Report pos_tagging

In this assignment I used the language pair English-Kazakh: data from Witchtower.

At first, I used just the initial pos_project.py script to project tags from English to Kazakh and then trained the model on it.

The accuracy of this model is just 25.95%.

I've tried to fill the gaps in tags with NOUN tag and it gave an improvement in accuracy: 42.83%

After that I tried to add another possible improvement: I estimated the probability of each tag that was mapped from English text to the Kazakh one and then sampled tags to fill the gaps in Kazakh text according to these probabilities.

Actually, it worked a bit worse than just using the NOUN tag for all gaps – accuracy: 41.52%

I saved the pos_project.py script with the last approach.