Instructions for the Homework #3

1. Starting with the Jupyter notebook for the Transformer-NMT model verify that you can get it running with the default English to Spanish translation example. Verify your implementation works and note the accuracy and loss achieved during training as your performance baseline.

2. Experiment with changing the following:

   a) Vary the depth of the network and compare results to the baseline
   b) Vary the model dimension (embedding vector length) and compare it to the baseline
   c) Vary the number of attention heads and compare to baseline
   d) Reduce the vocabulary in half and compare it to the baseline
   e) Remove the positional encoding and compare it to the baseline

   <u>For each experiment where you vary hyperparameters use at least three different values.</u>

   Explain the results you get in terms of your understanding of transformers.

3. Compare the baseline trained with **ReLU** to that using **GELU** as described in this paper (arXiv:1606.08415v4) & **Swish** described in this paper (arXiv:1710.05941v1). Note that GELU (Gaussian Error Linear Units) are what GPT-2 and GPT-J are using instead of ReLU. See you get better convergence or better loss during training and if any of the translations in your tests have changed. Both GELU and Swish are provided by Tensorflow.

4. Choose another Language of your choice and replace the English to Spanish training with English to <your choice>. Note that you will also need to alter the preprocessing steps appropriately. Demonstrate the new translator on some interesting examples and compare to Google Translate for the same test pairs.

5. Briefly summarize what you have learned from this exercise in a paragraph or two.