

Part 5 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 = $1e-5$

NER:

Num filters = 5

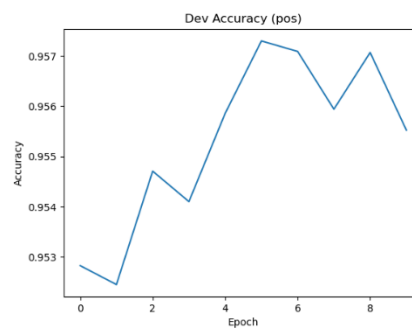
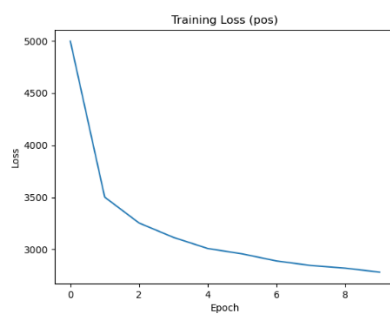
Filter width = 3

POS:

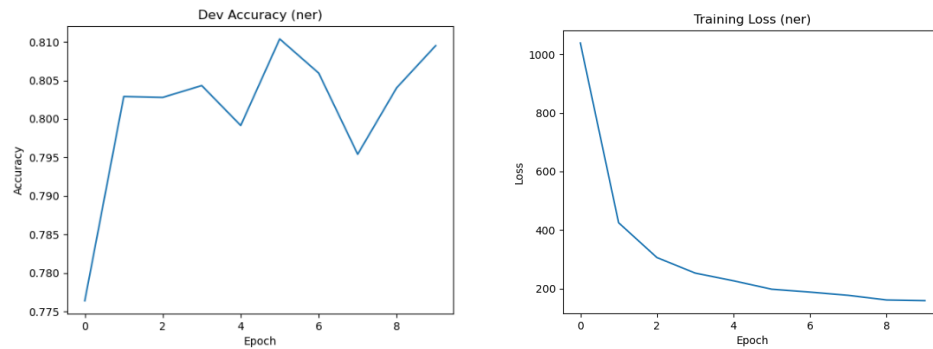
Num filters = 30

Filter width = 5

Pos:



NER:



In NER, the CNN significantly improved the results compared to subwords. In POS, the CNN results are similar to subwords.

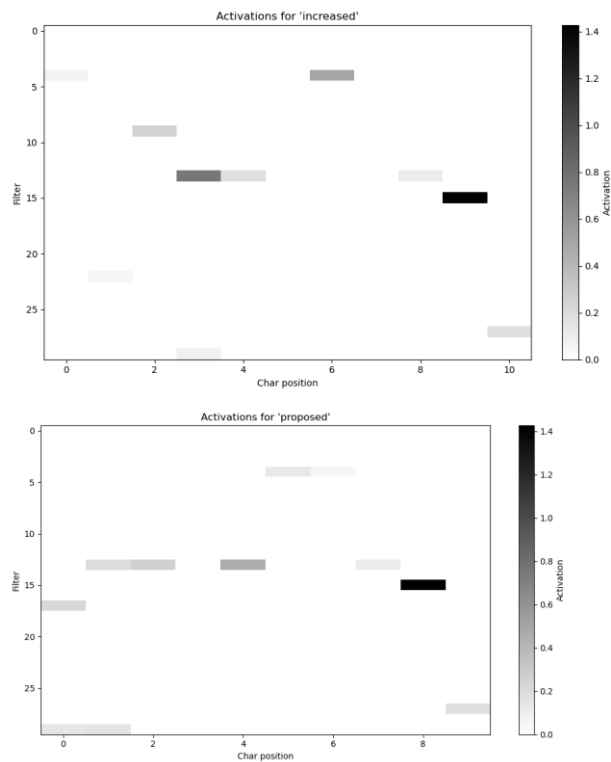
We ran different options of number of filters for each of the datasets but the change was significant only in NER with **5** filters

While in POS for varying numbers of filters the results were similar

Which probably indicates that in NER there are 5 features that best capture the labeling characteristics

We ran different combinations of window sizes, we did not see a significant effect for both datasets

In POS, there was a small improvement obtained for a window size of **5**



In this example, we can see that for the words *increased* and *proposed*, which is tagged **VBD** (in POS), we note that the significant weight of the prediction is based on the last letter, not the padding (which makes sense because verbs used in the past will usually end with **ed**). Apparently, filter number 15 detects endings.

We built this using a heatmap on the output of the convolution layer.