Thesis Proposal On

Natural Language Processing Based Machine Translation For Bangla-English Translation using GRU and Attention

Submitted by

Adity Barua

Roll : BKH1811034F

Session : 2017-2018

Thesis Supervisor

Tanvir Zaman Khan
Assistant Professor

Department of Information & Communication Engineering

Noakhali Science & Technology University

Noakhali – 3814

Bangladesh

Date : 29 August 2022

DECLARATION

This research proposal is submitted to the Department of Information & Communication Engineering, Noakhali Science & Technology University, Sonapur, Noakhali-3814, in partial fulfillment of the requirements for having a B.Sc degree in Information & Communication Engineering. So, I hereby declare that this research proposal has not been submitted elsewhere for the requirement of any kind of degree, diploma, or publication.

Adity Barua

Student of Information & Communication Engineering

Noakhali Science & Technology University

Roll No : BKH1811034F

Session : 2017-18

ACCEPTANCE

This research proposal is submitted to the Department of Information & Communication Engineering, Noakhali Science & Technology University, Sonapur, Noakhali-3814, in partial fulfillment of the requirements for having a B.Sc degree in Information & Communication Engineering. This thesis proposal will be evaluated under the course: Project and Thesis with course code ICE-4110.

Tanvir Zaman Khan

Assistant Professor

Department of Information & Communication Engineering

Noakhali Science & Technology University

Noakhali – 3814, Bangladesh

ABSTRACT

The term "Machine Translation," or "MT," refers to a branch of language computation that focuses on the idea of converting speech or text between different languages.The technique of automating translation from one language to another is presently in a new phase of Machine Translation (MT) research due to the quick development of various neural network architecture.In this study, we will investigate neural machine translation (NMT) employing an attention-based mechanism, Gated Recurrent Unit (GRU), and encoder-decoder modules to translate Bangla to English and vice versa.Combining different models or models' outputs in a neural network can help us to create a new model with potentially much superior outcomes.In order to create a better model, we shall integrate models in this proposal.The combination methods include Weight Averaging, Ensemble, and Checkpoint Ensemble.It demonstrates a possible study direction to improve Bangla-English NMT.

Keywords: Machine Translation(MT),Neural Machine Translation, GRU, Attention, Encoder-Decoder, Ensemble, Checkpoint Ensemble , Weight Averaging.

Table of Contents

Chapter 1

INTRODUCTION

1.1 Background Information

Natural language processing(NLP) is the automatic processing of natural language. It aims to develop computers with computer science, natural languages, information technology, and expert systems that comprehend and react to text or voice data—and answer with text or speech of their own—in a manner resembling that of humans.It belongs to artificial intelligence (AI).MT is a sub-field of computational linguistics that investigates the use of software to translate text or speech from one language to another.Bangla is the official, national, and most widely spoken language of Bangladesh and the second most widely spoken of the 22 scheduled languages of India. Bengali is the fifth most spoken native language and the seventh most spoken language overall in the world, with about 300 million native speakers and another 37 million who speak it as a second language.The fifth most widely used Indo-European language is Bengali[2][3].

NLP is used to make the machine intelligent. Every day, language processing becomes more sophisticated. Many studies defined the architecture for natural language processing , but some deal with the improvement of English to Bangla language translation.

There are some teams who worked to develop Bangla-English Machine Translation.However, there are currently few studies with complicated phrase structures and repeated sentence meanings.The use of a machine for translation is essential since it will be able to adapt to changes on its own as new data is added to the model.Additionally, a machine is capable of handling both multidimensional and multi-type data. Machine translation has the potential to save significant amounts of time by eliminating the need to spend time looking up definitions, which will enhance productivity[4].

1.2 Problem Statement

We all know , Natural language is not fixed. Every now and then language undergoes some small changes.New vocabulary (names, compounds,new word creation) is added to the language. Machine translation sometimes fails to translate properly. There are example for inappropriate translation below:

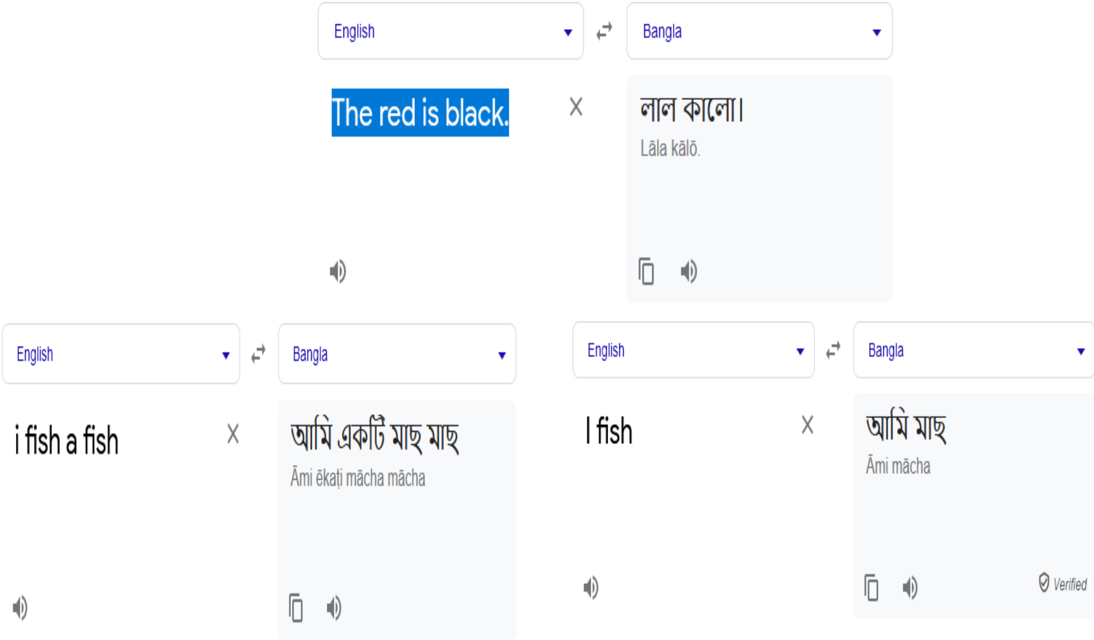

Figure: Example of inappropriate translation

This particular paper deals with NMT of English – Bangla languages. NMT is a type of MT methodology that applies a neural model to anticipate the potential series of words, often in the form of an entire sentence. Unlike statistical machine translation, which consumes more memory and time, NMT, a neural machine translation, continuously trains those parts to maximize performance.

1.3 Objectives

The aim is to design an architecture of English to Bangla Machine Translation system with Neural Machine Translation and use Combination technique to get a better model.

I. To model a MT system using Encoder-Decoder Model.
II. To design MT system architecture with GRU and Attention.
III. To use “Combination Technique” on the proposed model improve overall–
    a) Performance
    b) Training Time
    c) Decoding Time

Chapter 2

LITERATURE REVIEW

2.1 Literature review

A. Machine Translation

Machine translation is the use of artificial intelligence to mechanically translate text from one language to another without the need for human intervention.[1]The whole meaning of the text in the native tongue is communicated in the target language using modern machine translation, which goes beyond mere word-to-word translation.It examines every aspect of the text and finds the relationships between the words.The technique of automatically translating phrases from one particular language to another particular language using any natural language processing and machine learning technology is known as MT[6].

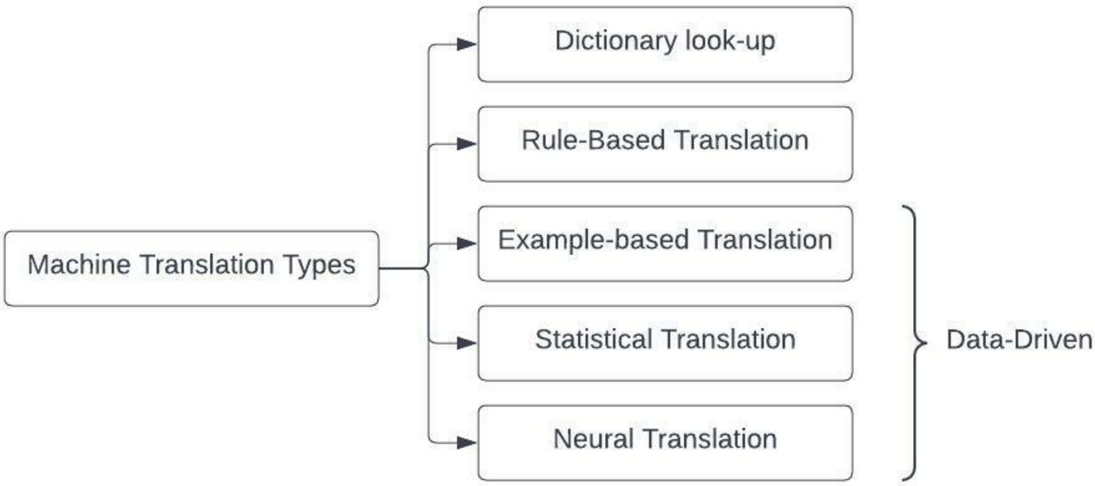

Figure: Machine Translation Types

B. Neural Machine Translation

NMT is essentially a kind of MT that makes use of Artificial Neural Networks (ANN) to predict the likelihood of a set of words.It is merely the modeling of complete phrases into a single integrated end-to-end entity.NMT is distinct from a statistical method based on language that makes use of separately created subcomponents.The basic method of neural machine translation is the use of embedded and continuous spatial representations, which are vector-based representations of words and internal states.Compared to the phrase-based model, the NMT model has a clearer and more complete structure.Instead of different language models, translation models, rearrangement models, etc., there is only one direct model that translates each entity, i.e., word or sentence at a time.Even so, in order for this sequence anticipation to work, the entire source dataset and the entire target sequence dataset must already have been produce.For accurate translations, the NMT model employs machine learning and deep learning models.The majority of word sequence modeling was initially carried out using an Recurrent Neural Network (RNN).The ANN uses a Bidirectional-RNN(Bi-RNN)[11][12], referred to as an encoder, to encode the results of a second RNN, referred to as a decoder, which is used to predict the words in the target language.It is challenging to compress larger or longer inputs into a single vector when using iterative ANN.This problem can be solved by employing an attention mechanism that enables the decoder to focus on particular input elements as it generates each phrase of output.Up until it encounters sentences with longer structures, neural machine translation operates without any problems.Technically speaking, the information bottleneck problem with the fundamental encoder-decoder model exists.This issue arises because we cannot predict the length of the phrase that will be provided as input when the model has to compress all the information of the source language into a vector of defined lengths.Because of information bottleneck problems, the translation model cannot handle sentences with large corpus sizes.Attention is here for our rescue.The fundamental tenet of the attention architecture is that we can connect to the encoder at any point during the decoding process in order to narrow our attention to a specific section of the source text.Technically speaking, we obtain a weighted sum of the encoder level values and use them to our advantage in the decoder component[6].

C. Attention-Based Mechanism

By choosing attention mechanisms that specialize in sentence subparts during translation, Neural MT (NMT) performance is increasingly being improved[10].The neural network includes the ocular mechanism.Verify that the supplied piece is crucial at each decoding step.With this setting, the encoder shows an illustration of every source token rather than compressing the entire source into a single vector.

D. Sequence to Sequence Models

Sequence to Sequence models, sometimes referred to as seq2seq models, are a special class of recurrent neural network architectures that are frequently used to address challenging language-related tasks, such as text processing, question answering, chatbot creation, and machine translation.In sequence-to-sequence learning (Seq2Seq)[9], models are trained to translate sequences from one domain to sequences in another.To avoid vanishing gradient problems, this is done using recurrent neural networks, more frequently Long Short-Term Memory (LSTM) or GRUs[8].The partial derivative of the loss function eventually approaches a value near to zero and vanishes as the network's layers increase, at which point the value of the product of derivatives starts to decline.This issue is known as the Vanishing Gradient Problem[7].

Chapter 3

DESIGN AND METHODOLOGY

3.1 Introduction

Uncertainties and translation divergences are difficult for any machine translation system to resolve (TD).We're now trying to come up with a solution that will address these issues and result in a top-notch translation.To build a system , the most important initial step is to make the basic architecture of the system and the methodology . It will simplify the implementation of the system. So, in this chapter, the design and the methodology will be explained.

3.2 System Architecture and Design

A. Encoder-Decoder Model

In NMT , Encoder-Decoder approach means,"A proper translation is produced by the decoder using the fixed-length representation that the encoder obtains from a variable-length input text."[15]

a) Encoder Related Step :

- ➢ Each word in an input sentence is encoded.
- ➢ The encoded statement will be sent to GRU.

b) Decoder Related Step :

- ➢ Each word in a target sentence is encoded.
- ➢ Then,get the weighted sum of the encoder outputs by using the attention layer in the system.
- ➢ Finally, put together the results that obtained from the prior two steps. This final tensor is sent to the GRU layer of the decoder.
- ➢ Output of this GRU layer is sent to a high-density layer.Word with a high probability ,considered to be the next word in the sentence.

B. Gated Recurrent Unit

GRU(Gated Recurrent Unit) is used to solve the Vanishing Gradient Problem[8]. GRU uses 2 gates.They are- update gate and reset gate.

C. Attention

Attention is used to solve the alignment problem of the encoder-decoder model[13].

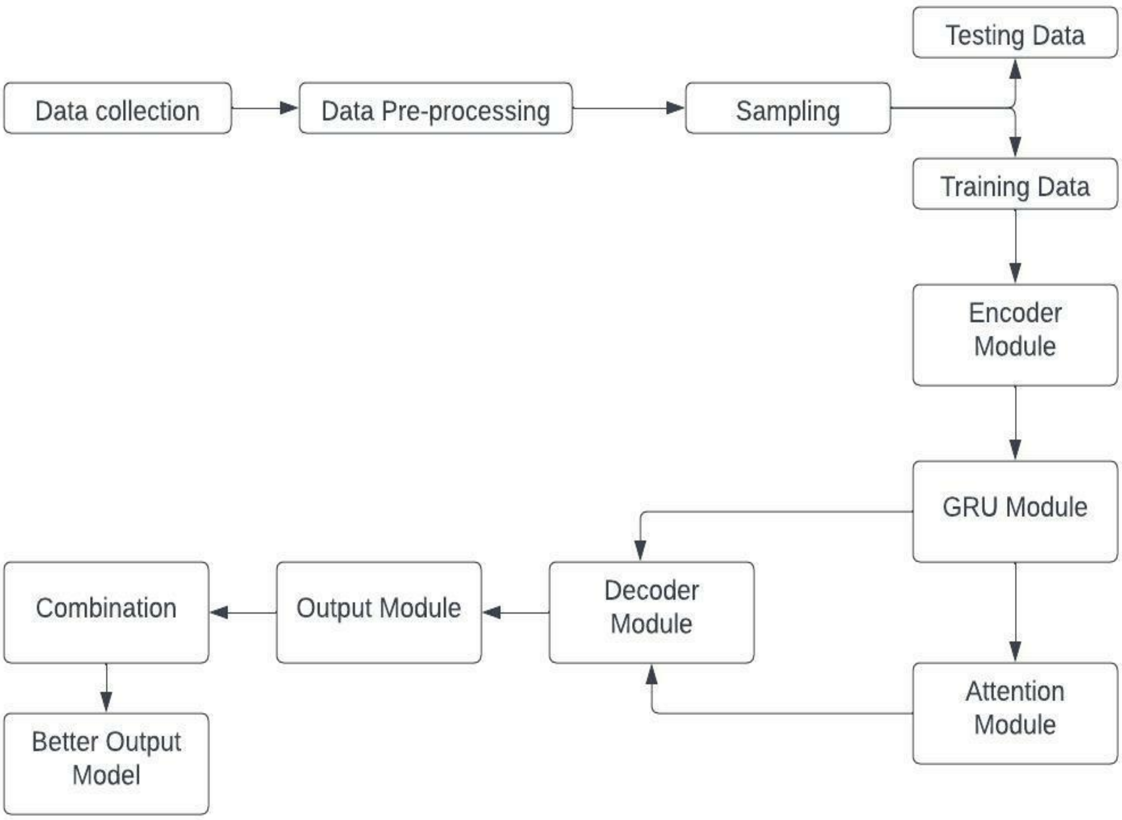

Figure : Basic Block Architecture

D. Combination

By combining different models or outputs of MT, we can significantly improve the output of MT. There are some different ways to improve the performances.Main 3 ways described below.

- Ensemble

  Ensemble means “Combination of Different Models”. In Ensemble, We can train our data with 3 or 4 Models(or more models). Then collect all the output layers. Finally , combine the output layers .

The performance of the ensemble is better , the reason is , here we can get outputs of all models, then take the best one for generating final output.But needs more training time and decoding time.

- Checkpoint Ensemble

Checkpoint Ensemble uses only one model to train the data. After training, we will save the best checkpoints. It is mainly ensemble models of different checkpoints. So , We will get better performance and training time(Only one model has to be trained). But Decoding time is the same as Ensemble.

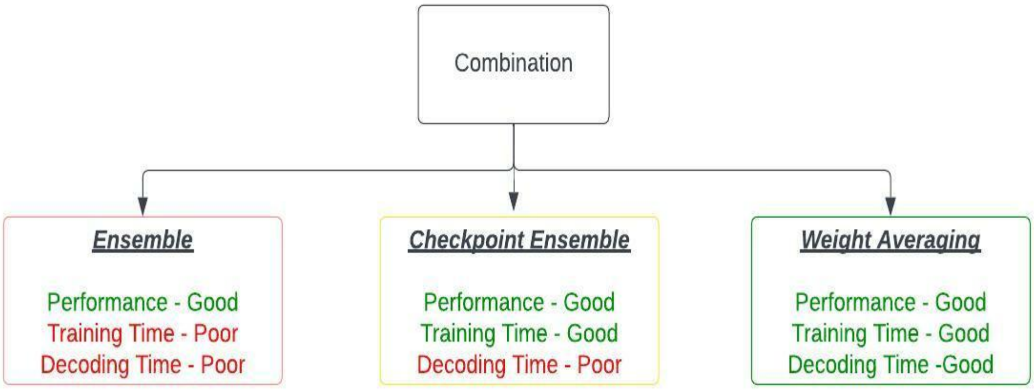

Figure: Combination

- Weight Averaging

Weight Averaging is an updated version of Checkpoint Ensemble. It works if we only use one model. Weight averaging also saves the best checkpoints .Before decoding the end points , It averages the weights . So , It overcomes the problem of decoding time from previous combination techniques.

3.3 Methodology

A. Data collection

Datasets will be gathered for machine learning algorithm training.Parallel sentences in English and Bangla make up the bulk of our research dataset.To train and test the system, we will require some corresponding Bangla sentences for each English sentence.The dataset will be created manually and partially from online sources.English and Bangla sentences can be no longer than seven and eight words, respectively.Following the creation of the dataset, data preprocessing is done, during which the data is prepared in a way that makes it useable.

B. Sampling

There will be about 4500 entities in the collection.Each entity will have a sentence in English and a sentence that has been translated into Bangla.The dataset is divided into training and testing halves in an 80:20 ratio.

C. Preprocessing

To prepare the data for training and evaluation, we tokenized sentences in both Bangla and English.There are certain text preprocessing techniques taken in order to normalize the dataset.All punctuation is removed from sentences and all letters are changed to lowercase.Before tokenization, the characters that do not correspond to English or Bangla letters are also removed.It includes any language, including identifying numbers, dates, times, percentages, abbreviations, emails, URLs, and money.

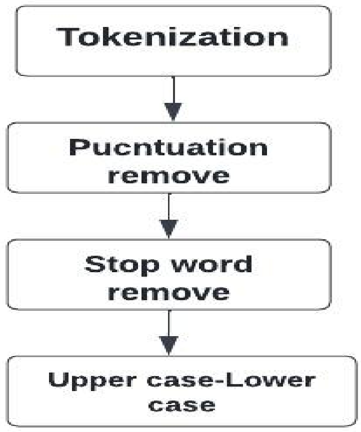

Figure: Preprocessing

The separation of tokens is then accomplished by applying a series of criteria using regular expressions.Such methods are crucial for tokenization since, frequently, problems like full stops (dots) being used in the middle of sentences make tokenization challenging when utilizing basic whitespace as delimiters.

D. Training of Neural Machine Translation

The goal is to translate into English-Bangla. There are basically 3 main models [14] .Also There will be a combination of these 3 models , to get better performance , training speed and Decoding Speed by using word probability, check-point ensemble,weight averaging.

E. Tools

For model design prototyping, Python and Anaconda Jupyter Notebook are used. The design model of the neural network is developed with Tensorflow-Keras python package distribution.

3.4 Requirements

- Processor: Intel(R) Core (TM) i5-1035G1 CPU @ 1.00GHz 1.19 GHz
- OS: Windows® 10
- RAM: 8.00 GB
- System type: x64-based processor

- Python Version: 3.7 (atleast)

Chapter 4

EXPECTED OUTPUTS / RESULTS

4.1 Expected Outcome:

The GRU with attention technique will perform better, especially in machine translation processes, by identifying the information in an input that is most relevant to completing a task.The dataset will be created for the research contains a wide variety of things, including phrases of various lengths and a sizable vocabulary.In relation to the neural machine translation model.The accuracy numbers will increase as the dataset size does as well.

- ➢ The GRU with attention technique will perform better
- ➢ The accuracy numbers will increase as the dataset size does as well.
- ➢ The Combination technique will increase the performance ,training time and decoding time.
- ➢ Lastly, Create a model which can be used for other languages with proper training.

Chapter 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

The GRU with attention approach identifies the information in an input most pertinent to accomplishing a task and gives better performance especially in machine translation processes. With increase in size of the dataset the accuracy figures will increase too. For future enhancement the quality of the dataset can be improved and the resources to train the NMT model can be modified so that faster model training and testing can take place.

5.2 Future Work

In the future the work can be implemented to the inner method of the existing work.With more data, the models will perform better. So there is a scope to test these models with more data. There is a scope to test these models with a different kind of source then the proposed model will perform more accurately.It can be used for other language translation.

Chapter 6

REFERENCE

[1] "What is Machine Translation? Guide to Machine Translation." *AWS*, https://aws.amazon.com/what-is/machine-translation/

[2] "BANGLA." *Association of Community Language Schools*, https://www.communitylanguageschools.org/services.

[3] Ali, Sadeq. "Bengali language." *Wikipedia*, https://en.wikipedia.org/wiki/Bengali_language

[4] Siddique, S., Ahmed, T., Rifayet Azam Talukder, M., & Mohsin Uddin, M. (2020). English to Bangla Machine Translation Using Recurrent Neural Network. *International Journal of Future Computer and Communication*, *9*(2), 46–51. https://doi.org/10.18178/ijfcc.2020.9.2.564

[5] Gupta, P., & Joshi, B. K. (2022). *Natural Language Processing based Refining Hindi to English Machine Translation*. *Icaaic*, 849–854. https://doi.org/10.1109/icaaic53929.2022.9792969

[6] Singh, J., Sharma, S., & J, B. (2022). *Natural Language Processing based Machine Translation for Hindi-English using GRU and Attention*. *Icaaic*, 965–969. https://doi.org/10.1109/icaaic53929.2022.9793214

[7] Hochreiter, Sepp, and ZHANGWu YUEHan ZHANG. "*The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions | International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems.*" *World Scientific*, https://www.worldscientific.com/doi/abs/10.1142/s0218488598000094.

[8] Hu, Y., Huber, A., Anumula, J., & Liu, S.-C. (2018). *Overcoming the vanishing gradient problem in plain recurrent networks*. *Section 2*, 1–20. http://arxiv.org/abs/1801.06105

[9] "Sequence to Sequence Learning with Neural Networks." *NIPS papers*, https://proceedings.neurips.cc/paper/2014/hash/a14ac55a4f27472c5d894ec1c3c743d2-Abstract.html.

[10] Gupta, P., & Joshi, B. K. (2022). *Natural Language Processing based Refining Hindi to English Machine Translation*. *Icaaic*, 849–854. https://doi.org/10.1109/icaaic53929.2022.9792969.

[11] Kim, H., & Lee, J. H. (2016). A recurrent neural networks approach for estimating the quality of machine translation output. *2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2016 - Proceedings of the Conference*, 494–498. https://doi.org/10.18653/v1/n16-1059

[12] Bapna, A., Chen, M. X., Firat, O., Cao, Y., & Wu, Y. (2018). Training deeper neural machine translation models with transparent attention. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018*, 3028–3033. https://doi.org/10.18653/v1/d18-1338

[13] Alkhouli, T., Bretschner, G., & Ney, H. (2018). On the Alignment Problem in Multi-Head Attention-Based Neural Machine Translation. *WMT 2018 - 3rd Conference on Machine Translation, Proceedings of the Conference*, *1*, 177–185. https://doi.org/10.18653/v1/w18-6318

[14] He, Q., Huang, G., Cui, Q., Li, L., & Liu, L. (2021). Fast and accurate neural machine translation with translation memory. *ACL-IJCNLP 2021 - 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 3170–3180. https://doi.org/10.18653/v1/2021.acl-long.246

[15] Cho, K., van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder–decoder approaches. *Proceedings of SSST 2014 - 8th Workshop on Syntax, Semantics and Structure in Statistical Translation*, 103–111. https://doi.org/10.3115/v1/w14-4012