

Development of an English to Yorùbá Machine Translator

Safiriyu I. Eludiora

Obafemi Awolowo University, Department of Computer Science & Engineering, Ile-Ife, 220005, Nigeria.
Email: sieludiora@oauife.edu.ng or safiriyue@yahoo.com

Odetunji A. Odejobi

Obafemi Awolowo University, Department of Computer Science & Engineering, Ile-Ife, 220005, Nigeria.
Email: oodejobi@oauife.edu.ng or oodejobi@yahoo.com

Abstract—The study formulated a computational model for English to Yorùbá text translation process. 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.

Index Terms—Yorùbá Language, simple sentences, orthography, experimental subject respondents, Human Expert, Africa

I. INTRODUCTION

Yorùbá is one of the major languages spoken in Africa. Other languages in this category include Fulfulde, Hausa, Lingala, Swahili, and Zulu. *Yorùbá* has a speaker population of about 30 million (South West Nigeria only) according to 2006 population census conducted by the National Population Commission of Nigeria [1]. *Yorùbá* language has many dialects but all speakers can communicate effectively using the standard *Yorùbá* (SY)

which is the language of education, mass media, and everyday communication [2].

Yorùbá is a tonal language with three phonological contrastive tones: High (H), Mid (M) and Low (L). Phonetically, however, there are two additional allotones or tone variances namely, rising (R) and falling (F) [3] and [4]. The *Yorùbá* alphabet has twenty-five letters comprising eighteen consonants and seven vowels. There are five nasalised vowels in the language and two pure syllabic nasal vowels [3] and [5].

Yorùbá has a well-established orthography which has been in use for over ten decades (around 1843). *Yorùbá* is relatively well studied when compared with other African languages and there is literature on the grammar of the language. The present work is one of the works that have examined the machine translation systems in the context of the text to text translation technology.

A. Machine Translation Evaluation Techniques

Machine translation systems output can be evaluated by considering numerous dimensions: the intended use of the translation, characteristics of the MT software and the nature of the translation process. There are various means for evaluating the performance of machine translation systems. The oldest is the use of human judges to assess a translation's quality. Though human evaluation is time-consuming, it is still the most reliable way to compare different MT systems developed using different translation approaches such as rule-based and statistical approaches. Automated means of evaluation include Bilingual Evaluation Understudy (BLEU), National Institute of Standards and Technology (NIST) and Metric for Evaluation of Translation with Explicit Ordering (METEOR) [6].

“Reference [7]” explains that “machine translation at its best automates the easier part of a translator's job, the harder, and more time-consuming part usually involves doing extensive research to resolve ambiguities in the source text, which the grammatical and lexical exigencies of the target language require to be resolved”. Such research is a prelude to the pre-editing in order to provide input for machine-translation software such that the output is not meaningless.

In certain applications, for example, product descriptions written in a controlled language, a dictionary-based machine-translation system has produced satisfactory translations that require no human intervention [8].

B. Turing Tests of Intelligence

Alan Turing considers the question, 'can a machine think? It is meaningless to deserve discussion. He proceeded to reduce the problem to the test of intelligence in which, any device that passes this test can be considered to be "intelligent". It is important to note that Turing believed that it is possible to develop a computing device that will achieve intelligence and the ability to think. He gives criteria that such computing device or machine must meet. This criterion is called the "Turing Test of Intelligence" [9].

Underlying the E-YMT system evaluation was the Turing test for intelligence. The purpose of this evaluation was to compare the translations from the system (English to Yorùbá machine translator) with literate Yorùbá speakers in terms of translation accuracy. The E-YMT system performance evaluation was based on the word syllable accuracy, sentence orthography, and syntax correctness accuracy. The E-YMT system and ESRs translated sentences were compared with those of the Human Expert translated sentences (speaker of Yorùbá and professional translator).

Section I introduces the study, related work are discussed in section II. The system design framework is presented in section III. Section IV describes the software design and implementation. The system performance evaluation and results discussion are presented in sections V and VI.

II. RELATED WORK

In this section, the contribution of different researchers are discussed. The methodologies used and the results got were considered while reviewing the studies.

"Reference [10]" experiments the machine translation process of English text to Yorùbá text. The study provided a machine translator that can translate simple sentences. The simple sentences can be basic subject verb object (SVO) or modified SVO. Rule-based approach was used and context free grammar was used for the grammar modelling. Re-write rules was designed and implemented for the two languages.

Translation processes for translating English ambiguous verbs are proposed by [11]. A machine translation system was developed for this purpose. Context-free grammar and phrase structure grammar were used. The rule-based approach was used for the translation process. The re-write rules were designed for the translation of the source language to the target language. The MT system was implemented and tested. For example, *Ade saw the saw, Ad ér íayìn nàà* [11].

"Reference [12]" experiment the concept of Yorùbá verbs' tone changing. For instance, *Ade entered the*

house, Adéwọ il é In this case, the dictionary meaning of *enter* in Yorùbá is *wọ*. This verb takes low tone, but in the sentence above it takes mid-tone. The authors designed different re-write rules that can address possible different Yorùbá verbs that share these characteristics. The machine translator was designed, implemented and tested. The system was tested with some sentences.

"Reference [13]" research on split verbs as one of the issues of English to Yorùbá machine translation system. The context-free grammars and phrase structure grammar are used for the modelling. Authors used rule-based approach and design re-write rules for the translation process. The re-write rules are meant for split-verbs' sentences. The machine translator can translate split verbs sentences. For instance, *Tolu cheated Taiwo, Tolú ré Táwòjẹ*.

"Reference [14]" propose the alternatives for the use of He/she/it => *Ó* of the third personal plural of English to Yorùbá machine translation system. Yorùbá language is not gender sensitive, the authors observed the problem that does arise when the identity of the doer/speaker cannot be identified in the target language. Authors proposed different representations for he/she/it. *Kìnrin* was proposed for *he*, *B̀̀rin* was proposed for *she*, and *̀̀kan* was proposed for *it*.

"Reference [15]" propose a rule-based approach for English to Yorùbá Machine Translation System. There are three approaches to machine translation process. The authors reviewed these approaches and considered rule-based approaches for the translation process. According to Authors, there is limited corpus that is available for Yorùbá language this informs the rule-based approach.

"Reference [16]" propose system that can assist in the teaching and learning of *Hausa, Igbo, and Yorùbá*. The study considers human body parts identification, plants identification, and animals' names. The English to Yorùbá machine translation and Yorùbá number counting systems are part of the main system. The model was designed to build a system for the learner of the three languages. It is on-going research work.

"Reference [17]" propose a web-based English to Yorùbá machine translation system. The Authors considered a data-driven approach to design the translation process. The context-free grammar was considered for the grammar modelling. The Yorùbá language orthography was not properly used in that study.

"Reference [18]" propose a hybrid approach to translating English to Yorùbá. The paper only itemised the steps the authors will take in the development of the proposed system. The study is on-going.

"Reference [19]" propose web-based English to Yorùbá machine translation system for noun phrase. The research work conducted by "Reference [10]" considered simple sentences translation.

"Reference [20]" propose English to Yorùbá machine translation system for noun phrase. According to the authors, rule-based approach and automata theory are used to analysis the production rules. The system is able to translate some noun phrases.

“Reference [21]” propose four methods to evaluate some system sponsored by DARPA and some other external systems. The MT systems are compared with human translators’ translations. The four methods used are comprehension test, quality panel, pre-test adequacy and pre-test fluency. The MT systems that were evaluated by “Reference [21]” used different approaches (statistical, and rule-based) for the development of their systems. The systems’ outputs are compared with one another. “Reference [10]” used three different metrics: syllable accuracy, word orthography, and syntax accuracy.

“Reference [22]” propose a manual and automatic evaluation of MT systems. The evaluation was based on some European systems. The systems were developed using statistical and rule-based approaches. The manual evaluation metrics used are based on adequacy and fluency. The automatic evaluation of the systems are based on *bilingual evaluation understudy* (BLEU) approach. “Reference [22]” reported that there are no significant differences in the performance of the systems using the BLEU. The authors established that using short sentences will produce good results than long sentences. In “Reference [10]” study, short sentences are used and the system was manually evaluated.

III. SYSTEM DESIGN FRAMEWORK

Generally, the MT system can translate between one natural language (SL) text and another natural language (TL) text. The translation may be from speech-to-speech, speech-to-text, text-to-speech, and text-to-text. The diagram in Figure 1 illustrates the translation paradigms.

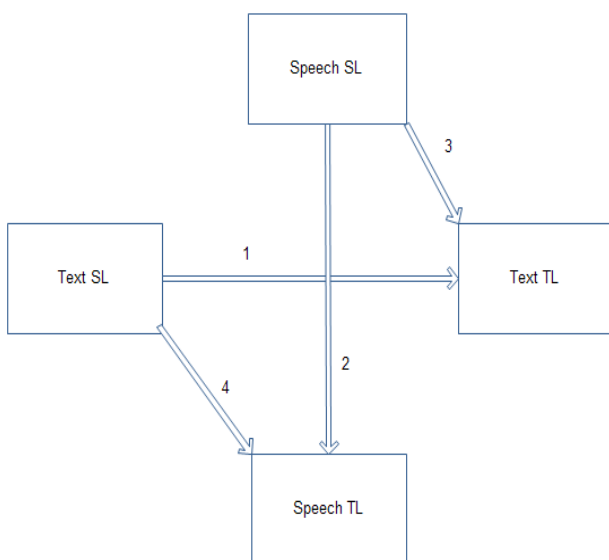


Fig. 1. The machine translation paradigms

A. The Conceptualisation of E-Y MT System

Figure 2 illustrates the conceptualisation of the English to Yorùbá (E-Y) translation process. The conceptualisation includes three individuals namely; (1) an English speaker, (2) a translator and (3) a Yorùbá

speaker. The English speaker is assumed to understand English only and his aim is to communicate his idea to a Yorùbá speaker. The Yorùbá speaker understands Yorùbá only. The job of the translator is to communicate the idea of the English speaker to a Yorùbá speaker. The E-Y MT system developed mimic human translator.

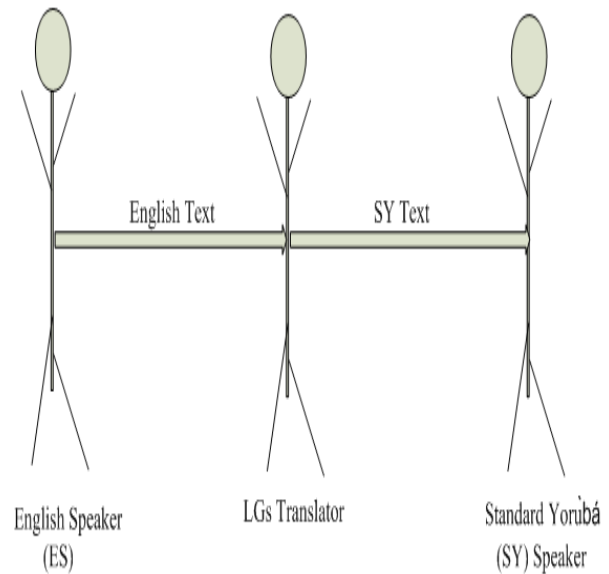
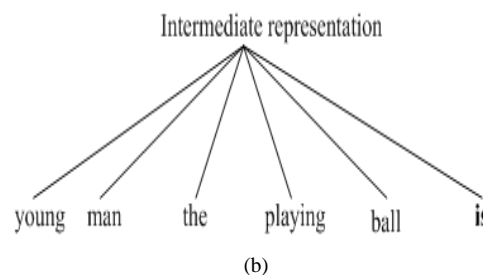
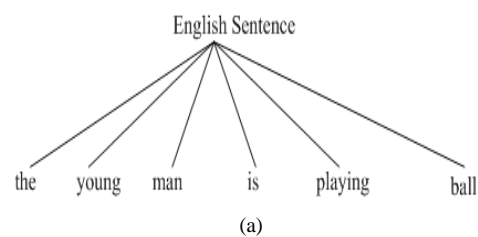


Fig. 2. Conceptualisation of English to Yorùbá Machine Translator

B. Abstraction of English to Yorùbá Machine Translator Model

Figure 3 explains the step by step English to Yorùbá translation process. Figure 3(a) is the English (source language) sentence to be translated. Figure 3(b) is the intermediate representation of Figure (a), this shows how the sentence is being re-arranged. The word “is” is removed in figure 3(c). Figure 3(d) is the target language. The word ‘playing’ is two words (*ń* and *gbá*) in the target language, they can be written together as *ńgbá* or separately as *ń gbá*. The word *ń* represents the present continuous tense.



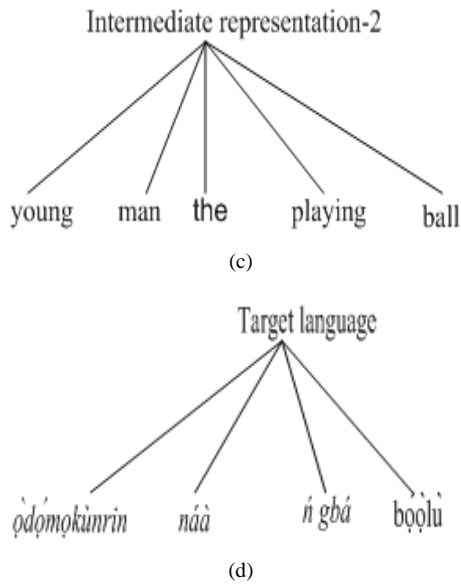


Fig.3. English to Yorùbá translation process

C. Re-write Rules' Design

The English and Yorùbá re-write rules for the above sentences are presented below. Word(s) swapping do occur in the target language. Rules 2) and 3) are swapped in the target language. Yorùbá is head first unlike the English language that is head last.

- 1) S -> NP VP
- 2) NP -> DET N|N DET
- 3) NP -> ADJP N|N ADJP
- 4) ADJP -> ADJ
- 5) NP -> N
- 6) NP -> PRN
- 7) NP -> PP
- 8) NP -> PRE NP
- 9) VP -> V
- 10) VP -> V NP

Figure 4 and 5 explain the state diagrams of the entire translation process. The machine translator can accept modified subject and object. There is swapping of words in the TL. The determiner and noun can be swapped. The adjective and noun can be swapped in the Yorùbá language.

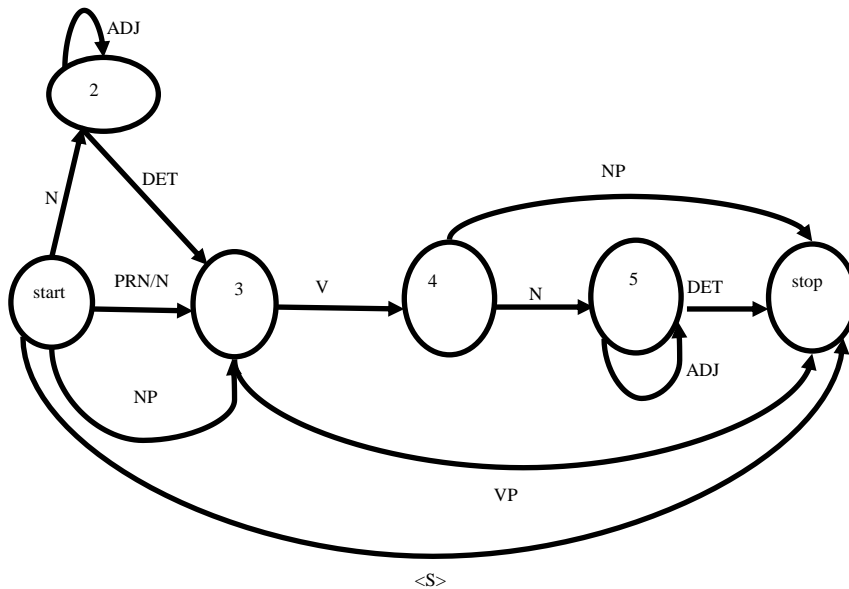


Fig.4. State diagram for the English translation process

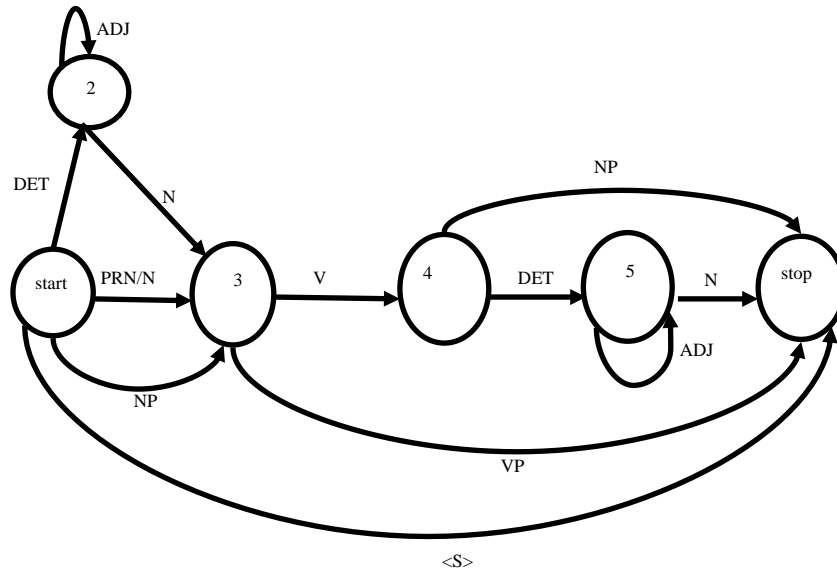


Fig.5. State diagram for the Yorùbá translation process

D. Parsing

The re-write rules of the two languages were designed and tested using NLTK tool. The sentences were analysed based on their POS. Re-write rules were designed to translate different sentences. An example is used to proof the re-write rules as shown in figure 6.

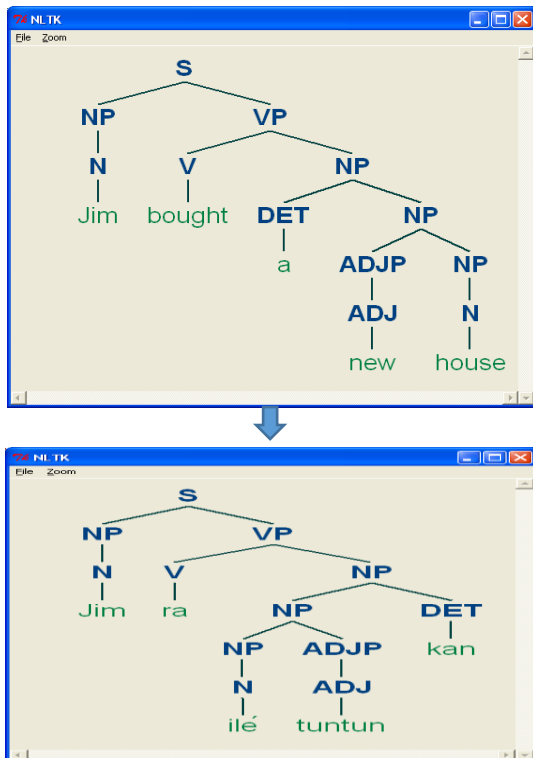


Fig.6. Sample of the tested re-write rules using NLTK

E. The E-Y MT System Database Design

This section discusses the E-Y MT system library design. This library can be described as lexicon or

database of the E-Y MT system. There is need to structure the library from inception to facilitate easy access. Sentences had been collected from the home environment for the system's library or database. The library is designed to store different words collected from the sentences. The POS are verbs, nouns, adjectives, adverbs, pronouns and prepositions. The words from the SL sentences are transcribed to the equivalent TL and placed side by side. The lexemes are organised in two ways in the database. A verb in the SL can be regular and irregular. If it is irregular, the word may deviate from the root word, e.g., 'go' and 'went'. In the case of Yorùbá, what changes is *ní* (present continuous) and *ti* (past tense) that can accompany the root words. For example, *go lo*, *going ní lo*, *has gone ti lo*, and *went lo* as shown in figure 7. Nouns, adjective, preposition etc are non-inflection (regular). Noun is one to one as shown in figure 8.

| English | Yoruba |
|-------------------------------|--------|
| 203 donate donates donated | dá |
| 204 do does did | se |
| 205 drag drags dragged | wé |
| 206 draw draws drew drawn | yá |
| 207 drink drinks drank | mu |
| 208 drive drives drove driven | wá |
| 209 drop drops dropped | bó |

Fig.7. Verb database

| | English | Yoruba |
|----|----------------|-----------|
| 40 | machine | ẹ̀rọ |
| 41 | back | ẹ̀hìn |
| 42 | necklace | ẹ̀gbàgrùn |
| 43 | Prize | ẹ̀bùn |
| 44 | bird birds | ẹ̀yẹ |
| 45 | egg | ẹ̀yìn |
| 46 | load | ẹ̀rù |
| 47 | animal animals | ẹ̀rànko |

Fig.8. Noun database

IV. SOFTWARE DESIGN AND IMPLEMENTATION

This section discusses software design and implementation,

A. Software Design

In this section, software design which includes; system flowcharts is reported. The flowcharts described the procedures for the software development. The flowcharts explain step by step software coding. Tagged words had been stored in the database. New sentences could be typed, the system checks if the words are in the database, if they are, the system sends the sentence to the parser module otherwise the translation process is terminated. The system Administrator can add new word(s) to the database. The re-write rules designed for the two languages are used to determine the correctness of the sentence. The system parses the sentence if it agrees with the rules, otherwise, the translation is terminated. Sentences are translated using the SVO sentence structure pattern. The system considers words that are needed to be swapped. The system re-groups the words and translates the sentence to the TL. The system generates value errors when it cannot translate the sentence. New sentences can be typed when the reset button is pressed, otherwise, the application can be closed. The translation processes involve all the steps mentioned in the flowcharts shown in Figure 9 below.

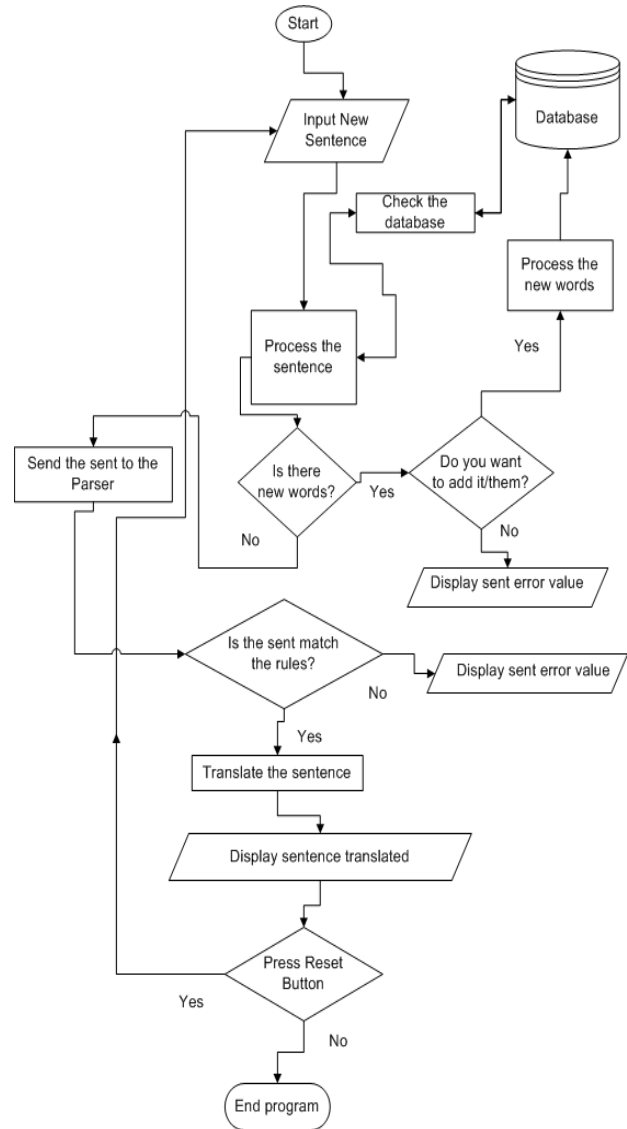


Fig.9. The E-YMT system flowcharts

B. Software Implementation

In this section, software implementation is reported. Natural language tool-kit (NLTK), Python, PyQt4, and file-based custom database are the tools used for the software implementation. The NLTK (is a version of python programming language) was used for the sentence parsing. The NLTK was used to determine the correctness of the re-write rules implemented. The phrase structure grammar model was used for the designing of the rules. The PyQt4 (a version of Python) was used for the implementation of the graphical user interface (GUI). The GUI has three planes. The first plane is where the English text can be typed. The second plane is the word for word transcription. The third plane is the translated sentence(s).

C. The E-Y MT System Outputs

The user can type the SL sentence through the first plane. The user clicks the translate button. If some of the words are not in the library, the system generates errors and informing the user whether she or he wanted to add new words to the library or database. Though only the administrator can add the new words to the various lexemes' databases. Figures 10 and 11 are the sample sentences of the E-Y MT system inputs and outputs. Figure 10 is the basic subject verb object (BSVO) sentence. Figure 11 is a modified subject verb object (MSVO) sentence. It has a subject, a verb and an object (the object was modified).

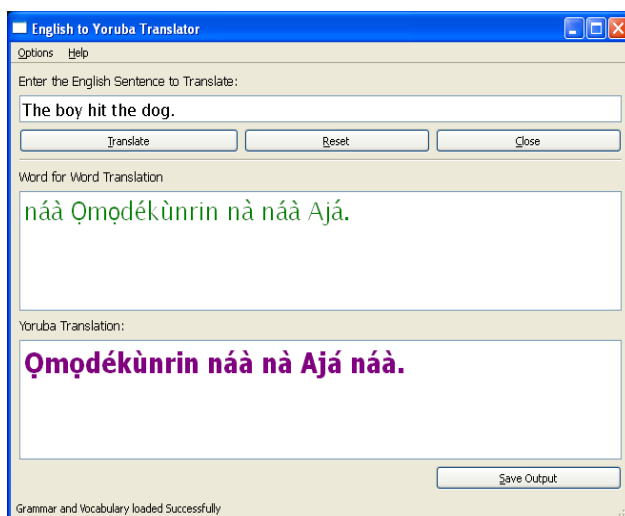


Fig.10. System sample outputs

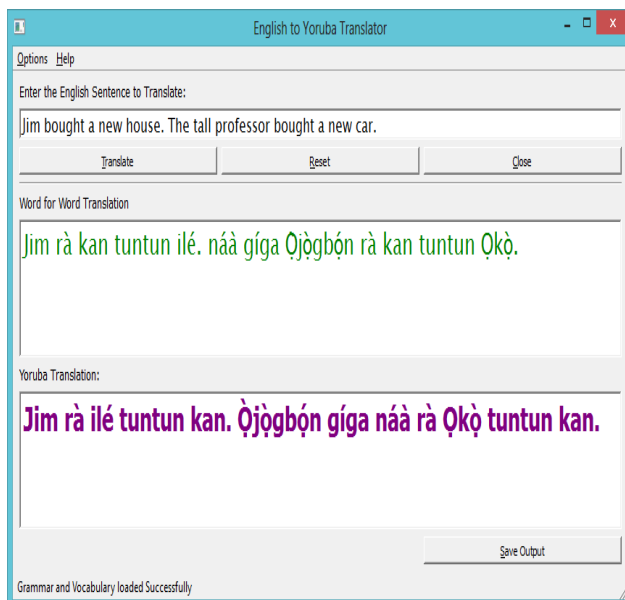


Fig.11. System sample outputs

V. EVALUATION

The evaluation process involves the questionnaire design, detailing Expert and Experimental subject respondents (ESRs) profiles, and questionnaire administration.

A. Questionnaire Design

The sentences in the instrument (questionnaire) are subdivided into two: basic SVO (BSVO) and modify SVO (MSVO) sentences. The BSVO sentences are basic sentences (subject, verb, and object) designed to test the Experimental Subject Respondents (ESRs) on their understanding of the translation of basic sentences. The subject and object can be noun or pronoun while the verb could be present or past. Ten BSVO sentences are used for the questionnaire. The sentence such as, 'Ade eats the food' is an example of a sentence in the BSVO instrument.

The MSVO words are designed to test the understanding of ESRs. For example, 'The tall Professor bought the big house'. The word tall tells more about the Professor and big tells more about the house. Ten MSVO sentences are used for the questionnaire. The sentences are designed to determine the understanding of the ESRs in terms of word syllable, sentence orthography, and syntax of translated *Yorùbá* sentences. The Appendix shows the questionnaire details.

B. Expert and Experimental Subject Respondents' Profiles

The age range, sex, hometown, state, educational level, knowledge of *Yorùbá* orthography and knowledge of machine translation system are filled in the questionnaire. The returned instrument provided the needed information about each ESR. The ESRs are postgraduate, graduate and undergraduate students. The majority of ESRs are males and hailed from Oyo, Ogun, Ondo, Osun, Ekiti and Lagos states. Their ages range between 20 and 50 years. Their *Yorùbá* orthography knowledge ranges between adequate and excellent. Some have used machine translation system before and some have not. The ESRs and Human Expert are literate *Yorùbá* speakers.

The human expert (HE) has a Ph.D. in *Yorùbá* and is a trained translator. He teaches *Yorùbá* and linguistics. The human expert translations are used to moderate the score of the machine, and ESRs translations during the evaluation.

C. Questionnaire Administration

The instrument comprises twenty (20) sentences administered to thirty (30) (ESRs). The distribution was done within the Obafemi Awolowo University, Ile-Ife, Nigeria. This environment was suitable because literate *Yorùbá* speakers are needed to do the translations. The questionnaires were distributed among *Yorùbá* ethnic group only. Fifteen ESRs returned the questionnaires. The evaluation was carried out on 20 translated sentences for fifteen ESRs, excluding the machine translated sentences.

D. The E-Y MT System Evaluation

We evaluated the system by considering some SVO sentences. We considered the ESRs and the machine translated sentences. The expert translations are used to control the choice of words between the ESRs and the machine. The three criteria used for the evaluation are word syllable accuracy, word orthography, and sentence syntax correctness. Word syllable accuracy means that a word in a sentence is properly tone marked and under-dotted. For example, a word *Tolú* has two syllables, *To* has one syllable with mid-tone on *o* and *lú* has one syllable with a high tone on *u*. We evaluated each word using this approach and we scored the ESRs and the machine based on their ability to identify these syllables. The word orthography accuracy was evaluated by considering each word's orthography within the sentence. Tone marks of some words change within the sentence. For example, *Kínl é Kínl é wọ il é wọ* (enter) change tone mark to *wọ* (enter) in this sentence. All these and some other Yorùbá writing styles were considered during the evaluation. The ESRs and the machine translated sentences were scored. The sentence syntax correctness was evaluated based on the sentence structure of the target language (TL). We considered the position of each word in a sentence. A well-positioned word in a sentence attracts good score.

VI. RESULTS DISCUSSION

This section discusses the results of the evaluation based on three criteria: word syllable accuracy, word orthography accuracy, and sentence syntax accuracy. The BSVO and MSVO translated sentences are evaluated.

A. BSVO Sentences

The translated sentences 1 to 10 are BSVO sentences. The results of the evaluation are discussed based on the word syllable, word orthography accuracy, and sentence syntax correctness.

1) BSVO Words Syllable Accuracy

Figure 12 shows the results of word syllable accuracy for sentences 1 to 10 (BSVO). Syllable accuracy means using appropriate tone marks on a word.

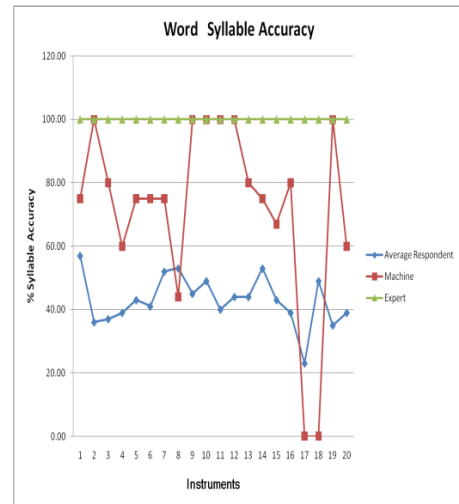


Fig.12. BSVO Word syllable accuracy

Here, each word is evaluated for its syllable accuracy. For example, the word *Tolú* has two syllables, that is, *To* (mid-tone) and *lú* (high tone). It is observed that the machine translator average score is higher than the experimental subject respondents (ESRs) average score. There were three instances where the machine translator score was equal to that of the Human Expert as it is shown in Figure 12, and no instances where the ESRs average score is equal to that of the Human Expert. The respondents' average score is 44.8 percent, the machine translator average score is 75.8 percent and the Human Expert average score is 100 percent.

2) BSVO Word Orthography Accuracy

Figure 13 shows the results of word orthography accuracy for sentences 1 to 10 (BSVO). The word orthography evaluation is considered in terms of tone and under-dots marking correctness. It is observed that the machine score is higher than that of the experimental subject respondents (ESRs). There are three instances where the scores of machine equal to that of the Human Expert as it is shown in the graph and no instance where the score of ESRs' average score is equal to that of the Human Expert. The respondents' average score is 68.7 percent, the machine translator average is 81.9 percent and the Human Expert average score is 100 percent.

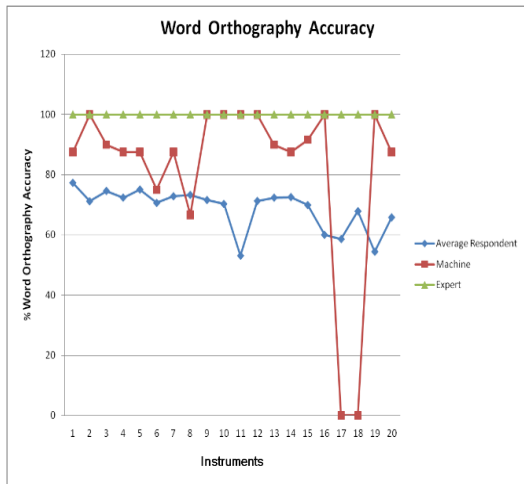


Fig.13. BSVO Word orthography accuracy

3) *BSVO Sentences Syntax Accuracy*

Figure 14 shows the results of syntax correctness for sentences 1 to 10 (BSVO). It is observed that the machine translator scores are the same with the Human Expert and the experimental subject respondents (ESRs) average scores are lower than that of the machine and the Human Expert. There is no instance(s) where the scores of ESRs equals to that of the Human Expert as it is shown in Figure 14. The ESRs average score is 86.5 percent, the machine translator average is 100 percent and the Human Expert average score is also 100 percent.

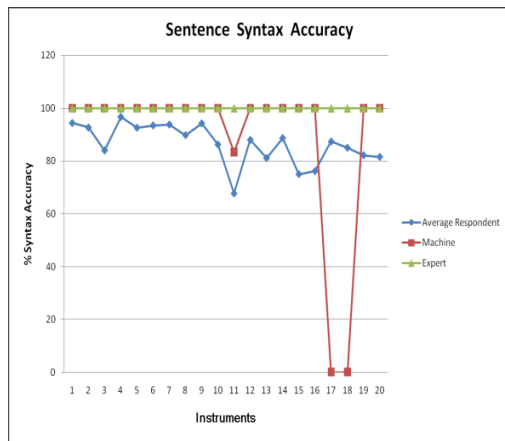


Fig. 14. BSVO Sentences syntax accuracy

B. *MSVO Sentences*

The translated sentences 11 to 20 are MSVO sentences. The results for the word syllable accuracy, word orthography, and syntax correctness are discussed.

1) *MSVO Word Syllable Accuracy*

Figure 15 shows the results of word syllable accuracy of sentences 11 to 20 MSVO. It is observed that the machine translator average score is higher than the

experimental subject respondents (ESRs) average score and lower than The Human Expert average score. There are three instances where the scores of machine are equal to that of the Expert and two instances where the machine translator scored 0 percent, the ESRs scored 21 percent and 45 percent as it is shown in Figure 15 and no instances where the score of ERSs' average score is equal to that of the Human Expert. The respondents' average score is 37.0 percent, the machine translator average is 86.7 percent and the Expert average score is 100 percent. However, it is noticed that respondents 5, 7, 9, 10, 13, and 14 did not tone mark the words. This is responsible for the performance recorded by the ESRs.

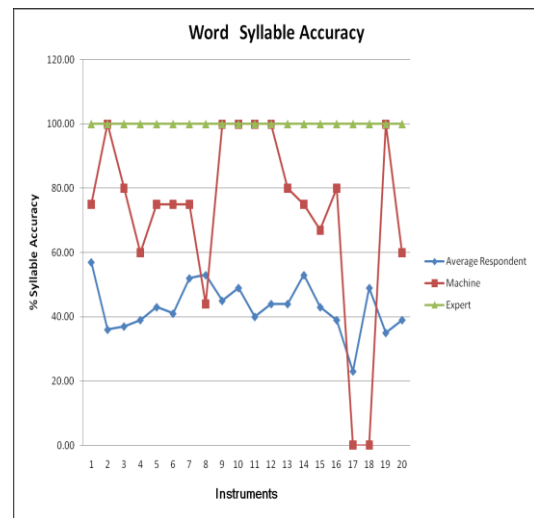


Fig.15. MSVO word syllable accuracy

2) *MSVO Word Orthography Accuracy*

Figure 16 shows the results of Word orthography correctness of sentences 11 to 20 MSVO. The result discussion focused on modified SVO sentences 11 to 20. It is observed that the machine score is higher than that of the experimental subject respondents (ESRs). There are four instances where the scores of the machine are equals to that of the Human Expert as it is shown in Figure 16 and no instances where the score of ERSs' average score is equal to that of the Human Expert. There are two instances where the ESRs scored 60 percent and 70 percent and machine translator scored 0 percent. The respondents' average score is 65.8 percent, the machine translator average is 87.5 percent and the Human Expert average score is 100 percent.

However, it was noticed that respondents 5, 7, 9, 10, 13, and 14 did not tone mark the sentences. This is responsible for the performance recorded by these ESRs. This reduces the ESRs average scores. Also, the machine was unable to translate sentences 17 and 18. The words used in the instrument are nouns only instead of adjective and noun. The machine could not use the re-write rules to translate the sentences.

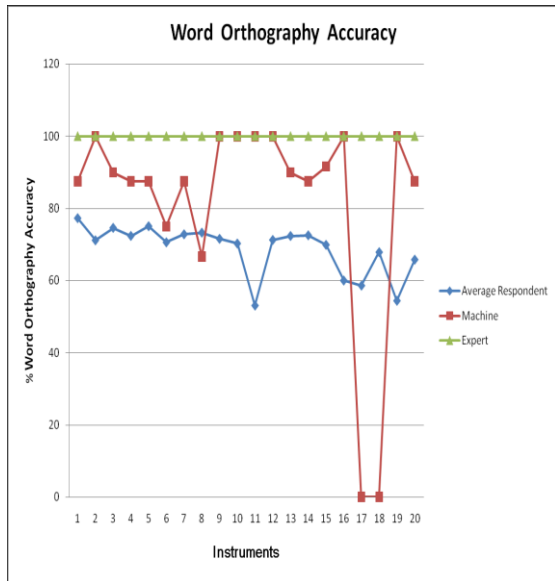


Fig.16. MSVO word orthography accuracy

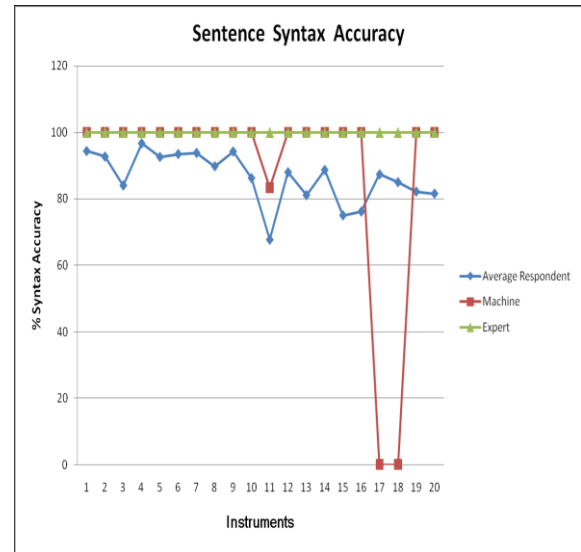


Fig.17. MSVO Sentence syntax accuracy

3) MSVO Sentences Syntax Accuracy

Figure 17 shows the results of syntax correctness of sentences 11 to 20 MSVO. It is observed that the machine score is higher than that of the experimental subject respondents (ESRs). It is observed that the machine translator scores are the same with that of the Human Expert except for three instances where machine translator scored below Human Expert. The experimental subject respondents (ESRs) average scores are lower than that of the machine translator and the Human Expert except for the two instances where the machine scores are zero (0) percentages. There are no instances where the ESRs scores are equal to that of the Human Expert as it is shown in Figure 17. The ESRs average score is 82.4 percent, the machine translator average score is 78.2 percent and the Human Expert average score is 100 percent.

VII. CONCLUSION

The English to Yorùbá machine translator was developed to make the Yorùbá language available to the larger audience. The study presented issues that affect the translation of English to Yorùbá and its underlining principles. The system was developed to enhance the learning of the Yorùbá language. It is user-friendly and allows learners to learn the language at ease. However, this work is a foundation for all other works, every other study will be integrated with this study. The evaluation focused on translation accuracy and the metrics considered are word syllable and orthography accuracy and sentences' syntax accuracy. The translation fluency evaluation will be performed in future.

REFERENCES

- [1] National population commission, 2006 census: URL: www.population.gov.ng (accessed: 25/06/2012).
- [2] Adewole, L. O. (1988) the Categorical Status and the Functions of the Yorùbá Auxiliary Verbs with Some Structural Analysis in GPSG, University of Edinburgh, Edinburgh.
- [3] Bamgbose A. (1966) A Grammar of Yorùbá Cambridge University Press, London
- [4] Cornell, D. and Ladd, D.R. (1990) Aspects of Pitch Realisation in Yorùbá Phonology, Vol 7, 1-29
- [5] Adewole, L. O. (1987) the Yorùbá Language: Published Works and Doctoral Dissertations, Hamburg: Helmut Buske.
- [6] Kishore, P., Salim, R., Todd, W., and Wie-Jing, Z. (2002) "BLEU: A method for Automatic Evaluation of machine translation". 311-318, Philadelphia, P.A. In proceedings of the 40th Annual meeting of the Association for Computational Linguistics (ACL). URL: <http://acl.ldc.upenn.edu/p/p02/p02-1040.pdf> (accessed: 20/09/2011).
- [7] Piron, C. (1994) Le défi des Langues (The language challenge) Paris L'Harmattan.
- [8] Muegge, U. (2006) fully automatic high quality machine translation of restricted text: A case study, In proceedings, translating and the computer, Aslib London URL: http://works.bepress.com/uwe_muegge/25 (Accessed: 23/08/2011).
- [9] Turing, A. M. (1950), Computing machinery and intelligence, journal of Mind, volume 59, pp 433-460
- [10] Eludiora, S.I. (2014) *development of an English to Yorùbá machine translation system*, unpublished Ph.D. thesis, Obafemi Awolowo University, Ile-Ife, Nigeria.
- [11] Eludiora, S. I., Agbeyangi, A. O. and Ojadiran, D. I. (2015a) Word Sense Disambiguation in English to Yorùbá Machine Translation System, Journal of Multidisciplinary Engineering Science and Technology, Berlin, Germany, vol 2, issue 7, 1814-1819.

- [12] Eludiora, S. I., Agbeyangi, A. O. and Fatunsin, A. (2015b) Development of an English to Yorùbá Machine Translation System for Yorùbá Verbs' Tone Changing, *International Journal Computer Application*, USA, vol 129, number 10, 12-17.
- [13] Eludiora, S. I., Okunola, M. A and Odejebi, O.A. (2015c): Computational Modelling of Yorùbá Split-Verbs for English to Yorùbá Machine Translation System, *International Journal of Advanced Research in Computer Science and Applications*, Bangalore, vol (3), issue no (4), 1-12.
- [14] Eludiora, S. I., Awoniyi, A. and Azeez, I. O. (2015d) Computational Modelling of Personal Pronouns for English to Yorùbá Machine Translation System, a paper presented at IEEE and The Science and Information Organisation (SAI) Intelligent Systems Conference 2015 (IntelliSys 2015) held in London, United Kingdom, November 10-11, 2015. 733-741
- [15] Agbeyangi, A. O., Eludiora, S. I. and Adenekan, D. I. (2015) English to Yorùbá Machine Translation System using rule-based approach, *Journal of Multidisciplinary Engineering Science and Technology*, Berlin, Germany, vol 2, issue 8, 2275-2280.
- [16] Odejebi, O. O., Ajayi, A. O., Eludiora, S. I., Akanbi, L. A., Iyanda, I. R. and Akinade, O. A. (2015) A web-based system for supporting teaching and learning of Nigerian indigenous languages, in *OAU TekCONF 2015 proceedings*, Nigeria, 350-360.
- [17] Akinwale, O.I., Adetunmbi, A.O., Obe, O.O., Adesuyi, A.T. (2015) Web-based English to Yorùbá Machine Translation, *International Journal of Language and Linguistics*, 3(3): 154-159.
- [18] Abiola, O.B., Adetunmbi, A.O. and Oguntimilehin, A. (2015) using a hybrid approach for English to Yorùbá text to text Machine Translation System (proposed), *International Journal of Computer Science and Mobile Computing (IJCSMC)*, vol 4, issue 8, 308-313.
- [19] Abiola, O.B., Adetunmbi, A.O., Fasiku, A.I. and Olatunji, K.A. (2014) a web-based English Yorùbá Noun phrases machine translation system, an international journal of English and Literature. Vol 5(3) 71-78.
- [20] Abiola, O.B., Adetunmbi, A.O. and Oguntimilehin, A. (2013) A computational model of English to Yorùbá Noun Phrases Translation System, *FUTA Journal of research in Sciences*, (1): 34-43.
- [21] White, J. S., O'Connell, T.A and Carlson, L.M. (1993) evaluation of machine translation, published in *HLT '93 proceedings of the workshop on Human Language Technology*, Association for Computational Linguistics, Stroudsburg, PA USA. 206-210.
- [22] Koehn, P. and Monz, C. (2006), *Manual and Automatic Evaluation of Machine Translation between European Languages*, in proceedings of the Workshop on Statistical machine translation, 102-121.

Authors' Profiles



Safiriyu I. Eludiora has Ph.D. in Computer Science from Department of Computer Science & Engineering, Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria in 2014. His Ph.D. research work was on machine translation. Precisely, he developed an English to Yoruba Machine Translator.

His students have developed text editors for the Nigerian three

official languages. He is currently working on Yoruba Word Processor System. He has published many articles in the field of machine translation, a sub-field of Computational Linguistics.

He is a member Nigerian Society of Engineers. He is a registered Computer Engineer. He is a member of Computing and Intelligent Systems Research Group (CISRG).


His research interest includes human language development, machine translations for local, national, and international languages. In addition, he has started working on the concept of Forensics Linguistics.



Odetunji A. Odejebi B.Sc. (Computer Engineering) First Class Honours; OAU, Ile-Ife, Nigeria, PhD(Computer Science) Aston University, United Kingdom, COREN Registered Engineer (Nigeria), CPN Registered.

Dr. Odejebi's research is in the area of Computing and Intelligent Systems Engineering with focus on speech and language engineering. Dr. Odejebi was a Visiting Scholar to the School of Information Technology and Electrical Engineering at the University of Queensland, Brisbane, Australia as a fellow of the United Nations University International Institute of Software Technology (UNU-IIST). He was a Commonwealth Scholar (British) at the University of Aston in Birmingham, United Kingdom. Dr. Odejebi also had visiting scholar positions at the Phonetics Laboratory of the University of Oxford, Oxford, England and The Centre for Speech Technology Research (CSTR), of the University of Edinburgh, in Scotland, United Kingdom. Dr. Odejebi was a Marie Curie Research Fellow on the Constraint Reasoning Extended to Enhance Decision (CREED) Project at the Cork Constraint Computation Centre (4C), of the University College Cork (UCC), Cork, Ireland (The Republic). He was a Research Academic in the same institution between 2008 and 2009. Dr. Odejebi is a consultant to a number of National and International organisations. For example, he served as a consultant to the African Languages Technologies Initiatives (Alt-I) on the Microsoft Vista Operating System Localisation Project. He is currently a Senior Lecturer In the Department of Computer Science and Engineering. Obafemi Awolowo University, Ile-Ife, Nigeria. He has many publications in Local and International Journals.

Appendix A Questionnaire Instruments



Department of Computer Science & Engineering
Obafemi Awolowo University, Ile-Ife.

Dear Respondent,

This questionnaire is meant to gather data required for the evaluation of a software developed as a result of a Ph.D Program in Computer Science Titled "Development of an English to Yorùbá Machine translation System". I am soliciting for your cooperation in responding honestly to the questions. The data provided will be used for research purpose and treated with utmost confidentiality.

1. Age: 12 - 15 16 - 24 25 - 30 >30 e.g. 20 years

2. Sex: M F

3. Home Town/State:

4. Educational Level:
 SSCE Undergraduate Graduate Post Graduate

5. Knowledge of Yorùbá Orthography (Writing) System:
 Weak Average Adequate Excellent

6. Please use accurate Yorùbá orthography: e.g. Ade bought the fish. *Ade ra eju náà*
 7. Have you used a Machine Translation System before

Yes No

Basic Subject Verb Object (SVO) Sentences

8. Tolu called the boy.

 9. She carried the load.

Figure A Questionnaire

Basic Subject Verb Object (SVO) Sentences

- Tolu called the boy.
Tolú pè Omòdekúnrin náà
- She carried the load.
Ó gbé erù náà
- The cat catches the mouse.
Ológbò náà tẹ̀ eku náà
- They checked the room.
Wọ̀n wò yàrá náà.
- She cleaned the table.
Ó nu tábilì náà.
- Ade closed the door.

Figure B Machine Translated Sentences

Basic Subject Verb Object (SVO) Sentences

- Tolu called the boy
Tolú pè omòkúnrin náà
- She carried the load
Ó gbé erù (náà)
- The cat catches the mouse
Ológinní rá eeku (náà) mú
- They checked the room
Wọ̀n yẹ̀ yàrá (náà) wò
- She cleaned the table
Ó nu tábilì

Figure C Human Expert Translated Sentences

| Respondent No | Sent No | Word1 | Word2 | Word3 | Word4 | Word5 | Ave Correctness |
|---------------|---------|--|--|--|--|--|-----------------|
| | | Under-dots & Tone Marks % Correctness | Under-dots & Tone Marks % Correctness | Under-dots & Tone Marks % Correctness | Under-dots & Tone Marks % Correctness | Under-dots & Tone Marks % Correctness | |
| | 1(8) | | | | | | |
| 1 | | 100% | 100% | 100% | 100% | 100% | 100% |
| 2 | | 50% | 100% | 0% | 100% | 0% | 63% |
| 3 | | 100% | 100% | 100% | 100% | 0% | 75% |
| 4 | | 100% | 100% | 100% | 100% | 100% | 100% |
| 5 | | 50% | 100% | 0% | 0% | 0% | 38% |
| 6 | | 100% | 100% | 100% | 100% | 100% | 100% |
| 7 | | 100% | 100% | 0% | 0% | 0% | 50% |
| 8 | | 100% | 100% | 100% | 0% | 0% | 75% |
| 9 | | 0% | 100% | 0% | 0% | 0% | 25% |
| 10 | | 0% | 100% | 0% | 0% | 0% | 25% |
| 11 | | 0% | 0% | 0% | 100% | 0% | 25% |
| 12 | | 100% | 0% | 0% | 100% | 0% | 50% |
| 13 | | 0% | 100% | 0% | 0% | 0% | 25% |
| 14 | | 0% | 0% | 0% | 0% | 0% | 0% |
| 15 | | 100% | 100% | 100% | 100% | 100% | 100% |
| Cum. Average | | | | | | | 57% |
| Machine | | 100% | 0% | 100% | 100% | | 75% |

Figure D Syllable Accuracy Analysis

| Respondent No | Sent No | Word1 | Word2 | Word3 | Word4 | Word5 | Ave Correctness |
|---------------|---------|--|--|--|--|--|-----------------|
| | | Under-dots & Tone Marks % Correctness | Under-dots & Tone Marks % Correctness | Under-dots & Tone Marks % Correctness | Under-dots & Tone Marks % Correctness | Under-dots & Tone Marks % Correctness | |
| | 1(8) | | | | | | |
| 1 | | 100% | 100% | 100% | 100% | 100% | 100% |
| 2 | | 50% | 100% | 0% | 100% | 0% | 63% |
| 3 | | 100% | 100% | 100% | 100% | 0% | 75% |
| 4 | | 100% | 100% | 100% | 100% | 100% | 100% |
| 5 | | 50% | 100% | 0% | 0% | 0% | 38% |
| 6 | | 100% | 100% | 100% | 100% | 100% | 100% |
| 7 | | 100% | 100% | 0% | 0% | 0% | 50% |
| 8 | | 100% | 100% | 100% | 0% | 0% | 75% |
| 9 | | 0% | 100% | 0% | 0% | 0% | 25% |
| 10 | | 0% | 100% | 0% | 0% | 0% | 25% |
| 11 | | 0% | 0% | 0% | 100% | 0% | 25% |
| 12 | | 100% | 0% | 0% | 100% | 0% | 50% |
| 13 | | 0% | 100% | 0% | 0% | 0% | 25% |
| 14 | | 0% | 0% | 0% | 0% | 0% | 0% |
| 15 | | 100% | 100% | 100% | 100% | 100% | 100% |
| Cum. Average | | | | | | | 57% |
| Machine | | 100% | 0% | 100% | 100% | | 75% |

Figure E Word Orthography Accuracy Analysis

| Respondent No | Sentences | Under-dot(%) | Tone Mark(%) | Syntax(%) | Semantic(%) | | Average Score(%) |
|---------------|-----------|--------------|--------------|-----------|---------------------|--------------------------|------------------|
| | | | | | Word Correctness(%) | Sentence Correctness (%) | |
| | 1 | 100 | 25 | 100 | 25 | 0 | 50 |
| | 2 | 100 | 0 | 100 | 0 | 0 | 40 |
| | 3 | 100 | 20 | 100 | 0 | 0 | 44 |
| | 4 | 100 | 0 | 100 | 0 | 0 | 40 |
| | 5 | 100 | 25 | 100 | 0 | 0 | 45 |
| | 6 | 100 | 20 | 100 | 20 | 0 | 48 |
| | 7 | 100 | 25 | 100 | 25 | 0 | 50 |
| 1(5) | 8 | 100 | 33.33 | 100 | 33.33 | 0 | 53.332 |
| | 9 | 100 | 20 | 100 | 0 | 0 | 44 |
| | 10 | 100 | 20 | 80 | 25 | 0 | 45 |
| | 11 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 12 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 13 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 14 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 15 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 16 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 17 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 18 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 19 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 20 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 21 | 100 | 75 | 100 | 50 | 0 | 65 |
| | 22 | 100 | 100 | 100 | 83.3333 | 0 | 76.66666 |
| | 23 | 100 | 100 | 100 | 60 | 0 | 72 |
| | 24 | 100 | 100 | 83.33 | 83.33 | 0 | 73.332 |

Figure F Syntax Accuracy Analysis

How to cite this paper: Safiriyu I. Eludiora, Odetunji A. Odejobi, "Development of an English to Yorùbá Machine Translator", International Journal of Modern Education and Computer Science(IJMECS), Vol.8, No.11, pp.8-19, 2016.DOI: 10.5815/ijmeecs.2016.11.02