INTERNATIONAL JOURNAL OF ICT AND MANAGEMENT
A Chronicle of Wisdom

Applying Rough Set Theory to Yorùbá Language Translation

Fagbolu O.O [A], Obalalu B.S [B], Udoh S.S [C]

[A] The Polytechnic Ibadan, Department of Computer Science, Ibadan, Nigeria, tolatalk2mii@yahoo.com

[B] Federal University of Agriculture, Department of Computer Science, Abeokuta, Nigeria, bobalalu@yahoo.com}

[C] University of Uyo, Department of Computer Science, Uyo, Nigeria.

udohss@yahoo.com

ABSTRACT

The purpose of this research is to design and implement Yorùbá Language Processor using Rough Set Theory (RST). Yorùbá is a dialect of West Africa with over 50 million speakers, Yorùbá Language is spoken by well over 30 million speakers as their first language, almost 10million speakers are found in other countries in Africa, about 5 million speak Yorùbá Language in the diaspora and about 5 million speak Yorùbá Language as their second language in Nigeria which includes Akpes, Jummu, Ebira, Ao, Awori, Yagba, Iworro, Wo, Ikale, Gbedde etc. Yoruba is a tonal language which is considered third most spoken native African language.

RST approach is used in Natural Language Processing, its vagueness was used to deal with the process of analyzing natural language, convert the Source Language (SL) English Language into its equivalent in Yorùbá Language (TL), RST is applied to make approximate matches between words and phrases to be translated into its appropriate translations in another language. RST is viewed as the theory of implementing vagueness and imprecision in NLP which is expressed by a boundary region of a set and not partial membership.

This research formulated a model translation for English phrases and words to be translated to Yorùbá equivalents, its design and implementation were done on mobile and web platforms and RST helped to improve the quality of translation, rate of data retrieval of NLP and organization of NLP with the inexact, uncertain and vague set of data in the Yorùbá words and phrases.

Keywords: Rough Set Theory, NLP, Vagueness, Phrases

I. INTRODUCTION

In the 21st century, it is reasonable to expect that some of the most important development in Science and Engineering would come through interdisciplinary research. Yorùbá Language Processor using Rough Set Theory (RST) cuts across Information Theory, Computer Science (System Design and Artificial Intelligence), Statistics and Lingual (Syntax and Semantics). Computers are applied in translating texts or speeches from one Natural language to another [1]. The design of machine translation system is directly affected by how the system translates a natural language (Source Language) to another natural language (Target Language). The design of Yorùbá Language Processor for foreign language users has one major aim which is to understand how certain Yorùbá words and phrases can be understood by visitors through electronic media.

This application can be CD-ROM based or on network (intranet and internet) or mobile platform such as smartphones which provides channels for communication between travellers (tourist) and their guide (tutor) in their respective destinations. Yorùbá Language Processor is a subset of Natural Language Processing which process and transforms a text (phrase or word) from a source language into a target language. It involves English language (lingua franca) to Yorùbá language (mother tongue). The app would be designed with an easy to read and pronunciation guide for all the phrases and words, it would also assist readers (learners) that need to communicate effectively and efficiently in order to solve visitors-dwellers predicament while visitors are abroad or places where different languages are spoken [2]. In the time past, the quality of communication and fluency was linked to one's mother's tongue while in the future; the quality of communication would be linked to digital apps for translation [3]. Thus, with the advent of internet which is becoming the most important source of information, tourist or prospective learners of any language can obtain all necessary words and phrases to communicate in Yorùbá while mobile phones would afford easier access at every point in time to intending learner.

Yorùbá Language Processor app must identify the inquisitiveness of a typical learner or tourist; promotes its ethical values to potential students, accept the enquiry of the readers of Yorùbá language, deliver the readers' or learners'request and support learners' use for non-Yorùbá tourists that wish to converse with Yorùbá speakers. Internet brings people together from any country in the world and reduces the distance between people in many ways [4]. The app will specifically use the internet and web facilities as its data transmission medium to bind different people together irrespective of their differences and distance. The web as a virtual environment helps learners and teachers of Yorùbá language to share a common interest by reducing the cost and increase the communication skills of intended tourists or learners. Today a web is frequently the first place teachers, learners or researchers go to conduct any research or find out any information. Any tourist to a Yorùbá nation would likely consult the web for a guide to have an enjoyable moment in Yorùbá speaking countries and the availability of an app for Yorùbá Language Translation will be an added value to Yorùbá nations and enhance better relationship between the visitors (learners) and dwellers. Prospective learner can use this app to learn a language that is geographically separated from them. It would enhance inter cultural existence of Nigeria nation during the compulsory programme of National Youth Service Corp (NYSC) with other non-Yorùbá speakers and will also portray and rebrand the images of Yorùbá nations well. Yorùbá Language Processor platform increases the opportunities of conversing in other languages apart from one's mother tongue, web enhancement feature will increase the speed and accuracy with which learners and teachers can exchange information and cost of learning are drastically reduced. It will provide wide range of phrases and words in Yorùbá language for any interested learner to read or study 24hours a day in a multi-layer domain for different situations and actions. If distance education is making it possible for people to learn skills and earn degrees no matter where they live or which hours they are available for study, so also is this app in translating English to Yorùbá language with spoken feature.

Yorùbá is a dialect of West Africa with over 50 million speakers. It is a member of Niger-Congo family of language and it is spoken among other languages in Nigeria, Togo, Benin and partly in some communities in Brazil, Ghana, Sierra Leone (where it is called Oku) and Cuba (where it is called Nago)[5]. Yorùbá is one of the three major languages in Nigeria and language being the principal means used by human beings to communicate with one another; it is spoken and considered as the third most spoken native African language. Yorùbá language has ancestral speakers who according to their oral traditions is Oduduwa (son of Olúdùmarè), the supreme god of the Yorùbá [6]. Yorùbá first appeared in writing during the 19th century and the first publications were a number of teaching booklets produced by John Raban in 1830 – 1832 and another major contributor to orthography of Yorùbá was Bishop Samuel Ajayi Crowther (1806 – 1891) who studied many of the languages of Nigeria [7], he wrote and translated some of the Yorùbá phrases and words. Yorùbá orthography appeared in about 1850 although with many inherent changes since then. In the 17th century Yorùbá was written in the Ajami script [8] and major development in the documentation of Yorùbá words and phrases were done by Anglican (CMS) missionaries that were working in places like Sierra Leone, Brazil, Cuba and they assembled the grammatical units in Yorùbá together which were published as short notes [9], in 1875 Anglican communion organized a conference on Yorùbá orthography. Johnson (1921) remarked that several fruitless efforts had been made to either invent new characters or adapt the Arabic, which was already known to Moslem Yorùbá. Finally, Roman character-based alphabets that were acquainted with Anglican (CMS) missionaries were adopted [10].

Yorùbá anthology can be traced to the publication of several Yorùbá newsprints in Lagos, Nigeria in 1920s such as Eko Akete in 1920 with Alaagba Isaac B Thomas as the editor, Akede Eko in 1922, Eletiofe in 1925 with E.A Akintan as the editor and many more which enhance the numerous usage of the language in the area of economic, political, diplomatic and cultural relations. Yorùbá Language Processor will allow intending visitors or learners of the language to tap into numerous advantages to be derived in its usage but nonexistence of English – Yorùbá corpus can inhibits.

Rough Set Theory is a formal approximation applied to deal with vagueness and uncertainty in the app and enables information retrieval, decision rule generation and improved performance, however it is unique in strength by providing an objective form of analysis [11], it will require no additional information, external parameters, models, functions, grades or subjective interpretations to determine set membership and can only work with information presented in the set.

II. STATEMENT OF PROBLEM

Yorùbá Language as an African Language is going into extinction; its cultural values, creed and heritage are being eroded daily. Yorùbá speakers are not as many as they were several decades back but the advent of this app that can teach, tutor or guide any prospective speakers or learner of Yorùbá Language and provide an avenue to add more words to Yorùbá diction which will make Yorùbá Language to compete well with its contemporary languages both in Africa and in diaspora.

III. OBJECTIVES

There is no research work without its difficulties but its inherent advantages that often outweigh the challenges bring its merits to society at large. The objectives of the research are to

Design a model from RST for Yorùbá phrases and words translation.

A. Implement the model in (A)
B. Evaluate the performance of implemented model

IV. METHODOLOGY

This paper formulated a computational model for English Language to Yorùbá translation process. It is designed, implemented and evaluated the performance of the model with digital Yorùbá corpus [12] so as to address the challenges of English to Yorùbá machine translator, digital resource are well

collected from all domains of human endeavors. The transfer architecture is employed in order to produce syntactic and semantic translation that would be both logical and meaningful rather than producing only lexical translation. The word or phrase that is to be translated is parsed, followed by target language mapping or selected-word from large corpus created.

The transfer architecture underpins the linguistic transfer model, thus, it re-frame rules for words and phrases which are divided into these tokens:

A. Source language parser which consists of rules for SL analysis (Syntactic/Semantic)
B. Transfer engine, this handles rules for source to target transfer and procedures covered between the lexical and structural ambiguity in the language formalism.
C. Target language generator, it consist of rules for generating the target language

The transfer approach first of all, divides phrase into words, tags each word, translate each word using the corpus and generate the translation using rules of Target Language (TL). Rough Set Theory (RST) is applied to deal with vagueness and uncertainty in the app which enables information retrieval, decision rule generation and improve performance, set of phrases (words) are represented in information table with a finite set of attributes which can be expressed as

$$S = \{A \cup B: x_i \in A, x_i \notin B, y_i \in B, y_i \notin A\}, \quad (1)$$

$$i = 1, 2, \ldots n$$

where **S** is a finite non empty set of words or phrases called Universe; **A** is a finite non-empty set of distinct words or phrases to be translated and **B** is a finite non-empty set of words or phrases with more than one meaning. Two words or phrases $\mathbf{x_i}$ and $\mathbf{y_i}$ in **S** are equivalent if and only if they have same values on all words or phrases to be translated, denoted by **E,** the equivalence is defined as $[x] = \{ y \in U \mid x \epsilon y \}$. The equivalence relation **E** induces a partition of U denoted by $S \mid E$, the subsets in $U \mid E$ for rough set theory are lower and upper approximations of C (word or phrase to be translated)

$$\text{apr}(C) = \{x \in U \mid [x] \subseteq C\};$$

$$\overline{\text{apr}(C)} = \{x \in U \mid [x] \cap C \neq 0\} \quad (2)$$

Adopted from Pawlak, 1991

V. DATA PRESENTATION

The translation concept reproduces the receptor language (source language) using the closest natural equivalent of the source language message, first in terms of meaning and secondly in terms of style [13], the translation which is word-for-word translation is the basis presented in the Natural Language Processing of Yorùbá and further captured phrase-to-phrase but sentence-to-sentence was not catered for because of several ambiguities such as different word order, etc., the major task in this research work is to developing large lexicon or corpus [12] [14] and the corpus will be used for word-for-word translation in Yorùbá Language Translation. Here are some of the samples on mobile and web platforms. Figure 1 shows translations of words in the corpus

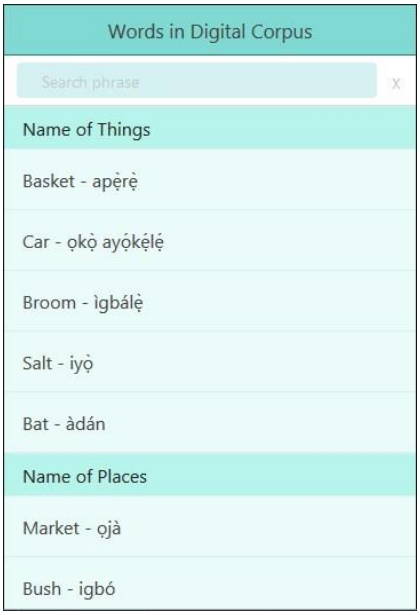

Fig. 1: Sample of several categories of words

This interface typifies one of the categorized main module "Words in Digital Corpus" by using touch inputs that correspond to real world actions like swiping, tapping, pinching and so on. Several other categories can be viewed. Once a word is selected, for example if "bat" is clicked, it will be displayed as shown in figure 2

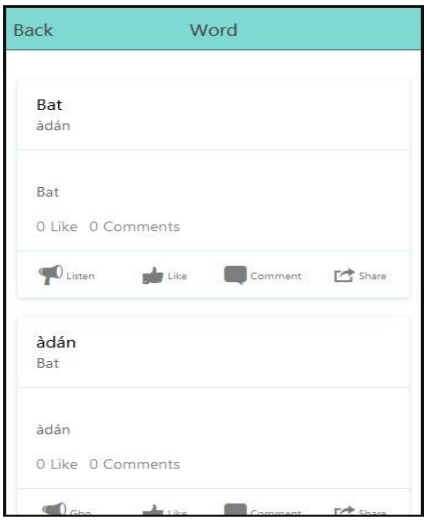

Fig. 2: Sample of a typical word

Below is a screenshot for the home domain of Yorùbá Language Processor on web platform where words or phrases to be translated are entered and their equivalent translations are produced using RST model and the developed corpus as in Figure 3.

Fig. 3: Typical home screen of web platform for Yorùbá translations

The screenshot below in figure 4a and 4b shows how words or phrases are translated and their translation into their equivalent target language.

Fig. 4a: Web platform of Yorùbá translating

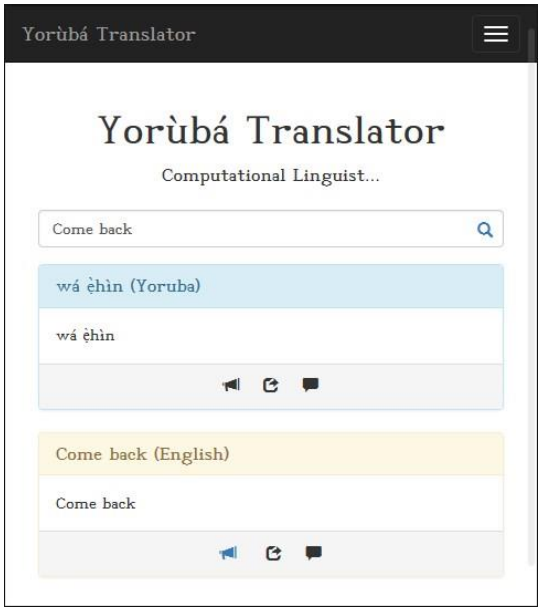

Fig. 4b: Web platform of Yorùbá translations result

VI. RESULT AND CONCLUSION

At the end, it was observed that the formulated computational model using Rough Set Theory translates English Language to Yorùbá with minimal errors and ambiguities and the set of words in the corpus is considered as objects of universe, RST will enhance the ways of picking an appropriate sample of word or phrase that matches. Similar words or phrases are based on equivalence relation in RST while the selection of appropriate translated word or phrase is partitioned as set of objects; each equivalence class contains similar queries and the indexing of the corpus are tagged based on Part OfA. Speech (POS).

This research formulated a model translation for Yorùbá phrases and words, the aim and objectives were achieved, its design and implementation were carried out on web and mobile platforms. During the course of this research, the Yorùbá language which is tending towards extinction was reawakened. This platform promotes indigenous African languages to native and non-native speakers, tourists and NYSC members from other ethnic groups in Nigeria that may want to associate with Yorùbá people during their visit or their service year.

In this research work cloud computing and lots of its facilities are being considered so as to obliterate several roadblocks to its wider implementation and develop Yoruba Corpus conveniently by availing large storage capability and reduce the length of time it takes to access or implement a fully functional Digital Corpus. Cloud computing will set new standard for rapid access and implementation which will speed up translation rate and quality with the availability of large storage capability.

It serves as one of the collective efforts to expand words, phrases and expressions in the Yorùbá language and make Yorùbá Language natural means of spoken and written communication for whosoever desires. Consequently, the language will be more popular, gain more value and prestige in addition to enjoying international acceptability and no one will denigrate it.

VII. CONTRIBUTION TO KNOWLEDGE

The findings of this research established a computational model by applying RST to Yorùbá Language Translations that:

Establishes mathematical theory of how units of grammar making Source Language are translated to units of grammar making Target Language;

Gives orderly and clear-cut definitions of English words and simple expressions in Yorùbá language, and

Provides means of teaching and tutoring Yorùbá language to non-indigenous learners and serves as a pedagogical tool with high reliability on electronic media. Cloud computing will enhance its design, avail wider implementation and increased performance if deployed.

VIII. REFERENCES

[1] Blank, D. (1998). *Definition of Machine Translation.* Retrieved from: http://www.macalester.edu/courses/russ65/definiti.htm

[2] Stephen, Howe, Kristina, & Henriksson. (2007). *Phrase books for writing papers and research in English*, London, Cambridge, The Whole World Company Press, England.

[3] Geere, D. (2009). *Talking digital phrasebook Travel Translator* Launches. Retrieved from: http//www.traveltranslator.htm

[4] Schneider, G., & Perry, J (2001). *Electronic Commerce*, Canada, Learning Inc.

[5] Bamgbose, A. (1965). *Yorùbá Orthography: A Linguistic Appraisal with Suggestions for Reform*. University Press, Ibadan.

[6] Biobaku, S. O. (1973). *Sources of Yorùbá History*, London, Oxford Clarence Press

[7] Oyenuga, S. (2007). *Learning Yorùbá* Retrieved from www.YorùbáForKidsAbroad.com

[8] Ogunbiyi, I. A. (2003). The Search for a Yorùbá Orthography since the 1840s: Obstacles to the Choice of the Arabic Script. Sudanic Africa. *A Journal of Historical Sources*, 14, 77–102.

[9] Adetugbo, A. (2003). *The Yorùbá Language in Yorùbá History*.

[10] Johnson, S. (1921). *The History of the Yorùbá. C.M.S*. Nigeria Bookshops, Lagos.

[11] Pawlak, Z. (1982). Rough sets. *International Journal of Computer and Information Sciences*, 11, 341-356

[12] Fagbolu O. O, Ojoawo A. O, Ajibade K. A and Alese, B. K.. (2015). Digital Yorùbá. *Corpus,International Journal of Innovative Science, Engineering and Technology*, *2*(8), 918-926.

[13] Eludiora, S. I. (2014). *Development of English to Yorùbá Machine Translation System*, Ph.D thesis, Obafemi Awolowo University, Ilé-Ifẹ̀, Nigeria.

[14] Fagbolu, O. O. (2015). *Development of web-enabled Yorùbá Phrasebook*, Ph.D thesis, Federal University of Technology, Akure, Nigeria.

AUTHOR'S BIOGRAPHY

Dr.Fagbolu holds B.Tech, M.Tech and Ph.D in Computer Science and has over 15 years professional and teaching experience with several scholarly acclaimed publications locally and internationally. He is a faculty member and researcher in one of the Nigerian Polytechnics, member of Nigerian Computer Society and other scholar bodies. His research interests are Quantum computing, Software Engineering and Natural Language Processing.

Mr.Obalalu is a reseach student in one of the Federal Universities in Nigeria, he holds Higher National Diploma and Post Graduate Diploma in Computer Science. He has contributed immensely to software exhibitions and development of several enterprise and customized software. His research interests AI, Software Engineering, HCI, Machine Learning and Natural Language Processing.

Mr.Udoh is a faculty member in the University of Uyo, Nigeria, he holds B.Sc, M.Tech and Ph.D Computer Science with several years of research in these domain - Artificial Intelligence, Artificial Neural Network and Fuzzy logic. He has numerous publications to his credit and belong to several scholarly acclaimed organizations within and outside the country.