Volume 9, Issue 9, September – 2024 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP1562

A Transformer-Based Yoruba to English Machine Translation (TYEMT) System with Rouge Score

Oluwatoki, Tolani Grace[1]
Department of Computing,
College of Science, Afe Babalola University, Ado Ekiti,Nigeria

Adetunmbi, Olusola Adebayo[2]; Boyinbode, Olutayo Kehinde[3]
Department of Data Science[2]; Department of Information Technology[3]
The Federal University of Technology, Akure, Nigeria

Abstract:- Automated translation systems for some indigenous Nigerian languages like the Yoruba, have historically been limited by the lack of large, high-quality bilingual text and effective approaches to modeling. This paper presents introduces an approach to bi-directional Yoruba-English text-to-text machine translation utilizing deep learning technique, specifically Transformer models. Transformer models, which utilizes self-attention mechanisms to improve translation quality and efficiency. The system was trained and evaluated on a newly curated Yoruba-English parallel corpus, which significantly augments existing resources. Experimental results demonstrate that the Transformer-based model performs translation accurately and fluently, achieving a ROUGE (Recall-Oriented Understudy for Gisting Evaluation) score improvement of 0.4649. This work not only advances the frontiers of Yoruba-English machine translation but also enriches a wider domain in the field of multilingual Natural Language processing (NLP) by addressing challenges associated with translating between languages with limited resources. Future studies include enhancing the available parallel corpus and exploring hybrid models that combine the strengths of both RNN and Transformer architectures.

Keywords:- *Yoruba, English, Transformer, Self-Attention, ROUGE Score, NLP.*

I. INTRODUCTION

A language is a medium of communication between two parties involving the use of symbols and signals that convey the same meaning and message to both sender and receiver. Languages serves to communicate information, request action, convey mood, emotions, give ideas and report experiences. Language serves a very crucial social function as a means of communication [23].

The major attribute that distinguishes humans from other animal is the ability to use language effectively. When you know a language, you have the ability to speak and be understood by others who know the language too. Knowing a language means understanding which sound or sign are included or excluded in the language [24] [25]. Humans will practically struggle tp convey thoughts, desires and beliefs as language forms the essential for societal and religion development [1]. Translation has defined by [2] and [26] is the process of converting a text or speech document in one language equivalent text/speech in another language. Every translation process involves: the source language (the language to be translated) and the target language (recipient language). In addition, [3] stated that translation entails replacing written content in one language (usually the source language) by equivalent content and textual materials in another language (target language). This brings two different languages and culture into contact, making it essential for a bilingual individual or machine to carry out the translation effectively.

NLP is a field that explore how data science interact with human language and it is experiencing rapid growth in various industries by allowing digital devices and computers to understand, generate and recognize text and speech by integrating computational linguistics [27]. Today, advancement in NLP along with enhanced data access and increased computational power are enabling various practitioners in many fields of endeavors such as finance, healthcare, media, human resources to obtain better results [28]. The foundation of NLP can be traced to a number of disciplines, which includes: Mathematics, Electrical Electronics Engineering, Computer Science, Artificial Intelligence, Robotics and Linguistics.

NLP involves using algorithms to identify and extract the rules of natural language by transforming unstructured language into a format a computer can comprehend through various techniques [29] such as syntax analysis techniques which include: Lemmatization, Morphological Segmentation, Word segmentation, Part-of-speech Tagging, Parsing, Sentence breaking and Stemming. Additionally semantics analysis methods include: named entity recognition (NER), word sentence disambiguation and Natural language generation, among others [30]. In broad term however, NLP tasks decompose language into smaller, fundamental components and examines the relationship between these components and how they are combined to convey meaning [4][31].

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP1562

- *Automated Translation (Machine Translation)*

Machine Translation primarily involves the use of computer software is used to convert text from one natural language like English into another such as French, Yoruba, Spanish and so on [32]. A translator needs to analyze and interpret all elements in the text by understanding how each word can affect the other [33]. Providing a friendly interface by which users can freely interact with the computer in human language is the primary objective of natural language processing.

- *Approaches Natural Language Processing and Machine Translation*

Several approaches are employed in natural language processing of indigenous machine translation systems, including are the following:

- *Direct Approach;*
- *Rule-Base Approach;*
- *Statistical Approach;*
- *Neural Networks Approach; and*
- *Hybrid Approach*

- *Direct Approach*

The Direct machine translation approach consists of directly substituting words from the source language with their equivalents in the target equivalent language**.** This approach is dictionary-driven because it makes use of two bilingual dictionary look-up through which translation takes place without any additional linguistic analysis and presentations [5] and [6].

- *Rule-Based Approach:*

A rule-based NLP model is a system that utilizes predefined rules to accomplish specific tasks like parsing, tagging, or extracting information from natural language texts or speech. These rules are typically crafted by human experts with linguistic knowledge and domain expertise. They can be grounded in various aspects of language, including syntax, semantics, morphology, phonology, and pragmatics. For instance, a rule-based model for sentiment analysis might employ a curated list of positive and negative words, phrases, or emoticons to determine the sentiment score of a given text [7],[8] and [9]

- *Stages in a Rule-Based Approach to NLP*

- ✓ Creation of Rules: Development of specialized linguistic rules tailored to the desired tasks, including syntax structure, semantic rules, pattern matching expressions and grammatical rules.
- ✓ Application of Rule: Apply the established rules to the input data to detect and capture corresponding patterns.
- ✓ Processing of Rule: Process the textual data according to the outcome of the aligned rules, enabling extraction of information and decision-making.
- ✓ Refinement of Rule: Continuously refine the rules through iterative processing to improve performance and accuracy. Informed by the feedback from previous analyses, update and modify the rules as necessary.

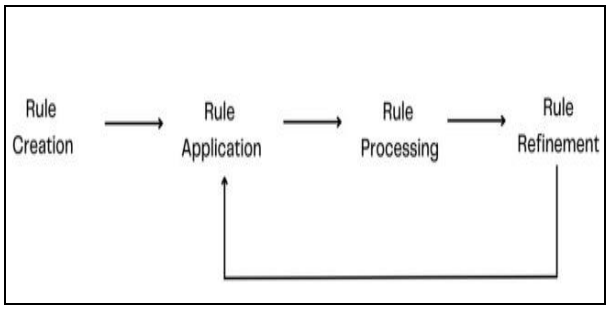

Fig 1: Stages in Rule-based NLP Approach

- *Statistical Approach:*

Statistical Machine Translation (SMT) systems are a type of corpus-based translation system that rely on parallel texts to calculate the parameters of the statistical models used in the translation process. Once the models have developed, they are used to deduce the translation of new sentences [34]

SMT is a method of machine translation that is defined by the application of machine learning techniques. Statistical Machine Translation (SMT) requires minimal human effort for translation and is characterized by the application of machine learning techniques. In this paradigm, translations are produced based on statistical models that analyze and generate language data. The goal of SMT is to basically translate written text from one language to another based on probability theory that is commonly applied in natural language processing tasks. It utilizes probability models to select the translation that maximizes P(e|f), which represents the likelihood of the output sentence e given the input sentence f

$$P(f_1^J) = \frac{P(e_1^I) \cdot P(e_1^I)}{P(f_1^J)}$$

$$\mathbf{T(}f\mathbf{)} = \hat{e}_1^I = \underset{e_1^I}{\mathbf{argmax}_e} P(f_1^J) = \underset{e_1^I}{\mathbf{argmax}_e} P(e_1^I) \cdot P(e_1^I)$$

Where P(e) is the **language model** and P(f |e) is the **translation model [35].** This process is known as **decoding**. [10] and [11].

- *Common Techniques in Statistical Approach*

- ✓ **N-gram Models**: Useful for language modeling and text generation, capturing the context in which words are used in sequences tasks.
- ✓ **Statistical Parsing**: Analyzes sentence structure using probabilistic grammar to determine the most likely parse tree.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP1562

✓ **Classification Algorithms**: Employed for tasks like sentiment analysis, where text is categorized based on learned probabilities from training data.

➢ *Neural Networks Approach:*

Neural Machine Translation NMT is a translation method that uses extensive artificial neural networks to forecast the probability of sequences of words, typically modeling entire sentences as a unified system [36]. An extension of NMT is the Deep Neural Machine Translation, utilizing a larger neural network that processes multiple layers rather than just one. Unlike traditional translation systems, NMT models are trained jointly in an end-to-end manner, optimizing the entire framework to enhance translation performance [12] and [13].

- *Key Features of Neural Network Approach*

✓ **Layered Architecture**: Neural networks consist of multiple layers (input, hidden, and output) that process data through interconnected nodes (neurons), allowing for the modeling of intricate relationships in language.
✓ **Learning Representations**: Neural networks automatically learn representations of words and phrases, capturing semantic meaning and contextual information without the need for extensive manual feature engineering.
✓ **End-to-End Learning**: Many neural network models are trained in an end-to-end fashion, by the means of the model learning to map input data directly to the output predictions, optimizing the entire process simultaneously.

- *Common Neural Networks Approaches:*

✓ **Feed forward Neural Networks**: This is a type of Neural network approach, where information flows in a single direction without cycles.
✓ **Recurrent Neural Networks (RNNs)**: Designed to process sequential data, to retain memory of the previous inputs, making them ideal for text generation and language modeling tasks.
✓ **Long Short-Term Memory (LSTM) Networks**: A variant of RNN that effectively captures long-range dependencies and mitigates issues like vanishing gradients.
✓ **Gated Recurrent Units (GRUs)**: Like LSTMs, GRUs use gating mechanisms to regulate information flow but have a simpler structure.
✓ **Convolutional Neural Networks (CNNs)**: CNN is applied to text classification, image processing, and for sentiment analysis to capture patterns in textual data.
✓ **Transformers**: the transformer employs the use of self-attention mechanisms to simultaneously process a complete sequence of data, rendering it effective for summarization, translation and other NLP tasks.

➢ *Hybrid-Based Approach*

The hybrid approach in machine translation (MT) combines multiple translation methodologies to leverage their respective strengths. This strategy aims to enhance translation quality, particularly in terms of fluency and adequacy, by integrating various techniques into a single system [14], [15] and [16]

- *Key Features of Hybrid-Based Approach*

✓ **Integration of Approaches**: The hybrid method utilizes both rule-based and statistical or neural techniques, allowing for a more comprehensive understanding of language nuances.
✓ **Quality Improvement**: By combining different approaches, hybrid systems seek to produce translations that are more accurate and natural-sounding, addressing the limitations of individual methods.
✓ **Compensation for Weaknesses**: The hybrid approach effectively mitigates the shortcomings of one method by incorporating the strengths of another, leading to better overall performance.

- *Common Hybrid Systems*

Some of the Common Hybrid Machine Translation Systems Include the Following:

✓ **PROMT**: Combines rule-based and statistical techniques for improved translation accuracy.
✓ **SYSTRAN**: Integrates various linguistic rules with statistical models to enhance translation quality.
✓ **Asia Online**: Utilizes a mix of approaches to cater to specific language pairs and contexts.

- *Categories of Hybrid Systems*

There are two main groups of Hybrid systems, namely:

✓ **Single-Engine Hybridization (SEH)**: Involves combining multiple methodologies within a single translation engine. This can include integrating rule-based elements with statistical or neural methods.
✓ **Multi-Engine Hybridization (MEH)**: Utilizes multiple translation engines, each employing different techniques, to produce a more robust output. The results from these engines are then combined to create the final translation

II. MOTIVATION

The dominance of the English language as a lingual franca is one of the factor threatening the indigenous languages with nearly all the Nigerian languages, including the three major ones – Hausa, Igbo and Yoruba are faced with challenges related to politics, training and language technology [37]. Many of our new generation children are losing sights of the cultural heritage. Many of them are not even aware of the use of idiomatic expressions and proverbs present in the language. Government establishments, media houses, the judiciary,

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP1562

educational sectors, place priority on the use of English language for conveying information during the working hours. In many primary and secondary schools, pupils and students are forbidden from speaking in their native language calling it a vernaculars. This is making our values present in our indigenous language going into extinction. The European and Asian languages, Chinese, Hindi, English, Russian, Indonesian, German, French and the likes have received a considerable number of research attention in past years. Some of the African researchers are also working on the African indigenous languages. In Nigeria however, a handful of researchers are now interested in ensuring the preservation of our indigenous languages like the Hausa, Igbo and Yoruba languages in recent years.

Many of these researchers adopted a single approach to machine translation of the English to Yoruba language. But these research works are yet to provide a satisfactory result in the area of shallow representation of words in the translations, accuracy and fluency of the systems. This research is motivated with the need to develop a bi-directional user-friendly translating system for Yoruba and English languages a deep learning approach with neural networks to extract patterns from the Yoruba and English languages in order to generate a better, more sensible and robust translation system for both indigenous and non-speakers of the Yoruba language.

A user interactive and friendly platform for Yoruba complex sentence (text) has been developed. It is a bidire ctional bilingual-translation system. This system seeks to enhance communication skills and bridge gaps between indigenous and foreign speakers of the Yoruba language across various fields.

III. LITERATURE REVIEW

In Nigeria, machine translation approaches of different kinds has been adopted and employed by researchers in the development of automatic translation system for some indigenous languages especially for English to Yorùbá language and vice-versa. Many researchers have worked in the area of Neural Machine Translation, most especially translation of English language to Yoruba Language. Such among them are:

[17] presented a work titled “Development of a RNN model for English to Yoruba Machine Translation” with the aim to creating a RNN model for English to Yoruba. The test and evaluation of the model was based on human and automatic assessments. The demonstrated fluency and good quality of translation, it was restricted to handle single tasks with a limited vocabulary.

In [18], titled “ HausaMT v1.0: Towards English Ha usa Neural Machine Translation” employed different datasets containing parallel corpus of Hausa language to develop a translations system by adopting Recurrent Neural Networks and Transformer encoder-decoder approach with Standard word-level tokenization and Byte Pair Encoding subword tokenization. Using the BLEU score evaluation metric, it was observed that the word-level tokenization performed better than the sub-word level. The system was limited to religion domain with limited datasets.

A work titled “Sentence Augmentation for Language Translation Using GPT-2” was developed by [19]. They aimed at exploring the use of GPT-2 to generate monolingual data to improve machine translation. Sentence generator based on GPT-2 was utilized in the system to produce additional data for the neural machine translation inorder to enable the maintenance of similar characteristics of the original dataset. The research lacked many recent properties available in GPT-3 model for augmenting sentences for language translation.

In [20], titled “Convolutional Recurrent Neural Network for the recognition of Yoruba Handwritten Characters” adopted a Convolution RNN in training the captured images after preprocessing. There a low character recognition accuracy by the system due to the presence of diacritics signs in the collected images.

[21] developed a work titled “Linguistically-Motivated Yoruba-English machine Translation” with the aim to analyze and provide linguistic descriptions of errors made by different models. Training of three-sentence model was done-SMT, BiLSTM and Transformer to translate from Yoruba to English language. There were incorrect words, wrong spellings, semantics and syntax errors, grammatical and word order errors as well as missing words were the major limitations of the system.

In [22], titled “Development of a XML-Encoded Machine-Readable Dictionary for Yoruba Word Sense Disambiguation” employed the use of Extensible Markup Language (XML) to transform a collection of Yoruba verbs and translation from an existing bilingual dictionary. The research was limited to monosyllabic words with limited words in the vocabulary.

[13] adopted the Neural Machine Translation model (Transformer model) in the work titled “Evaluating English to Nupe Machine Translation Model Using BLEU” to assess the effectiveness of machine translation systems that was used to translate English phrases to Nupe in comparison to the effectiveness English to Nupe language by human translators. The research was limited to evaluation metrics adopted by different indigenous machine translation systems.

In “[23], titled ‘Integrating Yorùbá cultural greetings into machine translation’ investigate the effectiveness of neural machine translation (NMT) systems in translating Yoruba greetings to English language an integral part of the Yoruba cultural heritage. Analyzing the performance of different multilingual NMT systems including Google, it was observed that the model struggles to accurately translate Yorùbá greetings into English.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP1562

English to Okun language machine translation was developed by [8] by adopting the rule-based approach of the prepositional phrases of the two languages because of the gradual extinction of the language due to the dominance of the English language..

IV. METHODOLOGY

This section outlines in details the methods and tools used in this study. It is divided into four sections: source text (Data collection and Description), data pre-processing, corpus exploratory analysis and the training process and the description of the transformer based model employed for the implementation of the system.

The system architecture is clearly illustrated as shown in figure 1.

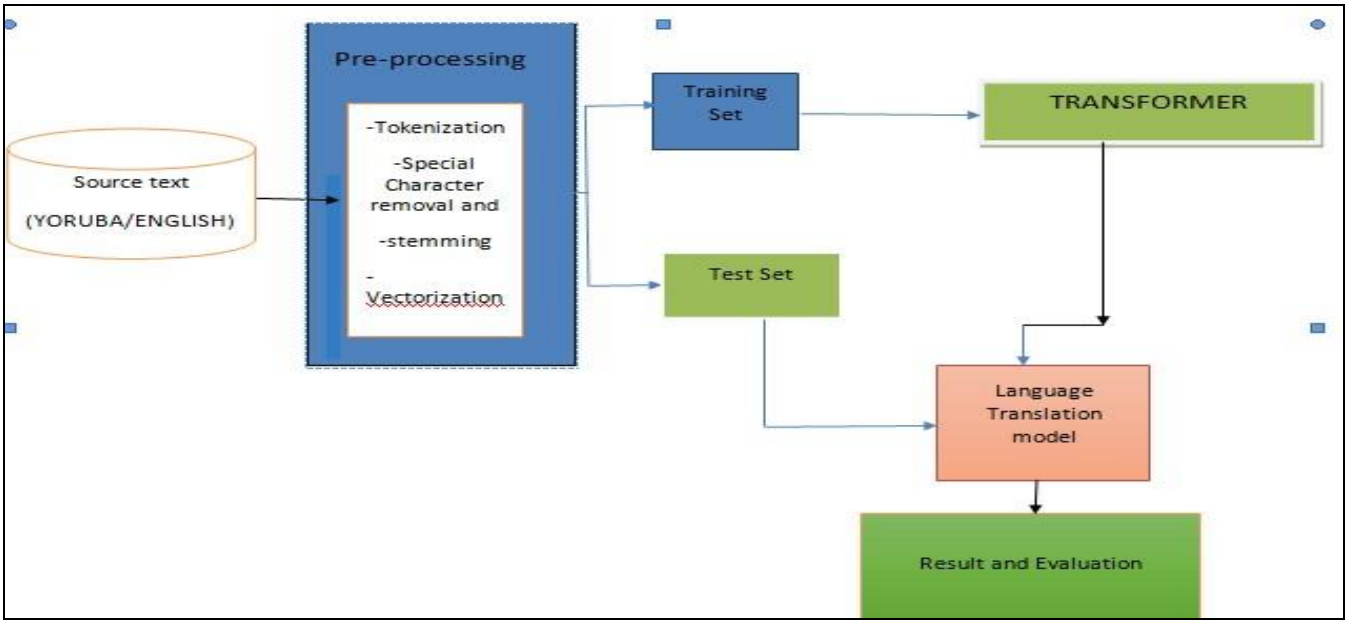

Fig 2: Architecture of the Proposed System

A. Source Text (Data Collection and Description):

This study aims to enhance the accessibility and practicality of Yoruba across various domains, thereby promoting its preservation and broader usage. The data utilized consists a bilingual generalized dataset with multiple domains of Yoruba and English long sentences sourced from Artificial Intelligence for Development-Africa Network (https://africa.ai4d.ai/). Having curated these sentences into a parallel corpus, the dataset has no null values. The source text is accepted by the system and then enters the p reprocessing stage.

B. Preprocessing Stage

The preparation of text deals with the method in NLP by which cleaning and preparing the text for the model building. This involves the removal of all forms of noise that may be represented as punctuations, symbols, special characters, emotions and numeric values. Also, missing values word redundancies, mistakes and inconsistencies are taken care of. Preprocessing helps to transform text into a more digestible form for machine learning algorithms to perform more accurately. The steps required for preprocessing are- tokenization, special character remover. Stemming, and lemmatization and vectorization.

- Tokenization breaks up unstructured raw data such as phrases, sentences, paragraphs or a whole document into chunks or smaller units like terms or individual words called tokens. In NLP, each word, punctuations or numbers are referred to as tokens which are located by word boundaries such as a space, full stop, question marks, special characters and symbols.
- Special Character and punctuation remover: there are about 32 main punctuations and special characters that must be taken care of in order to enjoy a clean and noise free database. These include space, apostrophe, hyphens, commas, semicolons, colons, exclamation points, brackets, parentheses, ellipsis, slash, backslash, Section markers, dashes, Bullets and numbering etc. Most of these elements are less necessary and their removal does not affect the performance and quality of the general system.
- Stemming: this is the procedure of minimizing a word to its root word. The presence of several variant of a single word in a language in a text corpus can result to data redundancy when developing an NLP model. Therefore, stemming normalizes root words text by the removal of repetitive words. Stemming is used in information retrieval and domain analysis with the use of Porter stemmer, Regular expression, Snowball and Lancaster stemming algorithms. In many instances,

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP1562

the removal of the last few characters by stemming, often leads to wrong spelling and meaning.

- Lemmatization: this is the process of grouping together words with the same root words but with different suffixes and prefixes. It basically converts words to it base from which can be found in the dictionary. This differs from stemming because lemmatization considers the context and converts the main word into its meaningful base form. Lemmatization is mostly useful in search engine optimization, biomedicine and in sentiments analysis. The commonly used lemmatization tools are Bio Lemmatizer, lemmatization API and Trinker/Textstem.
- Vectorization: which is also referred to as word embedding is the process by words are converted to numbers in order to extract distinct features out of the whole text for the model to be able to train on. Vectorization also known as Word embedding is essential because many machine learning algorithms and most deep learning architectures cannot effectively process string variables or plain text in their raw formats [38]. The commonly used techniques for word embedding or text vectorization are: Binary term frequency (One-Hot Encoding (OHE)), Bag of Words term frequency, (L1) Normalize term frequency, (L2) Normalize TF-IDF (Term Frequency- Inverse Document Frequency) and Word2Vec [39]. Figures 2 and 3 shows samples of the dataset before and after the removal of diacritics respectively.

	ID	Yoruba	English
0	ID_AAJEQLCz	A ṣètò Ìgbìmọ̀ Tó Ń Ṣètò Ìrànwọ́ Nígbà Àjálù l...	A Disaster Relief Committee was formed to orga...
1	ID_AASNedba	Ìrọ̀lẹ́ May 22, 2018 ni wọ́n fàṣẹ ọba mú Arákù...	Brother Solovyev was arrested on the evening o...
2	ID_AAeQrhMq	Iléeṣẹ́ Creative Commons náà	Creative Commons the Organization
3	ID_AAxlMgPP	Pẹ̀lú Egypt, Morocco àti Tunisia tí wọ́n ti lo...	With Egypt, Morocco and Tunisia out of the Wor...
4	ID_ABKuMKSx	Adájọ́ àgbà lórílẹ̀ èdè Náìjíríà (Attorney Gen...	The Attorney General of the Federation, Justic...

Fig 3: Sample of Dataset before Removal of Diacritics

	ID	Yoruba	English
0	ID_AAJEQLCz	A seto Igbimo To N Seto Iranwo Nigba Ajalu lat...	A Disaster Relief Committee was formed to orga...
1	ID_AASNedba	Irole May 22, 2018 ni won fase oba mu Arakunri...	Brother Solovyev was arrested on the evening o...
2	ID_AAeQrhMq	Ileese Creative Commons naa	Creative Commons the Organization
3	ID_AAxlMgPP	Pelu Egypt, Morocco ati Tunisia ti won ti lole...	With Egypt, Morocco and Tunisia out of the Wor...
4	ID_ABKuMKSx	Adajo agba lorile ede Naijiria (Attorney Gener...	The Attorney General of the Federation, Justic...

Fig 4: Sample of Dataset Without Diacritics

C. Corpus Exploratory Analysis

An exploratory analysis of the dataset was performed to study the patterns such as grammar, syntax, and semantics inherent in the source language and the target languages, which will inform the development of the proposed Yoruba-English MT system. Foremost, overall count of words in the corpus is **373348**, while the total number of corresponding sentence lines is **20108.** With **199443 Yoruba** and an estimate of **23985** unique words. The total number of the corresponding English word is **173905,** with **26354** number of unique words in English as shown in fig. 4

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP1562

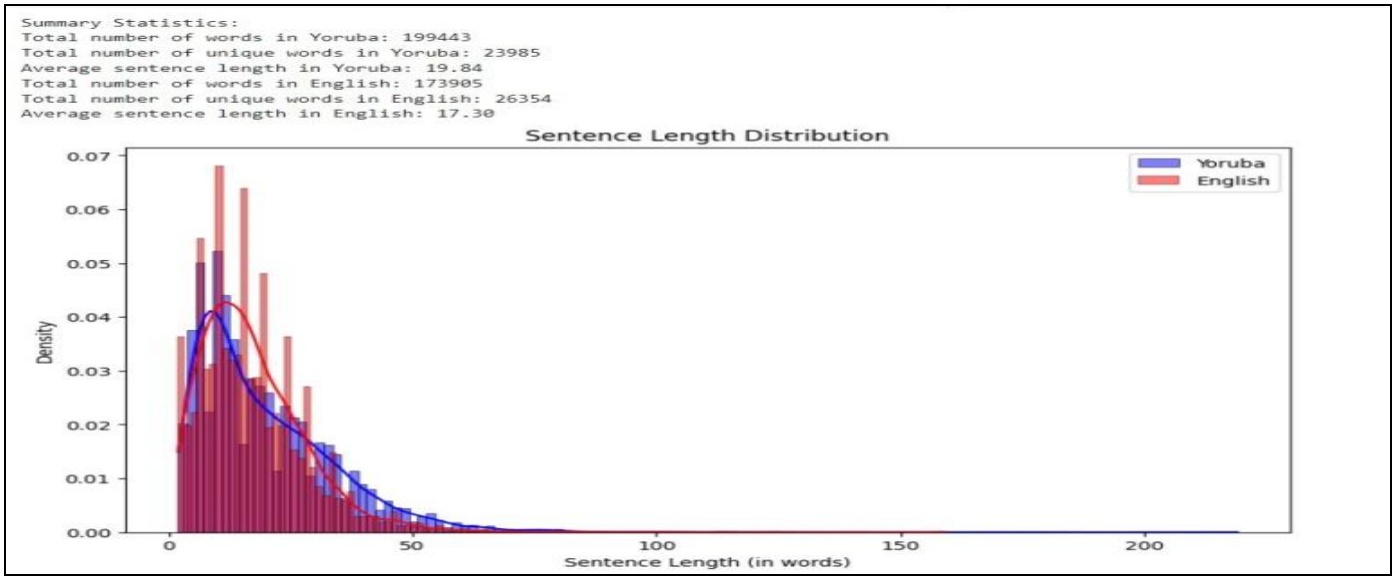

Fig 5: Yoruba-English Corpus Sentence Length Distribution

However, it could be inferred from the analysis that the source language has an unprecedentedly higher number of words than the target language in the corpus. The imbalanced dataset may nudge the proposed MT model to be biased towards the source language, hence potentially leading to poor translation performance. On the other hand, the target language has higher unique words than the source language (figure 5), which also implies that the target language's richer vocabulary can help disambiguate words with different meanings in the source language.

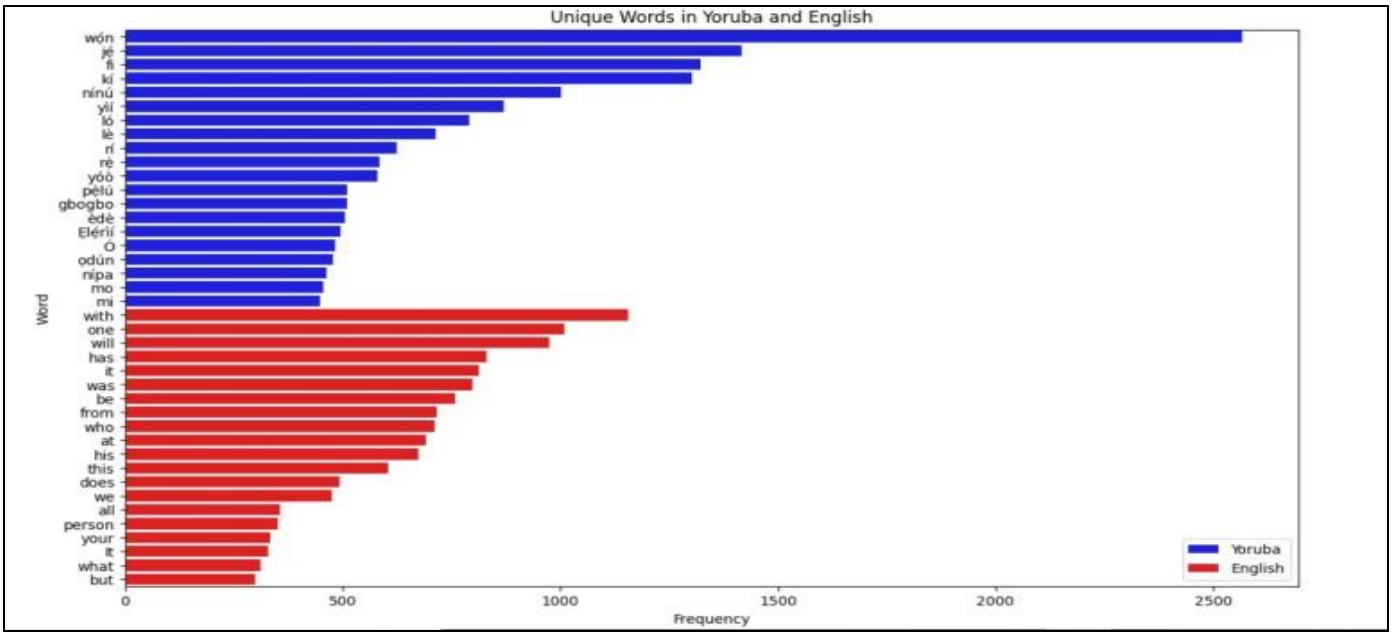

Fig 6: Unique Words in Yoruba and English

D. Training Process:

Pandas was used to go through the training dataset in order to get a good grasp of how the data actually looks. This was done by mounting the dataset from the required location on the personal computer (PC). We used the Helsinki NLP model, a pre-trained auto-tokenizer from transformer model and fine-tunning it using the Yoruba text. AdamW optimizer with scheduler to support the learning rate in case of noise from the dataset maximum length used during tokenizer while training the model on the dataset. This was used to enhance the performance and speed-up the convergence of our model. After the hyperparameter tunning, a most favourable outcome was achieved with a learning rate of 0.0001.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP1562

E. The Transformer Model

Figure 6 shows the architecture of the Transformer model

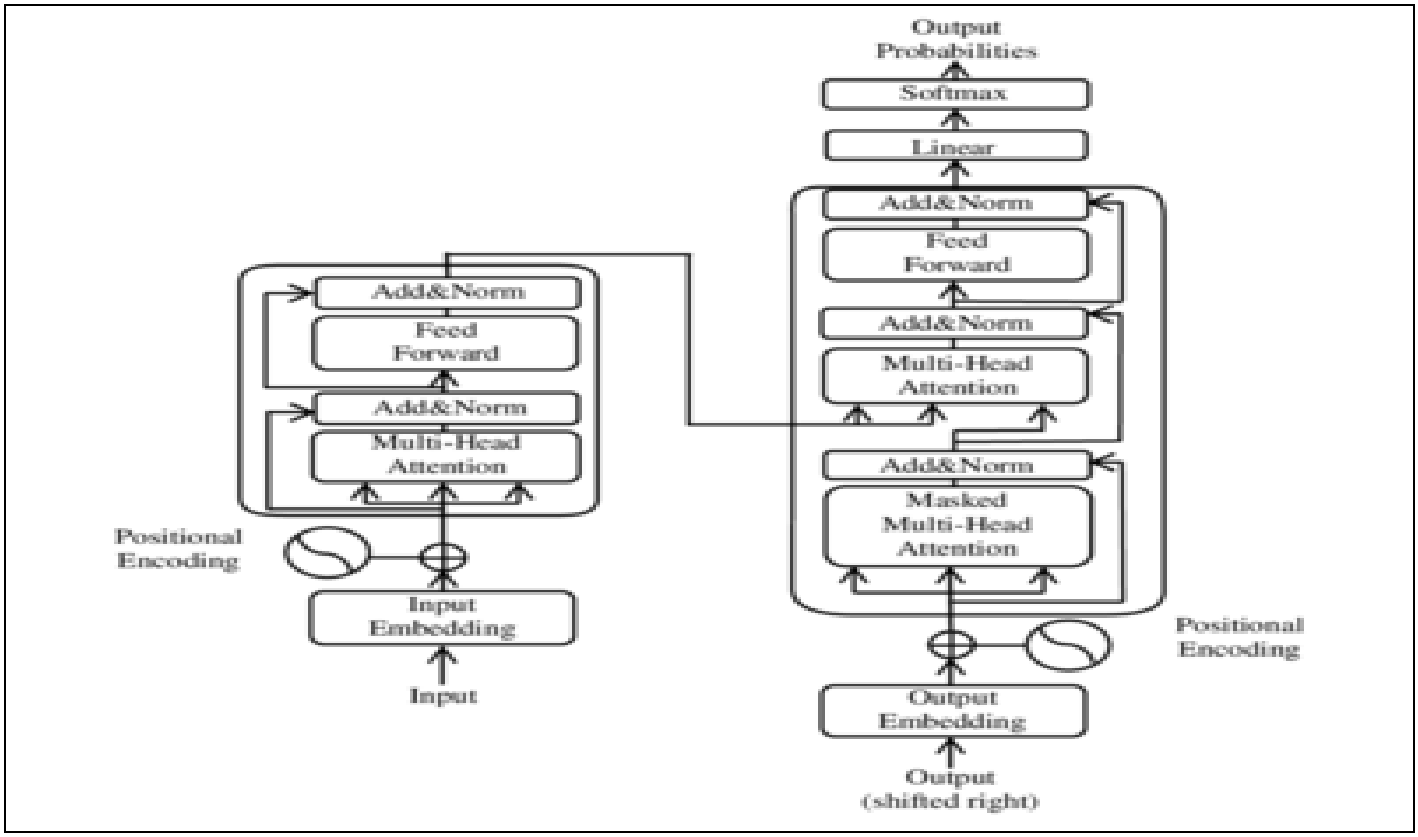

Fig 7: Transformer Architecture (Source: [40])

The Transformer model operates on sequences of tokens with variable-length inputs, as shown in equation (1), and a variable-length output, as indicated in equation (2). This is for the purpose of calculating the probability of the language model:

$$p(\mathbf{y}|\mathbf{e}) \quad (1)$$

The model is given by:

$$p_\theta(\mathbf{y}|\mathbf{z}= f_\theta(\mathbf{e})) \quad (2)$$

The encoder and decoder comprises of N identical layers, with each comprising of two sub layers

- A multi-head self-attention mechanism with hhh heads. It processes different linearly projected versions of the queries, keys, and values, generating h outputs in parallel that are then combined to produce the final result is implemented in the first sub layer.
- The second sub layer includes two linear transformations with a Rectified Linear Unit (ReLU) activation in between and fully connected feed-forward network:

$$FFN(y)= ReLU (w_1y + b_1)w_2 +b_2 \quad (3)$$

$$Multihead (Q,K,V) = Concat (head_1, head_2, \ldots, head_h) *W^O \quad (4)$$

Where $head_i$ = Attention (QW_i^Q, KW_i^K, VW_i^V)

W_i^Q, W_i^K , and W_i^Vdenote the projection matrices used for generating different subspace representations of the query, key and value matrices.

- **Query (Q):** it is a feature vector providing insight into what the target is in the sequence.
- **Keys (K):** This is a feature vector describing what contains in the element. It stands out to provide the identity of the elements and give attention by the query.
- **Values (V):** Process the sequence of input, each input element uses a value to know what to provide an average on.

Positioning encoding is used for injecting information into the transformer architecture and marks the specific location of objects in a sequence of words, and this is done by harmonizing the Sine and Cosine functions of the varying frequencies.

$$P (pos, 2i) = sin (pos / 10000^(2i / d model)) \quad (5)$$
$$P (pos, 2i+1) = cos (pos / 10000^(2i / d model)) \quad (6)$$

Where: pos is the position of a token within the sequence, i is the dimension within the positional encoding, and d model is the dimensionality of the model's embedding [41].

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP1562

V. EXPERIMENT

A. Learning Curve:

The model was trained using 40 maximum epochs with 32 batches per epoch and 80 looping. A loss of 0.7678 was recorded at the initial stage and 0.4561 average training loss was obtained at the end of the model training with 14000 iterations as shown in the learning curve figure 7. The learning curve shows that the model was able to learn each word in the source language with the corresponding target language which gave the system a better translation performance.

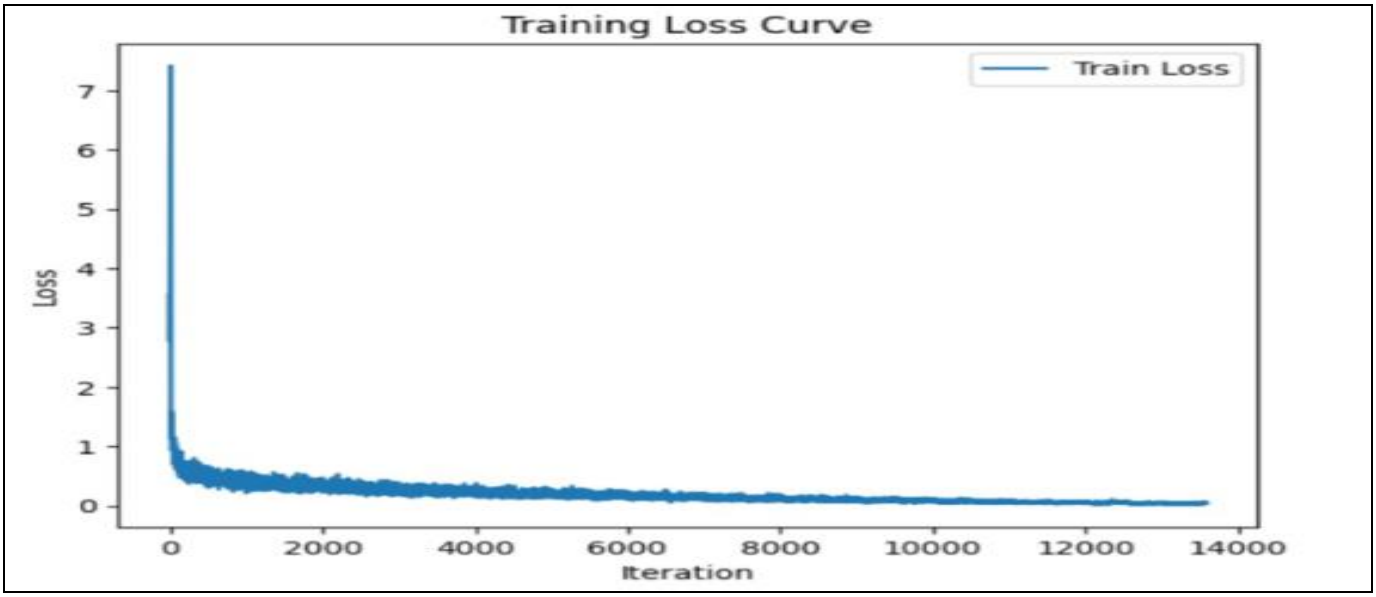

Fig 8: Final Model Learning Curve with 14000 Iterations

This shows a smooth training of the model. There was a reduction in Noise, overfitting with the average loss of 0.1712. Average Training Loss: 0.0450.

B. Evaluation:

Our Yoruba to English translation system, implemented using a transformer model, has yielded a ROGUE score of 0.4649. ROGUE-L measures the longest common subsequence between the translated text and the reference translations, capturing the fluency and relevance of translation. A score of 0.4649 indicates that a substantial portion of the relevant sequences was captured by the model but still has room for improvement in aligning with human-like translation quality.

C. TYEMT vs Google Translate

Table 1 shows the comparison of the system as compared to Google translate. It is observed the system and the Google translate are exactly the same translation while in some others, DBYEMT outperformed Google translate.

Table 1: TYEMT System Compared with Google Translate

Yoruba Words	TYEMT System	Google Translate
mofe lo si oja oko	I want to go to the farm market	I want to go to the farm market
mo n lo si ile iwe	I am going to school	I am going to school
Mo fe lo si igboro	I want to get over to the street	I want to go to the street
Kini oruko iya re?	What is your mother's name?	What is your mothe's name?
Mo ti so fun won pe , mo fe jeun	I have told them that, I want to eat	I told them, I want to eat
mo jade nitoripe mo fe lo siileijosin	I go out because I want to go to the ministry	I went out because I wanted to go to church
kini ore mi fe ra ni oja	What does my friend want to buy at the market?	What does my friend want to buy in the market?
Kini oruko iyawo re	What is the name of your wife.	what is your wife's name

➢ *The Sentence Translation User Interface*

Figure 7 shows the translation user interface of the system. It features two text areas: "Enter Yorùbá Text" and "English Translation." When the user types in text, they can click the submit button, after which the corresponding English sentence is shown in the "English Translation" field.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP1562

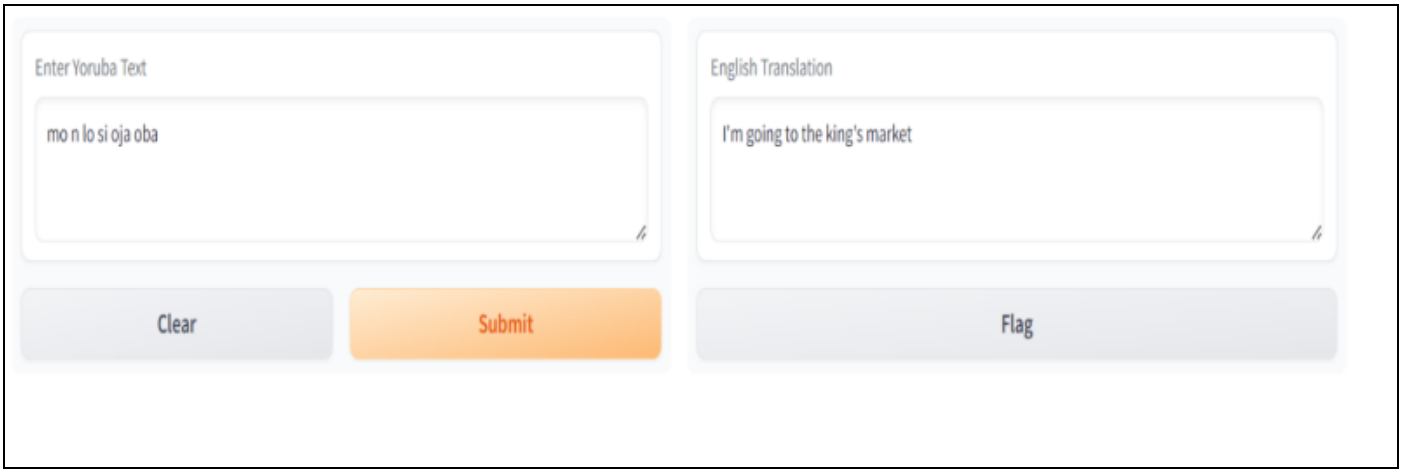

Fig 9: Sentence Translation User Interface

VI. CONCLUSION

This research work has demonstrated an effective, efficient and more accurate translation for the Yoruba to English language. This gives a possibility of establishing a continuous translation research of the Yoruba language. This is to enhance the learning of the language by both indigenous and foreign speakers thereby bridging the communication gap between business owners , foreign investors and to improve the learning of the language among the young folks in order to promote and preserve the language as well as the cultural heritage of the Yoruba people. However, the research can be extended to accommodate more indigenous languages such as Nupe, Efik, Igala,Igbo, Hausa and other minor lan guages inNigeria, also translations can be provided for ot her specific fields of human endeavours to accommodate their specific terminologies.

REFERENCES

[1]. Collins Online Dictionary, (2015). Pioneers in Language Reference for 200 years.
[2]. Kolawole, S. O. (2022). Translation Studies in Nigeria: Issues and Perspectives. Journal for Translation Studies in Nigeria (JTSN) , pp 17- 37.
[3]. Catford, J. C. (1965). A Linguistic Theory of Translation. Oxford University Press. 1-110.
[4]. Sas, V. (2019). Natural Lnaguage Processing (NLP), What it is and Why it Matters. https://www.sas.com/en_sg/insights/analytics/what-is-natural-language-processing-nlp.html. Retrieved 24/9/2024.
[5]. Venkateswara P.T. and Mayil M.G. (2013). Telugu to English Translation using Direct Machine Translation Approach. International Journal of Science and Engineering Investigations (IJSEI), pp 2532, ISSN:2251-8843.
[6]. Abiola, B.O., Adeyemo, O.A., Saka-Balogun, O.Y. and Okesola, F. (2020). A web-based Yoruba to English Bilingual Lexicon for Building Technicians. International Journal of Advanced Trends in Computer Science and Engineering. 9(1), pp 1-8.
[7]. Oyelami M.O., Famutimi, R.F. and Fadare, T.S. (2021). Development and Evaluation of an Android-based Yoruba Language Proverbs Preservatory and Repository System. International Journal of Computer Applications, 183(6), 9-15.
[8]. Esan, A., Sobowale, A., Adebiyi, T., Adio, M. and Toloruntomi, S. (2024). A rule-based Approach to English-Okun Prepositional Phrase Machine Translation. Dutse Journal of Pure and Applied Sciences (DUJOPAS), 10 (1c), 54-66.
[9]. Agbelusi, O., Matthew, O. O. and Aladesote I. O. (2024). Inclusive Mobile Health System for Yoruba Race in Nigeria. International Conference on Information and Knowledge System, 486, 255-264.
[10]. Fasakin T.G. (2017). An English to Yoruba Statistiacl Machine Translation system. M. Tech Thesis, Federal University of Technology, Akure.
[11]. Ayogu, I.I., Adetunmbi, A.O. and 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.
[12]. Adelani, D. I., Ruiter, D., Alabi, O. J., Adebonjo, D., Ayeni, A., Adeyemi, M. and Espana-Bonet, C. (2021). The Effect of Domain and Diacritics in Yoruba–English Neural Machine Translation. Proceedings of the 18th Biennial Machine Translation Summit Virtual USA. 61-75.
[13]. Sayuti, M. S.; U. S. Hassanand G. Danlami. (2023). Evaluating English to Nupe Machine Translation Model Using BLEU. Nigerian Journal of Engineering Science Research (NIJESR), 6(3), 1-7.
[14]. Ojo, A., O. Obe;A. Adebayo; and M. Olagunjoye. (2020). Development of English to Yoruba Machine translator Using Syntax-based Model. University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR), 6 (1):77-86.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP1562

[15]. Chinenyeze C.E. and Benntt E.O. (2019). A Natural Language Processing System for English to Igbo Language Translation in Adriod. International Journal of Computer Science and Mathematics Theory, pp 64-75.
[16]. Artur, N. and Tomaz, D. (2021). Adam Mickiewicz University English-Hausa Submissions to the WMT 2021 News Translation Task. Proceedings of the Sixth Conference on Machine Translation (WMT), 167–171.
[17]. Adewale, A. (2020). HausaMT v1.0: Towards English-Hausa Neural Machine Translation. 4th Widening NLP Workshop, Annual Meeting of the Association for Computational Linguistics, ACL, 1-4.
[18]. 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.
[19]. 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.
[20]. Adedara, I., Mageed, M.A and Silfverberg, M. (2022). Linguistically-Motivated Yoruba-English Translation. Proceeding of the 29th International Conference on Computational Linguistics. 5066-5075.
[21]. Adegoke-Elijah, A., Jimoh, K. and Alabi, A. (2023). Development of a XML-Encoded Machine-Readable Dictionary for Yoruba Word Sense Disambiguation. UNIOSUN Journal of Engineering and Environmental Sciences, 5 (1): 1-10.
[22]. Akinade, I., Alabi, J., Adelani, D. Odoje, C. and Klakow, D. (2023). Ku <Mask>: Integrating Yoruba Cultural greetings into Machine Translation. Proceedings of the First Workshop on Cross-Cultural Considerations in NLP (C3NLP), 1–7.
[23]. David, C. and Robert, H. R. (2024). The Encyclopaedia Britanical. Britannical.com/topic/language. Accessed 26/9/2024.
[24]. Titanium, M. (2021). Introduction To Language. https://www.coursehero.com/file/85286837/INTRODUCTION-TO-LANGUAGEdocx/.
[25]. Richard, N. (2019). Observations on What is Lang uage. https://www.thoughtco.com/what-is-a-language-1691218
[26]. Fabio, A. and Arnt, L. J. (2021) The Routledge Handbook of Translation Cognitive first Edition, 378-382.
[27]. Jim, H. (2024). What is NLP (Natural Language Pr ocessing)? https://www.ibm.com/topics/natural-language-processing. Retrieved 27/09/2024.
[28]. Diego, L. Y. (2019). Your Giude to Natural Language Processing (NLP). Towards Data Science.
[29]. Education Ecosystem (LEDU). A Simple Introduction to Natural Language Processing. Becominghuman.ai/a simple-introduction-to-natural-language-processingea66a1747b32. Retrieved 27/09/2024.
[30]. Andi, W. and Zixin, J. (1998). Word Segmentation in Sentence Analysis. Microsoft Reseach, 1-10.
[31]. Encyclopedia of Bioinformatics and Computaional Biology, (2019). Science Direct.
[32]. SYSTRAN by Chapsvision: Rule-Based Machine Translation Vs Statistical Machine Translation. https://www.systransoft.com /systran/translation. Retrieved 27/9/2024.
[33]. Akan, M. F. (2014). The Lingistic Overview of Arabic and Bangla: a Comparative and Contrastive Analysis. Bangladesh Research Foundation Journal, Dhaka, Bangladesh, 3(1), 103-110.
[34]. Sonali, S., Manoj, D., Prabhishek, S., Vijendra, S., Seifedine, K. and Jungeun, K. (2023). Machine Translation Systems Based on Classical-Statistical-Deep-Learning Approaches. Electronics, 1-29.
[35]. Joshua, A. M. (2015). An Overview of Statistical Machine Translation. ResearchGate, 1-14
[36]. Lucia, B. and Lubomir, B. (2020). Neural Machine Translation as a Novel Approach to Machine Translation. Research Gate, 499-508.
[37]. Imelda, U and Ima, E. (2020), Nigerian Languages and Identity Crries. Language and Semiotic Studies, 6(3), 96-108.
[38]. Chirag, (2021). Step by Step Guide to Master NLP. Word Embedding and Text Vectorization. https://www.analyticsvidhya.com/blog/2021/06/part-5-step-by-step-guide-to-master-nlp-text-vectorization-approaches/. Retrieved 28/9/2024.
[39]. Adem, A. (2021). Word Embedding Techniques: Word2Vec and TF-IDF Explained. Towards Data Science.
[40]. Vaswani, A., Shazeer, N., Parmar, N., Jacob, U., Jones, L., Aidan, N. G., Kaiser, L. and Illia, P. (2023). Attention is All You Need. https://arxiv.org/pdf/1706.03762.
[41]. Mehree, S. (2023). A Gentle Introduction to Positioning Encoding in Transformer Model. Machine Learning Mastery.
[42]. Artificial Intelligence for Development-Africa Network (https://africa.ai4d.ai/)