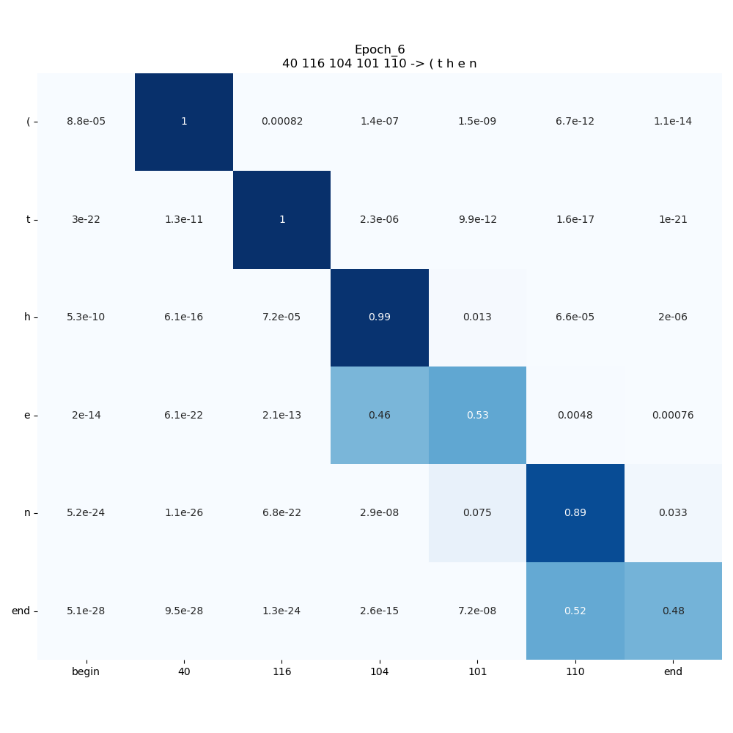
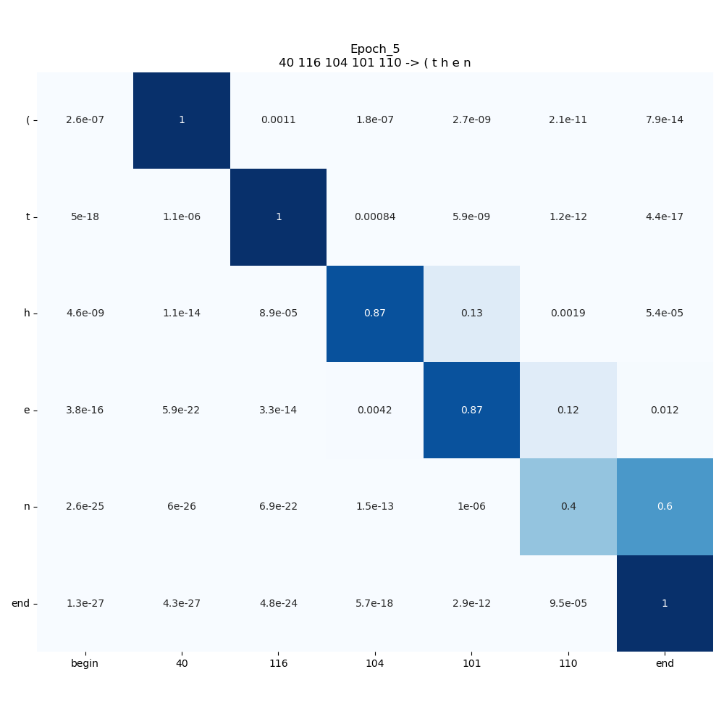
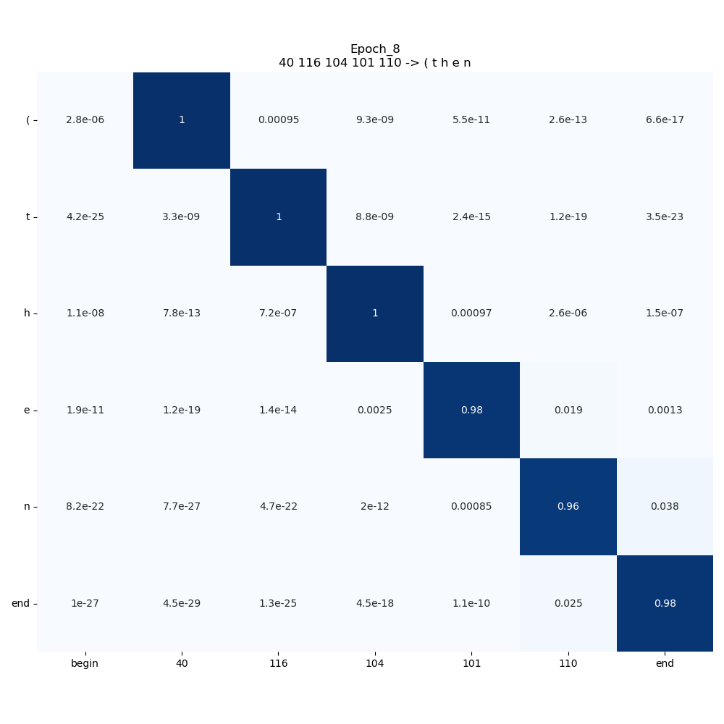
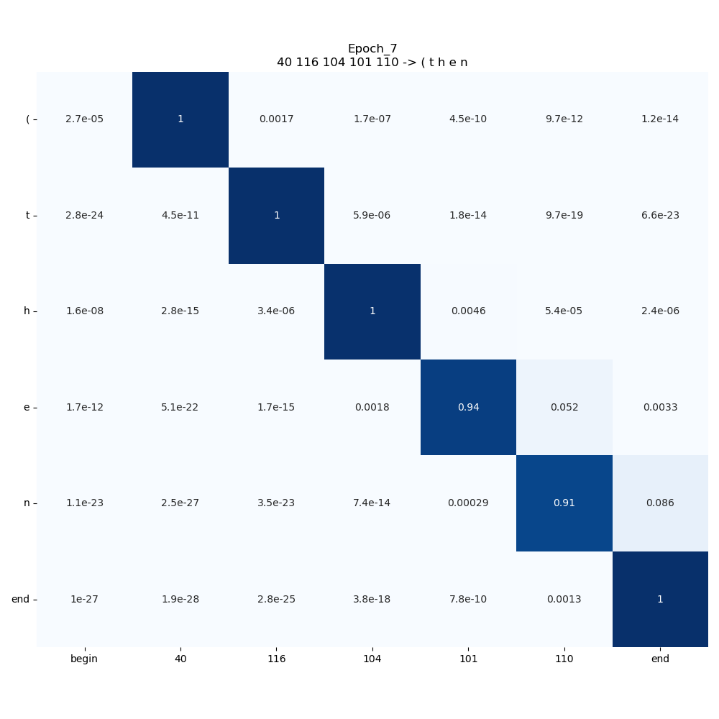
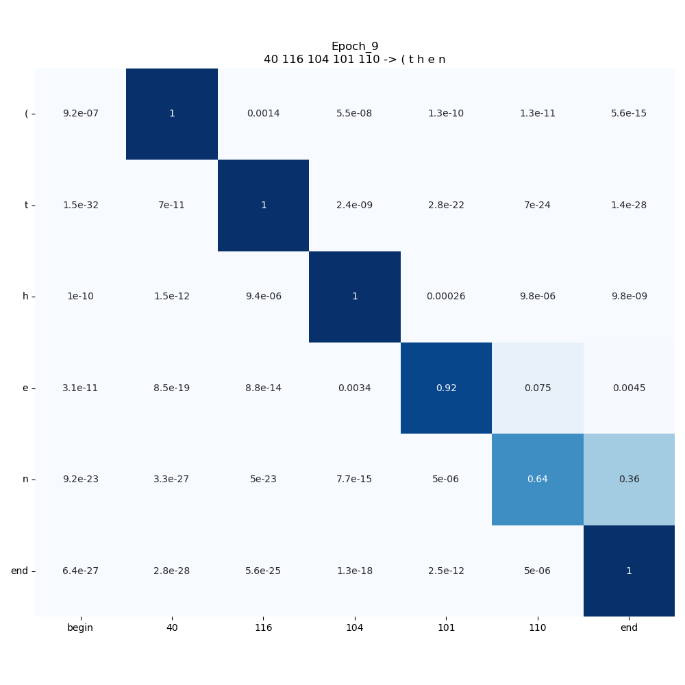
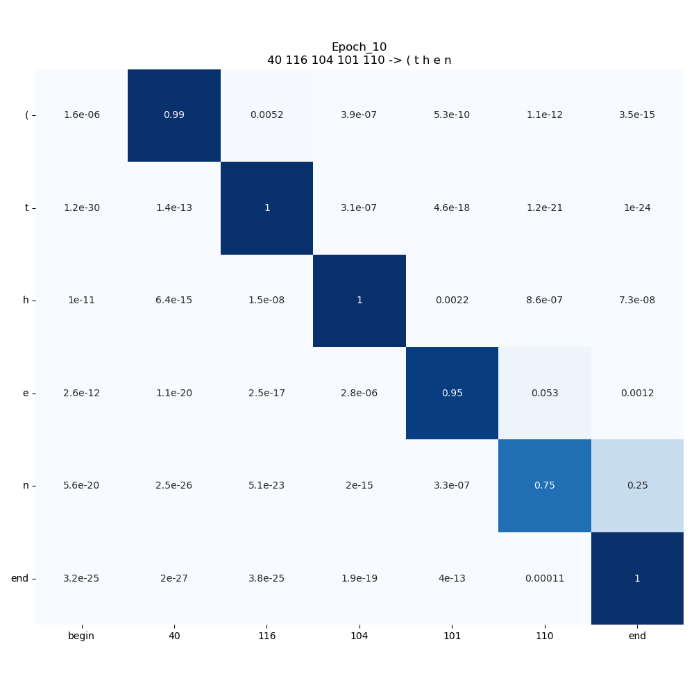
**Ex2 – Part 3 – Report**







We can see that at the beginning of the training the attention is mostly given to the end token (</s>) – this occurs because at the first steps our context vector doesn’t know how to align the words in the sentence.

As the training makes progress, we can see that the attention converges towards the diagonal of the heatmap – each token gets its appropriate attention to the translation.