

## Report

My model for the 2nd part is very bad. I'm not sure if this is my code or GPU is very crowded today, but I had to run my model with very small numbers. I didn't use the potential of the dataset which had over 13 000 sentences for the training dataset. I needed to use 1000 sentences to make it run. I lowered batch size to 2 and run it only with two epochs. I also found online that it might be helpful to add *fp16=True* to lower memory use with half-precision floating-point format.

Even though it was trained on very small data, it still has not terrible precision score 0.525597 and semi-good accuracy 0.820302. That's probably because it's guessing one most common chunk tags. In the model from part 1, the accuracy is 0.925612. Not only first part was trained on bigger dataset, but the model was specifically tuned for English language, as opposed to part 2 where I used multilingual model.

Stanislaw Lorys