Analysis outline

Our initial hypothesis was that the language surrounding pain and torture would progressively get more grim, moving through each circle of hell. Our graph translates our initial hypothesis and displays the trend where the amount of torture words used increases as the circles of hell get more tortuous. However, given the lack of complete and conclusive data, this trend is only a proof of concept and no formal conclusions can be drawn from it.

The graph displays the number of torture words divided by the total number of words in each circle, reflecting the percent/ratio of torture words in the particular circle. However, some circles do not have any torture words and some circles have incomplete data. The Y axis (vertical axis) represents the amount of torture words as a percent of the total words in the circle and the X axis (horizontal axis) represents the circles of hell.

[graph]

We further explored the use of torture words by implementing Wordnet, a large lexical database of English. Parts of speech (nouns, verbs, adjectives etc.) are grouped in synonyms, called synsets. We used these synsets to explore the richness of the expression of torture in the text. As we already had the torture words marked up, we used wordnet with the help of some python to find the appropriate synset for each torture word. We tagged the torture word elements with their individual synsets and number of distinct synsets. Then we performed an XSLT transformation to gain the “tortuous ratio.” By dividing the count of distinct synsets in the text by the count of torture words, we acquired the ratio - .854. A higher ratio, closer to 1, indicating a diverse and larger variety in vocabulary.