

Part 3 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

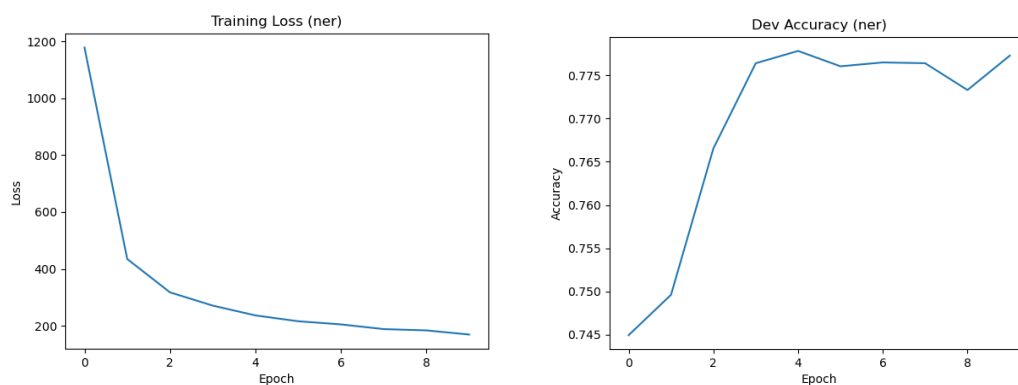
Weight decay = 1e-5

All words are converted to lowercase during preprocessing to ensure compatibility with the vocabulary and pretrained embeddings.

We initialized a random matrix with uniform initialization $\left(-\frac{1}{\sqrt{d}}, \frac{1}{\sqrt{d}}\right)$, $d = \text{embedding_dim}$) for all words in the vocabulary and for the representation of the padding and unknown words, each word that we found a representation of in the embedding file, we replaced with values from the file.

When we worked with the embedding we received, it improved a little, the main difference could be seen in the first few epochs and after that the results were similar to the results with the randomly initialized embeddings.

The ner plots:



The pos plots:

