

Yoel Jasner - 204380992
 Steve Gutfreund – 342873791

Part 3

Considerations:

1. **“How do you handle lower casing?”**

We’ve decided to convert all the data from the train set (and so we do for new examples like from dev set or test set) to lower case. We might have lost the advantage of words having capital letters (names, beginning of sentences etc.) but we are able to fully utilize the pre-trained embeddings.

2. **“What do you do with words that are not in the pre-trained vocabulary?”**

For words in the train set not found in the pre-trained vocabulary, we just add them to the vocabulary and add a corresponding embedding vector for each one of them.

In addition, we added the word *UNKNOWN* to the vocabulary so we can treat all the words from dev set we haven’t seen before by substituting them with *UNKNOWN*.

(And as explained in part 1, we added *START* and *END* so we can work properly with the window sliding.)

Throughout the epochs we updated the pre-trained embedding alongside the added embedding layers (for the words found only in the train set as explained above).

Parameters for NER:

epochs: 15
 learning rate: 0.01
 hidden-layer: 200

Parameters for POS:

epochs: 8
 learning rate: 0.01
 hidden-layer: 180

