Major Changes

There have been no major changes to my project.

Accomplishments

I’ve been training on 5000 sentences of a subset of languages with different properties: Finnish, Irish, Hebrew, Hindi, Indonesian, Japanese, Turkish and Chinese. Some of these languages are verb initial, some are noun-final, etc. I train on these languages combined with 50, 100, and 250 different English sentences to see how well they do against the monolingual English parser. From these results, we can see that as the number of English sentences increases, the languages seem to have a smaller effect, and the UAS scores all seem to be similar to that of the monolingual English parser. However, with fewer English sentences, there appears to be more variation between the scores, and some languages seem to help more than other languages.

Surprises

The effect of the 5000 sentences in another language does not appear to have as large of an effect on the output. Because of this, I plan on analyzing different parts of the tree to see if the different languages are better at predicting some results. Additionally, while the grammatical structures between Japanese and Turkish appear to be very similar, their UAS scores differ dramatically. This may be due to different grammatical decisions made by the treebankers.

Looking Ahead

I plan on finalizing the results and writing the final paper to talk about the different experiments I’ve tried.