Part 4

Refael Shaked Greenfeld 305030868,

Danit Yshaayahu 312434269

The subword units are more useful on our dataset because it is very small and almost 50% of the words appear only once so by subwords we can generalize our words also most of them are rare. In the other hand, we have no enough occurrences in order to fine tune our model on the new dataset. The contribution of the models are not complementary because if there is a word that has no pretrained vector and also it's suffix and prefix appear only once, both models will not succeed to learn this word. As shown below with the pretrained vectors, the NER mission improved by 1.5% while the POS mission got almost same accuracy as with the other trend.
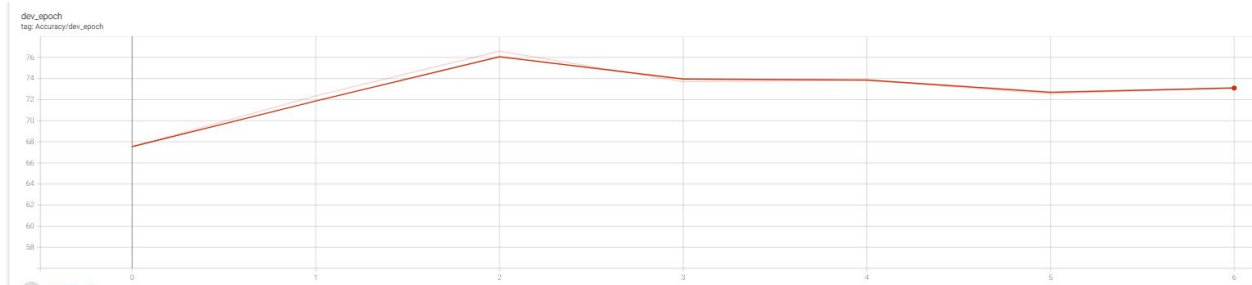
Our parameters of the best models:

POS:

- **Hidden dim size : 200**
- **Batch Size : 32**
- **Learning rate: 0.001**

NER:

- **Hidden dim size : 200**
- **Batch Size : 128**
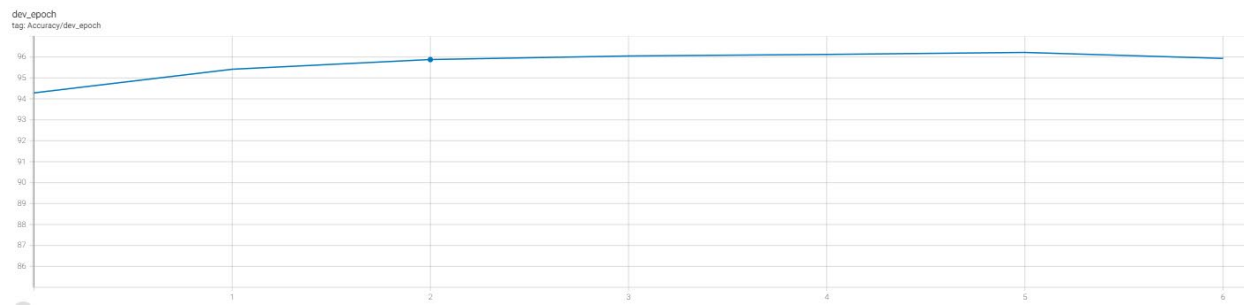- **Learning rate: 0.01**

Random vectors graph:
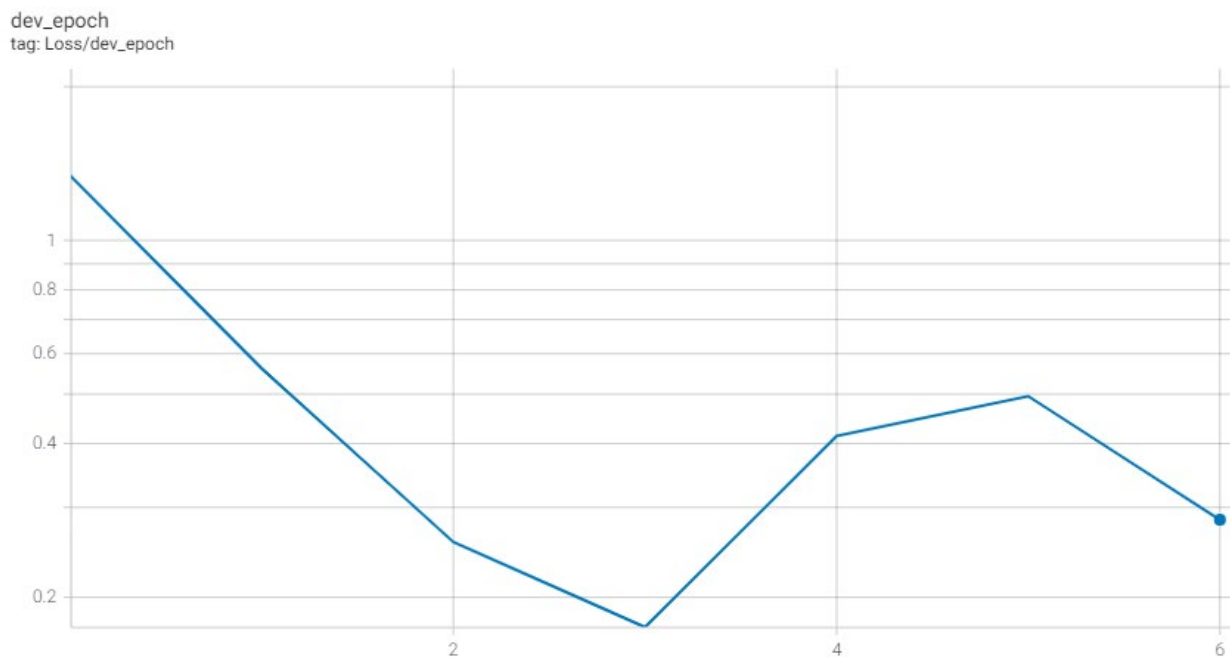
The Ner accuracy - Dev :
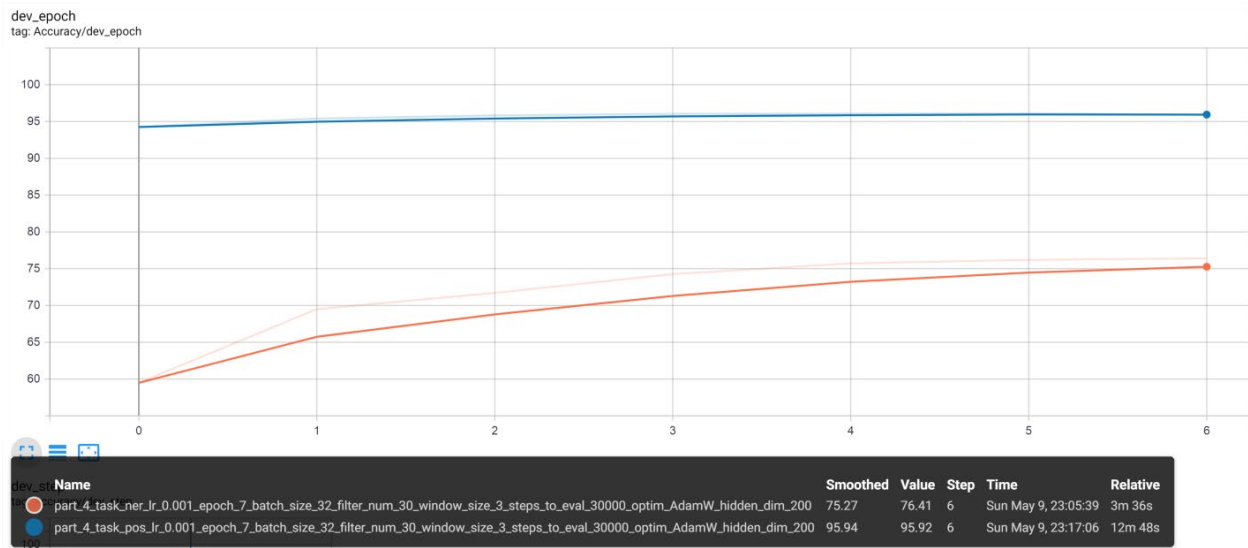

The Ner loss - Dev:

The POS accuracy - Dev:

dev_epoch
tag: Accuracy/dev_epoch

The POS loss - Dev:

dev_epoch
tag: Loss/dev_epoch

Pretrained vector garph:

POS and NER accuracy - Dev:

dev_epoch
tag: Accuracy/dev_epoch

Name	Smoothed	Value	Step	Time	Relative
● part_4_task_ner_lr_0.001_epoch_7_batch_size_32_filter_num_30_window_size_3_steps_to_eval_30000_optim_AdamW_hidden_dim_200	75.27	76.41	6	Sun May 9, 23:05:39	3m 36s
● part_4_task_pos_lr_0.001_epoch_7_batch_size_32_filter_num_30_window_size_3_steps_to_eval_30000_optim_AdamW_hidden_dim_200	95.94	95.92	6	Sun May 9, 23:17:06	12m 48s

POS and NER loss - Dev:

dev_epoch
tag: Loss/dev_epoch