

Neural machine translation has been lately established as the new state of the art in machine translation, especially with the Transformer model. This model emphasized the importance of self-attention mechanism and suggested that it could capture some linguistic phenomena. However, this claim has not been examined thoroughly, so we propose two main groups of methods to examine the relation between these two. Our methods aim to improve the translation performance by directly manipulating the self-attention layer. The first group focuses on enriching the encoder with source-side syntax with tree-related position embeddings or our novel specialized attention heads. The second group is a joint translation and parsing model leveraging self-attention weight for the parsing task. It is clear from the results that enriching the Transformer with sentence structure can help. More importantly, the Transformer model is in fact able to capture this type of linguistic information with guidance in the context of multi-task learning at nearly no increase in training costs.