# English to Hungarian Translation Model

**Name: Anika Duffus**

**Professor: Bálint Gyires-Tóth**

## Introduction

The general topic explored was "Language Translation". In the project, I developed a model that translates English to Hungarian text. Hungarian syntax is not as similar to English compared to other Western European languages like Spanish. Therefore, the project aims to determine how well a seq2seq model can perform this translation despite the difference in syntactic structure.

## Previous Solutions

The code for the project is based on an existing solution on keras.io for translating short sentences from English to French (<https://keras.io/examples/nlp/lstm_seq2seq/>). This solution used a basic character-level recurrent sequence-to-sequence model, instead of more commonly used word-level models.

## Dataset

Two datasets, approximately the same side, were used in this project. The first dataset is from ManyThings.org, which has various tab-delimited sentence pairs from the [Tatoeba Project](http://tatoeba.org/home) (<http://www.manythings.org/anki/hun-eng.zip>) which is a collection of sentences and translations. The second dataset is directly from Tatoeba.org (<https://tatoeba.org/en/downloads>). English was the input language and Hungarian was the target language.

## Proposed Method

The project uses Seq2Seq modelling, and LSTM encoder and decoder. The two dataset files were merged: 70% for training, 20% for validation and 10% for testing. Punctuation was removed from the data and the strings were made lowercase.

## Evaluation Method

The metric of interest is “accuracy”. The model’s accuracy was determined in two ways. The preliminary accuracy value was determined by the Model().evaluate method. The more relevant accuracy score for this domain was given using Bilingual Evaluation Understudy (BLEU). This is a score for comparing a candidate translation of text against one or multiple possible reference translations. This seems more appropriate given that there may be multiple ways to translate a given English sentence into Hungarian. The BLEU score for the entire test data (corpus\_bleu), and the score for each sentence in the test data (sentence\_bleu) were calculated.

## Results

The table below shows a sample of test data results:

|  |  |  |
| --- | --- | --- |
| Input Text | Target Text(s) | Decoded Text |
| kezdjünk | * lets begin * lets start * shall we begin | lets begin |
| táncolunk | * were dancing * shall we dance | were dancing |
| rendeljünk | * shall we order | were buying |
| mi kezdjük | * shall we start | were paying |
| imádja őt | * she adores him | you know hhes |
| felhívta őt | * she called him | he slapped her |

The model’s accuracy is 64%.

The **corpus\_bleu()** is the BLEU score for multiple sentences such as a paragraph or a document. The test data score was 0.0.

The **sentence\_bleu()** score for each sentence in the test data was 0.0.

The **sentence\_bleu() 1-gram** score, which evaluates just matching grams of a specific order such as single words, with weight on the first word of the sentence, was 0.0.

## Discussion

The corpus\_bleu() score indicates that there is a perfect mismatch between target test data and the decoded sentences. Since corpus\_bleu() calculates the geometric mean of individual n-gram scores, this score of 0.0 is unexpected since there are word matches present in the experiment. Nevertheless, this score is not extremely significant since the input text is not a collection of related sentences like a paragraph, but a set of sentences which have similar meanings.

Therefore, the sentence\_bleu() scores are more relevant. However, all the scores are 0.0 as well. Like with the corpus\_bleu() score, this is unexpected given there are word matches in some sentences. The scores for 1-gram were also zero. It could be that there was an error when using the BLEU score api since the sentence scores should be greater than zero given that there are first word and sentence matches present in the test data experiments.

Given more time, there could have been experiments to analyze the effect of the neural network's size on the accuracy and translation quality. These experiments were not conducted due to multiple technical issues when running the code.