Meagan Beckstrand
CS 410
Prof. Nyguen
5/27/24

Assignment 3

I primarily followed the minimal guidelines required for this assignment. There were a handful of features that I implemented, like checking for capitalization, stop words, as well as chunks and parts of speech frequently attributed to named entities. There were a lot of redundancies in my files, and I would absolutely change this had I gotten to a working version of my code earlier. Upon running with "python3 score.name.py dev.name output.txt" I got

- 43515 out of 51578 tags correct
- accuracy of 84.27
- 5917 groups in key
- 5639 groups in response
- 899 correct groups
- Precision of 15.94
- Recall of 15.19
- F1: 15.56

Instructions:

1. Run FeatureBuilder.py, "python3 FeatureBuilder.py train.pos-chunk-name <filename>.pkl"
   a. Outputs: "<filename>.pkl"
   b. This output stores the trained data classifier based on the features implemented in FeatureBuilder.py
2. Run AssignLabel.py "python3 AssignLabel.py test.pos-chunk <filename>.pkl"
   a. Output: "test.name"
   b. The output has the token and the predicted NER label attached, and this can be used to compare to the dev.name file.