Part 4

Refael Shaked Greenfeld 305030868,

Danit Yshaayahu 312434269

A pdf \_le called part4.pdf in which you describe the parameters of your best model for each of the tasks (NER and POS). If you had to make any choices when including the sub-word units, describe them also. For each of the conditions (subword-units, pre-trained) and (subword-units, no pre-trained), include two graphs showing the accuracy of the dev set for each of the task as a function of the number of iterations, and two graphs showing the loss on the dev set for each of the tasks as a function of the number of iterations. Write also a brief analysis of the results. For example { which of the pre-trained embeddings and the subword units is more useful, and why? are their contributions complementary? are the trends consistent across the different tagging tasks? why?

Analysis:

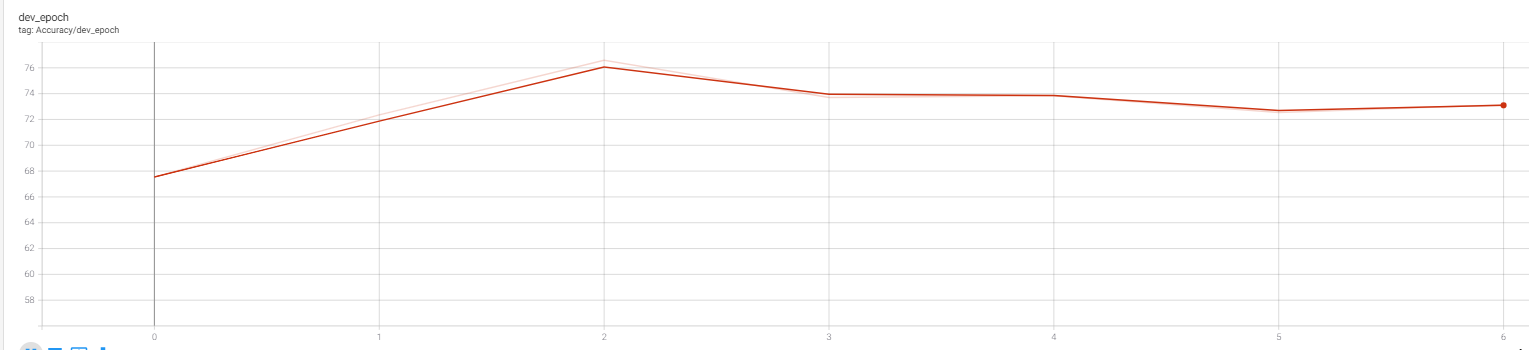
Our parameters of the best models:

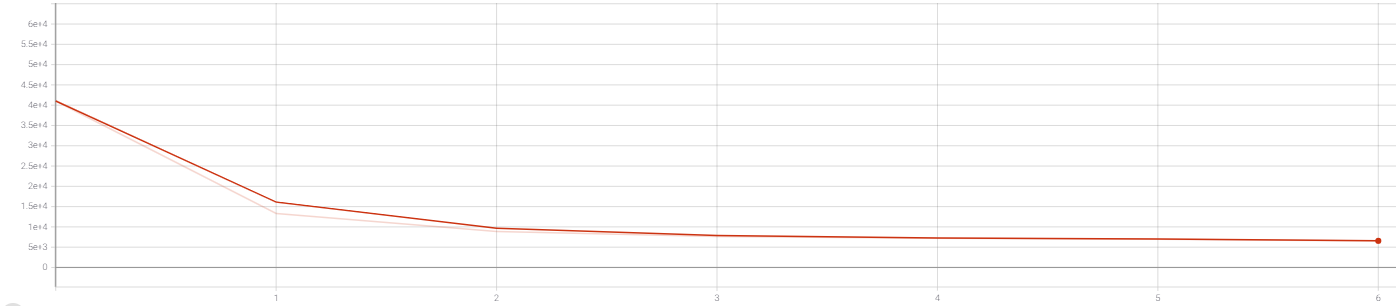
POS:

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

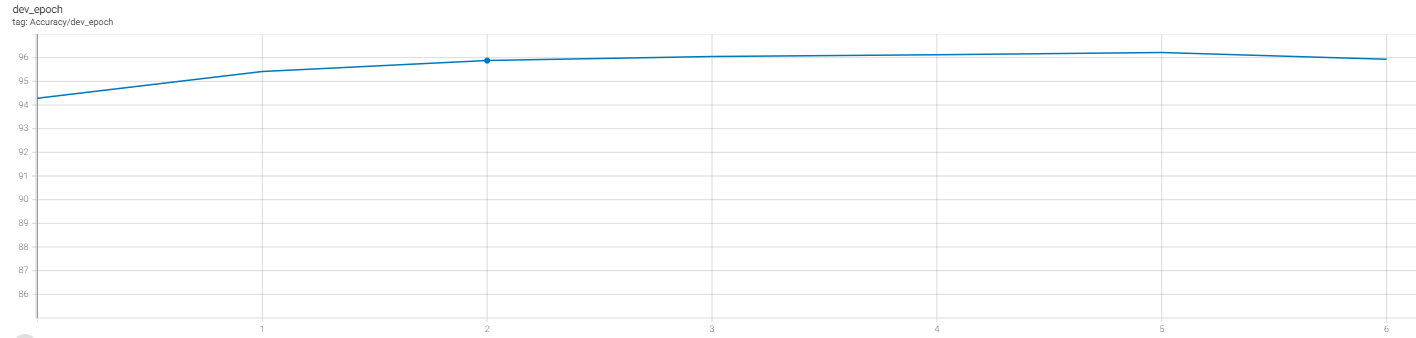
NER:

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

**The Ner accuracy - Dev** :

**The Ner loss - Dev**:

**The POS accuracy - Dev:**



**The POS loss - Dev:**

