Development of Bi-Directional English To Yoruba Translator for Real-Time Mobile Chatting

Adewale Olumide Sunday *adewale@futa.edu.ng*
Department of Computer Science, School of Computing,
Federal University of Technology,
Akure, Ondo-State, Nigeria

Agbonifo Oluwatoyin Catherine *ocagbonifo@futa.edu.ng*
Department of Information Systems, School of Computing,
Federal University of Technology,
Akure, Ondo-State, Nigeria

Olaniyan Julius *julaskky.julaskka@gmail.com*
Department of Computer Science, School of Computing
Federal University of Technology,
Akure, Ondo-State, Nigeria

Abstract

Machine translation (MT) is a subfield of computational linguistics that investigates the use of computer software to translate text or speech from one natural language to another. Translating between English language and Yoruba language comes with some computational complexities such as syntactic and grammatical differences in the language pair. This paper aims at exploring a multi-layer hybridized language translation approach, which combines the Corpus-based and Rule-based approaches of machine translation to generate its outputs. A parallel corpus was built with texts from English and Yoruba languages and stored in My Structured Query Language (MySQL) database. One hundred and forty seven computational rules were manually formulated and also stored in MySQL database for generating sentences in both languages. A di-bilingual dictionary was developed, one of which stored words in English with their corresponding Yoruba counterparts and their equivalent parts of speech while the other dictionary stored words in Yoruba with their corresponding English counterparts and their equivalent parts of speech. A real time mobile chatting interface was developed for users' interactions with themselves and the system. The research model was implemented using PHP for server-side scripting, JSON for data interchange and Java programming language for user interfaces accessible on users' mobile phones. The Java programming language was coded in Android Studio 3.0 Integrated Development Environment. Two hundred and eleven sentences from Contemporary English Grammar were used for system testing and the result shows 95% accuracy compare with Google Translate.

Keywords: Computational Linguistics, Corpus-based Machine Translation, Rule-based Machine Translation, English Language, Yoruba Language, Mobile Chatting.

1. INTRODUCTION

Natural language processing (NLP) also known as Computational Linguistics is a branch of computer science that aims to convert human language to a formal representation that can be easily manipulated by computers [1]. This computing field enables machine to read and understand the languages human being speaks.

One of the major tasks in NLP is Machine Translation, which is the central focus of this research work. Machine Translation, also known as Language Translation is defined as the transfer of text or speech from one human language to another. The language being translated is called the source language (SL) while the resulted language after translation is called the target language. It is a method of translating a particular language to another language in the form of a text or speech. Machine translation is still under investigation by linguists, philosophers, and computer scientists [2]. The translation is done by software and not by human being. There are three different approaches used in machine translation namely; Rule-based, Corpus-based and Hybrid. Rule-based approach in machine translation works by following sets of rules formulated by human translators to generate their output [3][4][5][6][7]. The formulated rules enable the software translator to generate correct output after translation without any major alterations in the meaning of the source language in the generated target language. Expert linguists are meant to create the rules because this approach involves profound linguistic expertise of both the source and target languages.

Three different approaches to Rule-based exist which include; Dictionary-based, Interlingua approach and Transfer-based [8][9][10][11][12]. Dictionary-based also known as direct approach is the simplest approach to machine translation because it is based on lexical transfer, or word-for-word (dictionary style) translation. In this model, words are simply translated as they occur [13]. Both transfer-based and interlingua-based machine translation approaches work in a similar way. To make a translation using either of these approaches, it is important to have an intermediate representation of the input sentence in order to generate the correct target sentence. However, in Interlingua-based machine translation, the intermediate representation must be independent of the languages involved, whereas in transfer-based machine translation, it has some dependence on the language pair involved.

Corpus-based systems use translation examples to infer the translation of the source text to generate their outputs [8][14][15][16]. There are two types of Corpus-based translation approach which are Statistical Machine Translation (SMT) and Example-based Translation. These approaches analyse a large number of original sentences or translated sentence pairs in order to discover which words or expressions in one language are most highly correlated with words or expressions in the other [17]. Statistical Machine Translation generates it output based on statistical models whose parameters are derived from text corpora while Example-based approach uses case based reasoning to generate its output. The third approach to machine translation is referred to as Hybrid Machine Translation. This approach combines the strengths and weaknesses of other but different approaches to machine translation to produce better results [18][13].

Nigeria is often refers to as the “Giant of Africa”, owing to its large population and economy with approximately 174 million inhabitants. Five Hundred and Twenty One languages have been spoken in Nigeria out which nine are now extinct. Hausa, Ibo, Yoruba, Ibibio, Edo, Fulfulde, and Kanuri are the major Nigerian’s indigenous languages, which are gradually going to extinction owing to dogmatic acceptance of the former colonial language, English, especially in the country’s urban communities. Yoruba is one of the three major languages spoken in Nigeria while the other languages are Hausa and Igbo which are referred to as regional languages [19][20][21].

Yoruba language is one of the mostly spoken languages in Nigeria with over 20 million speakers in south-western part of the country. English is a West Germanic language which has become a global Franca. English language is widely used as a means of communication among Nigerians. English language has been officially accepted in Nigeria and 53 other countries, such that it is the only recognised language in all domains: government and administration, education, science and technology, the media and others. English was the colonial language chosen as the lingua franca to facilitate both the cultural and linguistic unity among Nigerians. Communication in English language is more popular in Nigeria’s urban communities than it is in rural areas. Meanwhile urban communities comprise about three quarter of Nigeria population; this means that those who do not understand English language in urban communities are left out of information going on around them [22][23][24].

The Internet contains abundant amounts of useful documents in English language, which are inaccessible for most of the Yoruba speakers who do not understand English language. Therefore translation of those documents to Yoruba is necessary for making these useful online documents accessible for local use. Also, the world has actually turned to a global village and there is need for fast translation from English language which is a lingua franca to local languages such as Yoruba language. There is also fast growth in economic, political, technological, and cultural linkages that connect individuals, communities, businesses, and governments around the world [25][26].

According to the Nigerian Communication Commission (NCC), more than 2 million new mobile phone subscribers were registered between May and June 2015 making the total to rise from 144 to 146 million subscribers. In order to create textual rapport between these enormous Nigerian mobile phone users regardless of their linguistic dichotomy, there is need for a mobile-enabled bidirectional English/Yoruba (EY) translation system to allow for effective usage and dissemination [27].

The rest of the paper is organized as follows: In “Data and Methods”, the building of the system database and the EY translation model were fully discussed. Results and effectiveness of the developed system were shown in “Results and discussions”. And finally conclusion marks were provided in “Conclusion”.

2. RELATED WORKS

Zahum *et al. [7] developed* English to Bangla Phrased-Based Machine Translation, a phrase-based system that is capable of translating English text to Bangla text. Bangla, an Indo-Aryan language, is a language of South-Asia, which comprises the present day Bangladesh with nearly 230 million speakers. Phrase-based translation system based on Log-linear translation system was used as the baseline system.

All the available software toolkits for Machine translation system were used, which includes MOSES, GIZA++, SRILM, Mert, BLEU, NIST, TER, GMA sentence aligner. A transliteration system was in-corporated into the baseline system to handle proper nouns and Out-of-Vocabulary words (OOV). OOV words are words for which a system has no translation.

The drawback of the translation module incorporated into the system was that when names were being translated, the output translation consists of English words that are unknown words. Each translated name was XML marked up. Also, when OOV words were being translated, some words were not supposed to be translated. There was also a special component developed to handle preposition and grammatical differences between English and Bangla in a way that is accurate and efficient.

Aladesote *et al*. [20] used a Rule-based machine translation approach to develop a computational model of Yoruba Morphology Lexical Analyzer, which is also an essential part of language processing systems. The results obtained shows that the method adopted is satisfactory and would be of great importance to researchers who are working on Yoruba grammar. The system only analyses Yoruba morphemes.

Yetunde and Omonayin [6] used Statistical Machine Translation as a Translation Tool for Understanding Yoruba Language with an attempt to giving solution to the language barrier. Also, for the language modelling, SRLIM toolkit was used. Decoding was done using Moses, which is an SMT system that is also used to train translation models to produce phrase tables. Due to the approach used, the main resource needed which is the parallel corpora for English to Yoruba is not available, so it was created manually and this will affect the quality of the translation.

Google translate is a product of Google incorporation which has solved the problem of language barrier for over 64 languages. Google also developed a system for translating English language to Yoruba language using Statistical Machine Translation of corpus-based Machine Translation approach. The main limitation of this system is the lack of parallel corpora for English and Yoruba [5].

Abiola *et al.* [23] also attempted to solve the problem of language barrier by developing a Noun Phrase English to Yoruba machine translation, but the research work was limited to Noun Phrase which cannot translate a full English sentence to Yoruba language. Abayomi [22] developed and English to Yoruba Machine Translation System for Yoruba verbs Tone Changing.

Agbeyangi *et al*. [28] used rule-based machine translation approach to develop an English to Yoruba Machine Translation System. The system does not allow the creation of new additional words that are not present in the dictionary. In other words, the translation is strictly based on what is available in the dictionary, so words that are not contained in the dictionary are dropped.

Eludiora [15] developed a system which is an improvement over the system developed by Abiola et al. [23], in attempt to give a better solution. Rule-based approach was used in the research and computational rules were formulated. The system does not allow the creation of new additional words that are not present in the dictionary. In other words, the translation is strictly based on what is available in the dictionary, so words that are not contained in the dictionary are dropped.

Akinwale *et al*. [29] attempted the removal of language barrier among English and Yoruba language speakers by designing a web-based English to Yoruba Machine Translation system. It is an improvement on Abiola et al. [23]. The major drawback of this system is that it cannot translate complex English sentences to its Yoruba equivalent. It has no provision for proverbs semantically and its translation is just one directional.

3. SYSTEM DESIGN

This section discusses system architecture and mathematical model, system rule base and system modules.

3.1 System Architecture

Figure 1 shows the architectural design of the bi-directional EY translation system developed in this project. The architecture is made up of four major components namely; Store, Translator, Input and Output components. These components are fully explained as follows.

The input/output unit enables users to interact with each other textually on their mobile phones. This is where a user types his texts in either English or Yoruba language and sends them to the recipient while the recipient's textual response is also sent and displayed in the same interface. For example, let us say Mr. A wants to send an English sentence "I saw you in the market earlier today." to Mr. B., it is on this interface that the message would be typed and sent. Then the translated version of the sentence, which is "Mo ri e loja leekan loni" would be displayed on Mr. B's mobile phone still in the same interface.

The translator is the engine that performs the actual conversion of texts from one language to the other by interrogating the system dictionaries. Firstly, it checks the system's parallel corpus in the architecture to see if the user input exists in the database. If it exists, then the user's text is translated directly to the target text. This is known as direct translation. But in a case where the user's text is not found in the corpus, then the translator employs the rule-based approach of translation by first breaking the user input into variable chunks called tokens. This process is known as tokenisation. A token is discovered when a character or sequence of character delimited by a space or punctuation mark is encountered. The translator then checks whether a token is valid or not.

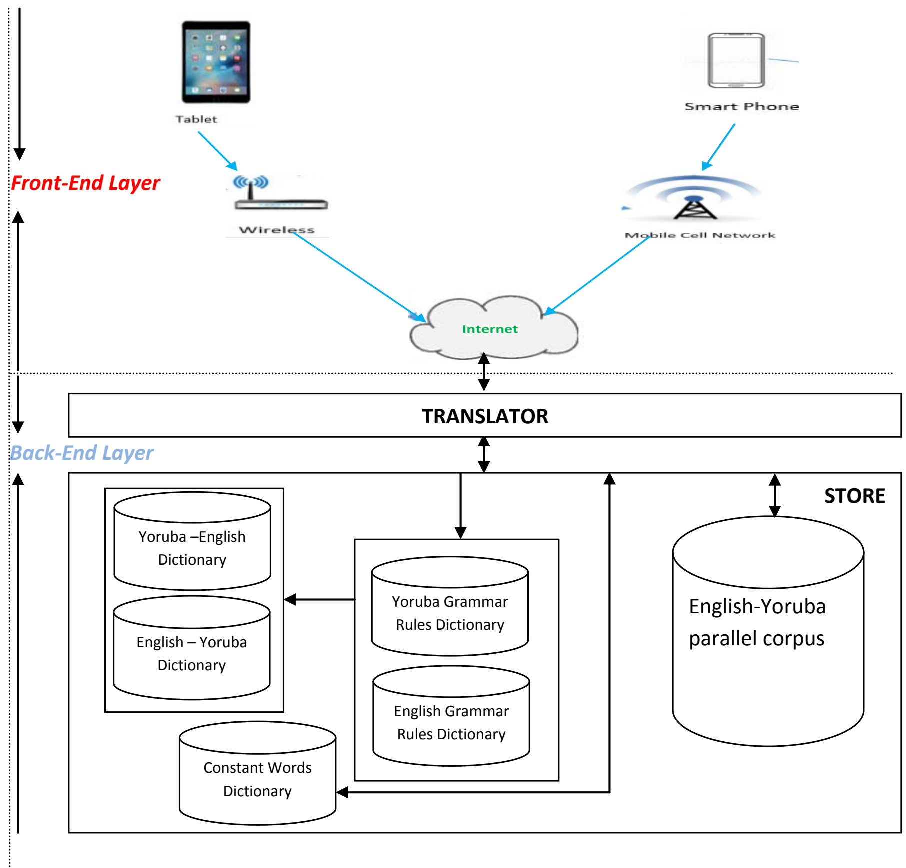

FIGURE 1: System Architecture of Bi-directional EY Translation.

It does this by comparing each token coming from the input stream with the valid ones stored in the system dictionaries. If a token is found invalid, then an error is flagged by the translator, else it assigns to each word or token its equivalent part of speech using the systems' rules dictionaries. This process is called tagging. The translator also ensures that words contained in user's sentence are reordered appropriately by using the rules dictionaries.

After word reordering, the translator replaces each token in the source language with its equivalent word in the target language using the system's English-Yoruba dictionaries and present the newly formed sentence to the user as the target sentence.

The store is an important component of the EY architecture where all the system dictionaries are kept. There are six dictionaries in the system architecture, which are: English-Yoruba parallel corpus, Yoruba rules dictionary, English rules dictionary, English-Yoruba lexicon, Yoruba-English lexicon and Constant words dictionary.

English-Yoruba Parallel Corpus: This is a dictionary of texts in English language together with their translated equivalents in Yoruba language with sentence level alignments stored in MySQL database. This dictionary is built from different sources namely; Bilingual New King James Bible, Bilingual Christ Apostolic Church Hymn Book, Newspapers, Advanced Learner Dictionary, etc.

Yoruba Rules Dictionary: The Yoruba rules dictionary is a MySQL database of all the rules used in sentence formation in Yoruba language. The rules depict the grammatical structures of all the obtainable Yoruba sentences. The rules were formed using the part of speech available in the language. Each rule is stored together with its translated equivalent rule in the target language that is used during parsing.

English Rules Dictionary: Like the Yoruba rules dictionary, the English rules dictionary is a complete collection of all the rules used in sentence formation in English language. The rules depict the grammatical structures of all the obtainable English sentences. The rules were formed using the part of speech available in the language. Each rule is stored in MySQL database together with its translated equivalent rule in the target language, which is used during parsing.

English-Yoruba Lexicon: The English-Yoruba lexicon is a comprehensive collection of words in English language and their equivalent meaning in Yoruba language stored in MySQL database. Each source word is stored together with its part of speech.

Yoruba-English Lexicon: The Yoruba-English lexicon like the English-Yoruba lexicon is a comprehensive collection of words in Yoruba language and their equivalent meaning in English language stored in MySQL database. Each source word is stored together with its corresponding parts of speech.

3.2 System Model

This section presents a bi-directional EY translation model that employed a multi-layered hybrid language translation approach combining both Corpus-based and Rule-based approaches of machine translation for effective and efficient bi-directional translation of English to Yoruba languages.

Let $T_1, T_2,\ldots,T_n$ be possible set of translated texts in the corpora database with $W_{T_1}, W_{T_2},\ldots, W_{T_n}$ as their respective weights, E be expected target sentence, W_E be expected target weight and F be final target sentence, then

$$F = \begin{cases} T_1 & \text{if } W_E/W_{T_1} < W_E / W_{T_i} \quad \forall\, i = 2,\ldots,n \\ T_2 & \text{if } W_E/W_{T_2} < W_E / W_{T_i} \quad \forall\, i = 1,3\ldots,n \\ \vdots & \\ T_n & \text{if } W_E/W_{T_n} < W_E / W_{T_i} \quad \forall\, i = 1,\ldots,n-1 \end{cases} \tag{1}$$

Let *S* be the input sentence in either of the languages, then the output(*F*) would be the sentence having a weight closest to the computed expected target weight(W_E) as selected from corpora database. That is, for a given sentence *S*, there may be several similar equivalent target sentences in the system database. Following equation 1, the weights of the target sentences are computed and are compared. The sentence with the highest weight is then selected as the final target sentence *F*. But when there is no related translation for a given input sentence, then control is transferred to the rule-based layer of the system.

3.3 Rules Extraction and Formation

A computational rule is manually formulated by extracting the part of speech of each word in the source sentence and aligning it with its equivalent parts of speech in the target sentence, and are stored both in MySQL database. The parts of speech are represented as: Noun (N), Pronoun(R), Verb (V), Adverb (A), Adjective (J), Determiner (D), Conjunction(C) and Interjection (I). Table 1 shows how rules are formulated for both language pair.

RNo	Source Sentence	English Rules	Yoruba Rules	Target Sentence
Rule1	That is the man whose car was stolen last Sunday.	D0V1D2N3D4N5 V6V7J8N9	N3D0D2N5V1V 6N9JV7D4	Okunrin naa toni okoayokele ton jigbe lojoisimi tokoja ni yen.
Rule2	My mother said she would not give me food until I finish my domestic chores	R0N1V2R3V4V5 V6R7N8C9R10V1 1R12J13N14	N1R0V2R3V4V 5V6R7N8C9R1 0N14J13R12	Mama mi sope oun ko ni fun mi lounje titi ma pari ise ile mi

TABLE 1: Sample Sentences and Rules.

The formation of rules, words replacements, words rearrangement and eventual generation of target sentence from a given input sentence is shown in Figure 1 and Figure 2 respectively.

First, consider Figure 1, where the input sentence is “That is the man whose car was stolen last Sunday.” The linguistic expert manually tokenizes the sentence to produce “that”, “is”, “the”, “man”, “whose”, “car”, “was”, “stolen”, “last”, and “Sunday” respectively. Next is manual tagging which is the process by which the expert manually labels each token according to the language grammars to produce - “that-D0”, “is-V1”, “the-D2”, “man-N3”, “whose-D4”, “car-N5”, “was-V6”, “stolen-V7”, “last-J8”, and “Sunday-N9” respectively. Bringing these grammatical tokens together, the linguist produces and stores D0V1D2N3D4N5V6V7J8N9 as the grammatical rule for and along with the input sentence, where D0V1D2N3D4N5V6V7J8N9 are Determiner with index position 0, Verb with index position 1, Determiner with index position 2, Noun with index position 3, Determiner with index position 4, Noun with index position 5,Verb with index position 6, Verb with index position 7, Adjective with index position 8, and Noun with index position 9 respectively. The index is the unique integer that identifies a given grammatical token.

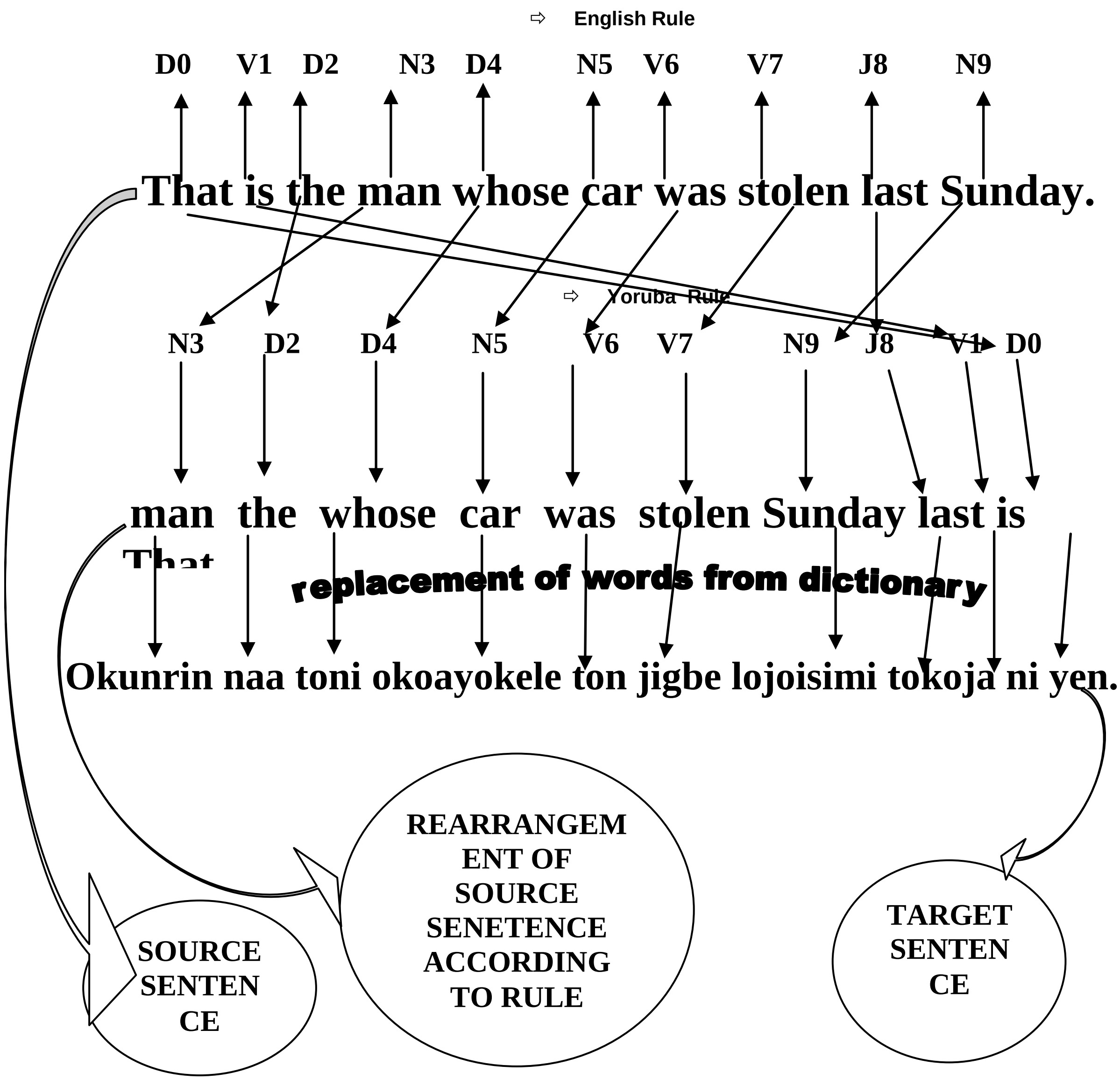

FIGURE 2: Formation of Rule 1.

Next is manual re-arrangement of the input tokens from the input sentence to produce - “man the whose car was stolen Sunday last is that.” using the grammatical structures of the target language. Then, the linguistic expert manually labels each token according to the language grammars to produce “man-N3”, “the-D2”,“whose-D4”, “car-N5”, “was-V6”, “stolen-V7”,“Sunday-N9”, “last-J8”, “is-V1”, and “that-D0” respectively. Bringing these grammatical tokens together, the linguist produces and stores N3D2D4N5V6V7N9J8V1D0 as the grammatical rule for and along with the target sentence, where N3D2D4N5V6V7N9J8V1D0 are Noun with index position 3, Determiner with index position 2, Determiner with index position 4, Noun with index position 5, Verb with index position 6, Verb with index position 7, Noun with index position 9, Verb with index position 1, and Determiner with index position 0 respectively.

The system does the actual translation by replacing each token with the stored equivalent target word as it occurs in the pre-processed input sentence. This produced – “Okunrin naa toni okoayokele ton jigbe lojoisimi tokoja ni yen.” from the pre-processed sentence – “man the whose car was stolen Sunday last is that", by replacing each word with its equivalent from the database using the formulated grammatical tokens – "N3D2D4N5V6V7N9J8V1D0."

Grammatical rule 2 was also formulated in a way similar to the formation of rule 1. Consider Figure 2, where the input sentence is "My mother said she would not give me food until I finish my domestic chores." The linguistic expert manually tokenizes the sentence to produce "My", "mother", "said", "she", "would", "not", "give", "me", "food", "until","I","finish","my","domestic" and "chores" respectively. Next is manual tagging where the expert manually labels each token in consonance with the grammatical structures of the source language to produce - "My-R0", "mother-N1", "said-V2", "she-R3", "would-V4", "not-V5", "give-V6", "me-R7", "food-N8", "until-C9","I-R10","finish-V11","my-R12","domestic-J13" and "chores-N14" respectively. Bringing these grammatical tokens together, the linguist produces and stores R0N1V2R3V4V5V6R7N8C9R10V11R12J13N14 as the grammatical rule for and along with the input sentence, where R0N1V2R3V4V5V6R7N8C9R10V11R12J13N14 are Preposition with index position 0, Noun with index position 1,Verb with index position 2, Preposition with index position 3, Verbwith index position 4, Verbwith index position 5, Verbwith index position 6, Preposition with index position 7, Noun with index position 8, Conjunctionwith index position 9, Preposition with index position 10, Verb with index position 11, Preposition with index position 12, Adjective with index position 13 and Noun with index position 14 respectively. The index is the unique integer that identifies a given grammatical token.

Next is manual re-arrangement (pre-processing) of the input tokens from the input sentence to produce - "mother my said she not would give me food until I finish chores domestic my" using the grammatical structures of the target language. Then, the linguistic expert manually labels each token according to the language grammars to produce "mother-N1","My-R0","said-V2","she-R3","not-V5", "would-V4","give-V6","me-R7","food-N8","until- C9", "I- R10","finish-V11","chores-N14","domestic-J13", and "my-R12" respectively.

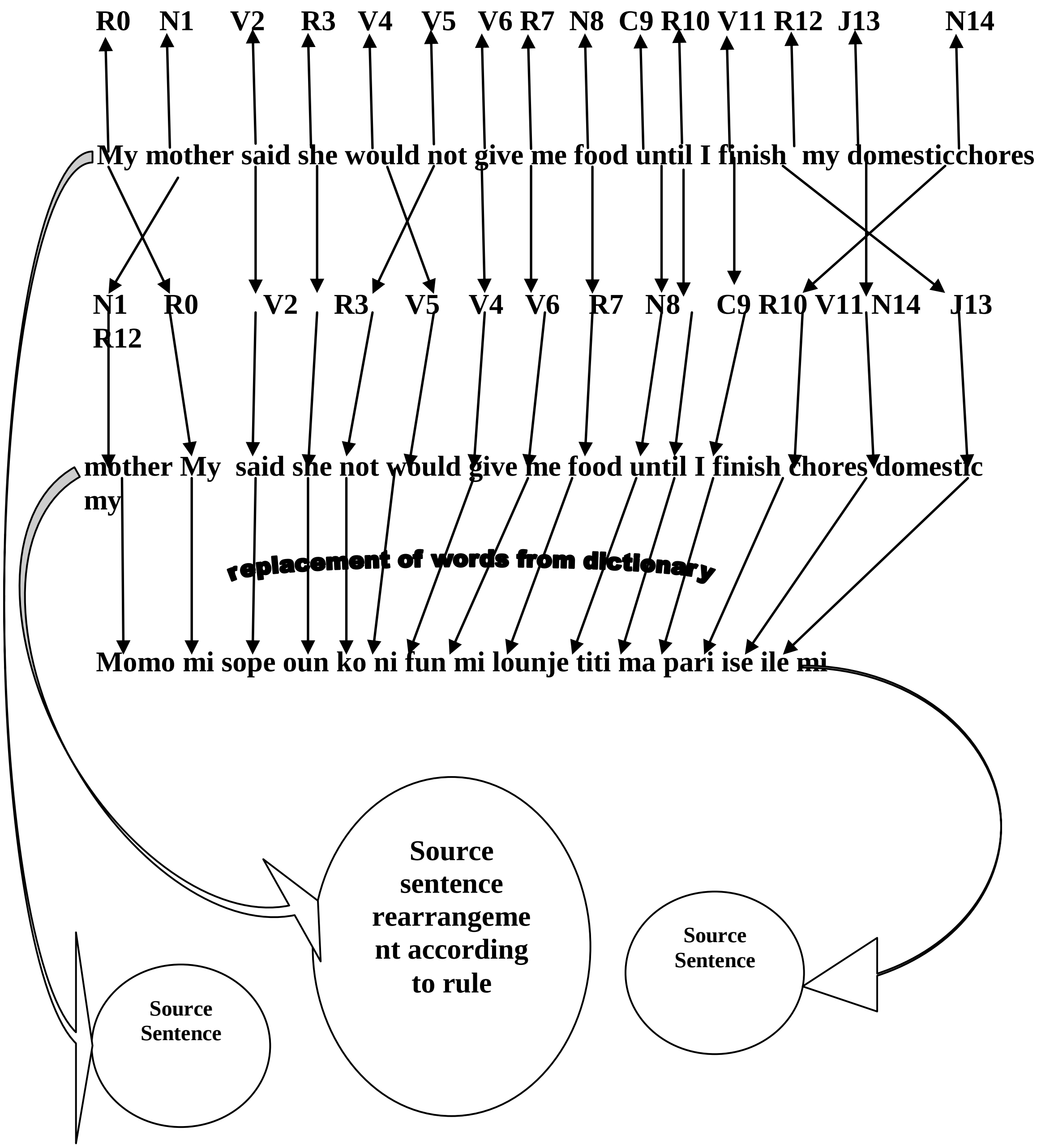

FIGURE 3: Formation of Rule 2.

By combining these grammatical tokens together, the linguist produces and stores N1R0V2R3V5V4V6R7N8C9R10N14J13R12 as the grammatical rule for and along with the target sentence, where N1R0V2R3V5V4V6R7N8C9R10N11J13R12 are Noun with index position 1,Preposition with index position 0,Verb with index position 2, Preposition with index position 3, Verbwith index position 5, Verbwith index position 4, Verbwith index position 6, Preposition with index position 7, Noun with index position 8, Conjunctionwith index position 9, Preposition with index position 10, Noun with index position 11, Adjective with index position 13 and Preposition with index position 12 respectively. The index is the unique integer that identifies a given grammatical token.

Using the manually generated rules, the system performs the actual translation by replacing each token with the stored equivalent target word as it occurs in the pre-processed input sentence. This then produced – “Momo mi sope oun ko ni fun mi lounje titi ma pari ise ile mi” from the pre-processed sentence – “mother My said she not would give me food until I finish chores domestic my”, by replacing each word with its equivalent target word from the database using the formulated grammatical tokens – N1R0V2R3V5V4V6R7N8C9R10N11J13R12.”

4. SYSTEM MODULES

This section discusses some of the modules that are integrated in the implementation of the system.

4.1 Registration Page

The user registration page as shown in Figure 4 enables users to register on the system’s chatting platform by entering their names, phone numbers, username, passwords and email addresses which are stored in the system database.

FIGURE 4: User Registration Page.

4.2 Translation Page

Figure 5 shows the translation page that enables a Yoruba speaker to chat with an English speaker mutually on their mobile phones. It is divided into two panes; the upper pane that keeps the chatting history and the lower pane that has an editable textbox where users can type their messages in either English or Yoruba languages. The chatting history in the upper section of the page is a mixture of translated messages in both English and Yoruba languages. The translation is done when the Send button below the editable textbox on the page is clicked.

Also, for any unsuccessful translation, an error is popped up on top of the translation window for user to see. The Add button provides a way of collating the sentences that could not be translated and store same in a temporary table for analysis by the system administrator, which he later adds to both the system’s corpus and rule-base respectively.

FIGURE 5: Chatting Page.

5. RESULTS AND DISCUSSIONS

In this section, the system evaluation is carried out by comparing the efficiency of the translator with Google translate which is still the readily available translator online for Yoruba language. Based on the rules formulated for this translator, some sentences were formed and tested on both translators. Table 2 is an extract of some of the sentences and the output of the two translators.

S/N	Input Sentences	EY-Translation Model Output	Google Translate Output
1	Can you pass me the salt please?	Se o le bami mu iyo naa jowo?	O le ṣe mi ni iyọ jọwọ?
2	Look at those lovely flowers	Wo awon ododo daradara wonyi.	Wo ni àwọn ẹlẹwà awọn ododo
3	Thank you very much for your letters	O se gaan fun awon leta re.	O ṣeun gidigidi fun awọn lẹta
4	Whose coat is this?	Talo laso yii?	Ti ndan ni yi?
5	Milk is very good for you	Wara dara gidigidi fun o.	Wara jẹ gidigidi dara fun o
6	Health and Education are very important	Ilera ati eko se pataki.	Ilera ati Education ni o wa gidigidi Pataki
7	Girls normally do better in school than boys	Awon omobinrin ma a nse daradara nile iwe ju awon omokunrin lo.	Girls deede ṣe dara ni ile-iwe ju omokunrin

8	A woman was lifted to safety by a helicopter	Oko-ojuofurufu kan lo gba obinrin kan la.	A obinrin ti a gbe to ailewu nipa a ọkọ ofurufu
9	A man climbing nearby saw the accident	Okurin kan to n pon ri ijamba naa.	A ọkunrin gígun nitosi ri ijamba
10	We use the general determiner any with a singular noun or an uncount noun when we are talking about all of those people or things	A n lo ditamina gbogbogboo pelu eyo oro-oruko kan tabi oro-oruko kolounka nigbati a ba n soro nipa gbogbo awon eniyan yen tabi awon nka won yen.	A lo gbogbo determiner eyikeyi pẹlu kan okan nọun tabi awọn ẹya uncount nọun nigba ti a ba wa ni sọrọ nipa gbogbo awọn ti awon eniyan tabi ohun

TABLE 2: Translation Samples.

The bi-directional EY-Translator and Google Translate were used to translate 211 sentences. Some of the translated sentences were used in formulating the rules for the bi-directional EY-Translation model while some were part of the system's parallel corpus.

From the result shown in Table 3, the partial translation refers to sentences translated that are grammatically correct but not perfectly translated. Partial Accuracy is the percentage of sentences translated that are grammatically correct with perfect translation. Wrongly translated are sentences that are not perfectly translated and grammatically correct in Yoruba language. Figure 6 shows the number of correctly translated and wrongly translated for EY-translator and Google translate.

Translator	Sentences Generated	Correctly Translated	Wrongly Translated	Partially Translated	Accuracy %
EY-Translator	211	201	8	2	95.3
Google Translate	211	7	201	3	3.3

TABLE 3: Translation Sample Results.

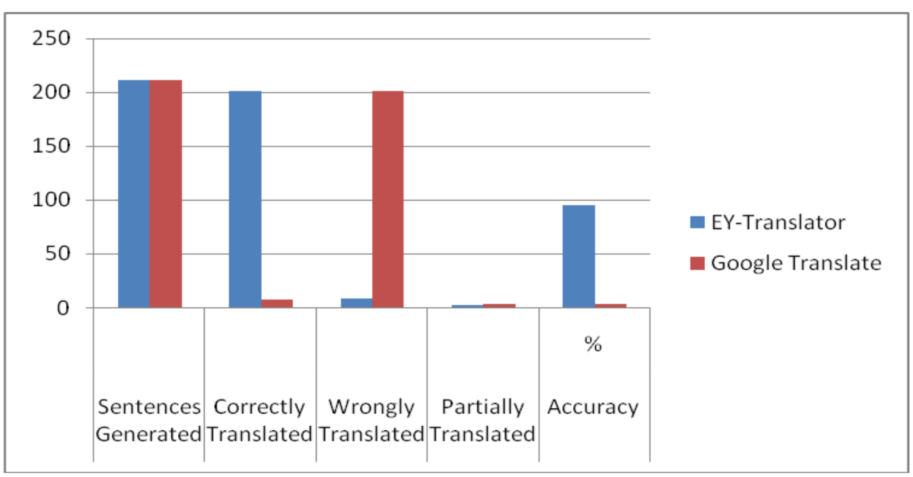

FIGURE 6: Evaluation Chart.

6. CONCLUSION AND FUTURE DIRECTION

One of the key challenges in language translation is the problem of not having a common language framework for developing a generic multi-language translation system, which this paper attempted to address to a large extent by employing a multi-layered hybridized language translation approach. This approach is relatively effective with essential features such as providing a robust platform for creating, maintaining and extending the computational rules, the system corpus and the language vocabularies thus enabling the linguistic experts to build their own language translation system of their choice with ease and without writing a single line of code.

In this paper, an hybridized multi-layered bi-directional language translator is implemented using PHP (Hypertext Pre-Processor) server-side scripting language, JSON (JavaScript Object Notation), Java programming language and MySQL (My Structured Query Language) as the backend database. It was tested on Tecno K7 running Android version 7.0 with 16GB internal memory. Two hundred and eleven sentences from Contemporary English Grammar by Jayanthi Dakshina Murthy were tested and gave accuracy of 95%, which was quite encouraging.

The future research work should focus on improving the system's computational rules in order to efficiently handle the language semantics and also for additional English and Yoruba bitexts in the system's bilingual corpus for efficient translations.

7. REFERENCES

[1] C. Ronan and W. Jason. "A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitasking Learning". *Appearing in Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland*, 2008.

[2] A. Mohammed. "Theoretical Overview of Machine translation". Proceedings ICWIT 2012 of the 45th Annual Meeting of the ACL, 2012, pp..25–32.

[3] J. G. Rubio. "Effective Deployment of Current Machine Translation Technology". LRC Best Thesis Award Winner, 2014.

[4] P. Senellart. "English as an international language". In English Teaching Forum Oxford Pergamon,2001.

[5] H. Stephen and P. S. Carmen. "Translation and the Internet: Evaluating the Quality of free Online Machine Translators", vol.17, pp.197-209, 2010.

[6] O. F. Yetunde and I. Omonayin. "Using Statistical Machine Translation (SMT) as a Language Translation Tool for Understanding Yoruba Language", EIE's 2nd Cont. Comp., Energy, Net, Robotics and Telecom, 2012.

[7] P. Zahum. "Phrase-Based Machine Translation System", Journal of Communication Monograph, vol.7, Issue 1, pp.13-36, 2009.

[8] O. Anthony. "Principles for mediation between cultures", *The Translator*, vol.7, Issue 2, pp. 297-321. 2013.

[9] S. I. Eludiora, (2014). "Development of a Machine Translation System for English to Yorùbá." Unpublished PhD Thesis, Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria, 2014

[10] S. I. Eludiora, T. Okunola and A. O. Odetunji. "Computational Modelling of Yoruba Split-Verbs for English to Yoruba Machine Translation System", International Journal of Advanced Research in Computer Science and Applications, vol.3, Issue 4, 2015.

[11] O. O. Fagbolu, B. K. Alese, O. S. Adewale, and A. O. Adetunmbi. "Android Platform for Machine Translation – A focus on Yoruba Language,"American Journal of Computation, Communication and Control, vol.5, No.1, 2018, pp.16-23.

[12] F. Henderson. "Giving a Voice to More languages on Google Translate, The official Google Translate blog," Available: http://googletranslate.blogspot.com/2010/05/giving-voice-to-more-languages-on.html, Oct. 23, 2010 [February 11, 2015].

[13] A. Mike, F. Taylor and P. Witt. "A meta-analytical review of the relationship between teacher immediacy and student learning," Journal of Communication Monograph, vol.71, Issue 2. 2004.

[14] M. Attia. (2014, May). "Handling Arabic morphological and syntactic ambiguity within the LFG framework with a view to Machine Translation," School of Languages, Linguistics and Cultures, The University of Manchester, UK.

[15] O. Chux "Language extinction". Internet: http://www.punchng.com/feature/gradually-nigerian-languages-are-dying/. May 23, 2013 [February 15, 2015].

[16] T. G. Dienes, H. Hild and G. Bakiri. "A Comparison of ID3 and Backpropagation for English Text-to-speech Mapping," Machine Learning, vol.18, Issue 1, pp.51-80, 2001.

[17] J. Daniel. "Speech and Language Processing," An Introduction to Natural Language Processing. Proceedings of the Eighth Workshop on Statistical Machine Translation, 2011, pp.460–467.

[18] M. Mahsa. "English - Persian Phrase - Based Statistical Machine Translation: Enhanced Models, Search and Training." Doctoral dissertation, Massey University, Albany (Auckland), New Zealand, 2012.

[19] J. Adeyemi. "English Language Dominance and the Fate of Indigenous Languages in Nigeria." International Journal of History and Cultural Studies (IJHCS) vol.2, Issue 4, pp.10-17. 2015

[20] I. Aladesote, O. E. Olaseni, A. O. Adetunmbi and F. Akinbohun. "A Computational Model of Yoruba Morphology Lexical Analyzer." International Journal of Computational Linguistics (IJCL), vol.2, 2011.

[21] B. A. Lawrence, O. A. Adebayo, K. A. Bonaface and A. O. Samuel. "Token validation in Automatic Corpus Gathering for Yoruba Language." FUOYE Journal of Engineering and Technology. 2017, pp.114-117.

[22] O. A. Abayomi. "Development of English to Yoruba Machine Translation System for Yoruba Verbs Tone Changing." International Journal of Advanced Research in Computer Science and Applications. Vol.3, Issue 4, April, 2015.

[23] O. B. Abiola, A. O. Adetunmbi, A. I. Fasiku and K. A. Olatunji. "A Web-Based English to Yoruba Noun-Phrases Machine Translation System." International Journal of English and Literature, 2014, pp.71-78, 2014.

[24] E. Adebimpe, O. Bolaji, O. Olatayo, O. Precious and O. Timileyin. "Development of Adjectival Phrase-Based English to Yoruba Machine Translator." International Digital Organization for Scientific Research IDOSR JOURNAL OF APPLIED SCIENCES, vol.3, Issue 1, pp.29-42, 2018.

[25] O. Adeoye. “A Web-Based English to Yoruba Noun-Phrase Machine Translation.” Master’s Thesis, Federal University of Technology, Akure, Nigeria, 2012.

[26] K. Aditi and S. Priti. “Review of Machine Translation Systems in India and different Translation Evaluation Methodologies.” International Journal of Computer Applications, vol.121, Issue 3, pp.17-18, 2015.

[27] The Communicator Magazine. “16 years of GSM in Nigeria,” Available: https://ncc.gov.ng/thecommunicator/index.php?option=com_content&view=article&id=1553&catid=13&Itemid=140, December, 2018 [May, 2019].

[28] A. O. Agbeyangi, S. I. Eludiora and O. A. Adenekan. “English to Yoruba Machine Translation System using Rule-Based Approach.” Journal of Multidisciplinary Engineering Science and Technology. Vol.2, Issue 8, Pp.7-11, 2015.

[29] O. I. Akinwale, A. O. Adetunmbi, O. O. Obe and A. T. Adesuyi. “Web-Based English to Yorùbá Machine Translation.” International Journal of Language and Linguistics. Vol.3, issue 3, pp 154-159, 2015.