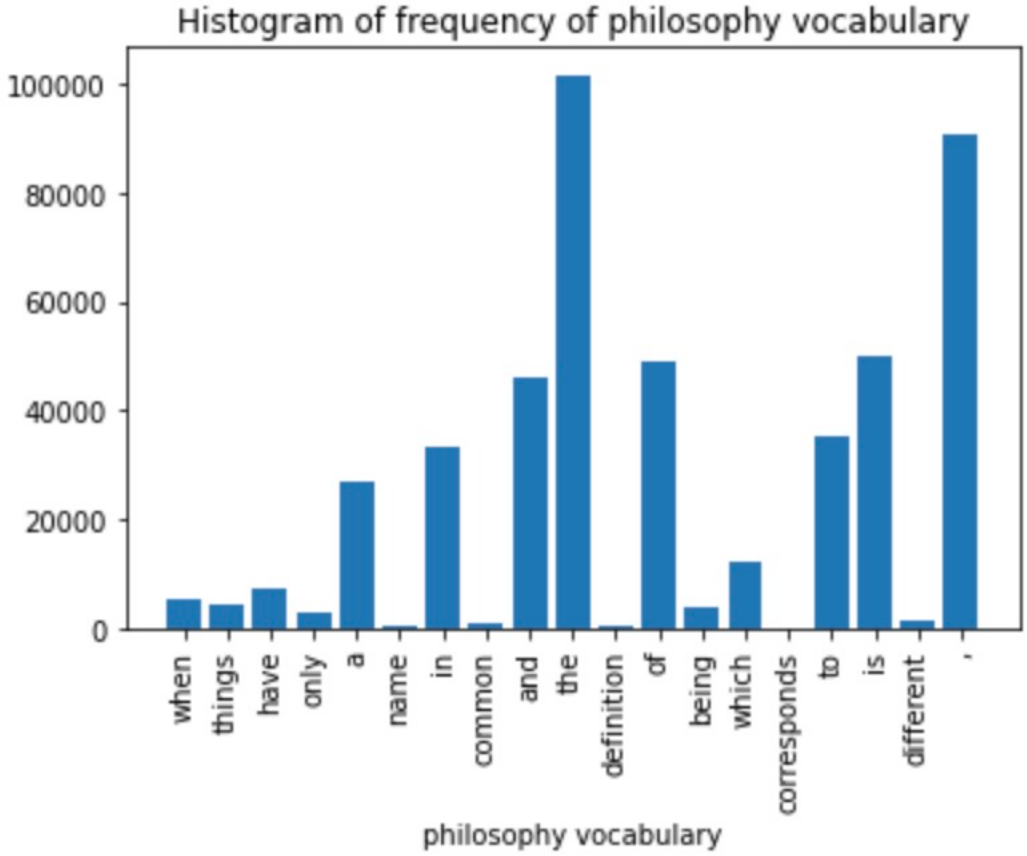
Report

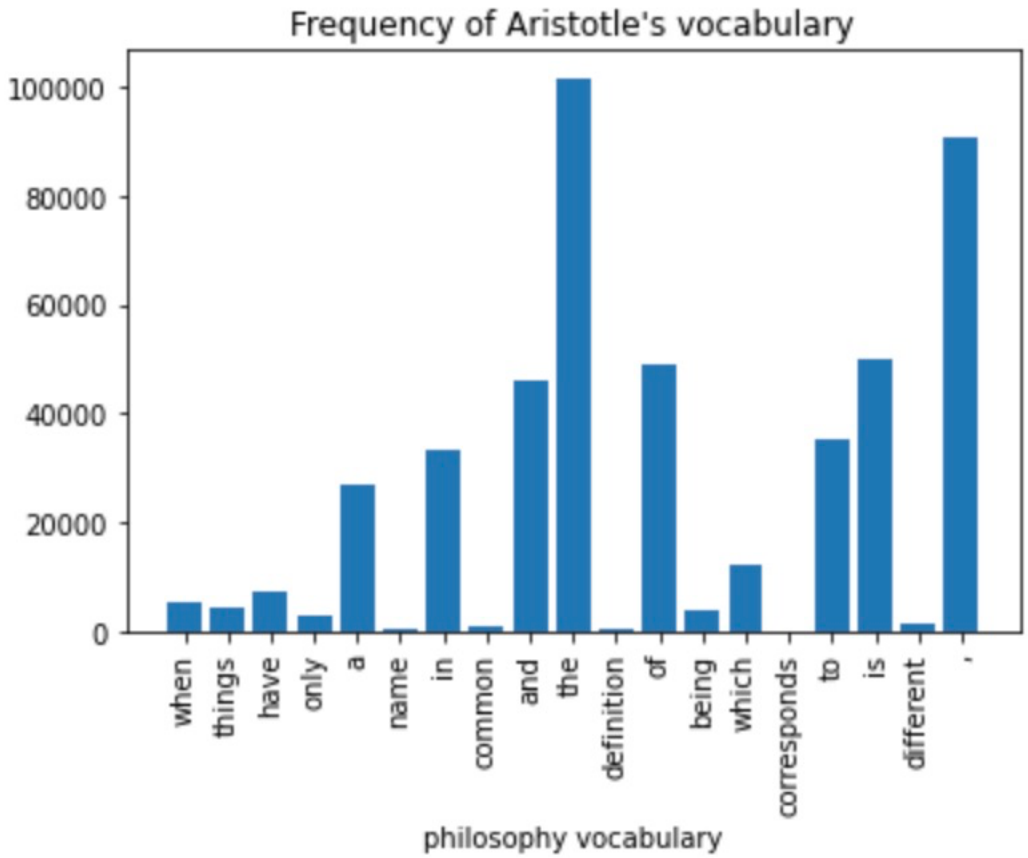
We have a database about philosophy which contains different sentence of philosophers from different school. As a starter who just read about philosophy, I decide to find some inner relationship or connection between different philosophers and make some prediction of their sentence structure. Thus, I firstly find out the most frequency vocabulary they were used before. And I use Markov Chain to generate their sentence and also use sentiment analyzer to distinguish the positive or negative of philosophers from different periods.

I. Frequency of Vocabulary

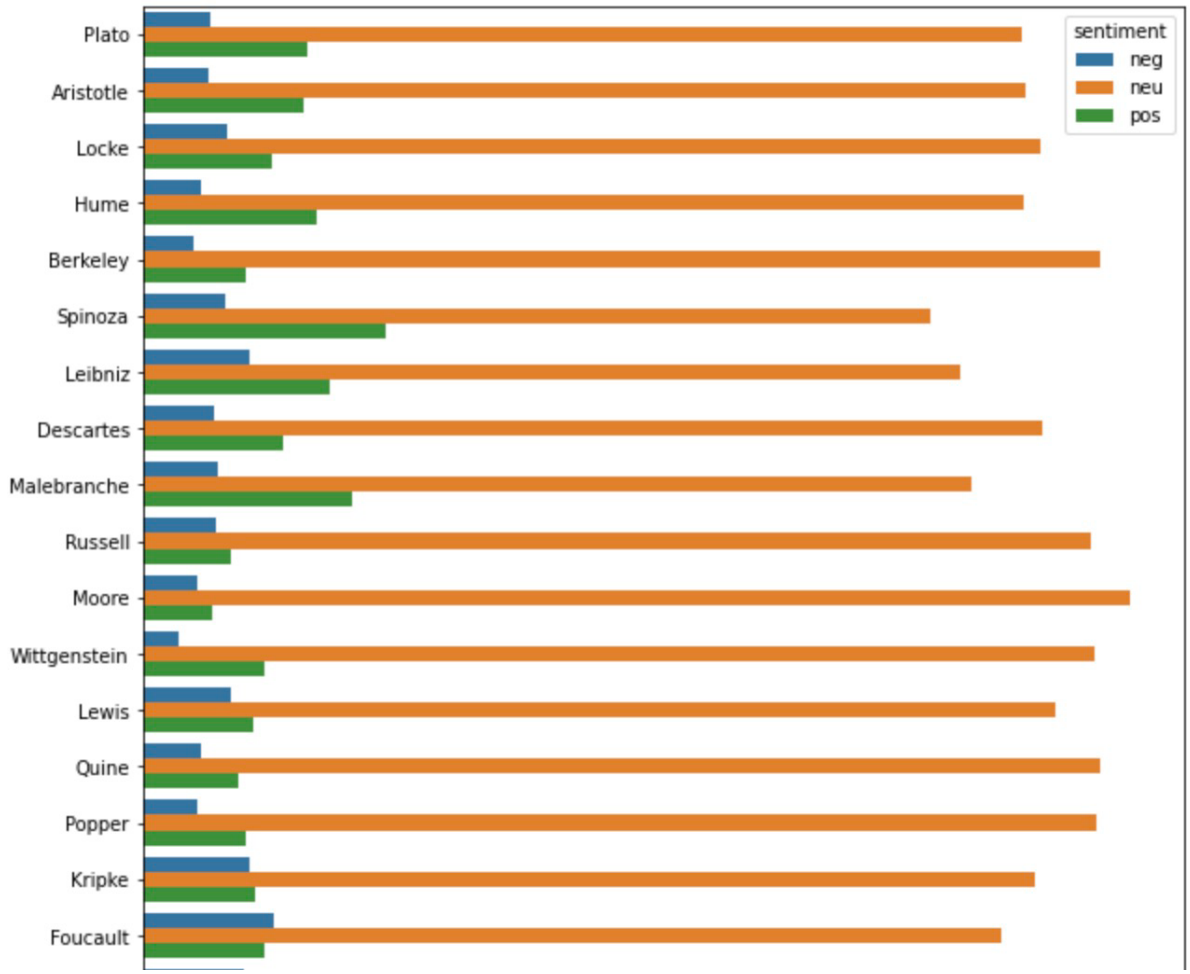
I would like to find out the most words of different philosophers. Thus, I tokens the text and get the roughly vocabulary they were used.

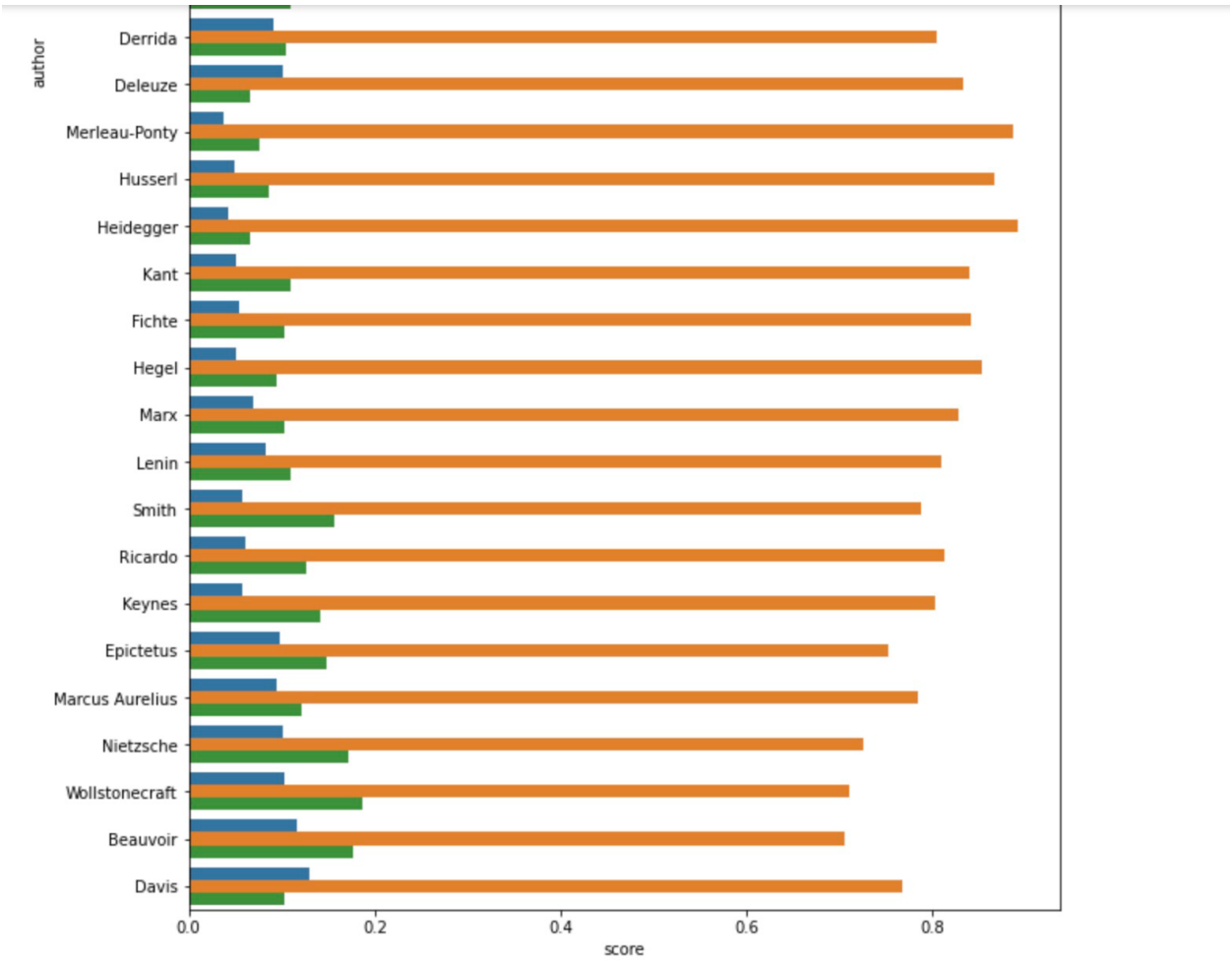


And with adjusting the frequency of words, we will get different results. The above graph is just a representative of frequency between 8000 and 8500. With the sentiment being adjusted to [9000 9050] and [10000 10050], we will get totally different graph. With the sentiment being adding on, we can see many philosophers will use words like “principle, leave”. Thus, it leads to two topics after this, which is frequency of some specific philosophers and how philosophers speaking trend of neu or positive or negative.



Here, I get the analysis of Aristotle's vocabulary as example. With tiny adjustment in the code, we are able to see frequency of different specific philosophers. And since Aristotle's vocabulary is pretty similar to normal words. I also simulate the result of some philosophers who lived in period which was closer to us. And their words become more abstract. Thus, it leads to the next analysis of how neu or pos those philosophers speak.


Plato
Aristotle
Locke
Hume
Berkeley
Spinoza
Leibniz
Descartes
Malebranche
Russell
Moore
Wittgenstein
Lewis
Quine
Popper
Kripke
Foucault
sentiment
neg
neu
pos



This is the scores after sentiment analysis after getting the samples of 100. I've tried 500 and 1000 samples. They all get the similar trends so that I just put the graph of using 100 samples above. If directly running codes of the whole files, it takes more than an hour to show simulation. But I think the continuously similar trend will have the representative meaning.

With the period changes, it is not hard to see that philosophers are more likely to speak for some specific class and have more speech that is not neu compared to many philosophers existed in elder period. This is closer to the background of more capitalism and communist and feminist become more familiar to public after WWII.

II. Markov Chain

I use Markov Chain to get text generator. I simply train model using markovify library on our database which said by Aristotle. With this method, we can simulate sentences of Aristotle of long or short by simply changing the last column of code to get the simulation of speech of Aristotle.

```
another line is to draw off liquid.
consequently, when moisture is engendered in greater abundance, so that it stands in the way.
near the city of thurium they say there is a meaning also in the pipes.
```

This is an example of 3 short sentence of Aristotle. With changing one column of code, we can get as much as we want, long or short sentence, of Aristotle.

Conclusion

After this, people can easily see and simulate different philosophers' speech and try to write or speak like philosophers better. And people can also change their neu or pos or neg to simulate for philosophers from different periods.