Part 3

Refael Shaked Greenfeld 305030868,

Danit Yshaayahu 312434269

In order to deal with miss-match between the training and dev data with the word vectors, we first had a slightly different preprocessor that lower case all the words when pooling the word vector. Also, we had a special vector 'unique' that we gave to all the unknown words, this way it got trained and it got some signal during the training and wasn't completely random. In part 4 we already added the extra signal of the suffix and prefix…

Regarding the improvement, with best hyper parameter tuning model 1 reached number higher in the best setup (78% in model 1 vs 76.2% in part 3) but in average the models were about the same, in the POS we got the almost same results (95.03 in part 3 cs 95.12 in part 3). We believe it is due the disadvantage of the unknown words that takes the model a bit down.

Our parameters of the best models:

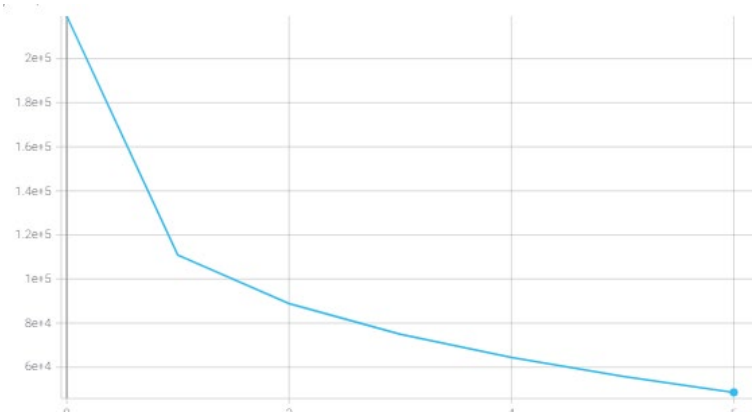
POS:

- **Hidden dim size : 200**
- **Batch Size : 128**
- **Learning rate: 0.001**

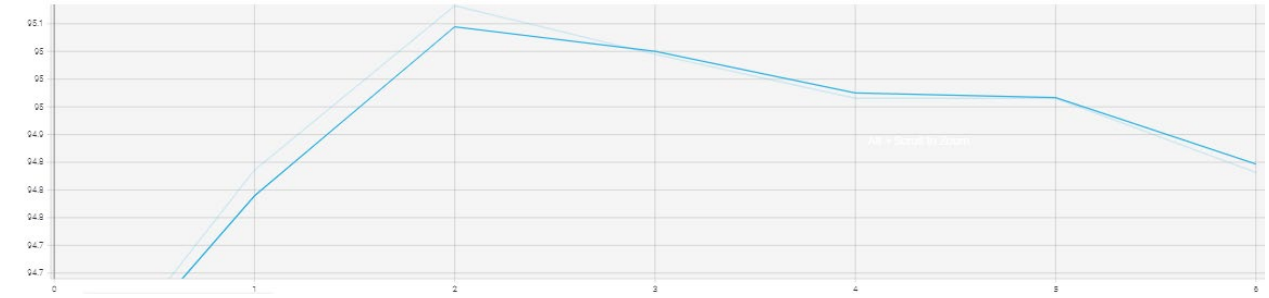
NER:

- **Hidden dim size : 200**
- **Batch Size : 32**
- **Learning rate: 0.001**

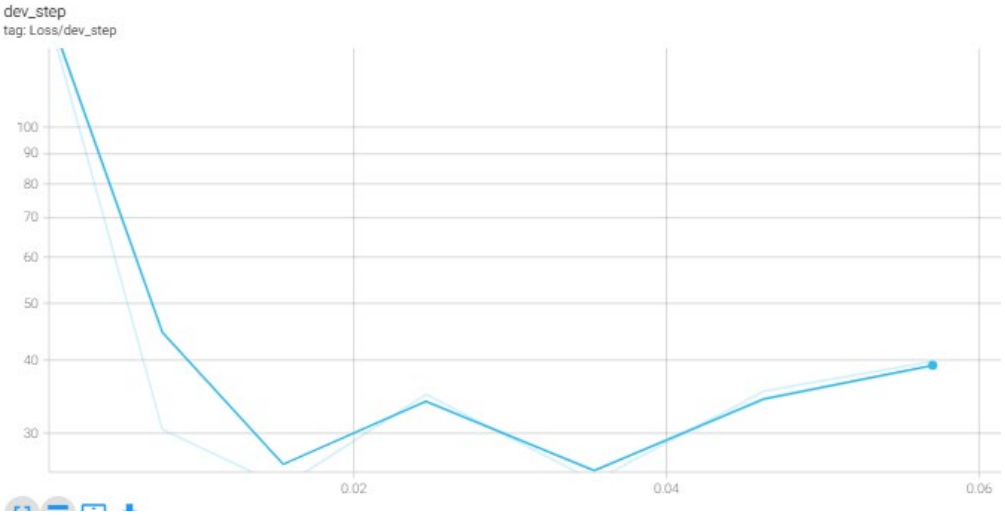
The POS Loss - Dev:



The POS accuracy - Dev



The Ner loss - Dev:



The Ner accuracy - Dev