**Title:** Positive movie reviews are more negative than vice versa

**Data:** Movie reviews (pos and neg)

**Description of method:** I want to look at the positive and negative classified movie reviews and see if positive movie reviews really contain more positive sentences than negative reviews and vice versa.

I am planning on using sentiment classification by Python NLTK. See: <http://text-processing.com/demo/sentiment/> for a demo. If this would not work, I could use another sentiment classifier.

**Alternatively (this was my first idea but I think the idea above is more interesting to talk about):**

I could look at each sentence in the entire (raw) WSJ-files and calculate the ratio between positive, negative and neutral sentences and compare those ratios to the ratios in the movie reviews to see in which texts the contrast is the biggest. In this case I would not only use the movie reviews files, but also at the WSJ-files.