Part 1 report:

Tamar Michelson 323805861

Shilo Avital 206487407

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

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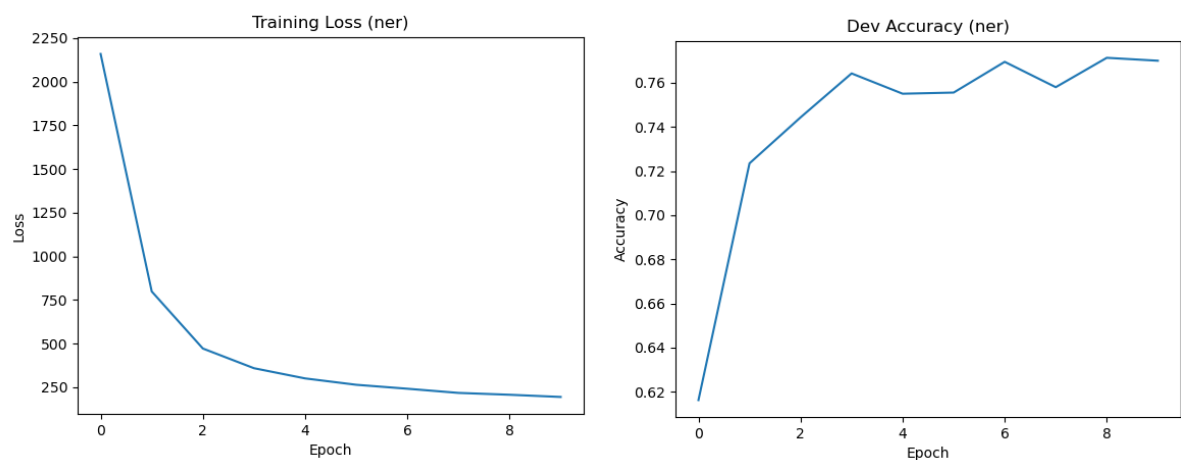Batch size = 32
Num epochs = 10
Learning rate = 1e-3
Weight decay = 1e-5

Considerations:
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: