IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11126

ENGLISH TO NUPE MACHINE TRANSLATION SYSTEM USING RULE-BASED APPROACH

Sayuti Musa Shafi'i [1], Hassan Suru [2], Abdulrahman Mohammed Saba [3]

Department of Information Technology, Continental Transfert Technique Limited, Abuja, Nigeria [1]

Department of Computer Science, Kebbi State University of Science and Technology Aliero, Nigeria [2]

Department of Computer Science, The Federal Polytechnic Bida, Nigeria [3]

Abstract: In Nigeria, the dominance of English as a medium of communication is worrying as minor indigenous languages are heading towards extinction. The main objective of this study is to develop a language framework that could take English input and translate it into Nupe. The theoretical framework was first considered. Due to the differences between the systems between English and Nupe and the lack of similar arrangements in English and Nupe languages, Transfer Rule-Based Machine Translation was used in the development of the system. It is used because it allows manual speech tagging, design and modelled of language translation systems. This process was designed using Unified Modeling Language (UML). UML is used to design the system flowchart, sequence diagram, use case diagram and class diagrams. A bilingual (dictionary) database is designed to store source language (English) and target language (Nupe) words with the Python programming language interpreter and test the translation grammar using a natural language application(NLTK). The results show that the system can provide translation of acceptable quality. The system can benefit many kinds of people as it allows them to process their translations quickly and easily.

Keywords: English, Nupe, Rule-based Approach, Machine Translation, Bilingual, tagging, Unified Modeling Language.

I. INTRODUCTION

A language is a medium of communication. Human language communicates ideas, emotions, feelings, desires, to co-operate among social groups, to exhibit habits etc., which can be translated along a variety of channels [8]. It is estimated that there are more than 250 established language families in the world, and over 6,800 distinct languages, many of which are threatened or endangered [13]. Access to information written in a language is very exciting because it helps to express our thoughts, desires and queries to the world around us, and the means of sharing information across languages is translation, therefore creating tools for translating from one language to another is a very crucial contribution to human development. Without interpretation, there is no communication, except between those who speak the same language, and many voices will not be heard without this critical function. Translation is critical for addressing the inequalities of information. A study conducted by Common Sense Advisory on behalf of Translators without Borders finds out that translation is critical for the public health, political stability, and social wellbeing of African nations [14]. According to [6], development depends on knowledge and knowledge is delivered through language and for knowledge to be properly shared, language must be shared. To share language, the language can either be learned or translated and both can be achieved using Natural Language Processing (NLP) and Machine Translation (MT).

The work of translation was originally carried out by human translators. At a point the supply of translation services could no longer keep pace with the demand for translated content. Moreover, human translation is costly, time consuming and inadequate for addressing the real-time needs of businesses to serve multilingual prospects, partners and customers. The inherent limitations of human translation made the search for an alternative means of translation paramount [18]. The search led to the discovery of what is known today as machine translation or computer assisted translation. Machine Translation is the use of computers to automate some or all of the process of translating from one language to another [17].

There are probably about 3.5 million Nupes, principally in Niger State. The Nupe language is also spoken in Kwara, Kogi and Federal Capital Territory. They are primarily Muslims, with some Christians and followers of African Traditional Religion. The Nupe people have several local traditional rulers. The Etsu Nupe (Bida) is not pure Nupe; his great grandfather from his father side is Fulani while the family of his mother was complete Nupe. His great grandfather from his father side came to rule Bida in the 1806. They were originally based at Rabah and only moved to Bida in the nineteenth century [12]. By language typology, Nupe, according to [9], belongs to the Kwa-language sub-group alongside

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11126

Gwari and Ebira etc. Nupe has had a written tradition and equally been studied as far back as the 1850s by missionaries like Bishop Samuel Ajayi Crowther and Henry Johnson. Benfield, especially produced 'Dictionary of Nupe Language' (volumes I and II) in 1914 and 1916. In fact, the whole Bible was first translated into Nupe language in 1953 [19]. It is widely believed that the ancestor of the Nupe was Ugban bn Nafi "who was said to have migrated from across North-East Africa to Nubia and then to Nupe land" [10]. The people now known as the Nupe are predominantly agrarian who subsequently came to live in the various Nupe lands as a result of the mass movement down south by various people during the desiccation of the Sahara (Ibid). The area where the language is predominantly spoken lies between the River Niger in the south and west and the Kaduna River in the north [4]. The Nupe people in their relation with other ethnic groups in Nigeria have been called names. The Hausa call them 'Banufe' (one man) or Nufawa' (all people); the Yoruba call them 'Takpa'; the Gwari, the 'Abawa'; the Igbira, 'Anupe' and the Kakanda, 'Anupecwayi'. The missionaries, western and Arab scholars of the 18th and 19th centuries, referred to the Nupe in their scholarly works as either 'Nefiu', 'Nife' or 'Nyffe' [11]. The Nupe language like many languages is not homogeneous it has from the past to the present developed into dialects: Nupe proper (or central), Ebe, Dibo (or zhitako), Kupa, Gbedeye or Gbedegi and Basa-nge.
The aim of the research is to develop a machine translation system, which translates English to Nupe language. The objectives of the research are: Search Literature, analyze English language syntactic rules and their Nupe counterparts, design a model for translation of English to Nupe language and implement the model.

II. RELATED WORKS

[2] developed Rule-based approach is a good approach for Machine Translation System used for language with lots of grammar which Yorùbá language is one. In this paper we present Rule-based approach to Yorúbà Machine Translation System. The popularity of Yorúbà language among the three main languages in Nigeria calls for the need to computerize the language. Transfer Rule-Based Machine Translation is use in the development of the System. It was used because it allows us to use manual tagging of the part of speech (POS). Rewrite rules was developed for the two languages (Yorùbá and English). The data was collected from home domain vocabularies. The re-write rule was verified using Natural Language Toolkits (NLTKs) and implement using python programming language.
[3] proposed a Rule-Based MT of Noun phrases from Punjabi to English. The approach used for this study was rule-based transfer approach. The steps involved are pre-processing, tagging, ambiguity resolution, translation and synthesis of words in the target language.
[1] developed English to Yorùbá Machine Translation System using rule-based approach. The motivation for the work was based on the need to contribute to knowledge in machine translations by experimenting with an African language. Text corpora were collected from home domain, context-free grammars were used to model the two languages, re-write rules and parse trees were also used. Automata theory was used to recognize the computational problem underlining the translation process.
[16] Developed an English to Yorùbá Machine Translator. The modelled translation process was designed, implemented and evaluated. This was with a view to addressing the challenge of English to Yorùbá text machine translator. This machine translator can translate modify and non-modify simple sentences (subject verb object (SVO)).
Digital resources in English and its equivalence in Yorùbá were collected using the home domain terminologies and lexical corpus construction techniques. The English to Yorùbá translation process was modelled using phrase structure grammar and re-write rules. The re-write rules were designed and tested using Natural Language Tool Kits (NLTKs). Parse tree and Automata theory based techniques were used to analyse the formulated model. Unified Modeling Language (UML) was used for the software design. The Python programming language and PyQt4 tools were used to implement the model. The developed machine translator was tested with simple sentences. The results for the Basic Subject-Verb-Object (BSVO) and Modified SVO (MSVO) sentences translation show that the total Experimental Subject Respondents (ESRs), machine translator and human expert average scores for word syllable, word orthography, and sentence syntax accuracies were 66.7 percent, 82.3 percent, and 100 percent, respectively. The system translation accuracies were close to a human expert.
[5] worked on Development of an English to Yoruba Machine Translation system. The research work carried out computational analysis of English to Yoruba texts translation process. Rule-based approach was used to carry out the research. The translator was modelled using context-free grammar and re-write rules, Parse Tree and Automata theory-based techniques and design of corresponding software using UML.
English to Yorùbá Statistical Machine Translation System was developed by [7]. The motivation for the work was due to the observation of Nigeria languages going towards total extinction. Phrase-based machine translation system was formulated to translate English to Yorùbá sentences and vice-versa. The system was implemented with Moses toolkits and evaluated with BLEU score.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11126

III. SYSTEM DESIGN FRAMEWORK

A. PROPOSED MODEL

Figure 1 is architectural diagram of the overall design of the proposed system. It has four main components, source text, preprocessing stage, lexical transfer, Target Language, Syntactic Generator, Target Text and Bilingual Lexicon.

- Source Text

This is the first system in the architecture, when the user enters the English-based script in the system; The system receives the source script and then enters the pre-processing stage.

- Pre-processing Stage
- At the preprocessing stage, the source text is first stored, after which it goes through preprocessing stage to determine the number of words in it. The system stops whenever a space is encountered, which signifies the end of a word and it eliminates the space automatically and moves to the lexical transfer stage.

- Lexical Transfer

The outputs from the preprocessing stage enter lexical stage. At this stage, the corresponding target text equivalent is assigned to each word of the input text.

- Target Language Syntactic Generator

The target text equivalent obtained for the source text in the lexical transfer phase enters this phase to produce the target text output.

- Target Text

This module displays the output of the source text in Nupe language.

Bilingual Lexicon: The bilingual-lexicon is the repository for English phrase and phrase Language building terminologies respectively. The architecture is clearly illustrated in Fig.1

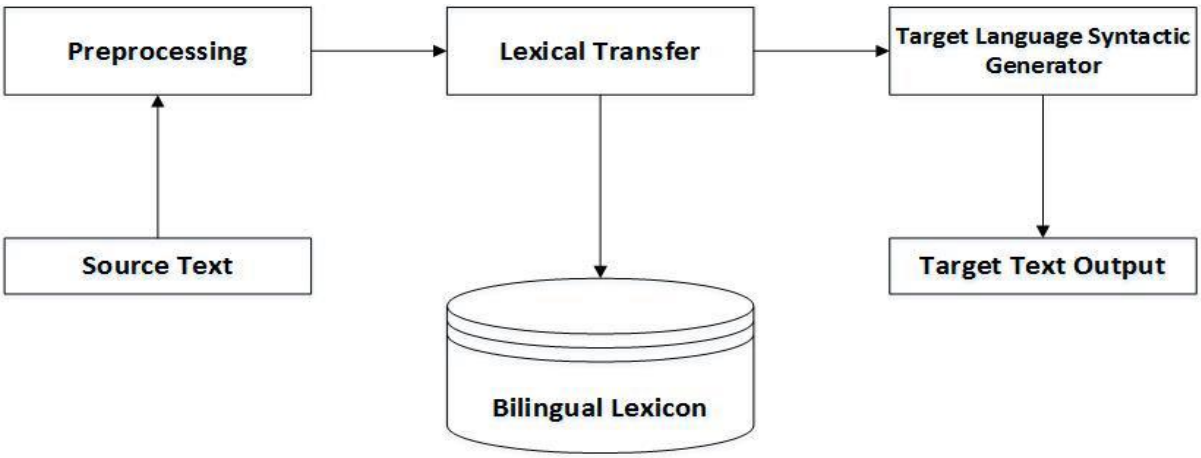

Fig. 1 Architecture of the System

B. RULE-BASED MACHINE TRANSLATION (RBMT)

Rule-based Machine Translation (RBMT) also known as `Knowledge-based Machine Translation', `Classical Approach' of MT is a general term that denotes machine translation systems based on linguistic information about source and target languages. Basically the linguistic information can be retrieved from (bilingual) dictionaries and grammars covering the main semantic, morphological and syntactic regularities of each language [15]. Rules play major role in various stages of the translation: syntactic processing, semantic interpretation, and contextual processing of language.

A noun phrase (NP) is a phrase in which a noun or pronoun is the head word, optionally accompanied by a modifier set. NP can be pre-modified or post-modified. If the modifier is placed before the noun, the NP is pre-modified. If the modifier is placed after the noun, then the NP is post-modified. English allows both forms of modification. Nupe allows only post-modification, modifiers are placed after nouns. A noun phrase consists of three parts, the head which is the principal part and other two optionally occurring parts. Possible modifiers of NP are:

Definite Articles (the)
Indefinite Articles (a, an)
Demonstratives (this, that, such, these, those, none, neither)
Quantifiers (few, every, several, all, no, some, any, more, most, less, enough)
Cardinal numbers (one, two, three …)
Ordinal numbers (first, second, third…)
Possessive Pronouns (my, your, their, our, his, her, its, mine, ours, yours…)

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11126

Pre- determiners (all, both, half, what, rather, such, quite…)
Nupe and English languages differ in the syntax of noun phrases. The placement of modifiers is not the same in the two languages.
Nouns means N
Definite Articles by DA
Indefinite Articles by IDA
Demonstratives by DEM
Cardinal Numbers by CDN
Quantifiers by QTN
Ordinal Numbers by ODN
Possessive pronouns by PPN
Pre-determiners by PDT
Where R1 to R20 are the rules numbers: R1 means rule 1…
For example, the little child means Egi Dzákàṇgi in Nupe language

NP ⟶ (Dart) (Adj) (N) (Rule 3)
⟶ The (Adj) (N)
⟶ The little (N)
⟶ The little child

The Nupe arrangement of the phrase is given as:

NP ⟶ (N) (Adj) (Dart)
⟶ Egi (Adj) (Dart)
⟶ Dzakagi (Dart)
⟶ Egi Dzakangi

PROPOSED SYSTEMS COMPONENTS

TABLE 1: RE-WRITE RULES

Rules	English Phrases	Examples	Nupe Phrases	Examples
R1	DA + N	The kitchen	N	katanjebo
R2	IDA + N	A House	N	Emi
R3	DA+ADJ+N	The beautiful house	N+ADJ+DA	Emi bologi
R4	IDA+ADJ+N	A red cap	N + ADJ	Fula dzuru
R5	DA+ADJ+ADJ+N	The beautiful white dress	N+ADJ+DA+ADJ	Ewo bokun e bologi
R6	IDA+ADJ+ADJ+N	A beautiful white dress	N+ADJ+DA+ADJ	Ewo bokun e bologi
R7	PPN + N	My horse	N + PPN	Doko mi
R8	PPN+ADJ+N	My red cap	N+ADJ+PPN	Fula dzuru mi
R9	PPN+ADJ+ADJ+N	My beautiful white shirt	N+ADJ+PPN+ADJ	Ewo bokumie bologi
R10	DEM + N	That house	N + DEM	Emi ga
R11	PDT + N	Few bachelors	N + PDT	Ekpa deegi
R12	PDT+PPN+N	All their houses	N+PPN+PDT	Emi a kpata
R13	CDN + N	Two doors	N + CDN	Kpako guba

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11126

R14	ODN + N	Second house	N + ODN	Emi bachi
R15	DA+CDN+N	The three doors	N+CDN	Kpako guta
R16	DA+ODN+N	The third building	N+ODN	Emi tachi
R17	DEM+CDN+N	That fourth house	N+CDN+DEM	Emi nechi ga
R18	PPN+ODN+N	My second horse	N+ODN+PPN	Doko bachi mi
R19	QTN + N	Many rooms	N + QTN	Kata zhi
R20	QTN+ADJ+N	Few beautiful houses	N + QTN + ADJ	Emi deegi ebolo

Fig. 2 shows the Noun phrase translation abstraction, from the source language to intermediate translation to the target language translation. Fig. 2(a) is the English (source language) sentence to be translated. Fig. 2(b) is the intermediate representation of Fig. 2 (a), this shows how the sentence is being re-arranged. The word "is" is removed in fig. 2(c). Fig. 2(d) is the target language.

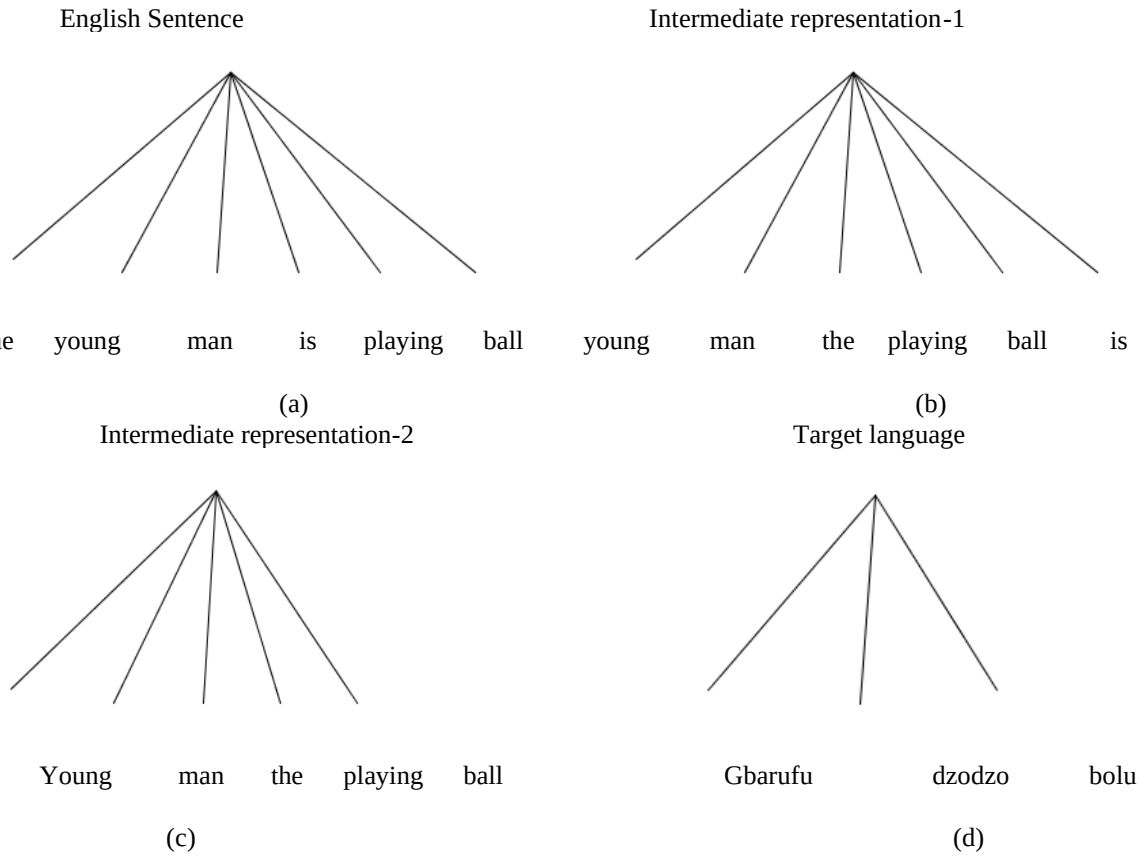

Fig. 2 English to Nupe translation process abstraction

IV. SYSTEM DESIGN AND IMPLEMENTATION

This section discusses software design and implementation,

A. SOFTWARE DESIGN

Unified Modelling Language (UML) diagrams of the object-oriented programming modelling methodology were used to model the behaviour of the machine translation. System modelling diagrams included Flowchart, data flow diagram, Use Case Diagram, Sequence Diagram and Finite State Transducer.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11126

The Flow Chart in Fig. 3 shows the sequential steps in the process of machine translation. The translation process starts with a user running the Machine Translation System and inputting an English phrase. The machine uses English language linguistic rules to decide if the supplied phrase is a valid English language. If invalid, the user has the opportunity of providing a grammatically correct phrase and if valid, a Nupe phrase is generated and displayed to the user. The output signifies the end of the translation process.

Fig. 4 is the Data Flow Diagram that shows the finer Machine Translation system processes. A user enters an English phrase and the system tokenizes it and assigns parts-of-speech tags to each token. The system checks to see if it is a valid English phrase and if invalid, the user has the option of rewriting the phrase. If the submitted phrase is a valid English phrase, any Articles in the phrase are removed and the remaining tokens are translated word for word using bilingual dictionaries. The resulting Nupe tokens are manipulated to make credible Nupe translations of English language which is eventually presented for viewing by the user.

The Use Case Diagram represents a user's interaction with the Machine Translation system and shows the relationship between the user and the various Use Cases in which the user is involved as indicated in Fig. 5.

The Sequence Diagram visually models the flow of logic within the system and facilitates both documentation and validation of logic as illustrated in fig. 6. The various logic components are represented as arrow labels in the Sequence Diagram. The user provides the system with an English language phrase which is then processed by Machine Translation system. If the phrase is a valid English word, a message confirming phrase validity is displayed to the user. Otherwise, an invalid phrase message is displayed for the user in which case the user will need to rewrite the phrase and submit it to the system for processing once more. If found to be a valid English language phrase upon processing, the phrase is translated into Nupe and translation subsequently displayed for the user.

4.1 FLOWCHART DIAGRAM

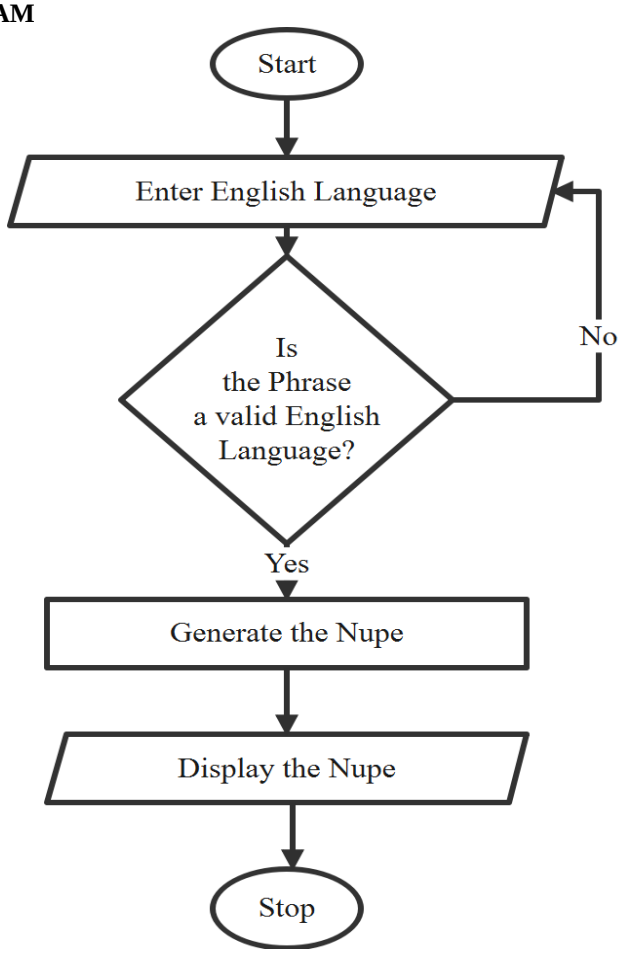

Fig. 3 Flowchart of Model for Translation of English to Nupe

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11126

4.2 DATA FLOW DIAGRAM

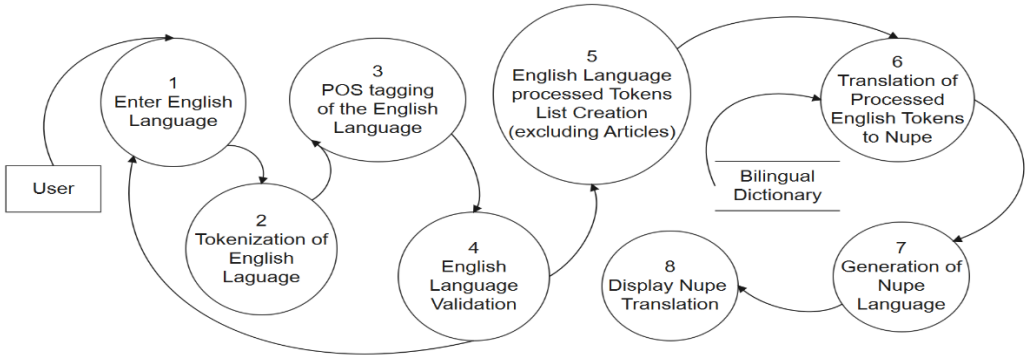

Fig. 4 Data Flow Diagram of Model for Translation of English Nupe Language

4.3 USE CASE DIAGRAM

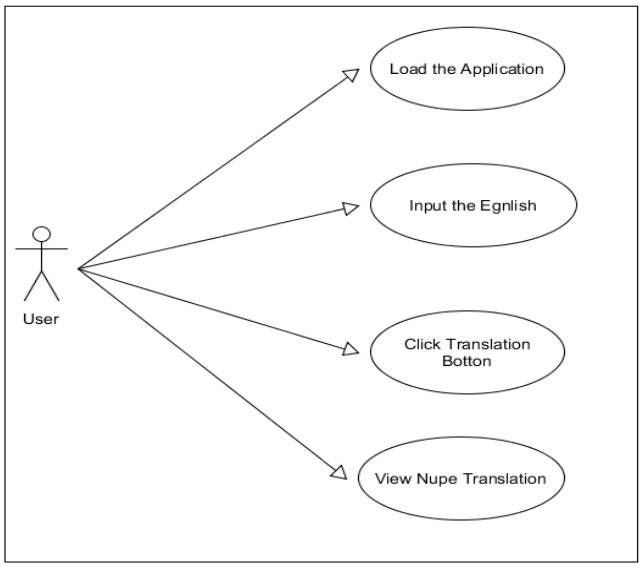

Fig. 5 Use Case Diagram

4.4 SEQUENCE DIAGRAM

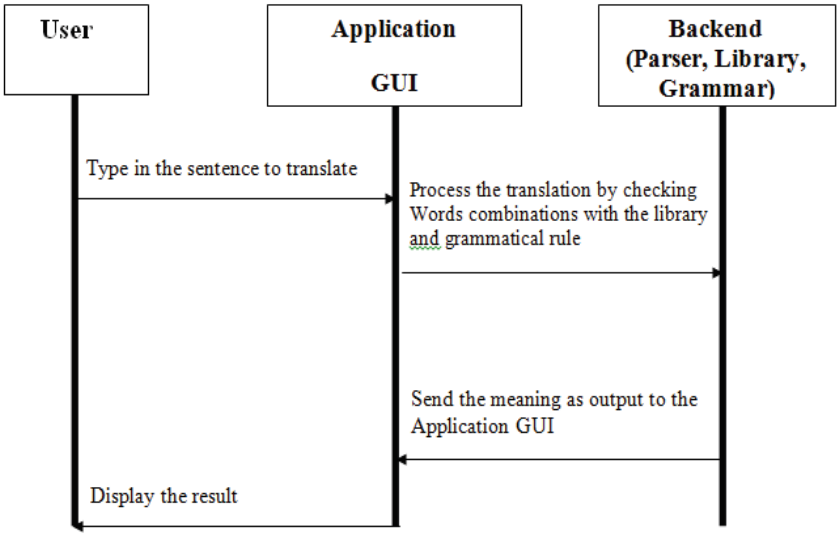

Fig. 6 Sequence Diagram of Model for Translation of English to Nupe Language

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11126

B. **SOFTWARE IMPLEMENTATION**

The data (corpus) for the research was from simple sentence spoken in the home environment. The sentences are further broken down and stored in pairs. Figure 5 below shows a sample of the stored data (corpus).

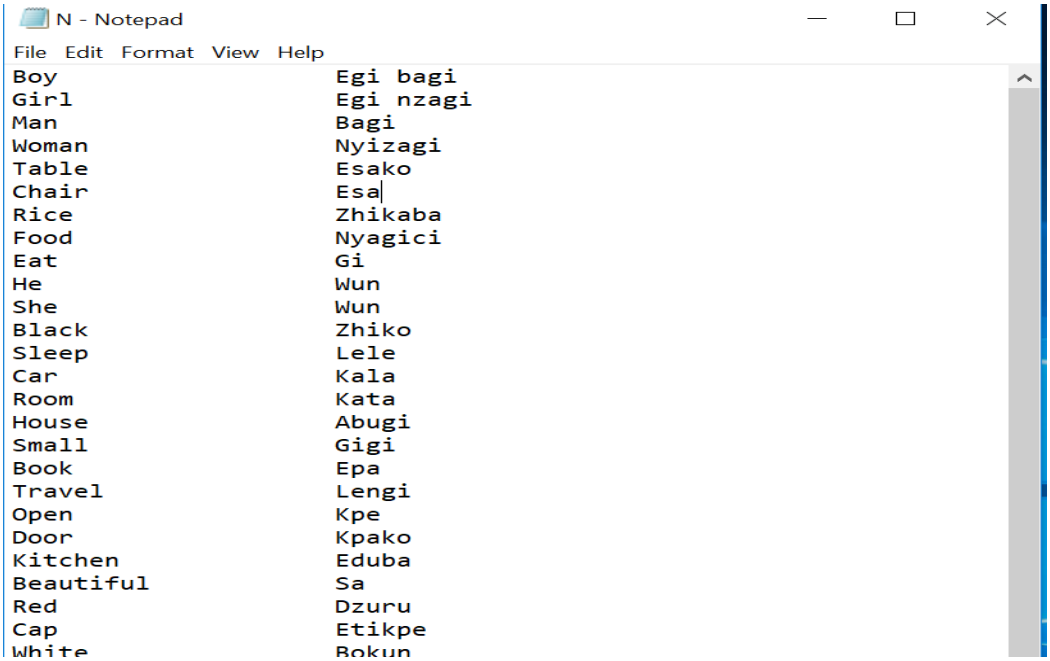

Fig. 7 Parallel corpus 1

The main tools used in this research are:

- Python programming language – this is the core programming environment for the application development.
- NTLK (Natural Language Toolkit) – this is a support kit for python programming language. Its features include: support for parsing, Part of Speech (POS) tagging, corpora design and analyses.
- JavaScript Object Notation (JSON): is a standard text-based format for representing structured data based on JavaScript object syntax. It is commonly used for transmitting data in web applications (e.g., sending some data from the server to the client, so it can be displayed on a web page, or vice versa)
- py2exe – this was used to compile the python codes (.py) to an executable file (.exe).
- NSIS – use basically in building the window installer for the application.

The requirements analysis and specifications of the software are as follow:

- To present a user friendly interface to the user;
- To give the user access to enter simple basic sentences in English language provided the sentence is within the domain covered;
- Parse the sentence to understand the structure;
- Translate and output the equivalent meaning of the sentences entered in standard Nupe language.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11126

Fig. 8 Sample I

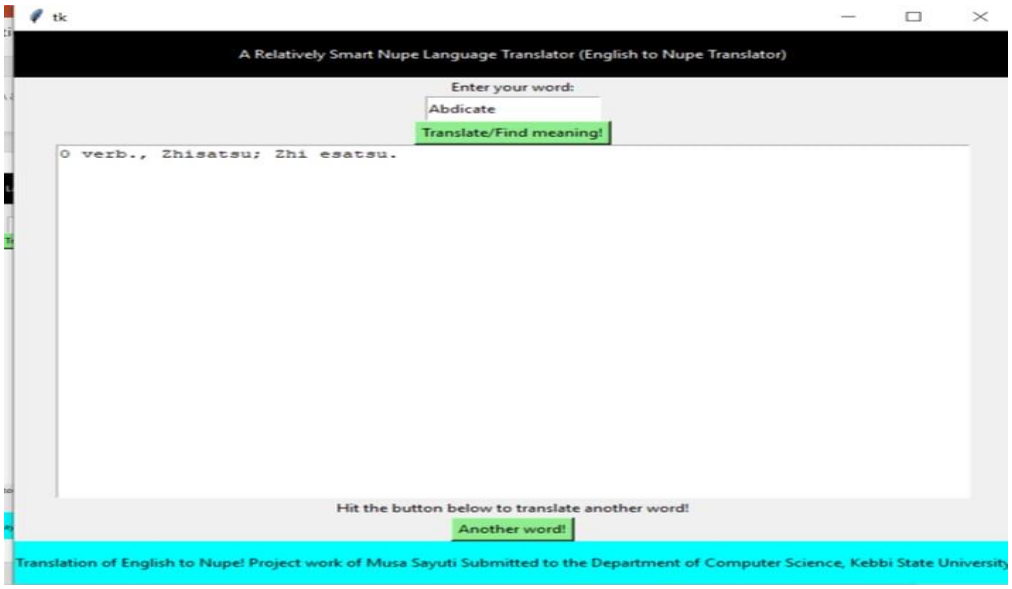

Fig. 9 Sample II

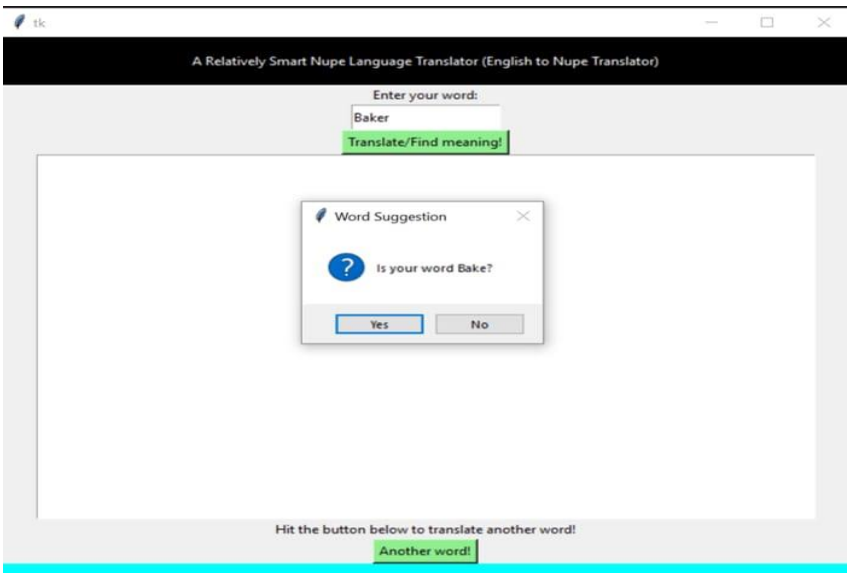

Fig. 10 Sample III

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11126

V. CONCLUSION

In conclusion, the software artefacts that translates English-Nupe were developed using rule-based machine translation, implemented and evaluated. Evaluation for Accuracy revealed that the majority of the translations contained all meaning represented in their corresponding English language. The accuracy can be improved by improving and extending the full form bilingual lexicon.

REFERENCES

[1] Agbeyangi, A. O., Adenekan, O. A., Lawal, O. O., & Durosinmi, A. E. (2015). Rule-Based Approach to Yoruba Machine Translation System. 4th iSTEAMS Research Nexus Conference, (pp. 55-62). Ilorin.

[2] Agbeyangi, A. O., Eludiora, S. I., & Adenekan, O. A. (2015). English to Yorùbá Machine TranslationSystem using Rule-Based Approach. Journal of Multidisciplinary Engineering Science and Technology, 2(8), 2275-2280.

[3] Batra, K. K., & Lahel, G. S. (2010). Rule Based Machine Translation of Noun Phrases from Punjabi to English. International Journal of Computer Science, 7(5), 409-413.

[4] Dunstan, E. (1969) Twelve Nigerian languages London: Longman.

[5] Eludiora, S. I., Afolabi, B. O., & Sanusi, A. A. (n.d.). (2013). Development of English to Hausa Machine Translation System. Pearl Research Journals. Retrieved from http://pearlresearchjournals.org/journals/rjet/index.html

[6] Fagbolu, O., & Obalalu, B. (2018). Smart Luganda language Translator. 7th International Conference on Natural Language Processing (NLP 2018) Dabai

[7] Fásakin, T. G. (2017). An English to Yoruba Statistical Machine Translation System. M. Tech Thesis, Federal University of Technology, Akure.

[8] Felix, A. S., & Dominic, O. N. (2014). Machine Translation of Noun Phrases from English to Igala using the Rule-Based Approach. West African Journal of Industrial & Academic Research, 11(1).

[9] Greenberg, J.H (1963) Languages of africa The Hague: mouton.

[10] Ibrahim, S. (1992). The Nupe and their Neighbours from the 14th Century. Ibadan: Heinemann Book Plc.

[11] Ismaila, D. (2002) Nupe in history (1300 to date) Jos: Olawale publishing company.

[12] Katcha, A. (1978). An exploratory demographic study of the Nupe of Niger State : the case of Sakpe village. Canberra: Australian National University.

[13] Keith, B., & Sarah, O. (2019). Language Family Definition and Examples. Retrieved from ThoughtCo.: https://www.thoughtco.com/what-is-a-language-family-1691216

[14] Nataly, K., Donald, A., DePalma, H., & Vijayalaxmi, H. (2012). The Need for Translation in Africa. Lowell, Massachusetts, United State of America: Common Sense Advisory, Inc.

[15] Osborne, M. (2012). MT History and Rule-Based Systems. University of Edinburgh, School of Informatics.

[16] Safiriyu, I. E., & Odetunji, A. O. (2016). Development of an English to Yorùbá Machine Translator. International Journal of Modern Education and Computer Science(IJMECS), 8(11), 8-19. doi:10.5815/ijmecs.2016.11.02

[17] Sani, F. A., Osuagwu, O. E., & Njoku, D. O. (2014). Machine Translation of Noun Phrases from English to Igala using the Rule-Based Approach. West African Journal of Industrial & Academic Research, 1(11), 17-26.

[18] Stephen, D. (2016). The Impact of Translation Technologies on the Process and Product of Translation. International Journal of Communication, 947-969.

[19] Wassagi, A. (2000) Pioneers of Nupe language. NDA Nupe News Magazine March edition, p.5.s