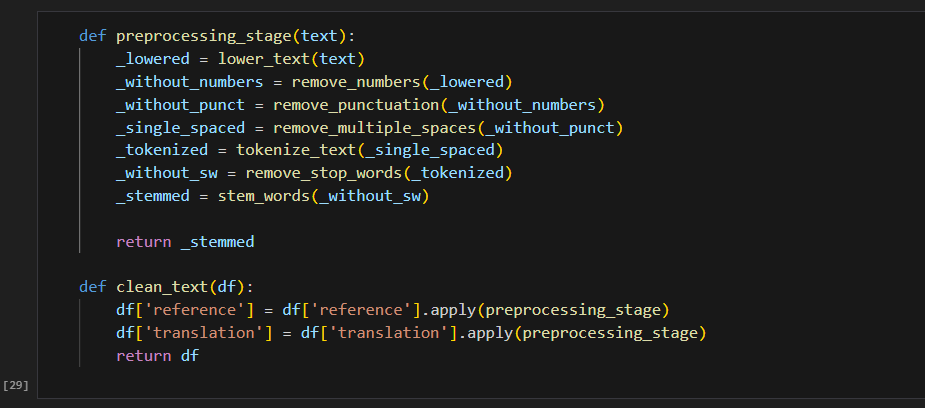
# Baseline: Dictionary based

The first step I took in this assignment was to read the paper provided in the task description. From this, I learned that I could use the model trained and uploaded by Dale et. al. on their GitHub repository. However, I also discovered that it would be more efficient to use the J-score, as described in the paper, for the final solution instead of the BLEU score, as it does not evaluate style and detoxification.

# Hypothesis 1

In order to preprocess the data, I created a notebook called "1.0-data-observation" and used functions from one of the labs to run on the dataset. However, I later realized that I should not preprocess translation text since it is what the model will generate as output. Therefore, it would be better to leave it "as it is".



# Hypothesis 2

As I began incorporating the ParaGeDi model, I encountered a roadblock - my laptop did not have a dedicated GPU, causing me to run out of computational power. This prompted me to move to the Kaggle environment to continue my work.

# Hypothesis 3

Despite my efforts, I was unable to make the solution work in Kaggle. As a result, I decided to use pretrained models from the "Hugging Face" library and fine-tune them, specifically using CondBERT as an example from the paper. However, this required me to have access to Hugging Face datasets and tokenizers, leading me to restart my preprocessing stage.

# Results

Overall, my approach using pretrained models from Hugging Face proved to be somewhat mediocre. In addition, it was a challenging process due to the need for specific metric of evaluating the toxicity and quite a lot of computational power. In the end, I was able to fine-tune 3 models and achieve satisfactory results.