Intelligent System for Learning and Understanding of Yoruba Language

Oyekanmi, Ezekiel Olufunminiyi
Department of MATS/STAT/CIS,
Achievers University,
Owo, Ondo State,
Nigeria

Oluwadare, Samuel Adebayo & Alese, Boniface Kayode*
Department of Computer Science,
The Federal University of Technology.
P.M.B. 704, Akure, Ondo State.
Nigeria
Email: *samoluwadare2013 {at} @gmail.com

Abstract—The analysis of Yoruba vowels has been the focus of several papers and has been a topic in numerous others. But while this body of work has led to a clear description of the way to learn Yoruba words, it has also helped in the analytical and theoretical interest of it. The development is based on the integration of ideas from computer aided education; computer mediated communication as well as techniques in artificial intelligence. The system is designed for access on a stand-alone PC. Various design and implementation issues with respect to components of the system are here discussed in detail. The system is developed to assist in the derivation of Yoruba equivalent of English words supplied as input. Since Yoruba is a tonal language, the system developed would in addition to giving the meaning, also pronounce the meaning. In developing the system therefore, recognition is given to some letters that are normally identified by placing dot under them (e.g) and caret (^) is used in place of the normal dot. This study illustrates how to process Yoruba language using computer. It would assist in the teaching and learning of Yoruba language.

Keywords-Yoruba; communication; vowels English words

I. INTRODUCTION

The ability to learn from past situations and events is the key component of Intelligent behavior and also a natural ability for a person who learnt by trial and error. Most of the things that make sense in our world today do so because of what we have learned about them rather than the intrinsic properties in the things themselves. Learning from experience, is not natural for computer system but however, the ability must be carefully programmed into the computer system. Thus, it makes it an intelligent machine, which means a machine that chooses and form opinion from among choices and understands a phenomenon beyond ordinary. The process involved is called Natural Language Processing.

Yorùbá is one of the major languages spoken in Africa. Other languages in this category include Arabic, Hausa, and Swahili. In Nigeria, Yorùbá is one of the three major native languages (Hausa and Igbo) spoken alongside English, which is the official language. In Nigeria, the homeland of Yorùbá lies between longitudes 2^0 30'and $6^0$30' East of the Meridian and Latitudes 6^0 and 9^0 North of the Equator [1].
Yorùbá is the native language spoken in the Southern-West of Nigeria like Lagos, Oyo, Ogun, Ondo, Ekiti, and Osun states of Nigeria. It is also spoken in some part of Edo, and Kogi states of Nigeria as well as in Central Togo, East Central part of the Republic of Benin and in Sierra Leone (where it is called Aku). It has twenty-five (25) alphabet, eighteen (18) consonants and seven (7) vowels sound. This paper creates avenue for self learning of Yoruba language regardless of your language or ethic.

II. REVIEW OF RELATED RESEARCH

Generally, research has been done in Computer Assisted Language Learning (CALL) and also in Intelligent Computer-Assisted Language Learning (ICALL) [2, 3]. Morphological usage of CALL had also been dealt with by various researchers [4, 5, 6, 7, 9, 9 and 10]. This paper does not attempt to review CALL and ICALL generally, but add to the existing knowledge in the field of computer language learning.

In [11] the motivation and prerequisites of a successful integration of ICALL tools into current Foreign Language Teaching and Learning (FLTL) practice is presented. The authors focused on (i) the relationship between activity design and restrictions needed to make natural language processing tractable and reliable, and (ii) pedagogical considerations and the influence of activity design choices on the integration of ICALL systems into FLTL practice. We profited from their insights while focusing on the task of supporting the (re-) learning of Yoruba, a mother-tongue language in some parts of Africa.

The TAGARELA framework developed in [10] was employed by [12] to develop online ICALL exercises for Russian. The system aims to teach basic grammar to learners of Russian, and this involves audio and video exercises that enable the observation of language situations outside the classroom and life-like listening practice. The system is internet-based, facilitating learning anytime and anywhere. The exercises

have fixed content, thus limiting learners to the content the developer put in the exercise and there is no avenue to automatically add to the knowledge base.

Research told us that GLOSSER is an early system that extensively utilizes a morphological analyzer in language learning [13]. The major components of this system include a morphological analyzer for French, a part-of-speech disambiguation system, a bilingual dictionary, and aligned bilingual corpora. The system provided intelligent assistance to Dutch students learning to read French. The focus, however, is the learning of vocabulary that needs to be acquired separately from reading exercises.

In recent years there has been an explosion of interest in using computers for language teaching and learning. It is predicted that the future of CALL will heavily rely on the ability of learners and instructors to find, evaluate, and critically interpret net-based information [1]. The insights gained from these works informed our research with respect to the history and future directions of CALL.

In spite of some interest in Computer-Assisted Language Learning for African languages, it is evident that more research needs to be done. From the literature reviewed, the focus of our study remains different from other studies reported in the following ways:

a. We focus on Yoruba, a mother-tongue language and commonly taught language but not generally known.
b. We target "re-learners", including learners who have only basic, passive abilities in Yoruba language both in group and individual.
c. We provide exercises derived from a natural language processing system, unlike in other learning systems where a morphological analyzer is used to analyze the learners' answers [5], or as aid in providing morphological knowledge or dictionary access [13, 14]. We utilized the morphological analyzer to develop exercises for learning.

III. OVERVIEW OF DESIGN

The design involved in the course of this project includes Analysis of the input; Speech segmentation; Text Segmentation; Word sense disambiguation and Syntactic ambiguity.

a. Analysis of the input

The linguistic knowledge in this learning system is derived from a morphological analyzer of Yoruba, which was developed using Natural Language Processing (NLP) techniques [15]. NLP techniques have been identified as instrumental in developing pedagogically sound language learning applications [16] and computationally tractable applications [11]. The morphological analyzer of Yoruba specifically utilized Finite State Automata [17, 18].

b. Speech segmentation

In most spoken languages, the sounds representing successive letters blend into each other, in this paper Yoruba alphabets were converted into sounds with the three Yoruba accent (*do re mi*) taken into consideration. In natural speech there are hardly any pauses between successive words therefore the location of those boundaries usually must be taken into account grammatically and constraints, as well as the context. Each successive word is segmented into syllable and the corresponding speech for the syllable is blended with the next until the last syllable within the given word.

c. Text Segmentation

Without text segmentation, there cannot be speech segmentation. In this section, the principle of garbage-in-garbage-out is adopted. Any Yoruba word needs to be text-segmented into its connotative meaning for correct pronunciation. The use of caret (^) was adopted for letters with dot sign at its bottom for easy recognition. Yoruba accent must also be used alongside to show the intonation of the syllabic word.

d. Word sense disambiguation

Many words have more than one meaning; and therefore the word which makes the most connotative sense in context is selected out of the alternatives.

e. Syntactic ambiguity

The grammar for natural languages is ambiguous, and there are often multiple possible parse trees for a given sentence. Choosing the most appropriate one usually requires semantic and contextual information. Specific problem components of syntactic ambiguity include sentence boundary disambiguation. This approach is used when dealing with a word which has many connotations.

IV. FLOW CHART OF THE SYSTEM

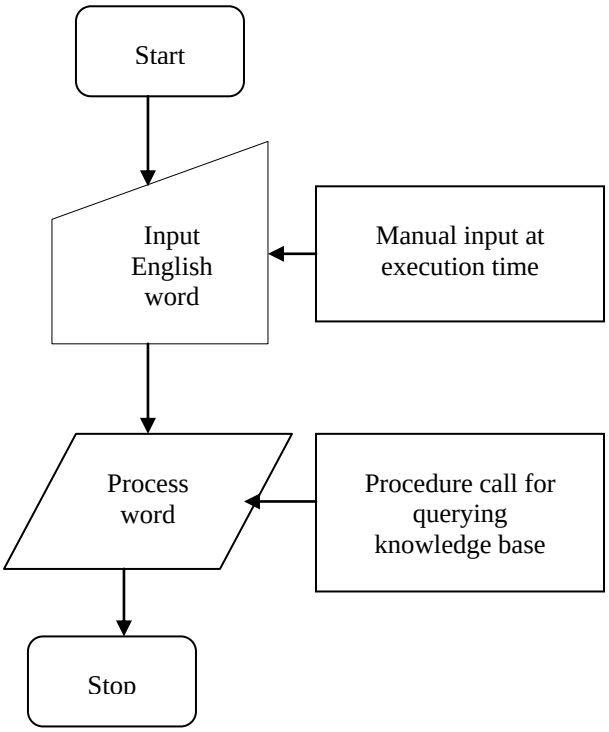

Figure 1. Input flow chart

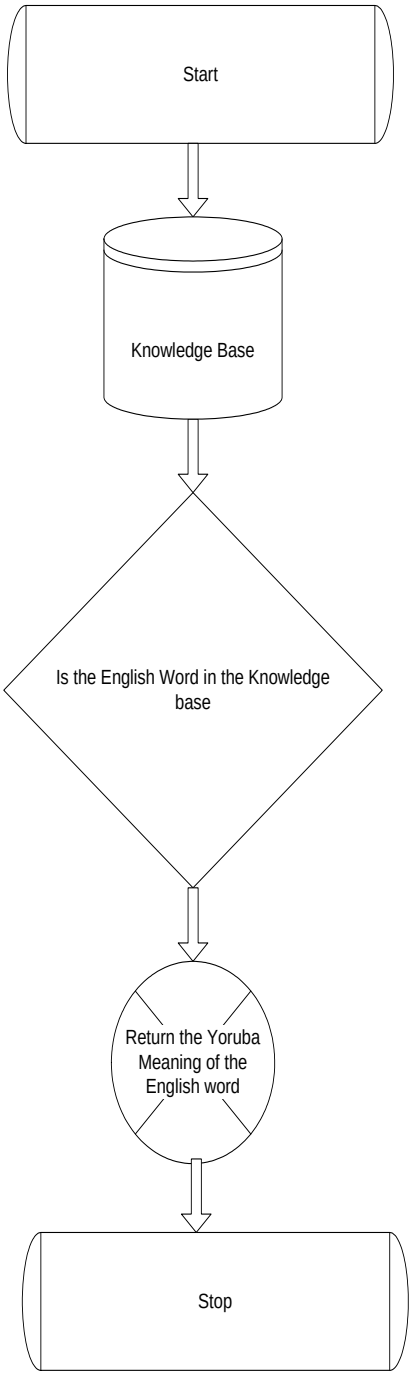

Figure 2: Query flow chart

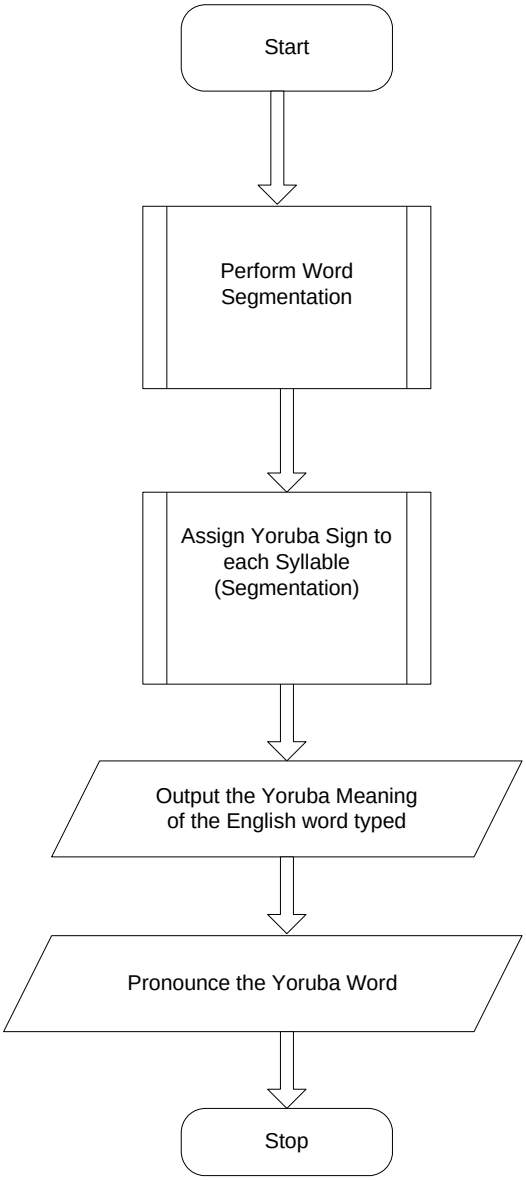

Figure 3. Output flow chart

V. DESCRIPTION OF THE SYSTEM

Intelligent System for Learning and Understanding of Yoruba Language (ISLUYL) is designed for a number of purposes. Some of which are: getting meaning to some English words in Yoruba language, updating Yoruba dictionary, allocation of

syllable sign and so on. Immediately the system is launched, it prompts the user for configuration in order to access and enjoy the software to its fullness.

a. Program modules

The program is divided into the following modules:

1. Meaning and learning module
2. Dictionary module
3. Update Dictionary module
3. Alphabet module
4. Word segmentation module

b. Meaning and learning module

This is the main module. It is inside this module that the user will be able to get Yoruba meaning to their typed word in English language. The supplied English-like word is translated into corresponding Yoruba meaning. User can listen to how the word is pronounced in Yoruba language, thereby helping the user to learn the word at a faster rate. The user can repeat after the system as much as possible by clicking ***Pronounce*** button. The pictorial view of this module is shown below.

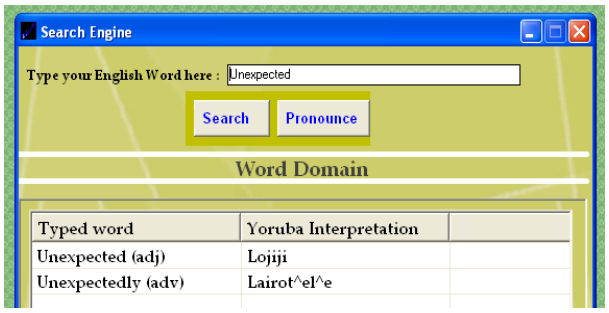

Figure 4: Translation interface

c. Dictionary module

Just like English dictionary, this is serving as mini-Yoruba dictionary where Yoruba meaning to the users' English word are being fetched and displayed to the user. This module is designed in such a way that the user can update the dictionary at any time, delete from the dictionary, edit or adjust the dictionary. The front view of this module is shown below.

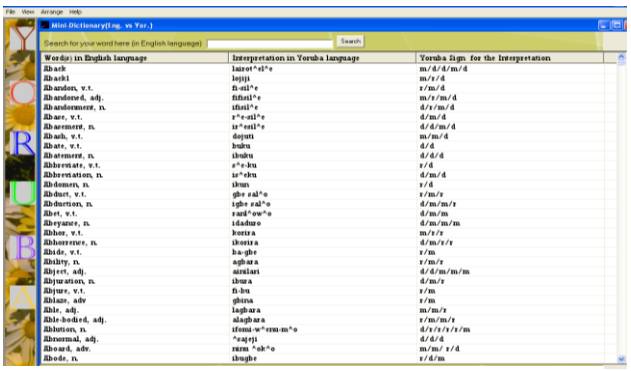

Figure 5. Dictionary interface

d. Update Dictionary Module

The novel of this work is displayed here. User can dynamically add words and its Yoruba meaning directly to the knowledge base and automatic speech for the word will be generated and can be tested for accuracy.

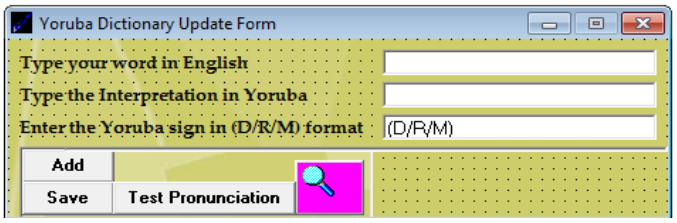

Figure 6. Dictionary interface

e. Alphabet module

In English language there are twenty-six alphabets with five vowels (a,e,i,o,u) and twenty-one consonants (b,c,d,f,g,h,j,k,l,m,n,p,q,r,s.t,v,w,x,y,z), likewise in Yoruba language, we have twenty-five alphabets with seven vowels sound (a,e,ẹ,i,o,ọ,u) and eighteen consonants. It is in this module that the user can learn how these alphabets are being pronounced. One interesting feature of this module is that the user can repeat the pronunciation of one alphabet at any convenient time. The interface is shown below

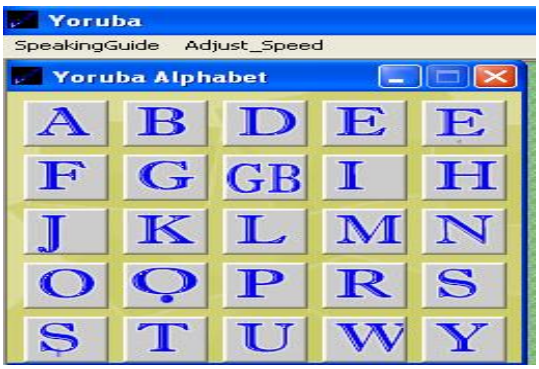

Figure 7. Yoruba Alphabet interface

f. Word segmentation module

This module performs morphological analysis of Yoruba words. In other words, it performs the breaking down of the Yoruba word into its simplest syllable(s).The interface for this module is shown below.

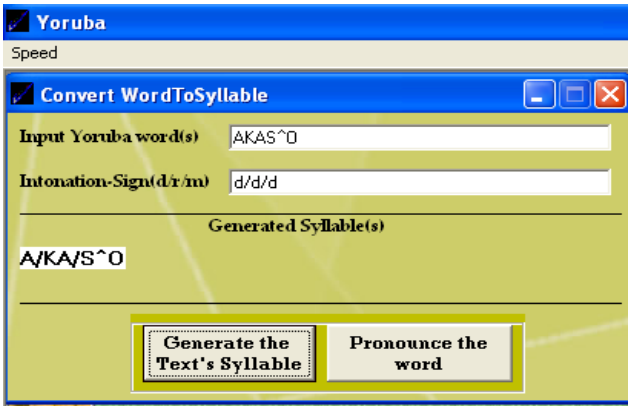

Figure 8. Segmentation interface

VI. CONCLUSION

In recent times, there has been call from stakeholders in the Nigerian educational sector on the need to adopt indigenous languages in teaching and learning. This call makes it imperative for teachers, learners and the society at large to evolve better and easier ways of learning indigenous languages most especially the three major ones namely Hausa, Igbo and Yoruba. In this paper, an intelligent system for learning and understanding of Yoruba language has been developed. Using personal computer, the system enables the user to learn at his/her own time and place. Further research is been carried out to make the system more portable by making it web-based.

References

[1] CIA World Factbook , 2001, http://www.cia.gov/cia/publucations/factbook.

[2] M. Warschauer and D. Healey, Computers and language learning: an overview. Language Teaching Forum, 31, 1998, pp 57-71.

[3] J. Gamper, and J. Knapp, A review of intelligent CALL systems. Computer Assisted Language Learning, 15(4), 2002, pp 329-342.

[4] G. Antoniadis. O. Kraif, T. Lebarbé, C. Ponton and S. Echinard. Modélisation de l'intégration de ressources TAL pour l'apprentissage des langues: La plateforme MIRTO. ALSIC. Apprentissage des Langues et Systèmes d'Information et de Communication, 8(2), 2005, pp 65-79.

[5] K. F. Shaalan, An intelligent computer assisted language learning system for Arabic learners. Computer-Assisted Language Learning, 18(1-2), 2005, pp 81-109.

[6] A. Blanchard, O. Kraif and C. Ponton, Mastering overdetection and underdetection in learner-answer processing: Simple techniques for analysis and diagnosis. Calico Journal, 26(3), 2009, pp 592-610.

[7] N. Nagata, Robo-Sensei. CALICO Journal, 26(3), 2009, pp 562-579.

[8] M. Dickinson, On morphological analysis for learner language, focusing on Russian. Research on Language & Computation 8(4), 2010, pp 273-298.

[9] O. Esit, Your verbal zone: An intelligent computer-assisted language learning program in support of Turkish learners. Computer Assisted Language Learning, 24(3), 2011, pp 211-232.

[10] L. Amaral and D. Meurers, Where does ICALL fit into foreign language teaching? Talk at CALICO. University of Hawaii, 2006, http://purl.org/net/icall/handouts/calico06-amaral-meurers.pdf

[11] L. Amaral and D. Meurers, On using intelligent computer-assisted language learning in real-life foreign language teaching and learning. ReCALL, 23(1), 2011, pp 4–24.

[12] M. Dickinson and J. Herring, Developing online ICALL exercises for Russian. Proceedings of the Third Workshop on Innovative Use of NLP for Building Educational Applications. 2008, pp 1-9. Columbus, Ohio.

[13] J. Nerbonne and D. Dokter, An intelligent word-based language learning assistant. Traitement Automatique des Langages. Special Issue on Multi-linguality, 40(1), 1999, pp 125-142.

[14] L. Amaral, Designing intelligent language tutoring systems: integrating natural language processing technology into foreign language teaching. 2007, Unpublished Doctoral dissertation, The Ohio State University, Columbus, Ohio, 2007.

[15] D. Jurafsky and J.H. Martin, Speech and language processing: An introduction to natural language processing, computational linguistics and speech recognition. 2nd ed. New Jersey: Prentice-Hall, 2009.

[16] J. Nerbonne, Computer-assisted language learning and natural language processing. In: R. Mitkov (Ed.) Handbook of computational linguistics (pp. 670-698). Oxford: Oxford University Press, 2002.

[17] K.R. Beesley and L. Karttunen, Finite state morphology. CSLI: Stanford, 2003.

[18] T. Hanneforth, fsm2 - a scripting language for creating weighted finite-state morphologies. In: C. Mahlow & M. Piotrowski (Eds.), State of the art in computational morphology (pp. 48-63). Heidelberg: Springer Berlin, 2009.