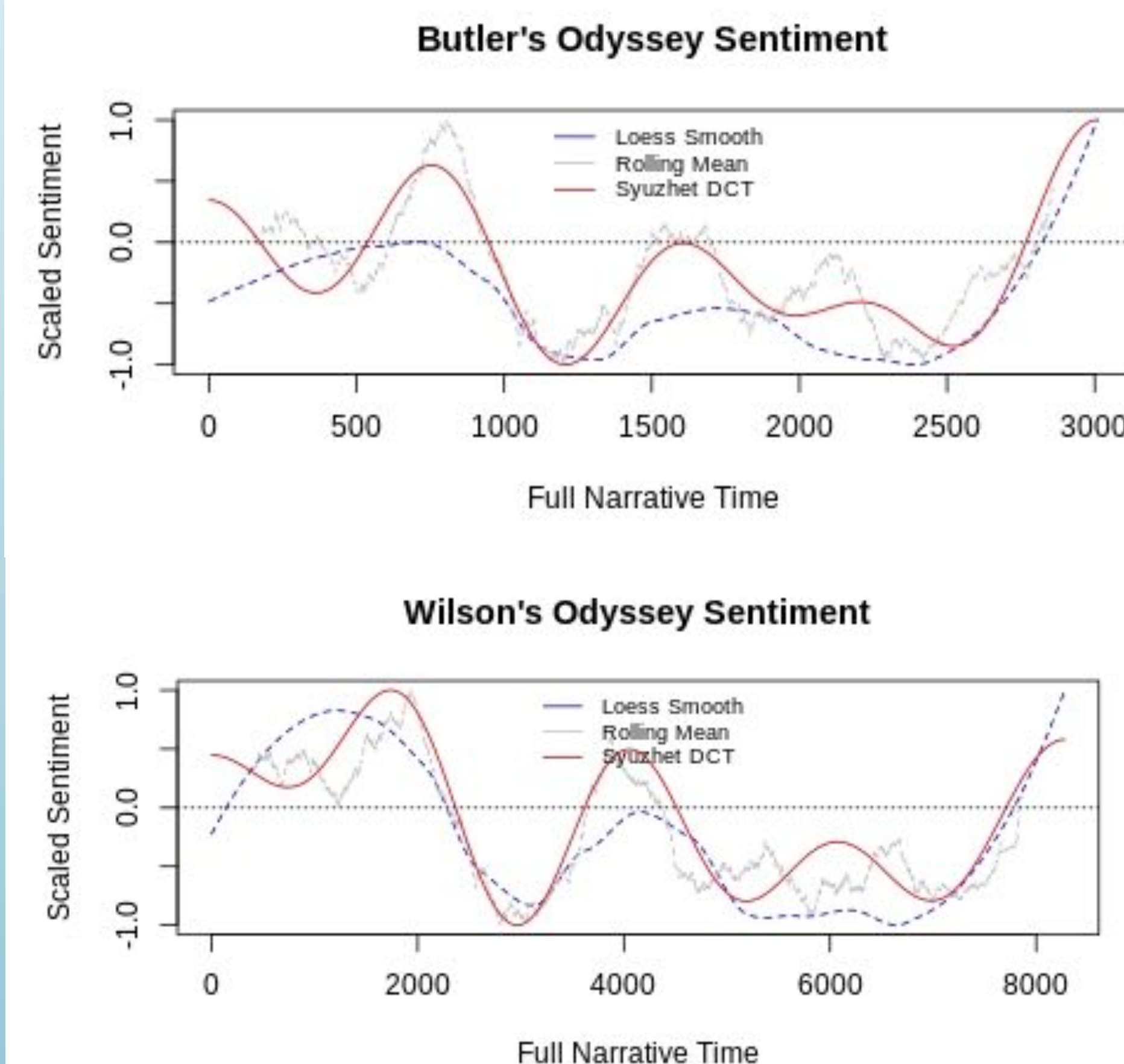
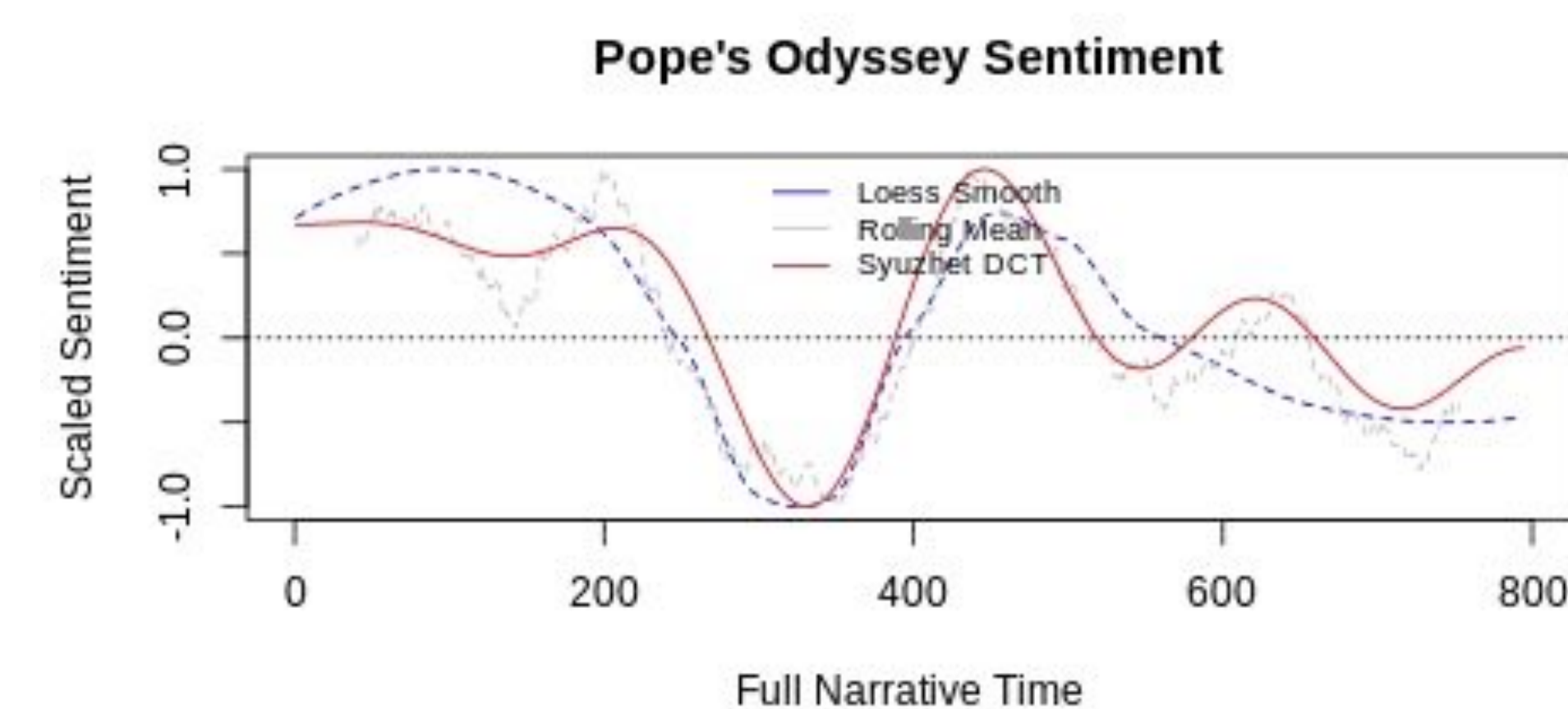
I chose translations by Alexander Pope, Samuel Butler, and Emily Wilson, published in 1725, 1900, and 2017 respectively. I wanted to explore how the translators' working in different time periods and having different life experiences might affect their interpretations of the text.

Methodology

The first thing I did was run the three translations through Syuzhet and look at the raw sentiment analysis. As you can see in the following graphs, Pope's translation is by far the least noisy, followed by Butler's, and then by Wilson's.



Though these graphs are certainly useful, the amount of noise can make it difficult to get accurate results. Thankfully, Syuzhet offers an option that simplifies the graphs, and allows for easier reading.



Once I had made the simplified graphs, I was able to pinpoint what are known as "cruxes," the specific high and low points of emotional valence. I focused on the Syuzhet DCT line, marked in red, for all three graphs, both for clarity's sake and for the fact that it appears to be closest to an average.

Interestingly enough, all three of the texts had approximately the same lowest point—in Book 11, when Odysseus sees the ghosts of dead comrades and his recently deceased mother in the House of Hades.

I found slightly more variation in the highest points. While both Butler and Wilson's translations mark the highest point as during Odysseus' time spent visiting the home of King Alcinoos, Pope marks it as when Odysseus visits the swineherd Eumaeus.

Close inspection of the above graphs reveals that all three of them dip slightly towards the beginning of the text, but only Butler's translation drops below the neutral line. When I pinpointed the cruxes, I found that while Butler and Wilson mark this low point in Book 4, where Telemachus visits Menelaus, Pope, who has the shallowest curve, marks it in Book 6, where Odysseus leaves Calypso.

Results

My initial hypothesis that the sentiment analysis would shift based on translator was correct—but only to a certain extent. While you can see from the graphs that different translations had different highs and lows, the overall arcs remained relatively similar. All three of the translations had roughly the same point of lowest valence, and while the points of highest valence had more variation, they all involved the theme of hospitality. Pope's translation had the most variations, with the most positive sentiment appearing at a different crux point (the visit to Eumaeus) and with variations at the beginning of the text.

Conclusion

Syuzhet proved to be extremely useful in tracing the differences between translators. Alexander Pope's work has different crux points than Butler and Wilson's, which correlates with how that his graph contains significantly less noise than the others. It's worth noting that his work was originally published in the early sixteenth century, while Butler worked in the eighteenth and nineteenth, and Wilson in the twenty-first.

Samuel Butler's translation has the most negative valence, being the only one to dip below average at the start of the text, and remaining there for the longest amount of narrative time.

I initially selected Emily Wilson's translation because of how recent it was, and the reviews I read online lead me to believe that it drastically differed from previous translations. In terms of Sentiment Analysis, Wilson's translation actually provided the closest thing I had to a middle ground between Butler and Pope. Of course, I only looked at the texts through the perspective of Sentiment Analysis, and I only chose three translations out of a large body of canonical work.

If I had the chance to do this project again, with the knowledge I have now, and significantly more time on my hands, I would do a few things differently. I would start by adding in other translations, to get a larger basis for comparison. Then I'd supplement my work with a variety of other tools, like Voyant, which includes features that note the frequency of words and their connections to each other, and could potentially shed light on the overall scores. I'd also go into a more in depth analysis of the texts themselves, particularly at the crux points I mentioned above, offering a human perspective to compare to the digital ones.

Acknowledgements

I would like to thank Professor Chun for his help with coding, Professor Elkins for her recommendations and her help with Syuzhet analysis, and Miles Shebar for helping me with the technical side of working with Wilson's translation.

References

- *The Odyssey* by Homer, translated by Samuel Butler
- *The Odyssey* by Homer, translated by Alexander Pope
- *The Odyssey* by Homer, translated by Emily Wilson
- Pope and Butler's translations came from Project Gutenberg
- <https://github.com/shaheene/OdysseyTrans>