Major Changes

There have been no major changes to my project.

Accomplishments

Since the last meeting, I have run more combinations of the languages. I’ve also started looking into the treebanks and calculating statistics such as average distance between the head and the dependent word, average number of children per node, etc. I’ve also looked more into the UD guidelines and analyzed different parses in the treebanks to see if they matched the guidelines correctly. Many of the treebanks were automatically converted from other types of treebanks, so it’s possible that some of them do not adhere to the guidelines.

Meeting Milestone / Revisions

Because the results of the experiments did not appear to have any patterns, I’ve decided to spend more time looking into the quality of the treebanks. Therefore, I think I will need to shift my milestones in order to better incorporate this analysis.

Surprises

Some combinations of the languages provided higher/lower results than expected. However, I think this might be related to the varying amounts of training data. For example, when I look at only German and English, I have a lower number of sentences to train on. It’s possible that the parser is performing worse solely due to the decrease in number of sentences.
Looking Ahead

I plan on continuing to analyze the different treebanks to see different properties and how they affect the scores of the parser. Additionally, I plan on adjusting the experiments to have the same number of total sentences to try and diminish the confounding variables in the experiment.