Research Journal of Engineering and Technology
Vol. 2 (1), pp. 12-32, September, 2016
ISSN 2467-8597
Review
http://pearlresearchjournals.org/journals/riet/index.html

Development of English to Hausa Machine Translation System

Safiriyu I. Eludiora*, Beatrice O. Afolabi and Abdulrahman A. Sanusi

Obafemi Awolowo University, Department of Computer Science and Engineering, Ile-Ife, 220005, Nigeria.

ABSTRACT

The Hausa language is one of the major indigenous languages in Nigeria. Hausa language speakers cover the Northern parts of Nigeria, West Africa and Africa in general. Hausa language used in media houses for news and in Schools for teaching. The development of English to Hausa machine translation (E-HMT) system will assist the interested learners to learn the language. Hausa is a tone language like Yorùbá. It was observed that the tone marks and diacritics are not used in writing unlike Yorùbá that the tone marks and diacritics must be used. The grammar structure of the two languages are subject verb object (SVO), but with some differences. Findings show that Google has developed E-HMT using the data driven approach. In the study reported in this paper rule-based approach was used to develop the E-HMT system. The E-HMT system can translate simple sentences. Home domain terminologies were used for the development of E-HMT system database. Some linguistics and computing theories and rules were used for the E-HMT translation process. The re-write rules were designed for the two languages. The re-write rules were tested using JFLAP. Unified Modeling Language (UML) was used to design the software. Java programming language was used for the implementation of the software. The E-HMT system could have been evaluated using Mean opinion score, but a good number of Hausa speakers in our community are not literate. The questionnaires were administered at Kaduna. We compared the E-HMT system outputs with the Hausa Google translator. The results show that E-HMT system performs better than Hausa Google translator.

Key words: Hausa, Ruled-based, re-write rules, ambiguity and Tagging.

*Corresponding author. E-mail: sieludiora@oauife.edu.ng or safiriyue@yahoo.com.

INTRODUCTION

In Nigeria, there are three major indigenous languages: Yoruba, Igbo, and Hausa. The languages spoken in Nigeria are not evenly distributed, for instance, in the South-West part of Nigeria, Yorùbá is largely spoken; Igbo is largely spoken in the South-East; while Hausa is largely spoken in North-West part of Nigeria. The dominance of the English language is quite overwhelming in Nigeria; this can be seen practically in all domains: government and administration, education, the media, the judiciary, science and technology to mention but a few. The English language is an official language in Nigeria.

A Brief Introduction To the Hausa Language

According to Bauer (2007), Hausa (usually pronounced as /ˈhaʊsə/) is the Chadic language (a branch of the Afro-Asiatic language family) with the largest number of speakers, spoken as a first language by about 180 million people, and as a second language by about 35 million or more, an approximate total of 102 million people (Parkvall, 2007). Native speakers of Hausa, the Hausa people are mostly to be found in Niger, in the north of Nigeria and Chad, but the language is used as a trade language across a much larger swathe of West Africa (Benin, Ghana, Cameroon, Togo, Côte d'Ivoire etc.), Central Africa (Chad, Central African Republic, Equatorial Guinea) and Northwestern Sudan, particularly amongst Muslims. Radio stations like BBC, Radio France Internationale, China Radio International, Voice of Russia, Voice of America, Deutsche Welle, and IRIB

broadcast in Hausa. Hausa is taught as a language at Universities in Africa and around the World. The language is the most commonly spoken language in Nigeria.

Hausa Dialects

Eastern Hausa dialects include Kanani which is spoken in Kano, Bausanchi in Bauchi, Dauranchi in Daura, Gudduranci in Katagum Macau and part of Borno and Hadejanci in Hadejiya. Western Hausa dialects include Sakkwatanci in Sokoto, Kutebanci in Taraba, Katsinanci in Katsina, Arewanci in Gobir, Adar, Kebbi, and Zamfara, and Kurhwayanci in Kurfey in Niger. Katsina is transitional between Eastern and Western dialects. Northern Hausa dialects include Arewa and Arawci. Zazzaganci in Zaria is the major Southern dialect. The Kano dialect (Kananci) is the standard.

Hausa Grammar

In this section, Hausa alphabets, and tones were discussed.

Hausa Alphabets

There are twenty-six (26) letters of the Hausa alphabets each of lower-case characters and upper-case characters out of which five are vowels and the rest are consonants. Hausa consonants: Bb Ɓɓ Cc Dd Ɗɗ Ff Gg Hh Jj Kk Ƙƙ Ll Mm Nn Rr Ss Tt Ww Yy Yy̓ Zz and Hausa vowels: Ii Aa Ee Oo Uu. The Hausa language uses a variation of the Latin alphabet, which is also used by English. There are several letters that can be written either as special characters (often called 'hooked consonants', because they appear to have small hooks on them) or as regular consonants with apostrophes. Hausa is a tonal language, meaning that different tones can change a word's meaning. There are three tones: the low tone, the high tone, and the falling tone. The Hausa language also distinguishes between long and short vowel sounds. In some dictionaries and language references, tones and vowel length may be marked with various diacritics over vowels. However, the diacritic marks are not used in the everyday written form of the language, so the best way to learn Hausa pronunciation is to listen carefully to the way it is spoken by native speakers (Bernard, 2013). Hausa is a tonal language. Each of its five vowels may have a low or a high tone. Grave and acute accents are typically used for representing tones. However, in everyday writing, tones are not marked (Bernard, 2013). In this research reported in this paper, tones are not marked.

MACHINE TRANSLATION (MT) APPROACHES

Statistical machine translation is a machine translation paradigm where translations are generated on the basis of statistical models whose parameters are derived from the analysis of bilingual text corpora (Weaver, 1955). Statistical machine translation (SMT) is characterized by the use of machine learning methods, for example, Hidden Markov Model for POS tagging. In less than two decades, SMT has come to dominate academic MT research and has gained a share of the commercial MT market (Weaver, 1955). Rule-based Machine Translation (RBMT) also known as `Knowledge-based Machine Translation', RBMT 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 (Bond, 2006; Osborne, 2012). Having input sentences (in source language), an RBMT system generates the output sentences (in the target language) on the basis of morphological, syntactic and semantic analyses of both the source and the target languages. This involves a concrete translation task (Bond, 2006; Osborne, 2012). We used rule-based approach for the research work that is being reported in this paper.

Hybrid machine translation (HMT) leverages the strengths of statistical and rule-based translation methodologies (Boretz, 2009). Several MT companies (Asia Online, LinguaSys, Systran, PangeaMT, and UPV) are claiming to have a hybrid approach using both rules and statistics. The approaches differ in a number of ways:

1. Rules post-processed by statistics: Translations are performed using a rule-based engine. Statistics are then used in an attempt to adjust/correct the output from the rules engine.
2. Statistics guided by rules: Rules are used to pre-process data in an attempt to better guide the statistical engine. Rules are also used to post-process the statistical output to perform functions such as normalization. This approach has a lot of power, flexibility, and control when translating, but complex. Hybrid machine translation is a combination of both statistical and rule-based translation methodologies (YeaseenurRahmanTahin, 2010).

RELATED WORKS

Eludiora (2014), proposed rule-based English to Yorùbá Machine Translation System. The system is able to translate English language texts to Yorùbá language. The context-free grammar (CFG) model within the context of Noam Chomsky phrase structure grammar theory was used to model the two languages. Automata theory was used to model the computational process underlining the translation processes. The mean opinion score was used

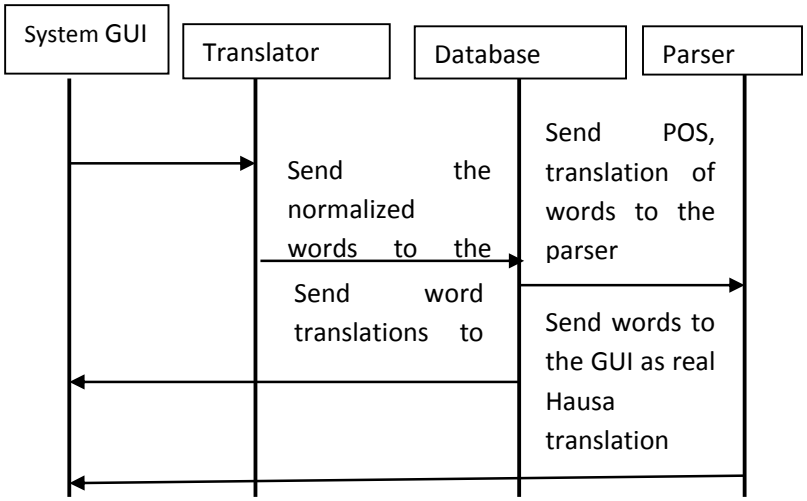


Figure 1. Sequence diagram for the system.

for evaluating the MT system. Batra and Lebal (2010) 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 preprocessing, tagging, ambiguity resolution, translation and synthesis of words in the target language. Alexandra (2009) proposed an Automatic Machine Translation (MT) in Broadcast News Domain, his paper describes the automatic translation system from Portuguese into English in Broadcast domain in which translation involve the creation of language model, standardization of training corpus, training of the system, phrase table filter, tuning and evaluation of the system. However, the system was tested using Google. Adeoye et al. (2014) work on a web-Based English to Yorùbá Noun Phrases MT System, Finite state Automata (FSA) was used whose operation was based on first set techniques that allow the parser to choose which production rule to apply based on first input word of an input phrase (Rule-Based).

SYSTEM FRAMEWORK DESIGN

In this section, we discussed the system, software, language resources, and database designs.

System Design

System design may refer to either 'all the activities involved in conceptualizing, framing, implementing, ultimately modifying complex systems'. System design usually involves problem-solving and planning a software or hardware solution. A sequence diagram has two dimensions: the vertical dimension shows the sequence of message/calls in the order that they occur; the horizontal dimension shows the object instances to which the messages are sent. The sequence diagram for the application captures the interaction between the user inputted texts and how the outputs are generated. The sequence diagram for the system is shown in Figure 1. There are four modules that communicate or interact with each other. The parser, database, translator and the system graphical user interface (GUI). The user interacts directly with the GUI by typing the sentence to be translated. The GUI will send the sentence to the translator. The translator will categorize the words and send it to the database to confirm if these words are in the database, if not it will request for the missing words. If their words are complete, it will send it to the parser. The parser will use the re-write rules to test the correctness of the sentence.

Language Resource Design

The language resource design section was subdivided into four: Data 1, Data 2, Data 3 and Data 4.

1. Data 1: Sentences were collected using home environment terminologies
2. Data 2: Lexical items were extracted from Data 1
3. Data 3: Data 1 and Data 2 were annotated with different parts of speech (POS) tags
4. Data 4: Parallel corpus used for the E-H MT system Database.

Basic Sentences

In data 1, sentences were collected using home

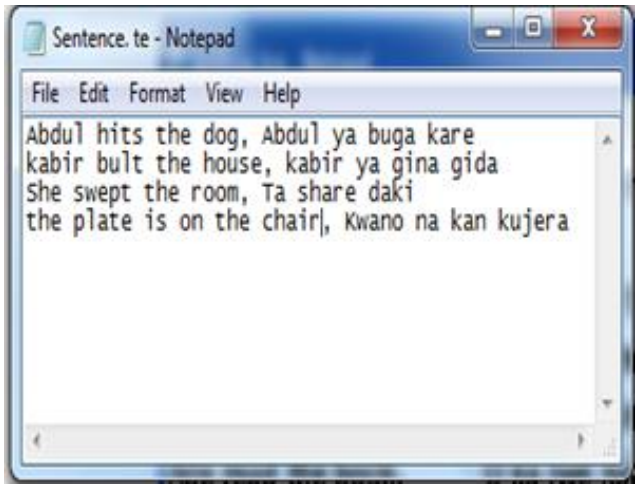


Figure 2. Parallel sentence corpora.

Table 1. A tag for English and Hausa word.

Terms (English)	POS	Terms (Hausa)	POS
He	PRN	Shi	PRN
Sweeps	Verb	Share	Verb
Car	Noun	Mota	Noun
The	Determiner	Ya	Determiner

environment terminologies. The sentences captured the different structures we had identified. That is Subject Verb Object (SVO) sentences.

The data used were the sentences collected from the home environment. The sentences were tagged accordingly. The sentences were from source language with its equivalent in the TL. Figure 2 was a sample of some sentences collected from home domain. Those sentences and others were used for the database. Words (lexemes) were extracted from those sentences and tagged according to their POS.

Parallel Corpora Design

The design of parallel corpora involved the POS tagging. The POS tagging was done manually.

Lexical Level Tagging

Table 1 shown the design of lexemes' (lexical items) POS tags. The POS tags design is meant for the two languages: the source and the target languages. The languages had the terms (words), and POS. These options were used for ease of words categorisation within a sentence.

Lexemes

In data 2, the lexical items were extracted from data 1. Different words from the two languages were used to form the lexicon/database that was used by the MT system. The lexemes formed the lexical items' database of the MT system. In the lexical database, words were stored according to their POS. Figures 3, 4, and 5 shown the lists of pronouns, verbs and nouns lexemes.

Lexical Items Annotation

In data 3, annotation of the lexical items (lexemes) of both languages was carried out. The lexemes were tagged based on their parts of speech (POS). Figure 6 shown the examples of POS tagging.

Tagged Sentences

Data 4 represents the structure of the MT system database as shown in Figure 7.

DATABASE DESIGN

The database is responsible for storing the English words and their Hausa translations based on their part of speech. A collection of words and their meaning can also be called corpora. The database contains seven tables each representing a part of speech. The parts of speech used in the domain of this work are noun, pronoun, verb,

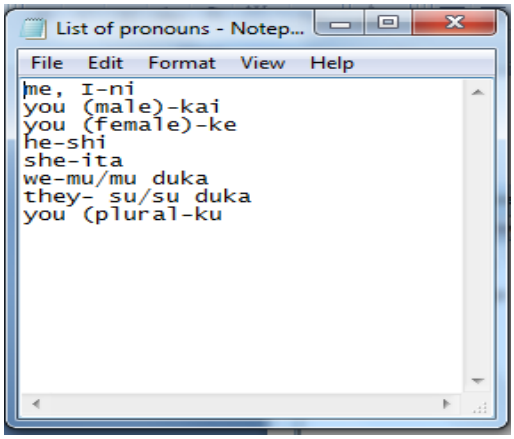


Figure 3. List of pronouns.

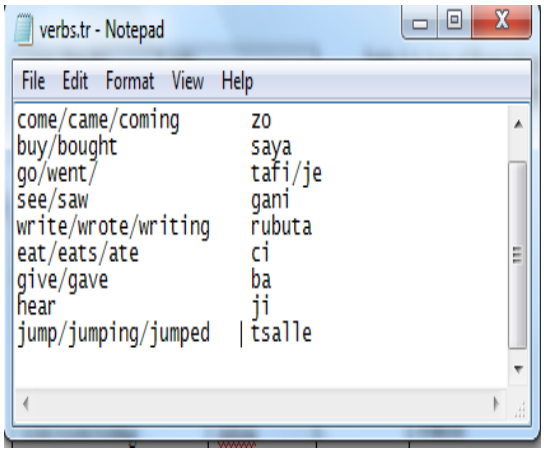


Figure 4. List of verbs.

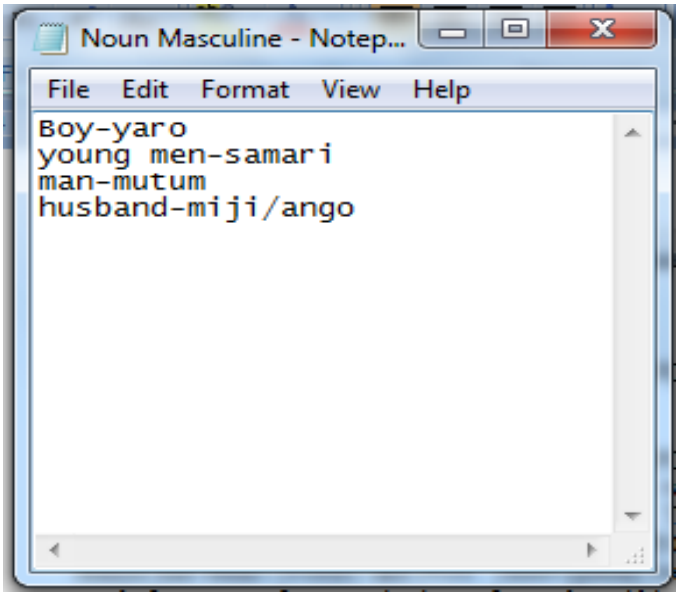


Figure 5. List of masculine nouns.

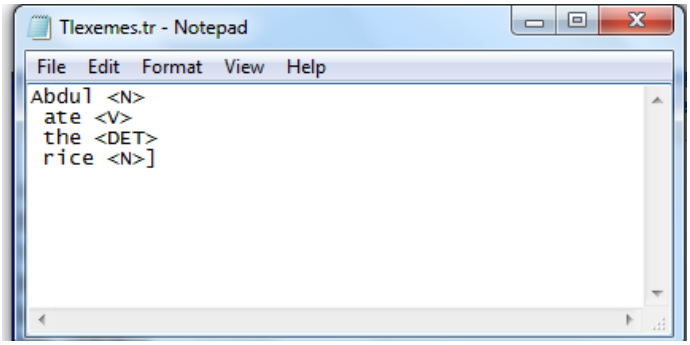


Figure 6. POS tagging.

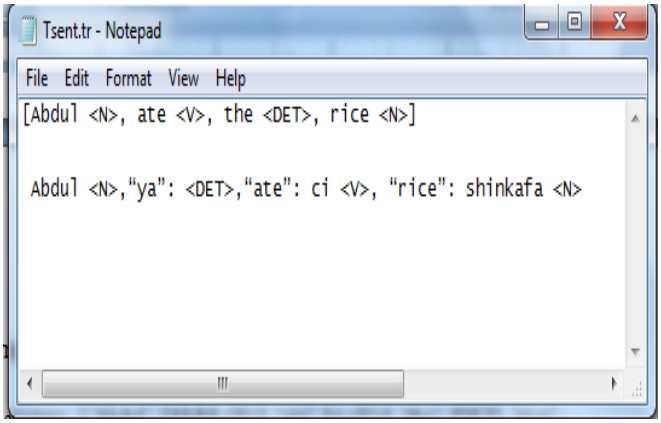


Figure 7. Sentences tagging.

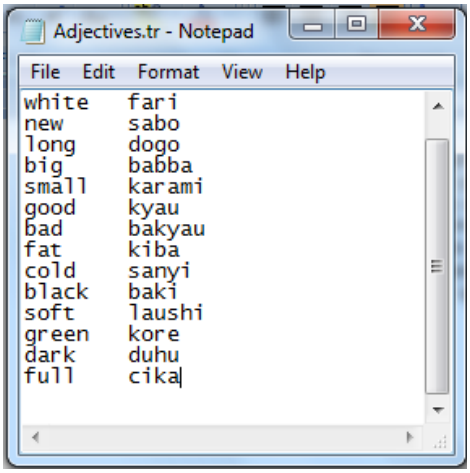


Figure 8. Adjectives database.

adjective, preposition, determiner and adverb. These are shown in Figures 8 to 10.

THE SYNTACTIC ANALYSIS OF THE E-HMT SYSTEM

Word swapping explains the structural differences between the English and Hausa languages. The two languages have SVO sentence structure but most of the time, the need for word re-ordering or swapping do arise. For example, `Ahmed ate the rice' (SL), 'Ahmed ya ci Shinkafan' (TL). Figure 11(a) illustrates the syntactic transfer of an English sentence to Hausa sentence.

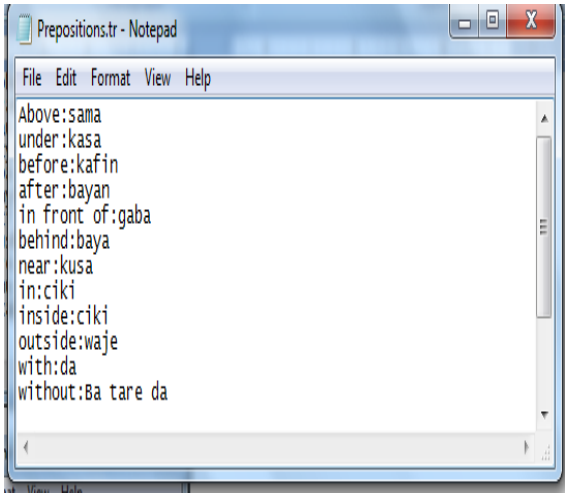


Figure 9. Preposition database.

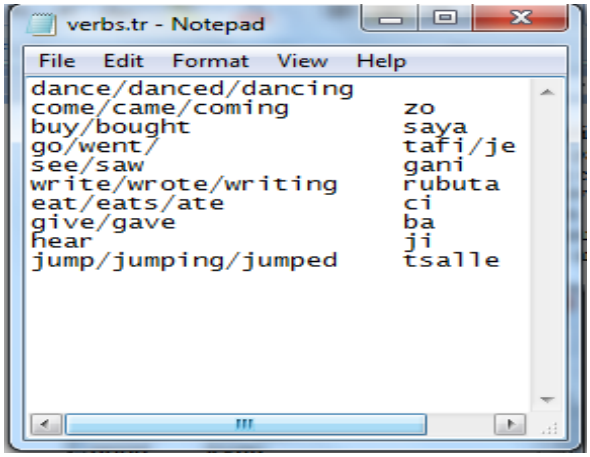


Figure 10. Verbs database.

Figure 11(b) was the first intermediate representation of the SL sentence being re-wording that is, Determinant “the” is swapped with “ate”. Figure 11 (c) is the second intermediate representation that shows semifinal translation. Figure 11(d), shows the representation of the TL in which appropriate re-write rules for the translation of sentences were considered.

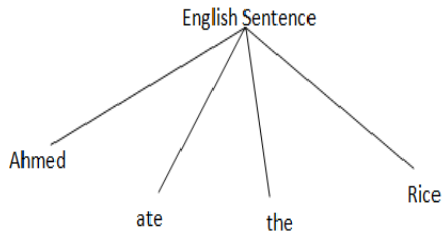
SOFTWARE DESIGN

A language translator is a form of the compiler that takes input sentence from a source language and maps it to sentence in the target language. The automata structure of the sentence is a context-free grammar (CFG). CGF is the most common way of modeling constituency. CFG = Context-Free-Grammar => Phrase Structure Grammar=> BNF = Backus-Naur Form. The idea of basing a grammar on constituent structure dates back to Wilhelm (1890), but not formalized until Chomsky (1956), and independently by Backus (1956). Grammar (G) is a mechanism to describe the language; it is used to describe how language may be formed. The automata structure of the English and Hausa sentences with respect to the context-free grammar can be described as a four-tuple grammar given as:

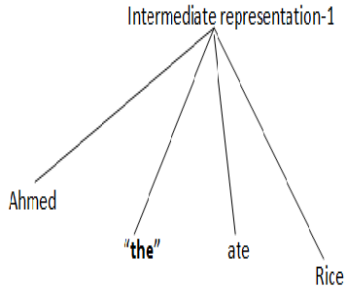
G = {T, N, S, P}. Where:

T: Is a finite non-empty set of terminal symbols. Each symbol corresponds to a concept in the language modeled by the grammar. Each terminal symbol cannot be further expanded or reduced to smaller forms hence they remain constant in an expression for the grammar. They are the leaf of a tree representation of an expression of grammar.

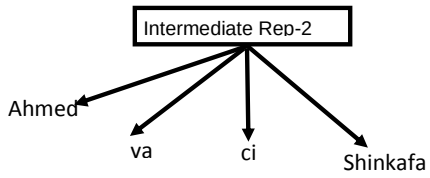
N: Is the finite non-empty set of the non-terminal symbol. Each symbol corresponds to a sub-expression in the



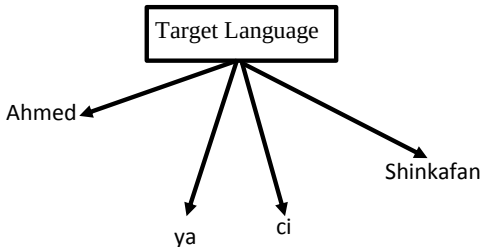
a)



(b)



C



(d)

Figure 11. (A, b, c and d) English to Hausa translation process.

language modeled by the grammar. Each non-terminal symbol can be further expanded or reduced to smaller forms by replacing it with a terminal symbol; hence, they form the variable in an expression for the grammar. They form the stem of a tree representation an expression of a grammar.

S: Is a distinguished non-terminal symbol called the start symbol. The start symbol corresponds to a complete expression in the language modeled by the grammar. The start symbol can be further expanded or reduced to smaller sub-expression using non-terminal. Hence, they form the statement or sentence in the program for the grammar. They form the root of a tree representation of an expression of a grammar.

P: Is a finite non-empty set of rewrite rules. It is also called the productions in the grammar. It is the rule that defines how non-terminal symbol can be replaced (or re-written) using terminal symbols. It is generally written as

Table 2. The production or re-write rules.

English production	Hausa production
<S> → <NP><VP>	<S> → <NP><VP>
<VP> → <V><NP>	<VP> → <V><NP>
<VP> → <V><PP>	<VP> → <V><PP>
<VP> → <V><ADJP>	<VP> → <V><ADJP>
<NP> → <DET><N>	< NP > → <N >
<PP> → <P><NP>	<PP> → <P><NP>
<ADJP → <ADJ><N>	<ADJP> → <ADJ><N>
<NP> → <PRN>	<NP> → <PRN>
<NP> → <POPRN><N>	<NP> → <POPRN>
<VP> → <V>	<VP> → <V>
<NP> → <N>	

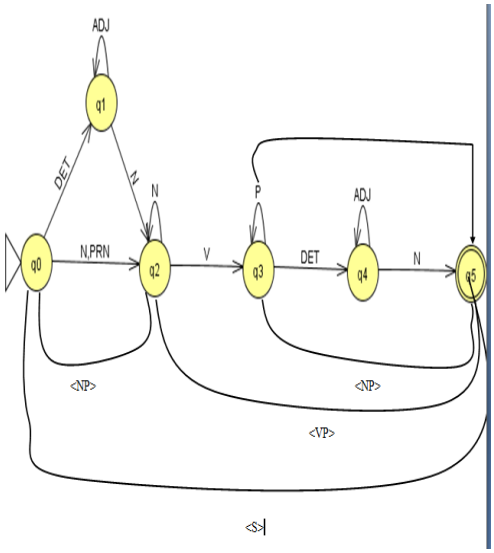


Figure 12. Transition graph for English sentence structure.

X → β, where X is a non-terminal and β is a sequence of terminals and non-terminals. Production is used to specify how a grammar transforms one string to another, thus, defining a language associated with a grammar.

T = {a noun, a pronoun, a verb, an adjective, a preposition, an adverb, determinant or article for example, “Abdul” as a noun, “ate” as a verb, “the” as a determinant, “inside” as a preposition, “blue” as an adjective}.

N = {NP = noun phrase, VP = verb phrase, PP = prepositional phrase, ADJP = adjectival phrase, also, DET (determinant), N (noun or pronoun), V (verb), P (preposition), ADJ (adjective), which will be replaced by corresponding terminal symbols}.

S = {the start which is N in Hausa but is either N or DET in English}.

P = {this is the rule of the sentence formation}

Table 2 shows the sentence formation rule, the decomposition and the arrangements of now- terminal symbols for English to Hausa. The sentence structures can also be represented as a transition graph. It is a graph that consists of three things:

- A finite set of state
- Input strings are non-terminal symbols
- A finite set of transition that shows how to move from one state to another depending on the input string.

For the purpose of this project, JFLAP is used to accomplish this task. Figures 12 and 13 show the transition graph of English and Hausa automata structures, respectively. Table 2 sentence formation rules, the decomposition and the Arrangements of Non-terminal Symbols from English to Hausa.

Re-write Rules Testing

It is possible to test the re-write rules using NLTK or JFLAP tools. The stored re-write rules correctness can be tested by using sentences. If the sentences are parsed

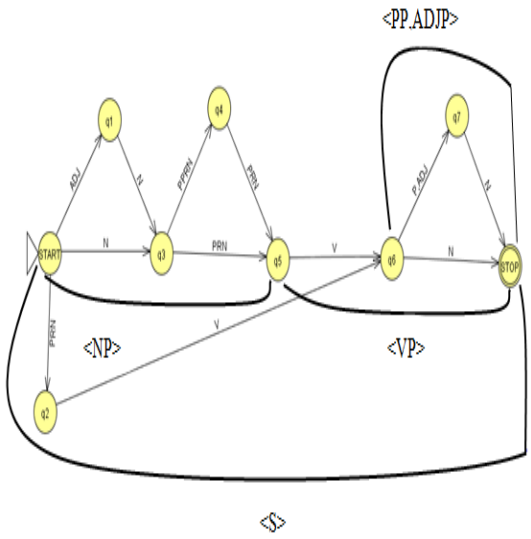


Figure 13. Transition graph for Hausa sentence structure.

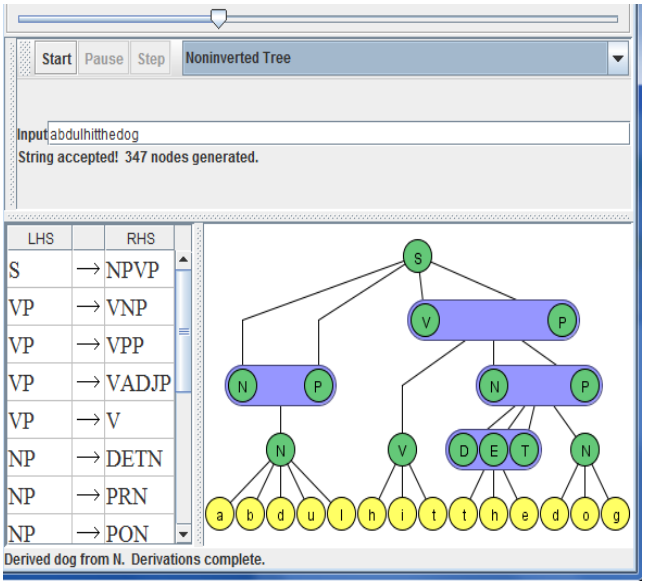


Figure 14. Parse tree for the English sentence.

using any of those tools indicate that the rules are corrected. Example 1 is the English sentence and its equivalent translation in The Hausa Language. We used JFLAP for the parsing of the sentence. Figure 14 and 15 showed the outputs generated 1. English sentence: “Abdul hit the dog” 1(a). Hausa translation: “*Abdul ya buga karen*”.

SOFTWARE AND SYSTEM IMPLEMENTATION

In this section, we discussed software and the system implementation.

Software Implementation

The class diagram uses classes and interfaces to capture details about the entities that make up your system and the static relationships between them. Class diagrams are one of the most commonly used UML diagrams, and they vary in detail from fully fleshed-out and able to generate source code to quick sketches on whiteboards and napkins. Figure 16 shows different classes used in the system. We had parser, translator, Application interface, result and Database/library classes. The classes were used based on the five modules we had in

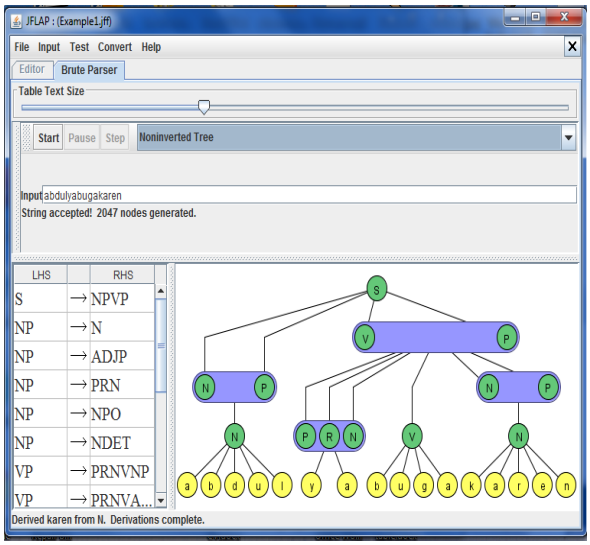


Figure 15. Parse tree for equivalent Hausa translation.

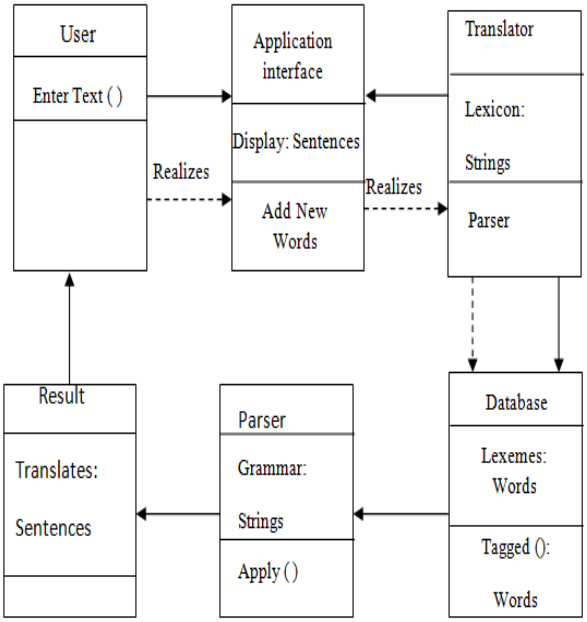


Figure 16. Class diagram of the system.

the software development.

Parser Class

This was a class that implemented the parsing of sentences. The parser parsed each sentence by using the designed re-write rules of both languages.

Translator

This was the class that implements the program. It linked data with the appropriate modules to produce the needed output.

Database Class

This class sets up the vocabulary and grammar structure. It is a module that implements the building of the Database. The system will read the library files (words) and generate the appropriate grammar structure, and use the stored data.

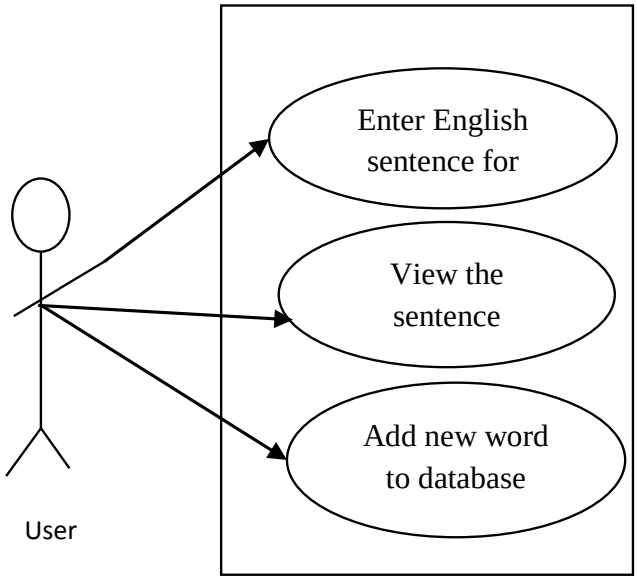


Figure 17. Use case diagram of the system.

Result Class

This was a class that implemented English to Hausa translations. The GUI displayed the translated sentence.

System Implementation

System implementation deals with the process of converting system specification to an executable system for the case of software, in the case of hardware; it involves the construction of components that will make up the system (gotten from the specification). The process of implementation involves the total steps taken from the analysis of the problem to the production of the executable file. NetBeans IDE provides an interface for writing Java codes, it is a package integrated into Java IDE for building the GUI, java compiler is used in compiling java codes to executable (.exe) file, install jammer is used to build the installer and flat file is used for the database.

Use Case Diagram

The main purpose of the use case diagram is to help development teams visualize the functional requirements of a system, including the relationship of ‘ actors’ to essential processes, as well as the relationships among different use cases. The use case for this system shows the basic things the user can do or will like to just from the interface level. Figure 17 shows the use case for this software. The use case diagram below shows the basic interaction and functionality a user can directly perform on the system. The user can load the application; enter text for translation and add a new word to the database.

System Outputs

The user typed the SL sentence (English sentence) through the first plane. The user clicked the translate button. If some of the words are not in the library, the user can inform the Administrator to add new words to the library or lexemes database.

The Graphical user interface (GUI)

The user interface is important in most application software because it provides an environment where the user can interact with the system. The test on any software starts at the interface level; failure at interface might lead to condemning the software. The interface contains textboxes, file menu, labels, and buttons. The label gives the description of the textboxes. The first label contains the description: “English” and the second label: “Translation” is then subdivided into “Word for word” and “Hausa Translation”. One of the textboxes is designed to accept the input that is; English sentence, the second display the rule based and the word for word translations, respectively of the inputted sentence in Hausa and they are not editable. File menu will allow the user to add to the list of words depending on the parts of speech to which the new word belong.

A button labeled “Translate” handles the event of the translation process and a button labeled “Clear” clears the boxes for next sentence translation. All these

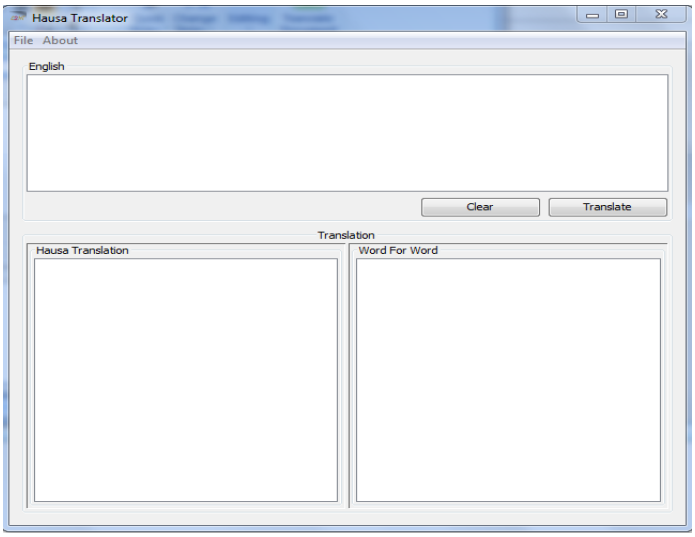


Figure 18. User interface.

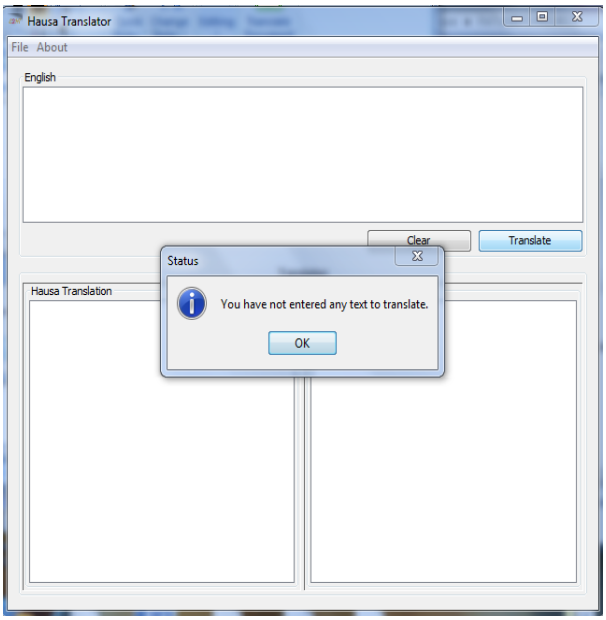


Figure 19. Empty texts.

elements were put together as simple as possible such that on seeing the interface, one would be able to predict that the software is used for translation. Figure 18 shows the user interface for the application, Figure 19 shows the message it will show when the user tries to translate an empty text, followed by the translation of sentences from Figures 20 to 22.

EVALUATION AND DISCUSSION

We compared the E-HMT system outputs with Hausa Google Translator as shown in the GUI in Figure 23. Some other examples were compared as shown in Tables 3 and 4. The E-HMT system does a better translation. Sentence 1a and 4a in Table 3 shows that the Hausa Google translator has no names for the words hits and blue. Hausa Language Professional can judge the two systems' outputs. Apart from the comparison, we designed the questionnaires for the evaluation of the system. The appendix shows the samples of the questionnaire. The English to Arabic machine translation system and the respondents have some of the translated sentences the same. The translated sentences vary

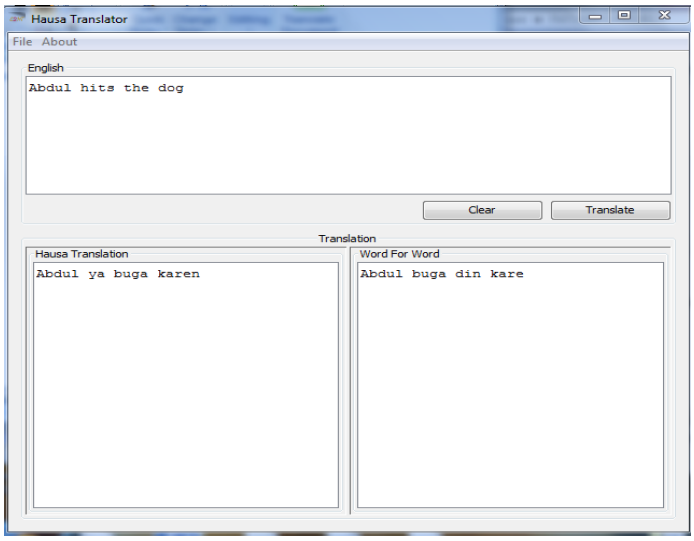


Figure 20. Translated text.

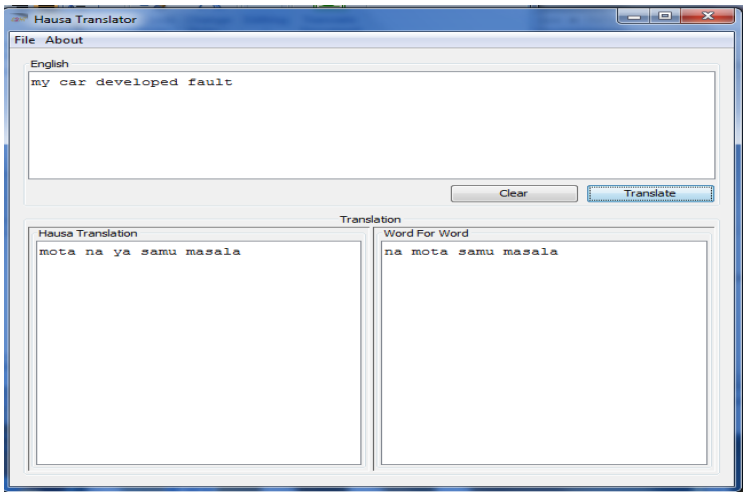


Figure 21. Add new word to database.

because of the dialects.

CONCLUSION

Research shows that there have been several attempts on the development of English language to several other Language texts of which The Hausa language is not an exception, but an attempt has not been made on the development of a translator that can translate English text to Hausa text only. English to Hausa machine translator is a system developed using rule-based machine translation approach. This involves accepting input text by the system, breaking it into a chunk of words, performing normalization on them, and then fetching their parts of speech from the data stored. The sequence or patterns of the parts of speech used by the system. Different sequence or patterns were programmed into the system. The patterns were developed in accordance with phrase structure grammar model also known as a rewrite rule. Each pattern has two parts: the English part; that checks for the validity of the input sentence and the Hausa part; that generate the translation of the system. English to Hausa machine translator has been developed as a standalone application for its effectiveness following the English and Hausa grammar rules.

The system has been tested with several home domain simple sentences and the outputs are correct. English to Hausa machine translator is opened for further development such as: expanding the scope by considering complex sentences that include; conjunction, interrogative, question mark, etc for a better

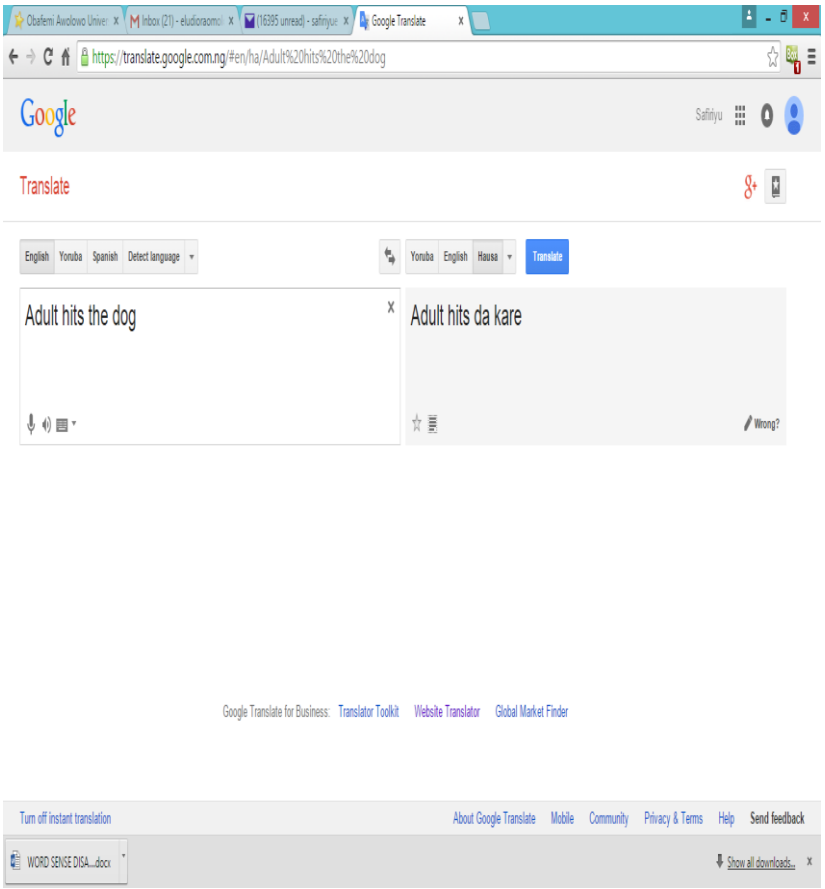


Figure 22. Hausa Google translator.

Table 3. Hausa Google translator.

S/no	
1	Abdul hits the dog
1a	Abdul hits da kare
2	My car developed a fault
2a	Ta mota ɓullo da laifi
3	The black man came yesterday
3a	Cikin bashi mutum ya zo jiya
4	The blue car hits the small boy
4a	Da blue mota hits da kananan yaro

Table 4. IFEMT (E-HMT system).

S/no	
1	Abdul hits the dog
1a	Adul ya buga karen
2	My car developed a fault
2a	Mota na ya samu masala
3	The black man came yesterday
3a	Baƙi mutumn ya Zo jiya
4	The blue car hits the small boy
4a	Shudi motan ya buga ƙaramar Yaron

understanding of Hausa language and voice can be incorporated in translation which will go a long way in improving the translator. Also, for the easy pronunciation of each word in a sentence/text, it is recommended that diacritics of the Hausa language be introduced in the development, although, even if the tone is not marked in everyday writing, it will go a long way in helping users of the translator in their pronunciation. However, E-HMT system could be used by the language (learner) beginners who are ready to learn the Hausa language. The E-HMT is the foundation for all other studies on The Hausa language in the Department. There are issues identified that we will work on. Different concepts of the two languages will be studied to resolve an issue like ambiguity.

REFERENCES

Adeoye OB, Adetunbi OA, Fasiku AI, Olatunji KA (2014). A web-based English to Yoruba noun-phrases machine translation system. Int. J. English Lit., 5(3): 71-78.

Alexandra G (2009). LF / INESC-1D Lisbon. Rua Alves Redol, 9, 1000-029 Lisbon, Portugal (ajag) 012f.inesc-id.pt.

Batra KK, Lehal GS (2010). Rule Based Machine Translation of Noun Phrases from Punjabi to English. Int. J.Computer Sci., 7(5): 409-413.

Bauer L (2007). هَوْسَ The Linguistics Student's Handbook, Edinburgh, Edinburgh University Press, Scotland, pp. ix+387.

Bernard C (2013). Hausa Grammatical Sketch, halshs-00647533, version 3 http://aboutworldlanguages.com/hausa (Access 2 December 2012)

Bond F (2006). Introduction to Machine Translation. www.cs.mu.oz.au/research/it/nlp06/materials/Bond/mt-intro.pdf (Access 02 March 2011).

Boretz A (2009). AppTek Launches Hybrid Machine Translation Software. Magazine, Speech Technology.http://www.speechtechmag.com/Articles/News/News-eature/AppTek-Launches-Hybrid-Machine-Translation-Software-52871.aspx (Access 04 August 2011).

Eludiora SI (2014). Development of an English to Yoruba Machine Translation System, Unpublished Ph.D. Thesis, Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria.

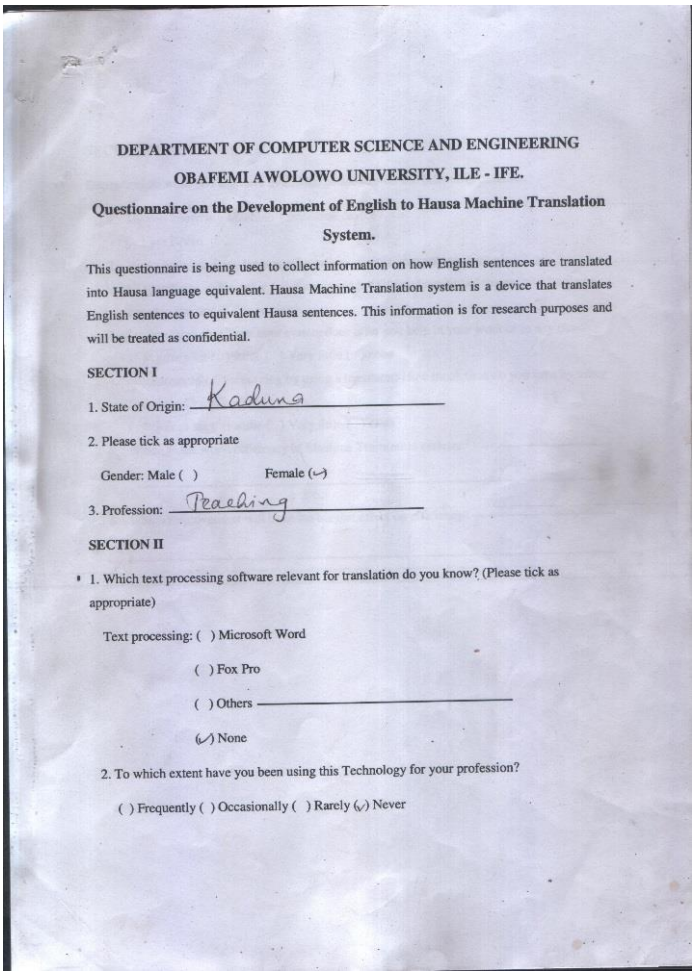
Osborne M (2012). MT History and Rule-Based Systems. School of Informatics, University of Edinburgh.

Parkvall M (2007). Världens 100 största språk 2007 (The World's 100 Largest Languages in 2007), in National encyclopedia.

Weaver W (1955). Translation, in Machine Translation of Languages. MIT Press, Cambridge.

YeaseenurRahmanTahin, EM (2010). Bootstrapping of a Rule-Based English, Bangla Machine Translation System Using Work Done for a Sister Language, Master's thesis, BRAC University,School of Engineering and Computer Science.

APPENDIX



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

OBAFEMI AWOLOWO UNIVERSITY, ILE - IFE.

Questionnaire on the Development of English to Hausa Machine Translation System.

This questionnaire is being used to collect information on how English sentences are translated into Hausa language equivalent. Hausa Machine Translation system is a device that translates English sentences to equivalent Hausa sentences. This information is for research purposes and will be treated as confidential.

SECTION I

1. State of Origin: Kaduna

2. Please tick as appropriate

Gender: Male ( ) Female (✓)

3. Profession: Teaching

SECTION II

1. Which text processing software relevant for translation do you know? (Please tick as appropriate)

Text processing: ( ) Microsoft Word

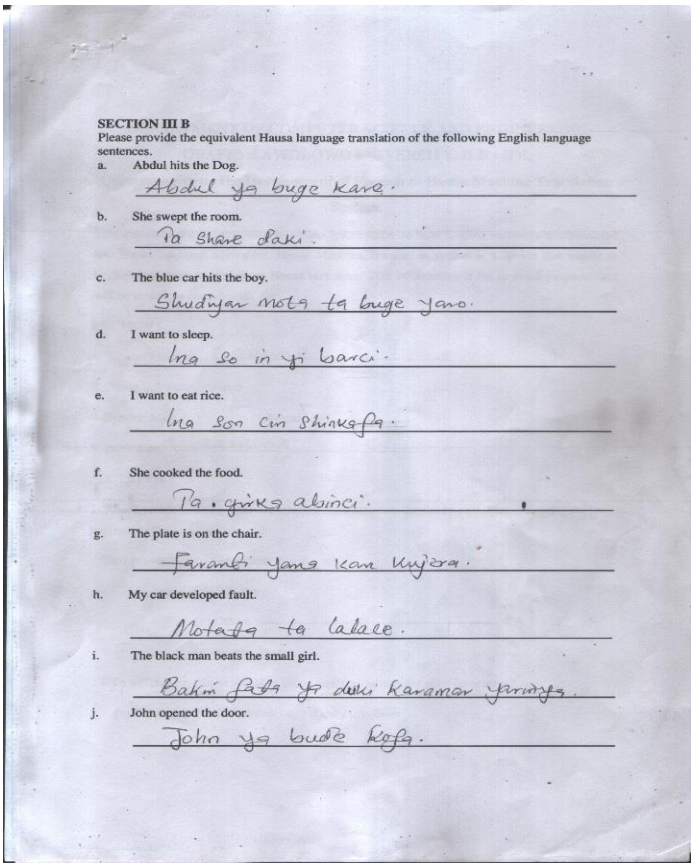
( ) Fox Pro

( ) Others _____

(✓) None

2. To which extent have you been using this Technology for your profession?

( ) Frequently ( ) Occasionally ( ) Rarely (✓) Never



SECTION III B

Please provide the equivalent Hausa language translation of the following English language sentences.

a. Abdul hits the Dog.
Abdul ya buge kare.

b. She swept the room.
Ta share daki.

c. The blue car hits the boy.
Shudiyar mota ta buge yaro.

d. I want to sleep.
Ina so in yi barci.

e. I want to eat rice.
Ina son cin shinkafa.

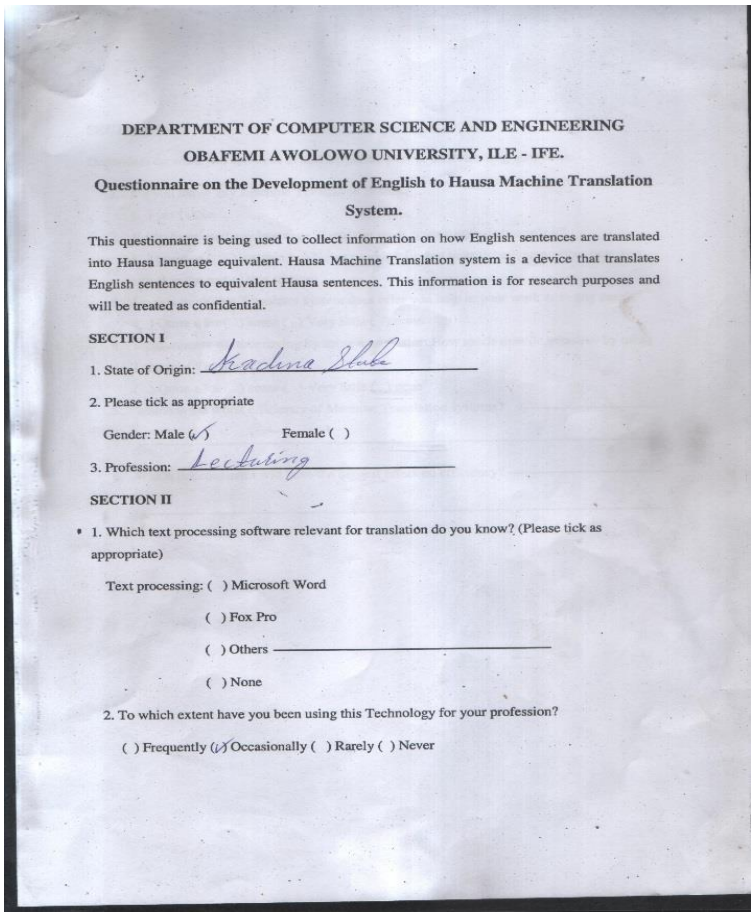
f. She cooked the food.
Ta. girka abinci.

g. The plate is on the chair.
Faranti yana kan kujera.

h. My car developed fault.
Motata ta lalace.

i. The black man beats the small girl.
Bakin fata ya duki karamar yarinya.

j. John opened the door.
John ya bude kofa.



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

OBAFEMI AWOLOWO UNIVERSITY, ILE - IFE.

Questionnaire on the Development of English to Hausa Machine Translation System.

This questionnaire is being used to collect information on how English sentences are translated into Hausa language equivalent. Hausa Machine Translation system is a device that translates English sentences to equivalent Hausa sentences. This information is for research purposes and will be treated as confidential.

SECTION I

1. State of Origin: Kaduna State

2. Please tick as appropriate

Gender: Male (✓) Female ( )

3. Profession: Lecturing

SECTION II

1. Which text processing software relevant for translation do you know? (Please tick as appropriate)

Text processing: ( ) Microsoft Word

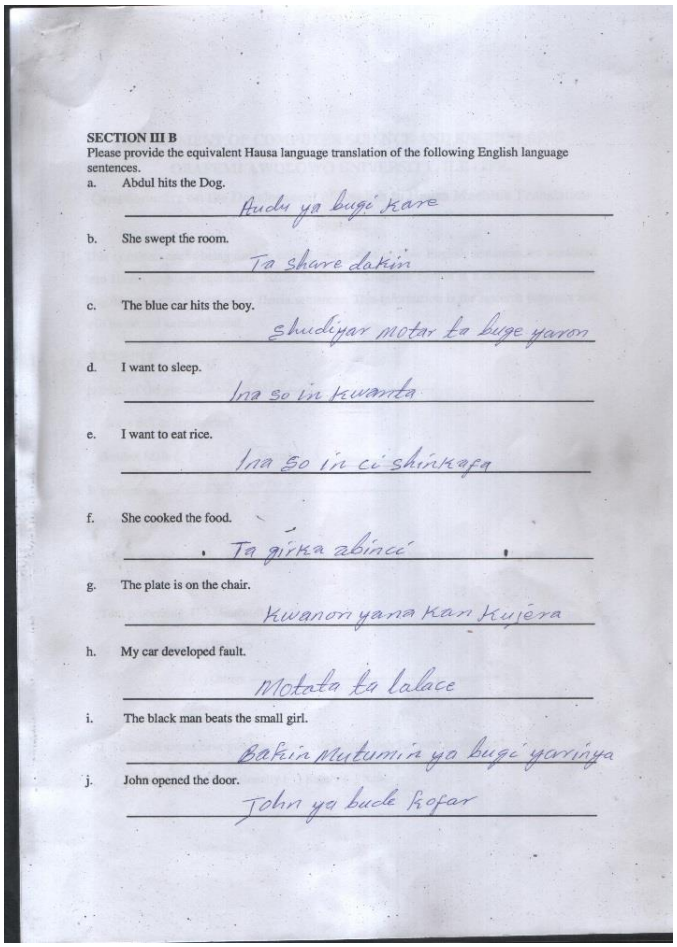
( ) Fox Pro

( ) Others _____

( ) None

2. To which extent have you been using this Technology for your profession?

( ) Frequently (✓) Occasionally ( ) Rarely ( ) Never



SECTION III B

Please provide the equivalent Hausa language translation of the following English language sentences.

a. Abdul hits the Dog.
Audu ya bugi kare

b. She swept the room.
Ta share dakin

c. The blue car hits the boy.
Shudiyar motar ta buge yaron

d. I want to sleep.
Ina so in kwanta

e. I want to eat rice.
Ina so in ci shinkafa

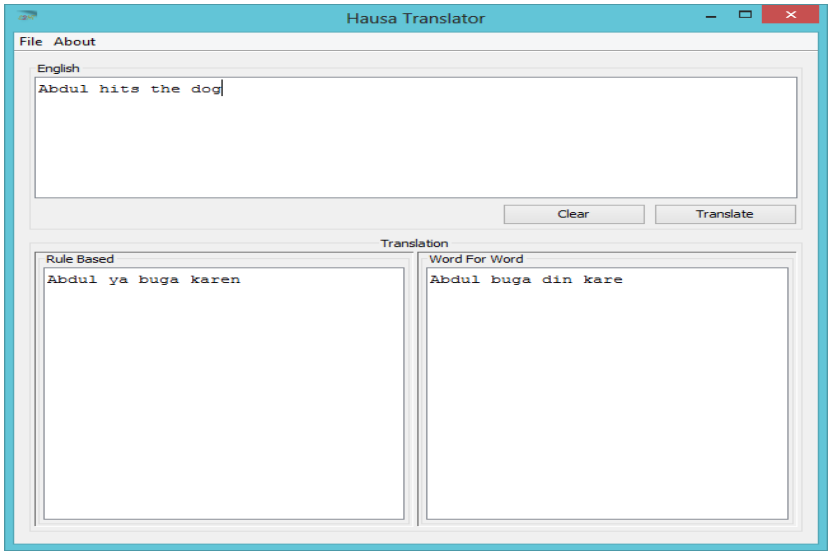
f. She cooked the food.
Ta girka abinci

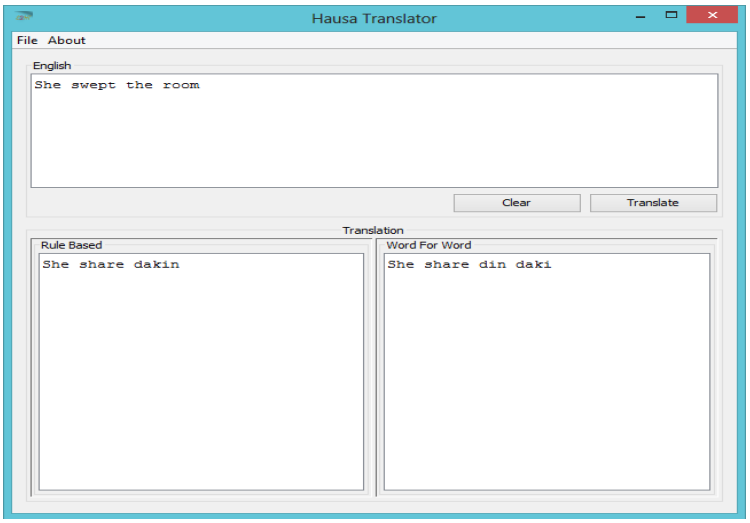
g. The plate is on the chair.
Kwanon yana kan kujera

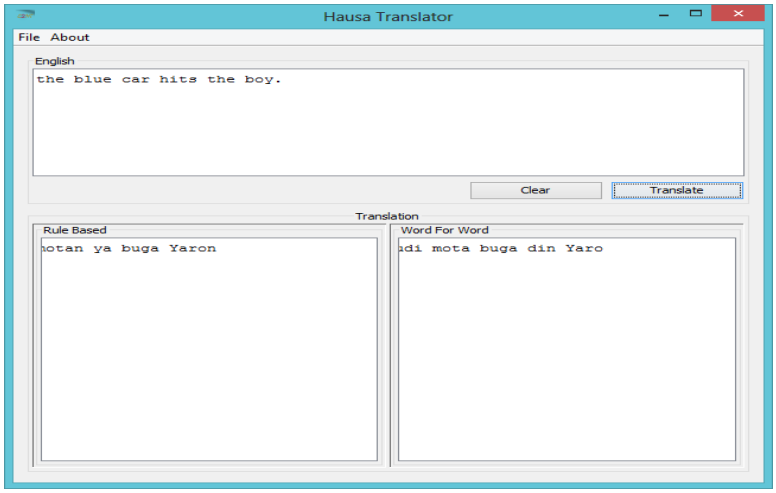
h. My car developed fault.
Motata ta lalace

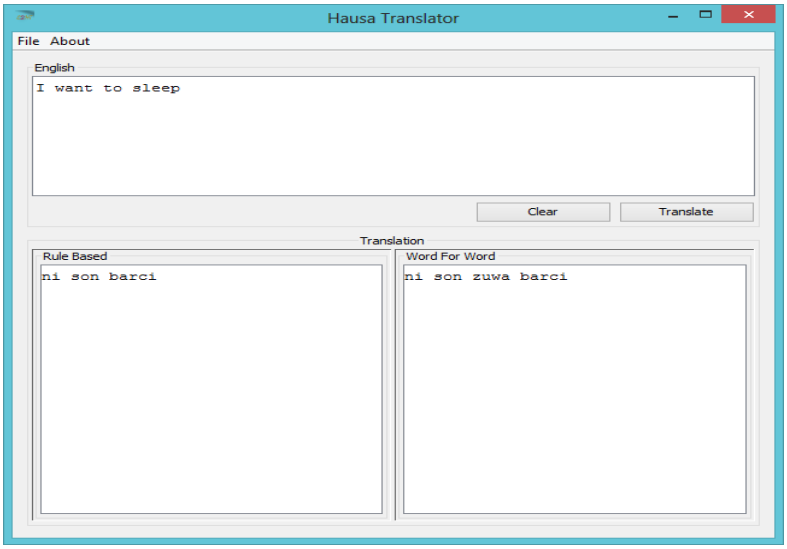
i. The black man beats the small girl.
Bakin mutumin ya bugi yarinya

j. John opened the door.
John ya buda kofar




Hausa Translator
File About
English
She swept the room
Clear
Translate
Translation
Rule Based
She share dakin
Word For Word
She share din daki


Hausa Translator
File About
English
the blue car hits the boy.
Clear
Translate
Translation
Rule Based
otan ya buga Yaron
Word For Word
idi mota buga din Yaro


Hausa Translator
File About
English
I want to sleep
Clear
Translate
Translation
Rule Based
ni son barci
Word For Word
ni son zuwa barci