e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

# LANGUAGE TRANSLATION

**Shivam Tiwari*1**

*1UG Student, Department Of Information Technology, B.K Birla College Of Arts, Science And Commerce (Autonomous) Kalyan, India.

## ABSTRACT

Language Translation can be utilized for translating text from English to Hindi. Language is core medium of communication and translation is core tool to understand the information in unknown language. Language translation helps the people to understand the information of unknown language without the help of Human translator. Transformer architecture is used for creating new model and this self-trained model is used to translate text from English to Hindi as the main objective is to translate text from English to Hindi.

**Keywords:** Natural Language Processing, Language, Translation, Communication, Hindi, Neural Machine Translation.

## I. INTRODUCTION

Translation is the communication of the meaning of a source-language text by means of an equivalent target-language text. Language translation also known as machine translation is the process of converting the written word from one language into another language in a way that is culturally and linguistically appropriate so it can be understood by its intended audience. People usually face many miscommunications while communicating with other culture language. Language Translation helps people to communicate, connect and grow. Also, the education system in India has made a change that all education can be provided in a regional language. Here Language Translation comes into the picture since till now all the books of higher education are printed in English language only. While translating books replacing technical words is a task which our system is going to overcome.

### OBJECTIVE OF STUDY

In the world, we have multiple and hugely diverse languages and scripts, hence scope and need for language translation are immense. As the world is becoming an integrated place due to remote job opportunities it becomes difficult for the person to establish communication with abroad localities. Also, consuming knowledge from different cultures fails due to language barriers. Education in regional language also requires translation of already available books into regional language. Main aim is to translate English to Hindi.

## II. LITERATURE REVIEW

### A. LIST OF MODELS

#### 1. CNN ENCODER-DECODER:

Each input element embedding is combined with its positional embedding (which signifies the position of the input element). Positional embeddings help the network to realize what part of input it is dealing with, currently. Encoder-Decoder. Both the encoder and decoder are CNN blocks along with a multistep attention mechanism with multiple 'hops' (Sukhbaatar et al., 2015). Each block consists of one-dimensional convolutions followed by a Gated Linear Unit (GLU) non-linearity (Dauphin et al., 2016). GLU is a gating function over the outputs of the convolutions. The multi-step attention mechanism suggests that the attention mechanism is applied to every layer inthe decoder. The attention of the first layer gives contextual information which is then given as an input to the next layer that considers this information while calculating the attention weights of the current layer.

#### 2. RNN ENCODER - DECODER:

The typical architecture encodes the sequence of source word embeddings to generate annotations for the source words. The encoder is typically a bi-directional RNN layer of LSTM or GRU units. The final state of the encoder is used to initialize the decoder. The decoder is also an RNN which generates one output token at a time. Each output token is predicted based on the decoder state, previous output word and the context vector. The context vector encodes source information required for predicting the words and is generated using an attention mechanism on the source word annotations.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

**3. ENCODER-DECODER WITH ATTENTION BASED MECHANISM:**

Language translation has seen a transformative shift with neural machine translation (NMT), particularly through the adoption of encoder-decoder models. These models, enhanced by attention mechanisms, have significantly improved translation quality and fluency. This review provides a synthesized overview of the evolution of NMT, underscoring the pioneering work by researchers like Bahdanau and their attention mechanisms, which revolutionized the field. It offers insight into the contemporary state-of-the-art in encoder-decoder models with attention for language translation.

**4. TRANSFORMER:**

The Transformer architecture represents a pivotal advancement in the field of natural language processing. Introduced by Vaswani et al. in 2017, it dispenses with recurrent neural networks (RNNs) and relies on self-attention mechanisms. Transformers have enabled more efficient and parallelized training, resulting in state-of-the-art performance on various NLP tasks. The architecture's key innovation is the self-attention mechanism, allowing the model to weigh the importance of different parts of the input sequence during both encoding and decoding. Transformers have since become the foundation of various pre-trained language models, such as BERT, GPT, and RoBERTa, and have revolutionized the way we approach tasks like language understanding and generation.

**B. EXISTING SYSTEM**

**1. GOOGLE TRANSLATE:**

Google Translate is a multilingual neural machine translation service developed by Google to translate text, documents and websites from one language into another. It offers a website interface, a mobile app for Android and iOS, and an API that helps developers build browser extensions and software applications.

**2. MICROSOFT TRANSLATE;**

Microsoft Translator is a multilingual machine translation cloud service provided by Microsoft. Microsoft Translator is a part of Microsoft Cognitive Services and integrated across multiple consumer, developer, and enterprise products; including Bing, Microsoft Office, SharePoint, Microsoft Edge, Microsoft Lync, Yammer, Skype Translator, Visual Studio, and Microsoft Translator apps for Windows, Windows Phone, iPhone and Apple Watch, and Android phone and Android Wear.

**3. AMAZON TRANSLATE:**

Amazon Translate is a neural machine translation service that delivers fast, high-quality, affordable, and customizable language translation. Amazon Translate lets you localize content for diverse global users and translate and analyze large volumes of text to activate cross-lingual communication between users.

## III. METHODOLOGY

**PROPOSED SYSTEM**

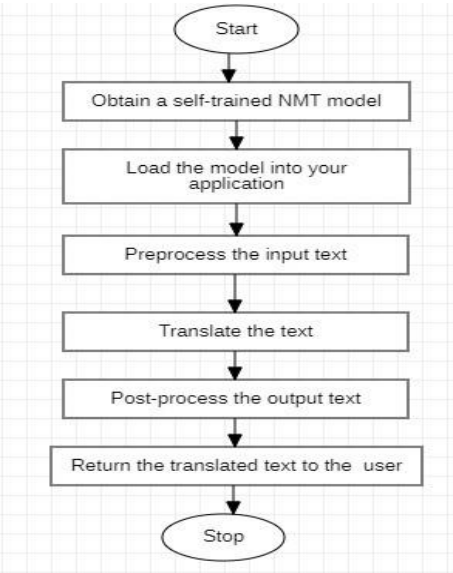


**Fig 1:** Program Flow

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

**A. ENCODER-DECODER WITH ATTENTION BASED MECHANISM:**

**Encoder**: The encoder network plays a vital role in extracting features from the input data. It generates a context ve-ctor, an internal state that encapsulates crucial information from the input. This context vector is then passed on to the decoder to ensure accurate pre-dictions.

**Decoder**: The decoder is responsible for inte-rpreting the context vector derived from the encoder's final cell and generating the output sequence. It employs a series of LSTM units, each producing an output and a hidden state. These- states are then passed on to the subsequent unit for further network processing.

**ATTENTION MECHANISM BASIC WORKING:**

Attention makes sequence models smarter by letting them pay extra attention to important stuff in their input. Imagine it's like you're reading a recipe, and you want to bake a cake. Instead of just following the recipe step by step, attention is like highlighting the most crucial parts, like the ingredients and the oven temperature. This way, when you're baking the cake, you know exactly where to focus your attention. So, attention in these models helps them highlight what's essential, making the whole process more flexible and efficient.

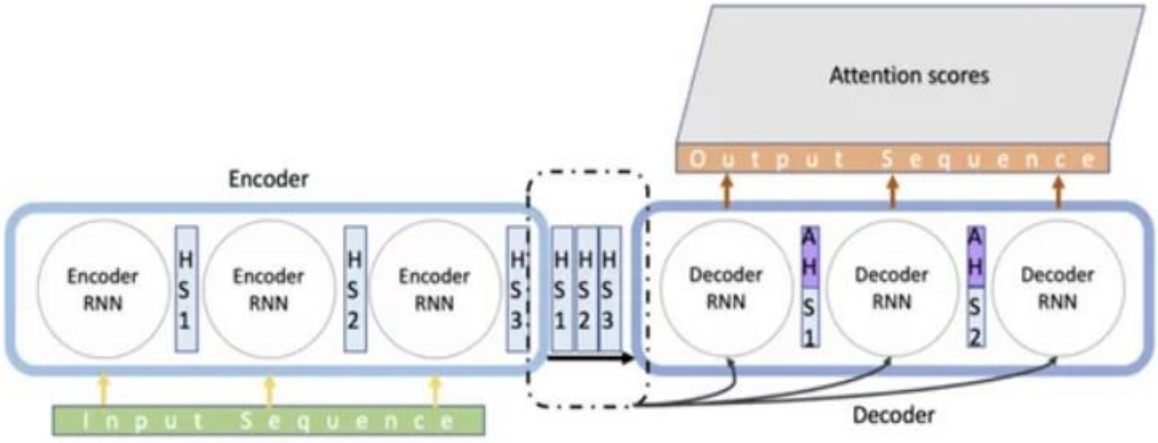


**Fig 2:** Basic Architecture Of Attention Based Model

The attention-based sequence-to-sequence model introduces a neat trick. It makes the output depend on select elements in the input. These chosen elements, which receive attention, become sort of rules that guide the model in generating the desired results. Although attention models demand substantial computational resources, they outperform traditional models. They also reveal how the model pays attention to the input when predicting the output. This insight helps in understanding precisely what the model considers for specific input-output pairs. Unlike the traditional model, which encodes input into a single fixed context vector, the attention model goes a step further. It creates a context vector tailored to each output time step, allowing the model to focus on specific important words and use them to train the decoder for accurate predictions.

**B. TRANSFORMER ARCHITECTURE:**

**Encoder**: The encoder is composed of a stack of N = 6 identical layers. Each layer has two sub-layers. The first is a multi-head self-attention mechanism, and the second is a simple, position2 Figure The Transformer architecture. wise fully connected feed-forward network. We employ a residual connection around each of the two sub-layers, followed by layer normalization. That is, the output of each sub-layer isLayerNorm(x + Sublayer(x)), where Sublayer(x) is the function implemented by the sub-layer itself. To facilitate these residual connections, all sub-layers in the model, as well as the embedding layers, produce outputs of dimension dmodel = 512.

**Decoder**: The decoder is also composed of a stack of N = 6 identical layers. In additionto the two sub-layers in each encoder layer, the decoder inserts a third sub-layer, which performs multi-head attention over the output of the encoder stack. Similar to the encoder, we employ residual connections around each of the sub-layers, followed by layer normalization. We also modify the self-attention sub-layer in the decoder stack to prevent positions from attending to subsequent positions. This masking, combined with fact that the output embeddings are offset by one position, ensures that the predictions for position i can depend only on the known outputs at positions less than i.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

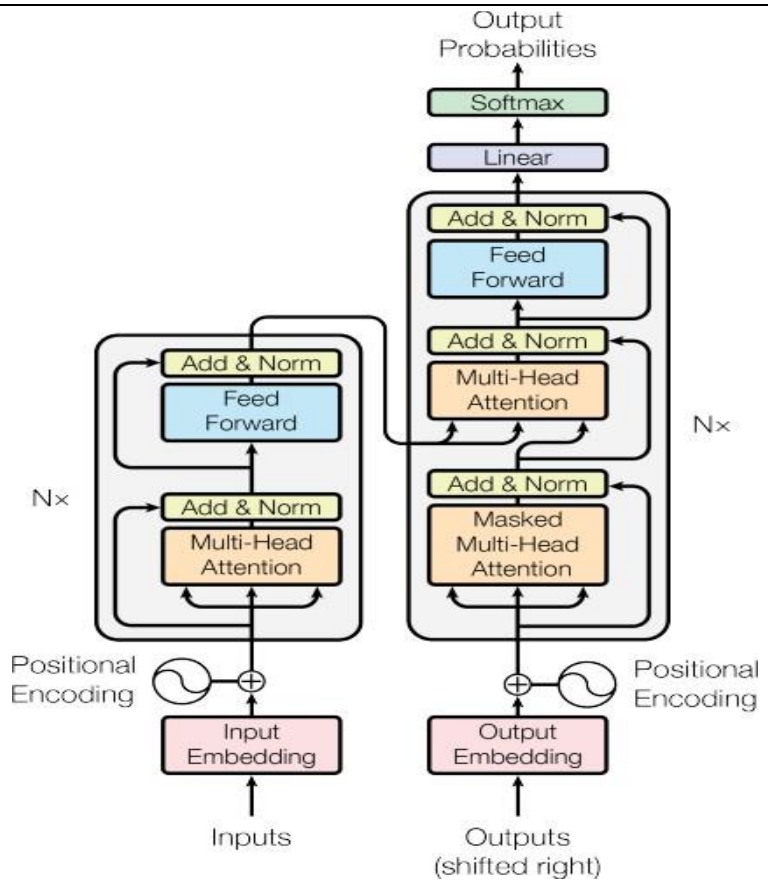


**Fig 3:** Transformer Architecture

**IMPLEMENTATION**

**Step 1: Dataset**

IIT Bombay English-Hindi Corpus is selected.

**Step 2: Pre-processing**

Tokenization: Each character and word is assigned a number, conventionally called character IDs and word IDs, respectively.

Padding: Since the sentences are dynamic in length, padding is done at the end of sentences to make them the same length. Creating word embeddings and positional encoding matrix.

**Step 3: Model Selection**

Evaluating different Natural Language Processing models to choose the best amongall.

Transformer architecture(Attention mechanism) is used.

**Step 4: Model Training**

To train model using English sentence and its parallel hindi sentence.

Here the model is trained using tokenize vectors not the actual words of language.

## IV. EXPERIMENT RESULT

**TRAINING DETAILS:**

Encoder decoder with attention:

Epochs 40

Loss 0.069

Time required: 50 hours

Transformer:

Epochs 40

Loss 4.757

Time required: 90 hours

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

**Table 1:** Comparing Translation Of Two Models

| English | Hindi (encoder-decoder with attention) | Hindi (Transformer model) |
|---|---|---|
| The website of the university opens in new window | विश्वविद्यालय में एक ऑनलाइन का अंतर्राष्ट्रीय जालपृष्ठ हूँ। | राष्ट्रीय प्रौद्योगिकी विश्वविद्यालय बाहरी वेबसाइट जो एक नई विंडो में खुलती हैं |
| Do you want to create this | क्या आप को अवसर चाहते हैं | क्या आप निश्चित हैं कि आप इस कार्य को सहेजना चाहते हैं |
| Agriculture is backbone of our country's economy | कृषि का देश है कि गरीबी हमारी अर्थव्यवस्था है। | हमारे देश की अर्थव्यवस्था में वृद्धि |
| Politicians do not have permission to do | राजनीतिज्ञों के पास करने को तैयार करने की अनुमति नहीं है | पुलिस की सुरक्षा की कोई शिकायत नहीं है। |
| load the saved file | तो यह तो सिर्फ | फ़ाइल ' 1' को लोड करें |
| I can do this kind of work | मैं इस काम के काम करती हैं। | मैं इस बारे में कुछ कर सकता हूँ |
| Download the application from website | डाउनलोड की सुविधा है। | वेब ब्राउज़र को डाउनलोड करें |
| Program execution is in progress | नीति का हल हो रहा है | प्रोग्राम के लिए प्रोग्राम |
| Program output is invalid | जिसका उत्पादन है | प्रोग्राम नाम 1 के लिए अवैध है |

## V. CONCLUSION

Language Translation also known as Machine Translation is the approach to translate one language into another. The project revolves around translating English text to Hindi text. Two models were evaluated to translate from English to Hindi: The encoder decoderwith attention and Transformer. The encoder-decoder model performed well in the case of direct, straight meaning sentences. The transformer model performed well incase of technical words translation. Both have alternate drawbacks and benefits.

Machine Translation is a useful technology and so is the current research topic in the Natural Language Processing domain. It is more challenging for translating regional languages due to its various semantics and structure.

1. Transformer model can be trained on a large dataset and then its translations can beevaluated.
2. Encoder-decoders with attention model can be trained on technical or domainspecific datasets.
3. Both the models can be used to actually translate the English textbooks into Hindi language.

## VI. REFERENCES

[1] Ilya Sutskever, Oriol Vinyals, Quoc V. Le, Sequence to sequence learning with theneural network, NIPS 2014.

[2] Sandhya Singh, R. Panjwani, P. Bhattacharyya, Comparing Recurrent andConvolutional Architectures for English-Hindi Neural Machine Translation, WAT@IJCNLP 2017.

[3] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, Illia Polosukhin, Attention is all you need, NIPS 2017.

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 24 May2019.

[5] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang , Michael Matena , Yanqi Zhou, Wei Li, Peter J. Liu, Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer, 2020.

[6] Middi Venkata Sai Rishita1, Middi Appala Raju and Tanvir Ahmed Harris, Machine translation using.