**Part 1 report:**

Tamar Michelson 323805861

Shilo Avital 206487407

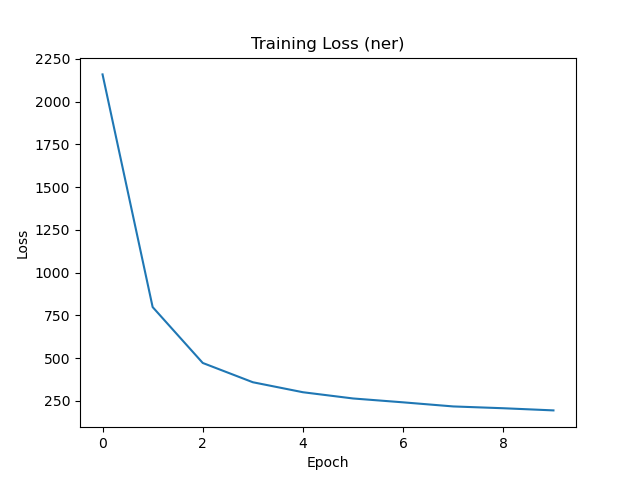
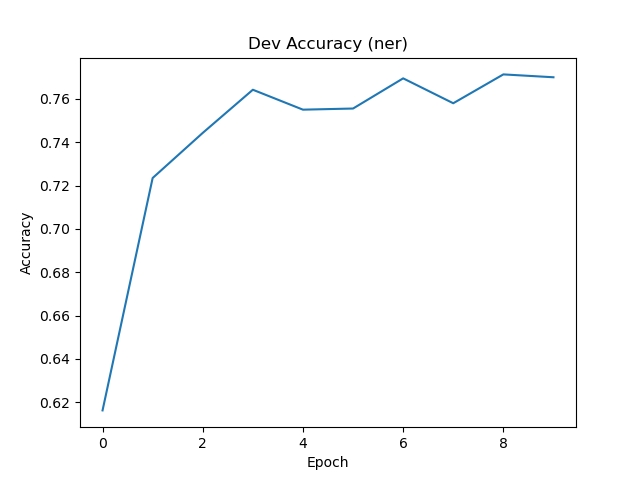
The parameters of the best model for both of the tasks (NER and POS):

Hidden dim = 100  
Batch size = 32  
Num epochs = 10  
Learning rate = 1e-3  
Weight decay = 1e-5

What do you do with a word that appears in the train set and not in the dev set?

which word embeddings will you assign to it? What vectors will you use for the words surrounding the first word in the sequence? And for the last word?

Handling Special Cases  
Words that appear in the dev sets but not in the training set are assigned the embedding for <UNK>, a special unknown token.  
At the beginning and end of sentences (to build the sliding window), we pad the sequence with <PAD> tokens. These have a fixed embedding of zeros.

The ner plots:

The pos plots: