A Basic Encoder-Decoder Model

Amitai Yacobi

June 6, 2022

Description

As we can see in the figures below, in each step of the decoder, the input symbol that gets most of the attention, is the one that its translation needs to be predicted in the current step.
In the early epochs we can see (especially in the last symbols of the sequence) that the attentions is still not presice because there are input symbols that the attention for them is divided over couple of output symbols, but in the rest of the epochs can see an improvment, until the attention in each decoder step is almost 1000% precise.
Another thing that we can notice is that in epoch 9, there is sort of a regression in the attention weights, and maybe we can assume that this is caused because the model got to an overftting.

Explanation

In the attention mechanism, as much as the model trains on the data, it can learn a better representation of the source sentence in every decoder step.
This is why we can see in the figures below, that the attention of the model becomes more and more relevant along the epochs.

Attention Plots

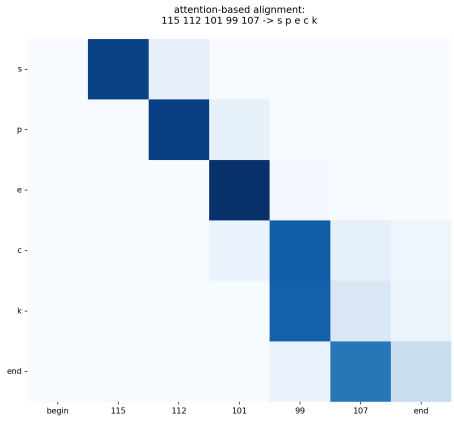

Figure 1: Epoch number 1

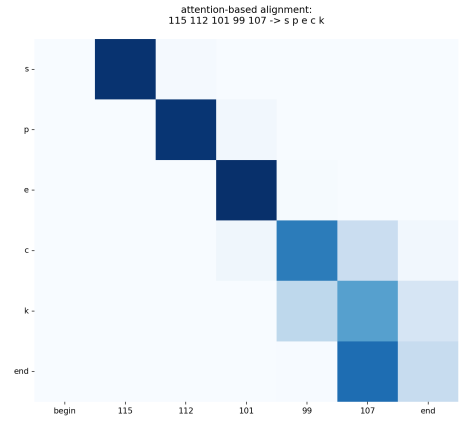

Figure 2: Epoch number 2

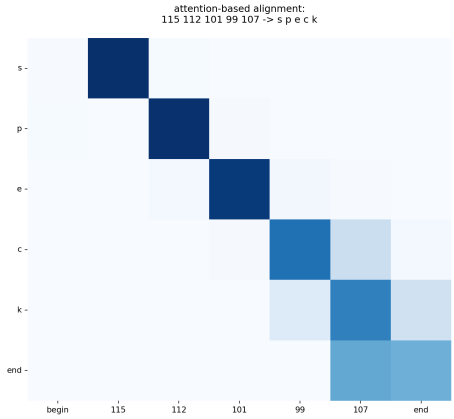

Figure 3: Epoch number 3

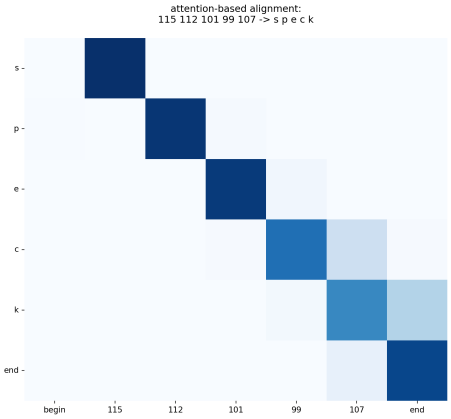

Figure 4: Epoch number 4

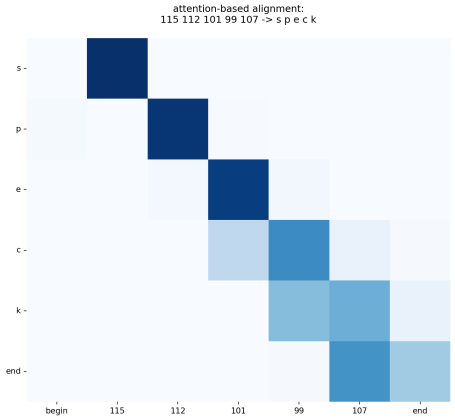

Figure 5: Epoch number 5

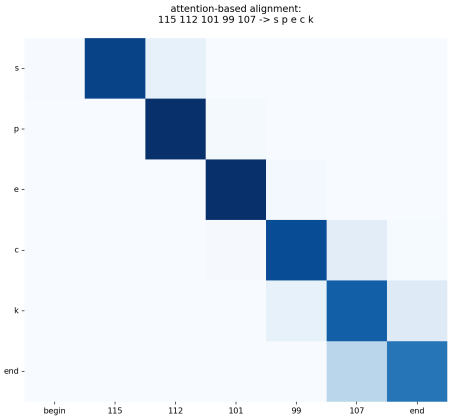

Figure 6: Epoch number 6

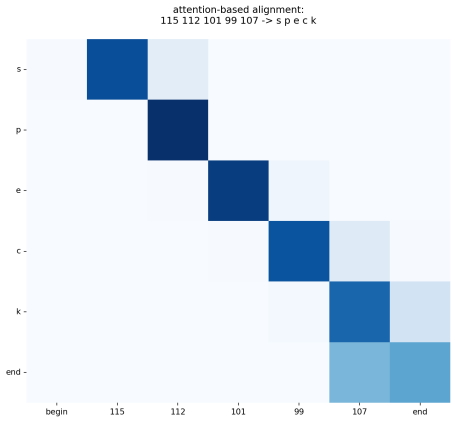

Figure 7: Epoch number 7

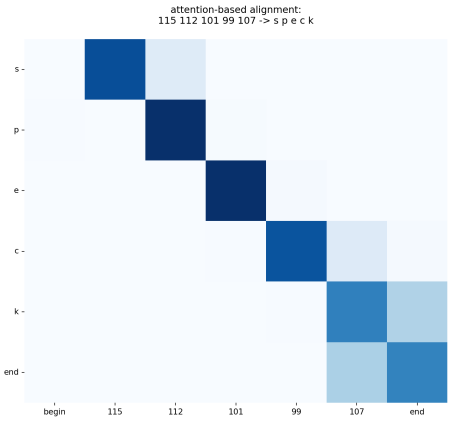

Figure 8: Epoch number 8

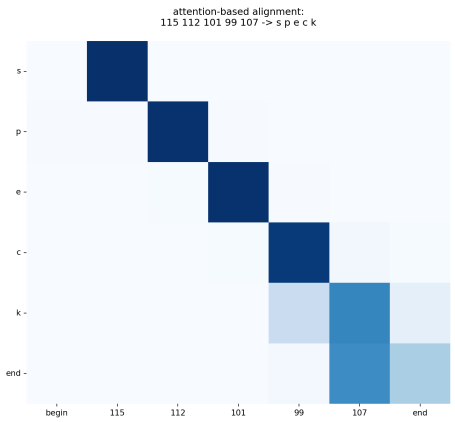

Figure 9: Epoch number 9

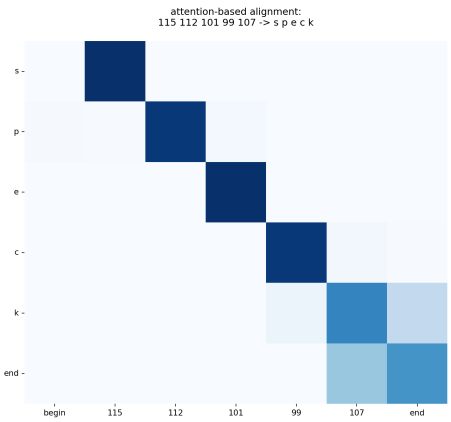

Figure 10: Epoch number 10