

Report - Homework #3

1 Parsing and Transformers

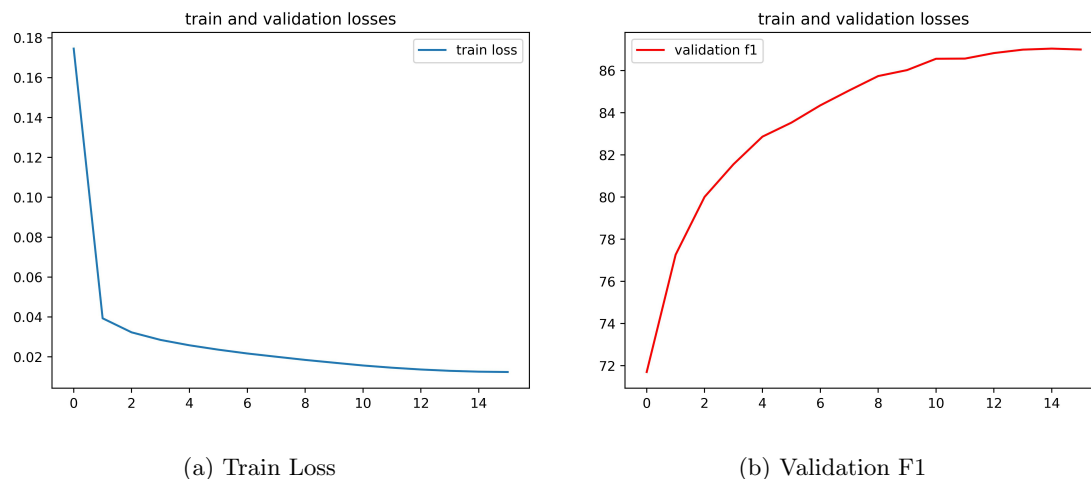


Figure 1: Learning Curves

On EVALB, the model is able to achieve an F1 score ("Bracketing FMeasure") of 90.12 for sentences of length 40 or shorter, with recall 88.90 and precision 91.38. The model also performs quite well for sentences of all lengths even though it's trained only on short sentences, reaching an F1 score of 88.66. Precision is always a little bit higher than recall since the correct syntax tree is more compact than we predicted, which make senses. Here I included the learning curves. As we can see, the validation F1 score monotonically increases as train loss decrease, showing that there's no overfitting.

Below is an example of the parsing result on the validation set.

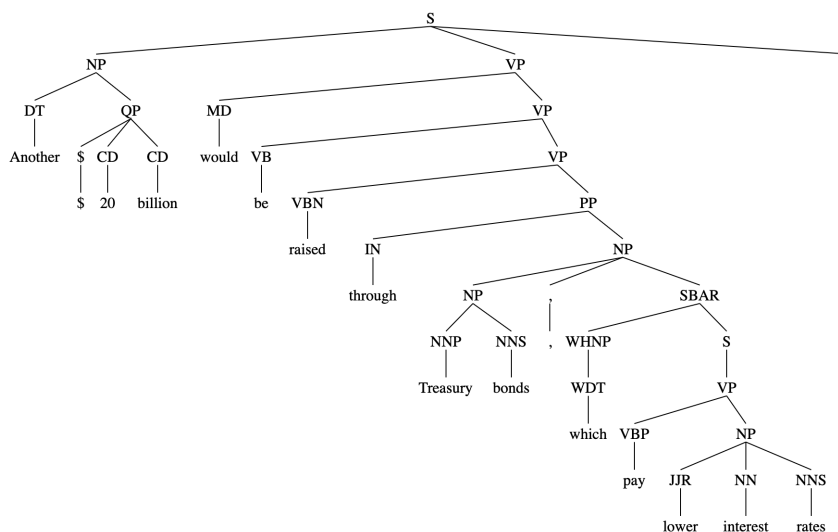


Figure 2: Tree