Project Milestone

Yucen Li (Lily)

Major Changes

There have been no major changes. However, I have realized that my initial research question of "how to linguistically determine which languages to train on" is rather broad. Over winter break, I plan on reading more papers to see what research has already been done in order to narrow down the topic.

Accomplishments

During this semester, I’ve been able to read all of the Universal Dependency documentation, so I have a better understanding of how the treebanks are set up. Reading this documentation has guided me to reading more about different grammatical structures, since they introduced me to topics in linguistics that I wasn’t completely familiar with.

I was also able to reproduce the results from Many Languages One Parser, which used a multilingual parser (MaLOPa) that trains on specific languages and then is tested on another language to measure improvement. Because I got similar scores on the languages as the results that the paper got, I am decently confident that I will be able to continue using MaLOPa to test different hypotheses regarding which languages are beneficial.

I have also read many related papers to crosslingual parsing in order to better understand the current algorithms and the status of current research.

Meeting Milestone

I have met the meeting milestone, as I have been able to reproduce results from the paper, and I have read papers to further my understanding of multilingual parsing.
Surprises

Because the code from MaLOPa was from 2015, the installation and running procedure was a little different, and I needed to change a few parts of the code to make sure that everything worked properly. The setup was more tedious than I expected because of all the minor modifications I needed to make.

Also, I learned that the formulas provided in papers for crosslingual parsing often aren’t complete and will omit parts that are understood to be common through all parsing algorithms. However, because I do not have a lot of background, it can be hard for me to immediately fill in the parts of the equation that are missing.

Revisions

I currently do not have any revisions to my milestones.

Resources

Next semester, I want to meet briefly with someone very familiar with the ML part of NLP so they can answer any technical questions I have about any of the algorithms.