INTERNATIONAL JOURNAL OF SOFT COMPUTING

MEDWELL PUBLICATIONS

# Development of an Adjectival Phrase-Based English to Yorùbá Machine Translator

Adebimpe Esan, Bolaji Omodunbi, Olatayo Olaniyan, Precious Odiase and Timileyin Olaleye
*Department of Computer Engineering, Federal University Oye-Ekiti, Ekiti State, Nigeria*

**Key words:** Machine translator, orthography, questionnaires, respondents, ADJP, JFLAP

**Corresponding Author:**
Bolaji Omodunbi
*Department of Computer Engineering, Federal University Oye-Ekiti, Ekiti State, Nigeria*

Page No.: 97-102
Volume: 15, Issue 4, 2020
ISSN: 1816-9503
International Journal of Soft Computing


**Abstract:** An Adjectival Phrase-based (ADJP) system was developed in this article for English to Yorùbá machine translation. The data for the developed system was extracted from locally spoken words and stored in a database. JFLAP was used to test the re-write rules and grammar using parse trees and Python programming language is the core programming language used in developing the system. The developed translator was evaluated by comparing expert's translated phrases to that of the developed translator and the experimental subject respondents using the Mean Opinion Score (MOS) technique based on word orthography. Results show that the expert's average accuracy was 100% while the respondent's was 76.3% and the developed machine translator's accuracy was 95.5%. In conclusion, the developed system's accuracy is close to the expert's and higher than that of the experimental subject respondent's.

## INTRODUCTION

Yorùbá language is gradually going into extinction because most speakers don't know how to write it despite that it is being taught in primary and secondary schools in Nigeria. In addition, the total dominance of English language over Yorùbá language in almost all human endeavour is a major challenge. Therefore, this call for the need of modern day processing tools like machine translators to make the language catch up with the technological growth the world is experiencing. Machine Translation (MT) is a subfield of computational linguistics that investigates the use of computer software to translate text or speech from one natural language to another[1]. It is an important part of natural language processing in artificial intelligence which accepts characters of source language and map to the characters of the target language to generate the words with the help of various rules and other learning process techniques[2].

The history of machine translation can be traced back to the pioneers and early systems of the 1950's and 1960's, the impact of the ALPAC report in the mid-1960's, the revival in the 1970's, the appearance of commercial and operational systems in the 1980's, research during the 1980's, new developments in research in the 1990s and the growing use of systems in the past decade[3]. Meanwhile, in the face of rapid globalization, the significance of machine translation cannot be overemphasized because it can be applied in various areas such as: e-Health, governmental organization, industries, e-Learning, speech translation, software localization and information retrieval[4]. It also help to translate a content quickly and provides quality output, thus, saving human

the stress and time of poring on translating books or looking for human translators. Hence, this research developed an Adjectival Phrase-based (ADJP) system for English to Yorùbá Machine Translation. The adjectival phrase was chosen because it provides important information about location, description of people and things, positions, relationships, time and ideas.

**Literature review:** Several approaches have been used in translating from one language to another. One of these approaches is the rule based approach which is commonly used for languages with lots of grammar. This approach was used by Eludiora *et al*.[5] to translate YoruÌbaì verb's tone changing words to English and the developed system was implemented and tested for twenty tone changing verbs within the home domain. Results show that the MT system can translate these tone change verbs correctly and efficiently at required response time. Meanwhile, another aspect of the rule based approach that allows manual tagging of the Part of Speech (POS) is the transfer-based approach. The approach was employed by Eludiora[6] to propose a machine translator that can only translate simple sentences using context-free grammar and phrase structure grammar. Generally, the rule-based approach uses Natural Language Toolkits (NLTKs) or JFLAp to verify the correctness of the rules of a language. This is confirmed by Agbeyangi *et al*.[7] where English to Yorùbá Machine Translator was developed using the Rule-Based Approach. In their research, rewrite rules were developed using Natural Language Toolkits (NLTKs) and implemented using python programming language. Results revealed that the system performance is close to the expert's opinion. Moreover, in analyzing the production rules of a rule-based approach, the automata rule can be employed. This is confirmed by Abiola *et al*.[8] where English to Yorùbá machine translator was developed for noun phrase. According to the researchers, using the rule-based approach, automata theory was employed in the analysis of the production rules and the developed system was evaluated using Nigerian daily news. Research revealed that the system translation accuracy using some phrases was 90%.

Furthermore, another approach to machine translation is the statistical approach which is a data-oriented statistical framework for translating text from one natural language to another. It relies on large parallel aligned corpora based on the knowledge and statistical models extracted from bilingual corpora. Statistical machine translation based Punjabi to English transliteration system was proposed by Kumar and Kumar[2]. They were motivated by the need to break communication barrier between Punjabi native speakers and those who do not understand Punjabi but English and results show that the system has 97% accuracy when tested on 2000 words and trained on about 15000 words. The major problem with the development of MT systems using the statistical method is the necessity of parallel corpus. Folajinmi and Omonayin[9] developed a Statistical Machine Translator (SMT) for English-Yoruba but lack of existing English-Yoruba parallel corpus necessitates its creation.

Furthermore, in order to improve the quality of a machine translation, two or more approaches can be combined to form a hybrid approach[4]. Using a hybrid approach, Haque *et al*.[10] developed an English to Hindi Transliteration system based on the Phrase-Based Statistical Method (PB-SMT). PB-SMT was modeled as adecision process and source context was modelled into the state-of-the-art log-linear PB-SMT. Results show that the system outperformed existing systems built using single approach.

## MATERIALS AND METHODS

The main tools used in developing the work include: JFLAP, Python programming language, NTLK (Natural Language Toolkit), PyQt5 and Py2exe used to compile the python codes (.py) to an executable file (.exe). The architecture of the developed system as shown in Fig. 1 include: system GUI that interfaces the user and the translator, translator which translates the words provided by users by fetching the corresponding translation provided by the Parser, Parser which automatically analyses texts according to the grammar and parses the words in the database and the database that stores the parallel corpus (data) used for the translation.

**Re-write testing:** The rules that guide the system design are:

**Rule 1:** An Adjectival Phrase (ADJP) consists of adjective and Noun Phrase (NP). In the case of target language noun (ỌO) comes before determiner (AIỌO). For example:

- SL: an<DET>old<ADJ>man<N>
- TL: Àgbàlágbà<ỌA>Ọkùnrin<ỌO>Kan<AIỌO>

**Rule 2:** A determiner must precede an adjective and a noun in SL but reverse is the case in the TL. For example:

- SL: The<DET> tall<ADJ> boy<N>
- TL: ọmọkùnrin<ỌO>gíga<ỌA>náà<AIỌO>

JFLAP was used to test the rewrite rules as shown in Fig. 2 and 3. The mode of translation is based on the grammar designed for both English language and Yorùbá language.

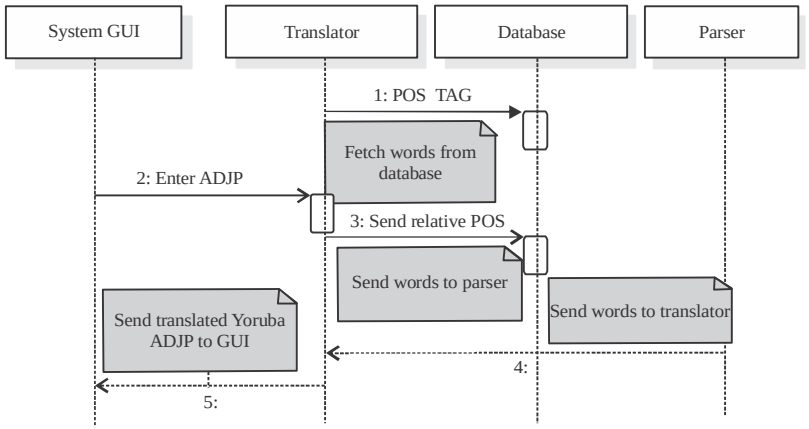


Fig. 1: Architecture of the developed translator

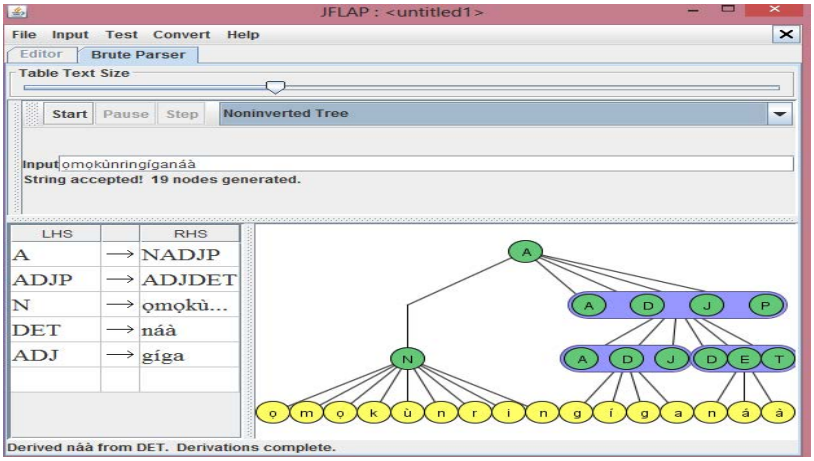


Fig. 2: Yoruba adjectival phrase rewrite test

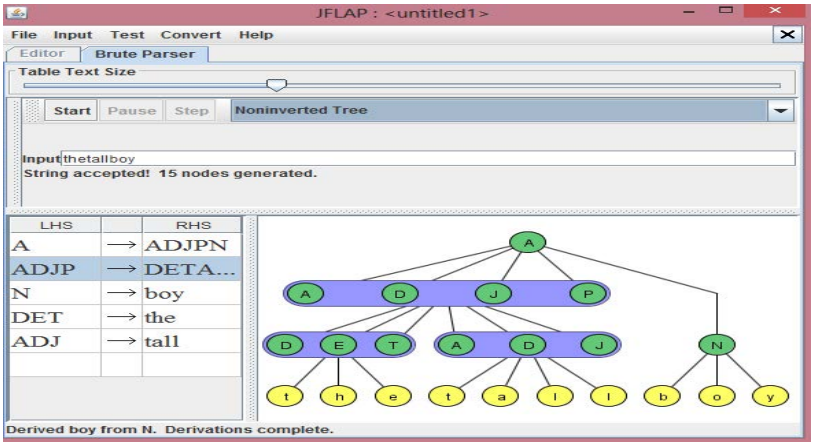


Fig. 3: English adjectival phrase rewrite test

**Adjectival phrase translation process:** The English ADJP translation process model described possible phrases that can be translated from the Source Language (SL) to the Target Language (TL) as shown in Fig. 4. The

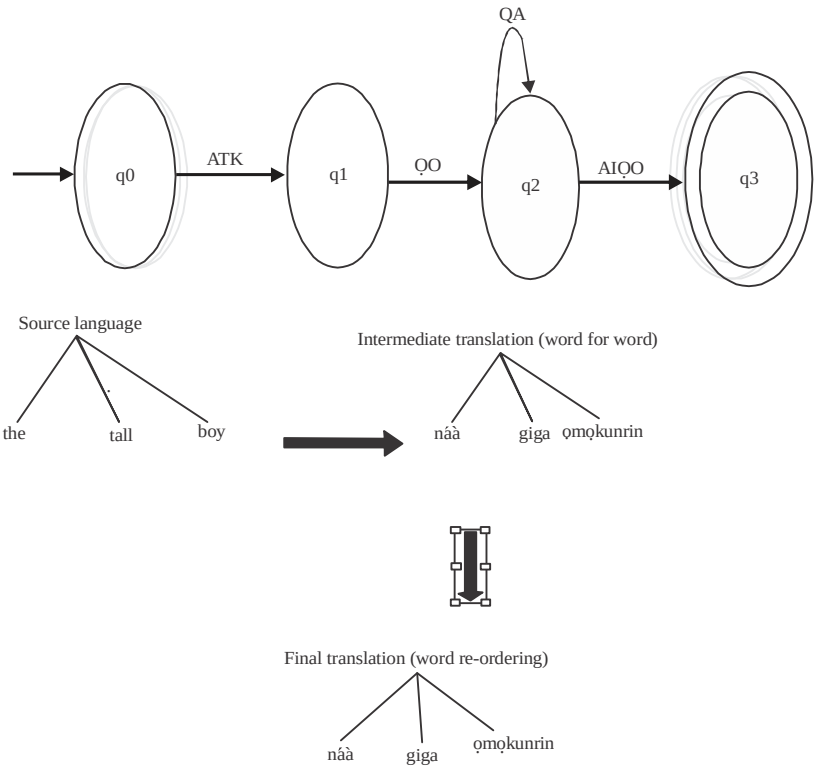


Fig. 4: Adjectival phrase translation process abstraction

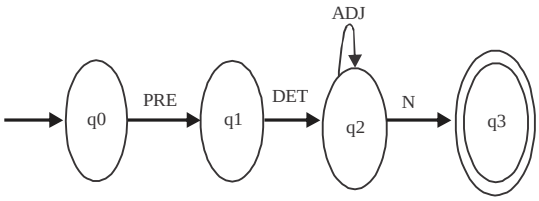


Fig. 5: State diagram for the English translation process

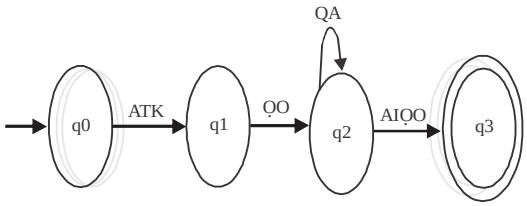


Fig. 6: State diagram for the Yorùbá translation process

ways translation of English adjectival phrase can be combined are: ADJDETN and PREDETADJN. Figure 5 is the state diagram for the Yorùbá language ADJP translation process while Fig. 6 shows possible combinations of adjectival phrases that can be accepted by the TL. They are: ATKỌOAIỌO and ATKỌOỌAAIỌO. One important thing to note is that, the noun (ỌO) and adjective (ỌA) swapped with the determiner. It shows that Yorùbá language is head first and English language is head last (Table 1).

Table 1: List of English pronouns and their Yorùbá equivalents

| English | Yorùbá |
|---|---|
| She/he/it | Ó |
| They | Àwọn |
| You | ìwọ/ìrẹ |
| We | Àwa |
| Them | wón |

**Design and implementation:** The Graphical User Interface (GUI) GUI has three planes, the first plane is where user enters the adjectival phrase, the second plane display input phrases word for word and the third plane displays the translated Yorùbá adjectival phrase. After an English phrase is input, the translator module of the code begins to execute. The phrase is broken into lexemes, it then tagged into different parts of speech. The translator module accepts input sentence from the GUI module then break it down and send it to the database module to confirm that the lexemes are in the database. However, if the lexemes are not in the database an error message will be generated. The translated sentence to target language is then displayed by the GUI. The lexemes are manually tagged and each word is categorised according to its parts of speech. The Natural Language Tool Kits (NLTKs)

was used as the parser module. The translation process is based on the phrase grammar rules built in the source code which implements the re-write rules. The machine translation system has the capability to translate sentences that contains an adjectival phrase from the English language to Yorùbá language in its textual form Table 2 and 3.

Table 2: List of English determinants and their Yorùbá equivalents

| English | Yorùbá |
|---|---|
| A | Kan |
| An | Kan |
| Some | Díè |
| The | Náà |

## RESULTS AND DISCUSSION

The sample of the output generated by the system is shown in Fig. 7.

**Evaluation of system:** The developed system was evaluated by administering questionnaires to respondents and the mean opinion score (Human Judgment) approach was used in determining the performance of the system (Fig. 8). The expert who is a professional translator translates the sentences from English language to Yorùbá language. The questionnaire designed has simple phrases that consist of adjectival phrases to test the experimental

Table 3: Analysis of results

| Phrases | Expert | Respondent average | Machine |
|---|---|---|---|
| 1 | 100 | 70 | 100 |
| 2 | 100 | 78 | 100 |
| 3 | 100 | 80 | 85 |
| 4 | 100 | 75 | 100 |
| 5 | 100 | 80 | 90 |
| 6 | 100 | 70 | 100 |
| 7 | 100 | 80 | 90 |
| 8 | 100 | 78 | 90 |
| 9 | 100 | 77 | 100 |
| 10 | 100 | 75 | 100 |
| Average | 100 | 76.3 | 95.5 |

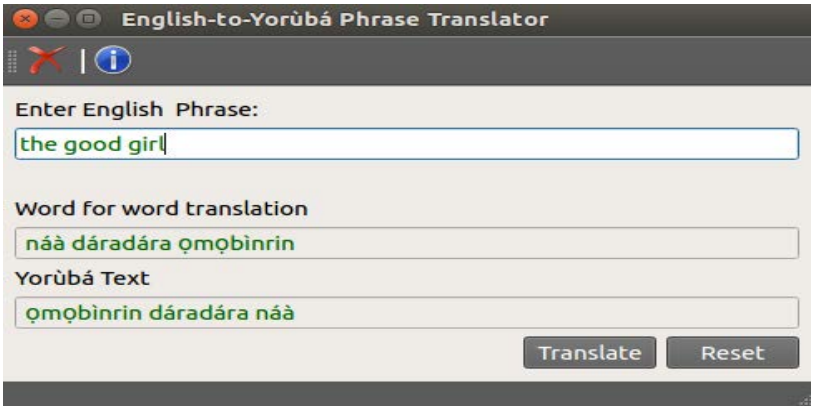


Fig. 7: System output

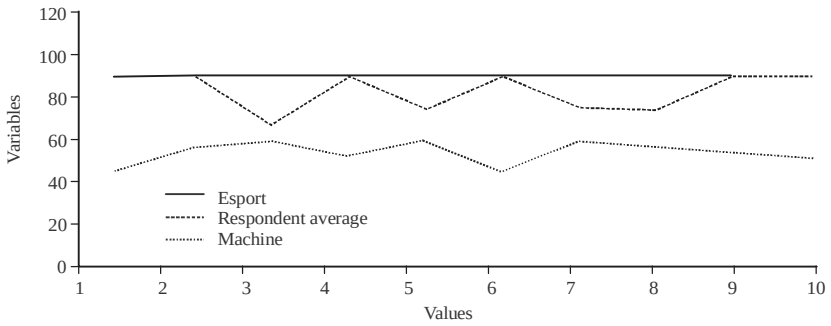


Fig. 8: Translated phrases orthography accuracy

subject respondent on the ability to translate simple sentences. The questionnaire has 10 simple adjectival phrases which were used in testing the respondent's translation accuracy based on Yorùbá language orthography and the syntax of the language which is described in term of tone marks and diacritics (dotted vowels and consonant). The questionnaires were administered in Ikole-Ekiti, Ekiti state, Nigeria. This area was chosen because there are literate Yorùbá speakers and the questionnaires were distributed among the Yorùbá speakers from the Yorùbá ethnic group.

The developed system was evaluated to determine its performance and this therefore demonstrates the quality and shortcoming of the developed system based on system accuracy using word orthography (tone marking and under dotting). From Table 3, the expert's percentage accuracy was 100 while the developed system has 95.5% accuracy and the result from experimental subject respondents is 76.3%. Figure 7 depicts that the machine correctness is close to that of the expert and more accurate than that of the average experimental subject respondents. This is in line with Eludiora where an English-Yoruba Translator was designed and the results 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, 82.3 and 100%, respectively. This thus establishes that system translation accuracies were close to a human expert (Fig. 8).

## CONCLUSION

An adjectival phrase-based system was designed in this research to translate English language to Yorùbá. The system was designed to enhance the learning of Yorùbá language with a user-friendly interface. Results show that the developed system was able to give accurate translations with appropriate tone-marks and under-dots because its accuracy is close to that of the Expert and higher than the experimental subject respondent's. It is recommended that future researchers work on translation of Adjectival phrases from English to other languages (Hausa and Ibo) in Nigeria and Neural Machine Translation.

## REFERENCES


01. Arnold, D., L. Balkan, S. Meijer, R.L. Humphreys and L. Sadlery, 1994. Machine Translation: An Introductory Guide. NCC Blackwell, London, England, UK., ISBN:9781855542464, Pages: 240.
02. Kumar, P. and V. Kumar, 2013. Statistical machine translation based Punjabi to English transliteration system for proper nouns. Intl. J. Appl. Innov. Eng. Manage., 2: 318-321.
03. Hutchins, J.W., 1994. Machine Translation: A Brief History. Academic Press, London, UK.,.
04. Oladosu, J., A. Esan, I. Adeyanju, B. Adegoke and O. Olaniyan *et al*., 2016. Approaches to machine translation: A review. FUOYE. J. Eng. Technol., 1: 121-126.
05. Eludiora, S.I., A.O. Agbeyangi and A. Fatunsin, 2015. Development of an English to Yoruba machine translation system for Yoruba verbs tone changing. Intl. J. Comput. Appl. USA., 129: 12-17.
06. Eludiora, S., 2014. Development of English to Yoruba machine translation system. Ph.D Thesis, Obafemi Awolowo University. Ife, Nigeria.
07. Agbeyangi, A.O., S.I. Eludiora and D.I. Adenekan, 2015. English to Yoruba machine translation system using rule-based approach. J. Multidiscip. Eng. Sci. Technol. Berl. Germany, 2: 2275-2280.
08. Abiola, O.B., A.O. Adetunmbi, A.I. Fasiku and K.A. Olatunji, 2014. A web-based English to Yoruba noun-phrases machine translation system. Intl. J. English Lit., 5: 71-78.
09. Folajimi, Y.O. and I. Omonayin, 2012. Using Statistical Machine Translation (SMT) as a language translation tool for understanding Yoruba language. Proceedings of the EIE's 2nd International Conference on Computing Energy, Networking, Robotics and Telecommunications, November 21-23, 2012, Covenant University, Ota, Nigeria, pp: 86-98.
10. Haque, R., S. Dandapat, A.K. Srivastava, S.K. Naskar and A. Way, 2009. English-Hindi transliteration using context-informed PB-SMT: The DCU system for NEWS 2009. Proceedings of the 2009 Conference on Named Entities Workshop: Shared Task on Transliteration, August 07, 2009, Association for Computational Linguistics, Suntec, Singapore, ISBN: 978-1-932432-57-2, pp: 104-107.