Volume 9, Issue 7, July – 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL767

Bilingual Neural Machine Translation From English To Yoruba Using A Transformer Model

Adeboje Olawale Timothy[1]
Department of Mathematical and Computing,
Koladaisi University Ibadan, Nigeria

Adetunmbi Olusola Adebayo[2]
Department of Computer Science,
Federal University of Technology, Akure. Nigeria.

Arome Gabriel Junior[3]
Department of Cybersecurity,
Federal University of Technology, Akure. Nigeria

Akinyede Raphael Olufemi[4]
Department of Information Systems,
Federal University of Technology, Akure. Nigeria.

Abstract:- The necessity for language translation in Nigeria arises from its linguistic diversity, facilitating effective communication and understanding across communities. Yoruba, considered a language with limited resources, has potential for greater online presence. This research proposes a neural machine translation model using a transformer architecture to convert English text into Yoruba text. While previous studies have addressed this area, challenges such as vanishing gradients, translation accuracy, and computational efficiency for longer sequences persist. This research proposes to address these limitations by employing a transformer-based model, which has demonstrated efficacy in overcoming issues associated with Recurrent Neural Networks (RNNs). Unlike RNNs, transformers utilize attention mechanisms to establish comprehensive connections between input and output, improving translation quality and computational efficiency.

Keywords:- *NLP, Text to Text, Neural Machine Translation, Encoder, Decoder, BERT, T5.*

I. INTRODUCTION

The existence of man on the surface of the earth is identified as the existence of communication. In the beginning, God created man in His image and since then there has been a communication between man and God and between men (Genesis 1:27-29). This simply shows that communication is the basis of human animal existence. Communication is the means of sending and receiving information, ideas, thoughts, and feelings between individuals or groups. It involves the transmission and reception of messages through various channels such as speech, writing, gestures, body language, and even through technological means like emails or social media. This is the reason why the desire to communicate is essential in man and this is exhibited even by babies, who in their own way express their needs and feelings by crying, laughing, groaning, etc., even animals communicate to one another, while machines on the other hand communicate at their lowest levels using discrete values of 1s and 0s called bits i.e., binary digits.

Communication can be categorized into two: verbal/oral and non-verbal. Non-verbal communication, which predates verbal communication, does not rely on words but instead employs gestures, cues, vocal tones, spatial relationships, and other non-linguistic elements to convey messages. It is often used to express emotions such as respect, love, dislike, or discomfort and tends to be spontaneous and less structured compared to verbal communication.

Verbal/oral communication involves the use of spoken words or language to convey information. It requires organizing words in a structured and grammatically correct manner to effectively communicate ideas to the audience. Verbal communication offers several advantages over non-verbal communication, including efficiency in time and resource utilization, support for teamwork, prompt feedback, flexibility, and enhanced transparency and understanding between communicators.

Nigeria is a culturally diverse society with a multitude of languages [22]. Estimates suggest there are between 450 to 500 languages spoken throughout the country. While English serves as the official language, Nigeria has also designated three regional languages: Hausa, spoken by approximately 20 million people in the North; Yoruba, spoken by about 19 million in the West; and Igbo, spoken by around 17 million in parts of the South-East. These three languages are considered major languages.

Additionally, Nigeria recognizes about 500 other languages, some of which hold varying degrees of official status within their regions and serve as lingua franca, such as Efik, Ibibio, and Fulani. Among these languages, Nigerian Pidgin serves as a national lingua franca.

In the Nigerian context, "majority" languages refer specifically to Hausa, Yoruba, and Igbo, while "minority" languages encompass those spoken by smaller ethnic groups. These distinctions are formally acknowledged and hold significant roles in various sectors, particularly in education within the country.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL767

English Language has since continued to gain so much prominence in the country that its dominance has stifled the growth (and even led to the extinction of some) of the 529 indigenous languages in Nigeria [18]. English holds a dominant position across all sectors in Nigeria, including government, education, media, judiciary, and science and technology. Government officials often refrain from using indigenous languages even in informal settings to avoid perceptions of tribalism. Both at the national and state levels, English remains the primary language for debates and official records, despite efforts to promote indigenous languages. Due to Nigeria's linguistic diversity, English plays a central role in national unity and administrative functions. Legal proceedings, particularly in the Supreme Court and High Courts, are primarily conducted in English, with lawyers presenting arguments and judges delivering judgments in the same language.

However, the prevalence of the English language in Nigeria, particularly in the southwestern region, has resulted in the diminishing use of our esteemed Yoruba language, notably among its speakers. This shift has profound cultural and linguistic implications for national identity, leading to many children being unable to communicate in their native tongue. As a result, many children are unable to speak their native language. Language and culture are inseparable, and the decline in the use of Yoruba among younger generations has contributed to the erosion of our cultural heritage. Consequently, younger people are increasingly losing touch with the core values and traditions of their cultures. This shift is also evident in their fashion choices, which now mirror those of the predominant language speakers.

Research conducted by [19] showed that using the mother tongue in primary education significantly benefits learning outcomes. In a six-year primary education project, it was observed that pupils taught in Yoruba performed better compared to those taught solely in English. It was noted that "the child learns better in his mother tongue, which is as natural to him as his mother's milk." However, if English continues to serve as Nigeria's national and educational language, as well as its lingua franca for unity in a diverse and multilingual country, its pervasive use in media, legislation, and throughout society could pose a threat to indigenous languages. This trend suggests that indigenous languages are not merely at risk but are gradually declining [20].

To break the language barrier, there arise the language translator. Language translators are essential tools that enable communication between individuals or groups who speak different languages. Initially, human language translators are there to translate languages. However, human language translators face several challenges and problems, despite their invaluable role in facilitating communication across languages. Some of these challenges include accuracy, precision, ambiguity etc. Neural machine translation (NMT) has revolutionized the field of translation due to several key advantages it offers over traditional approaches.

Neural machine translation encompasses various forms: text to text, text to speech, speech to speech, and speech to text. Natural languages are categorized into high-resource languages (HRL) and low-resource languages (LRL) based on the availability of textual and speech data on the web [17]. High-resource languages, such as English, benefit from abundant data resources that facilitate machine learning and understanding. Similarly, many Western European languages like French and German, as well as Chinese, Japanese, and Russian, fall into this category. In contrast, low-resource languages lack sufficient data resources, making them less studied, resource-scarce, and less commonly taught. These languages are often characterized by limited computational tools and sparse textual and speech data online [21]

The Yoruba language falls under the category of low-resource languages (LRL), facing several significant challenges. One of the primary issues is the difficulty in employing alignment techniques across different levels (word, sentence, and document) for annotation, creating a cohesive dataset, and collecting raw text. These tasks are essential for any natural language processing (NLP) endeavor or mapping technique [17].

Furthermore, the lack of adequate textual material poses a critical challenge, as lexicons form the backbone for many NLP tasks. Developing an effective lexicon for LRLs like Yoruba remains particularly challenging due to limited textual resources. Additionally, the morphology of LRLs such as Yoruba is dynamic and continuously evolving, leading to a diverse and expansive vocabulary. This variability complicates the development of a robust framework for morphological pattern recognition, given the presence of multiple roots and complex word formations.

This research is set to propose a text-to-text neural machine translator for English language to Yoruba language to address the challenges of low resource language as Yoruba language, extinction of Yoruba Language, inability of Yoruba children to speak Yoruba language, vanishing of Yoruba culture, and the aforementioned challenges using a transformer model. The Transformer model is a type of neural network [15] initially recognized for its exceptional performance in machine translation. Today, it has become the standard for constructing extensive self-supervised learning systems. In recent years, Transformers have surged in prominence not only within natural language processing (NLP) but also across diverse domains like computer vision and multi-modal processing. As Transformers advance, they are progressively assuming a pivotal role in advancing both the research and practical applications of artificial intelligence (AI) [16].

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL767

II. LITERATURE REVIEW

Many researchers have worked in the area of Neural Machine Translation, most especially translation of English language to Yoruba Language. Such among them are:

In their work titled "Web-Based English to Yoruba Machine Translation" [24], the researchers aimed to achieve English-to-Yoruba text translation using a rule-based method. Their approach primarily utilized a dictionary-based rule-based system, considered the most practical among various types of machine translation methods. However, a limitation of this approach was its inability to accurately translate English words with multiple meanings, influenced by their grammatical context.

In their study titled "Development of an English to Yorùbá Machine Translator" [1], the researchers aimed to create a machine translator that converts English text into Yorùbá, thereby making the Yorùbá language more accessible to a wider audience. The translation process employed phrase structure grammar and re-write rules. These rules were developed and evaluated using Natural Language Tool Kits (NLTKs), with analysis conducted using parse tree and Automata theory-based techniques. However, the study noted that accuracy decreases as the length of sentences increases.

In [2], titled 'Machine to Man Communication in Yoruba Language,' the research aims to establish communication between humans and machines using a Text-To-Speech system for the Yoruba language. The study involves text analysis, natural language processing, and digital signal processing techniques. An observed outcome of the research is that pronunciation accuracy decreases as sentence length increases.

[3] in “Development of a Yorúbà Text-to-Speech System Using Festival”. The objective of the research is to develop a speech synthesis for Yoruba language. Speech data were recorded in a quiet environment with a noise cancelling microphone on a typical multimedia computer system using the Speech Filing System software (SFS), analysed and annotated using PRAAT speech processing software. However, the speech produced by the system is of low quality and the measure of naturalness is low.

In [4], titled 'Token Validation in Automatic Corpus Gathering for Yoruba Language,' the research aims to create a language identifier that identifies potentially valid Yoruba words. The approach involves using a dictionary lookup method to collect words, which is implemented as a Finite State Machine using Java. However, the system's capability is currently restricted to gathering monolingual Yorùbá language data.

In [5], titled 'Developing Statistical Machine Translation System for English and Nigerian Languages,' the research aims to create a translation system between English and two Nigerian languages: Igbo and Yorùbá. The study employs a phrase-based Statistical Machine Translation approach, which involves grouping source language words into sequences, translating each phrase into the target language, and optionally reordering them based on target language models and distortion probabilities. However, the research faces challenges such as insufficient data, orthographic errors, and high error rates during system compilation.

In [6] titled Japanese-to-English Machine Translation Using Recurrent Neural Networks system. The research objective is to translate Japanese language to English language. A bidirectional recurrent neural network (BiRNN) was used as the encoder with the forward hidden state of an RNN that reads the source sentence as it is ordered and the backward hidden state of an RNN that reads the source sentence in reverse. However, there was no complex sentence translations.

In "[7], titled 'Arabic Machine Translation Using Bidirectional LSTM Encoder-Decoder,' the research aimed to develop a model primarily based on Bidirectional Long Short-Term Memory (BiLSTM). The objective was to map input sequences to vectors using BiLSTM as an encoder, and subsequently use another Long Short-Term Memory (LSTM) for decoding the target sequences from these vectors. The input sentences were sequentially embedded into the BiLSTM encoder word by word until the end of the English sentence sequence. The hidden and memory states obtained were fed into the LSTM decoder to initialize its state, and the decoder output was processed through a softmax layer and compared with the target data. However, the system faced limitations due to being trained on a limited number of words in the corpus.

In "[8], titled 'Using Bidirectional Encoder Representations from Transformers for Conversational Machine Comprehension,' the goal was to apply the BERT language model to the task of Conversational Machine Comprehension (CMC). The research introduced a novel architecture called BERT-CMC, which builds upon the BERT model. This architecture incorporates a new module for encoding conversational history, inspired by the Transformer-XL model. However, experimental findings indicated that Transformer-based architectures struggle to fully capture the contextual nuances of conversations.

In "[9], titled 'Development of a Recurrent Neural Network Model for English to Yorùbá Machine Translation,' the aim was to create a recurrent neural network (RNN) model specifically for translating English to Yorùbá. The model underwent testing and evaluation using both human and automatic assessment methods, demonstrating adequate fluency and providing translations of decent to good quality. However, the research had limitations: it could only handle single tasks and a limited vocabulary size, and it lacked the capability to expand its vocabulary extensively.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL767

In [10], titled 'Design and Implementation of English to Yorùbá Verb Phrase Machine Translation System,' the research aimed to create a system for translating English verb phrases into Yorùbá. However, the study faced several limitations: the evaluation method was time-consuming, the system's engine lacked robustness, which hindered the addition of new rewrite rules, and the system showed low responsiveness to queries due to a limited database size.

In "[11], titled 'A Novel Technique in the Development of Yorùbá Alphabets Recognition System Through the Use of Deep Learning Using Convolutional Neural Network,' the objective was to create a deep convolutional network named YORÙBÁNET for detecting Yorùbá alphabets. The model was implemented in the Matlab R2018a environment using a custom framework. The training dataset consisted of 10,500 samples, with 2,100 samples reserved for testing. Training the model involved 30 epochs, with 164 iterations per epoch, totaling 4,920 iterations. The entire training process took approximately 11,296 minutes and 41 seconds. This research focused solely on Yorùbá alphabet recognition and did not involve sentence translations.

In "[12], titled 'Sentence Augmentation for Language Translation Using GPT-2,' the research aimed to explore the use of GPT-2 for generating monolingual data to enhance Machine Translation. The study utilized a sentence generator based on GPT-2 to produce additional data for neural machine translation (NMT), aiming to maintain similar characteristics to the original text. However, the research did not assess the effectiveness of the more recent GPT-3 model in augmenting sentences for language translation.

In "[13], titled 'Linguistically-Motivated Yorùbá-English Machine Translation,' the research aims to analyze and provide linguistic descriptions of errors made by different models. Three sentence-level models—SMT, BiLSTM, and Transformer—were trained to translate from Yorùbá to English. However, these models exhibited shortcomings such as missing words, incorrect words or spellings, semantic and syntactic errors, as well as issues with grammar and word order.

In "[14], titled 'Convolutional Recurrent Neural Network (CRNN) for the Recognition of Yorùbá Handwritten Characters,' the collected data underwent preprocessing steps including grayscale conversion, binarization, and normalization to eliminate distortions from the digitization process. The convolutional recurrent neural network model was then trained using these preprocessed images. However, the model exhibited low recognition accuracy for characters containing under dots and diacritic signs.

III. METHODOLOGY

This research proposed to translate English language text into Yoruba language text using transformer model. Figure 1 shows the architecture of a transformer – model.

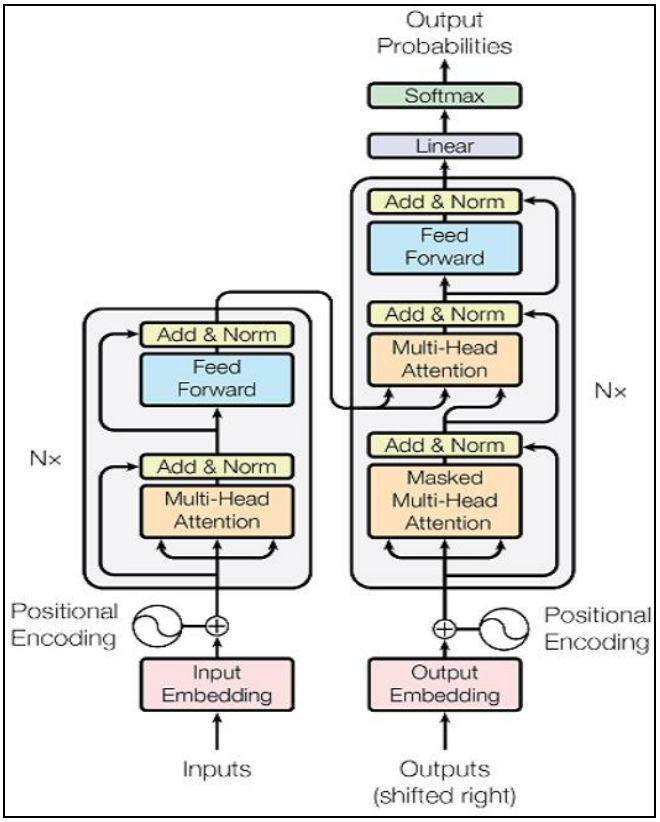

Fig 1 The Transformer - Model Architecture

The Transformer model is divided into two macro blocks, which are Encoder and Decoder.

➢ *Encoder*

It is responsible for language learning and identification. The encoder encodes the input sequence and passes it to the decoder. The encoder block 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 fully connected feed-forward network. The research will make use of Bidirectional Encoder Representations from Transformers (BERT) for the encoder. Bidirectional Encoder Representations from Transformers (BERT) is a language representation model, designed to pre-train deep bidirectional representations, with the goal of extracting context-sensitive features from the input text. It is a neural network-based technique for language processing pre-training.

- *Dataset selection*

The selection of datasets was a critical step in our methodology. The system will make use of a parallel corpus obtained from MENYO-20k and odunola/yoruba-english-pairs (https://huggingface.co/datasets/odunola/yoruba-english-pairs) to have a robust Yoruba corpus. This is the source language in string (English language), which is expected to be translated by the model.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL767

- *Text representation and modeling techniques*

In this section, the research explores the text representation and modeling techniques to be employed in this research.

The input (raw text) is transformed into embedding vector that can easily be understood by the BERT algorithm. The input (string) will be converted into set of tokens, the token goes through token embedding where it is vectorize which have contextual meaning.

Consider the embedding of the $i-th$ word in the sequence, denoted by

$$word_{input} = x_i \in R^d \quad (1)$$

Where x_i is the word in the sequence, R^d is the transformation matrix and $word_{input}$ is the transformed matrix.

For example, lets consider the two sentences: Sentence 1: He is a good boy. Sentence 2: Is he a good boy

Step 1: Apply word Piece Tokenization. Also, CLS and SEP tokens will be added. CLS is added to the start of the input sentence and SEP tokens to the end of the input token.

Step 2: Token embedding ($\boldsymbol{TE}$) : Lookup the 768 dimension vector dimension. This is the input identification. (1, 5,768). Where 1denotes the first sentence, 5 denotes the number of words in sentence 1 and 768 denotes the BERT model transformation.

Step 3: Segment Embedding ($\boldsymbol{SE}$): This is much important when there are multiple sentences in the input sequence. Sentence 1=0 and Sentence 2= 1

Step 4: Apply Position Embedding ($\boldsymbol{PE}$): This determines the index position of the individual token in the input sequence. We want the model to treat words that appear close to each other as "close" and words that are distant as "distant". Once Position Encoding is computed, it will be reused for every sentence during the training and interference.

Positional Embedding is calculated for time set to even and odd respectively in equation 2 and 3:

$$PE\,(POS,\, 2i) = \sin \frac{POS}{1000^{2i/dmodel}} \quad (2)$$

$$PE\,(POS,\, 2i+1) = \cos \frac{POS}{1000^{2i/dmodel}} \quad (3)$$

Where PE = position encoding, POS = position of the sentence, i = index position in the input sequence, d_{model}= dimension of the vector.

This is the sum of position, segment and token embedding to form the input for the BERT model

$$BERT_{input} = PE + \text{TE} + \text{SE} \quad (4)$$

Where PE is the positional encoding, TE is the Token encoding and SE is the segment Encoding

$BERT_{input}(\ seq,\ d_{model})$ where seq= sequence of length and d_{model} is the size of the embedding vector.

From figure 1, It is observed that the Add and Norm takes the input of Multi-Head attention and a residual connection from the positional encoding. The residual connection is done to ensure that there is a strong information signal that flows through deep networks. This is required because during back propagation, there is vanishing gradient, so to prevent that, we induce small strong signal from the input to different parts of the network.

Residual connections carry over the previous embeddings to the subsequent layers. As such, the encoder blocks enrich the embedding vectors with additional information obtained from the multi-head self-attention calculations and position-wise feed-forward networks.

After the Multi Head attention, the result will be added to the embedded input using the residual connection.

$$v_{(seq,\, d_{model})} = x + MultiHead_{(Q,V,K)}(x) \quad (8)$$

Where x is the input embedded and $MultiHead_{(Q,V,K)}(x)$ is the resultant multi head attention process

The next is to perform a layer normalization and this is calculated by :

$$\text{Layer Norm}\,(v_{(seq,\, d_{model})}) = \gamma \frac{v-\mu}{\sigma} + \beta \quad (9)$$

Where μ= mean, σ = standard deviation, γ= gamma (multiplicative), β =beta (additive) and will initially set to 1 and 0.

The Transformer model utilizes "Add & Norm" blocks to facilitate efficient training. These blocks incorporate two essential components: a residual connection and a Layer Normalization layer. The residual connection establishes a direct path for the gradient, ensuring that vectors are updated rather than entirely replaced by the attention layers. This helps with gradient flow during training. On the other hand, the Layer Normalization layer maintains a reasonable scale for the outputs, enhancing the stability and performance of the model. Unlike batch normalization, layer normalization aims to reduce the effect of covariant shift. In other words, it prevents the mean and standard deviation of embedding vector elements from moving around, which makes training unstable and slow (i.e., we can't make the learning rate big enough). Unlike batch normalization, layer normalization works at each embedding vector (not at the batch level).

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL767

Normalization converts features into similar scale, this will improve the performance of the system.

Next is the Feed Forward Network (FFN) comprises two dense layers that are individually and uniformly applied to every position. The Feed Forward layer is primarily used to transform the representation of the input sequence into a more suitable form for the task at hand. This is achieved by applying a linear transformation followed by a non-linear activation function. The output of the Feed Forward layer has the same shape as the input, which is then added to the original input.

The FFN takes a vector v (the hidden representation at a particular position in the sequence) and passes it through two learned linear transformations, (represented by the matrices W_1 and W_2 and bias vectors b_1 and b_2 . A rectified-linear (ReLU) activation function applied between the two linear transformations.

$$FFN_{(v)} = max(0x\, W_1 + b_1\,)\, W_2 + W_2 \quad (11)$$

Where W_1 and W_2 and b_1 and b_2 are bias vector

After the Multi Head attention, the result is added to the embedded input using the residual connection.

$$v_{(seq,\, d\, model)} = x + MultiHead\ _{(Q,V,K)}(x) \quad (12)$$

Where x is the input embedded and $MultiHead\ _{(Q,V,K)}(x)$ is the resultant multi head attention process.

The next is to perform a layer normalization and this is calculated by :

$$\text{Layer Norm } (^{v}{}_{(seq,\, d\, model)}) = \gamma\, \frac{v-\mu}{\sigma} + \beta \quad (13)$$

Where μ= mean, σ = standard deviation, γ= gamma (multiplicative), β =beta (additive) and will initially set to 1 and 0.

The encoder outputs for each word a vector that not only captures its meaning (the embedding) or the position but also its interaction with other words of the multi head attention. The output of the encoder is another matrix that has the same dimension as the input matrices in which the sequence of embedding. The output of the encoder will be passed to the next encoder and the adjacent decoder.

➢ *Decoder*

The goal of the decoder is to translate English language to Yoruba language. The output of the Encoder (value and key) is joined to the output (Query) of the masked multi head attention in the Decoder phase. The output (shift right) is the trained language, however, this starts with the start of sentence (SOS).

The research will make use of T5-Base on the decoder side of the transformer. Text-to-Text Transfer Transformer (T5). T5-Base is the standard or baseline version of T5. It strikes a balance between model complexity and efficiency, making it widely used and well-suited for many NLP tasks. It offers a good trade-off between computational cost and performance and serves as a solid starting point for most applications. T5-Base strikes a good balance between computational efficiency and task performance, making it a widely used variant.

The output will be embedded as done in the encoding stage

$$T5_{output} = PE + \text{TE} + \text{SE} \quad (14)$$

Where PE is the positional encoding, TE is the Token encoding and SE is the segment E, $T5_{output}(\, seq,\, d_{model})$ where seq= sequence of length and d_{model} is the size of the embedding vector.

The $T5_{output}(\, seq,\, d_{model})$ will be divided into four, one will be sent to the Add and Norm phase and three will be sent to the Multi-Head Attention. The three will be Query (Q), Key (K) and Value (V).

The attention between the Q,K and V is calculated by

$$Attention\ (Q,K,V) = \ softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)v \quad (15)$$

The goal here is to make the model casual. It means the output at a certain position can depend on the words on the previous position. The model must not be able to see the future words.

$$head_i = Attention(QW_i^Q\, , KW_i^k\, , VW_i^v\,) \quad (16)$$

With the new head, the next is to concatenate the heads by :

$$MultiHead(Q,\, K,\, V\,) \ = \ Concat(head_i,\ \ldots, head_h)W^O \quad (17)$$

Where Q=Query, K= Key and V=Value

After the Multi Head attention, the result is added to the embedded output using the residual connection

$$T5_{output}\ \ _{(seq,\, d\, model)} = x + MultiHead\ _{(Q,V,K)}(x) \quad (18)$$

Where x is the input embedded and $MultiHead\ _{(Q,V,K)}(x)$ is the resultant multi head attention process

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL767

The next is to perform a layer normalization and this is calculated by :

$$\text{Layer Norm } (T5_{output\,(seq,\,d\,model)}) = \gamma \frac{v-\mu}{\sigma} + \beta \quad (19)$$

Where μ= mean, σ = standard deviation, γ= gamma (multiplicative), β =beta (additive) and will initially set to 1 and 0.

Next is the multi-Head attention, this can be called a cross- attention because the output of the encoder (key and value) is joined with the output of the masked multi head attention, which is the query. The output of the encoder is represented by $v_{(seq,\,d\,model)}$ and the query (output of the masked multi head) is represented as $T5_{output\,(seq,\,d\,model)}$ The attention between the $v_{(seq,\,d\,model)}$ and $T5_{output\,(seq,\,d\,model)}$ is calculated by

$$Attention\left(v_{(seq,\,d\,model)}, T5_{output\,(seq,\,d\,model)}\right) = softmax\left(\frac{T5_{output\,(seq,\,d\,model)}\,K^T}{\sqrt{d_k}}\right) v_{(seq,\,d\,model)} \quad (20)$$

Where $v_{(seq,\,d\,model)}$= output of the encoder, $T5_{output\,(seq,\,d\,model)}$= query, $d_k = \frac{d_{model}}{h}$ where h is the head

Linear will be used to map $seq,\ d_{\ model}$ into another sequence by vocabulary size. It tells for every embedding the position of the word in the vocabulary.

Softmax (seq, vocabulary size) will calculate the loss function, that is cross entropy loss. The research will select a token from the vocabulary corresponding to the position of the token with the maximum value. The research will use Beam search strategy to select at each step the top B words and evaluate all the possible next words, keeping the top B most portable sequence.

IV. EXPERIMENTS

This research will be evaluated using the following standard metrics such as such as BLEU score, Precision, Recall, F1 score, mean opinion score, comparative mean opinion score etc.

V. CONCLUSION

The research work is expected to develop an efficient, effective and more accuracy translation system for English language to Yoruba language to enhance the learning of Yoruba language, enhance the education standard of Yoruba people and to promote Yoruba culture.

REFERENCES

[1]. Eludiora, S. I., & Odejobi, O. A. (2016). Development of an English to Yorùbá Machine Translator. International Journal of Modern Education and Computer Science, 8(11), 8.

[2]. Akintola, A., & Ibiyemi, T. (2017). Machine to Man Communication in Yorùbá Language. Annal. Comput. Sci. Ser, 15(2).

[3]. Iyanda, A. R., & Ninan, O. D. (2017). Development of a Yorúbà Textto-Speech System Using Festival. Innovative Systems Design and Engineering (ISDE), 8(5).

[4]. Adewole, L. B., Adetunmbi, A. O., Alese, B. K., & Oluwadare, S. A. (2017). Token Validation in Automatic Corpus Gathering for Yoruba Language. FUOYE Journal of Engineering and Technology, 2(1), 4.

[5]. Ayogu, I. I., Adetunmbi, A. O., & Ojokoh, B. A. (2018). Developing statistical machine translation system for english and nigerian languages. Asian Journal of Research in Computer Science, 1(4), 1-8.

[6]. Greenstein, E., & Penner, D. (2015). Japanese-to-english machine translation using recurrent neural networks. Retrieved Aug, 19, 2019.

[7]. Nouhaila, B. E. N. S. A. L. A. H., Habib, A. Y. A. D., Abdellah, A. D. I. B., & Abdelhamid, I. E. F. (2017). Arabic machine translation using Bidirectional LSTM Encoder-Decoder.

[8]. Gogoulou, E. (2019). Using Bidirectional Encoder Representations from Transformers for Conversational Machine Comprehension.

[9]. Esan, A., Oladosu, J., Oyeleye, C., Adeyanju, I., Olaniyan, O., Okomba, N., ... & Adanigbo, O. (2020). Development of a recurrent neural network model for English to Yorùbá machine translation. Development , 11(5).

[10]. Ajibade, B., & Eludiora, S. (2021). Design and Implementation of English To Yor\ub\'a Verb Phrase Machine Translation System. arXiv preprint arXiv: 2104.04125.

[11]. Oyeniran, O. A., & Oyebode, E. O. (2021). YORÙBÁNET: A deep convolutional neural network design for Yorùbá alphabets recognition. International Journal of Engineering Applied Sciences and Technology, 5(11), 57-61.

[12]. [12] Sawai, R., Paik, I., & Kuwana, A. (2021). Sentence augmentation for language translation using gpt-2. Electronics, 10(24), 3082.

[13]. [13] Adebara, I., Abdul-Mageed, M., & Silfverberg, M. (2022, October). Linguistically-motivated Yorùbá-English machine translation. In Proceedings of the 29th International Conference on Computational Linguistics (pp. 5066-5075).

[14]. [14] Ajao, J., Yusuff, S., & Ajao, A. (2022). Yorùbá character recognition system using convolutional recurrent neural network. Black Sea Journal of Engineering and Science, 5(4), 151-157.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL767

[15]. [15] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention Is All You Need.(Nips), 2017. arXiv preprint arXiv:1706.03762, 10, S0140525X16001837.
[16]. [16] Xiao, T., & Zhu, J. (2023). Introduction to Transformers: an NLP Perspective. arXiv preprint arXiv:2311.17633.
[17]. [17] Magueresse, A., Carles, V., & Heetderks, E. (2020). Low-resource languages: A review of past work and future challenges. arXiv preprint arXiv: 2006.07264.
[18]. [18] Ajepe, I., & Ademowo, A. J. (2016). English language dominance and the fate of indigenous languages in Nigeria. International Journal of History and Cultural Studies, 2(4), 10-17.
[19]. [19] Fadoro, J. O. (2010). Revisiting the mother-tongue medium controversy. Montem Paperbacks, Akure.
[20]. [20] Mishina, U. L., & Iskandar, I. (2019). The role of English language in Nigerian development. GNOSI: An Interdisciplinary Journal of Human Theory and Praxis, 2(2), 47-54.
[21]. [21] Bibi, N., Rana, T., Maqbool, A., Alkhalifah, T., Khan, W. Z., Bashir, A. K., & Zikria, Y. B. (2023). Reusable Component Retrieval: A Semantic Search Approach for Low-Resource Languages. ACM Transactions on Asian and Low-Resource Language Information Processing, 22(5), 1-31.
[22]. [22] Omoniyi, A. M. (2012). SOCIO-POLITICAL PROBLEMS OF LANGUAGE TEACHING IN NIGERIA. Advisory Editorial Board, 152.
[23]. [23] Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: state of the art, current trends and challenges. Multimedia tools and applications, 82(3), 3713-3744.
[24]. [20] Mishina, U. L., & Iskandar, I. (2019). The role of English language in Nigerian development. GNOSI: An Interdisciplinary Journal of Human Theory and Praxis, 2(2), 47-54.
[25]. [21] Bibi, N., Rana, T., Maqbool, A., Alkhalifah, T., Khan, W. Z., Bashir, A. K., & Zikria, Y. B. (2023). Reusable Component Retrieval: A Semantic Search Approach for Low-Resource Languages. ACM Transactions on Asian and Low-Resource Language Information Processing, 22(5), 1-31.
[26]. [22] Omoniyi, A. M. (2012). SOCIO-POLITICAL PROBLEMS OF LANGUAGE TEACHING IN NIGERIA. Advisory Editorial Board, 152.