

Computational Morphological Analysis of Yorùbá Language Words

S. I. Eludiora, O. R. Ayemonisan

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

Article Info

Article history:

Received Nov 21, 2017

Revised Jan 10, 2018

Accepted Feb 12, 2018

Keyword:

Computational morphology

English language

Words formation

Yorùbá language

ABSTRACT

Nigeria official languages are English, Yorùbá, Igbo and Hausa. The focus of the study reported in this paper is to develop learning tool that can assist learners to learn the Yorùbá language using its alphabets. The study is critical to Yorùbá language, because of its endangerment. There is need to introduce different learning tools that can mitigate its extinction. A Yorùbá word perfect system was developed to assist people in learning the Yorùbá language. English and Yorùbá words formation are experimented using computational morphological approach (word formation). The theoretical framework considered Finite state automata (FSA) to realise different ways of combining the consonants and vowels to form word. Two to five letter words were considered. The system was designed and implemented using UML tools and python programming language. The system will teach the users on how the words are formed, and the number of syllables in each word. The user need not to know how to tone mark word before he/she can use the system. Any word typed will be analysed according to its number of syllables. This approach produces representatives of all parts of speech (POS) of the two languages. It produces corpora for the two languages.

Copyright © 2018 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

S. I. Eludiora,
Department of Computer Science and Engineering,
Obafemi Awolowo University,
Ile-Ife, Nigeria.
Email: eludioramolola@gmail.com

1. INTRODUCTION

Firstly, can other African languages be learnt using this approach? In a way can Igbo and Hausa (the other two Nigerian languages) be learnt using this approach? Yorùbá, Igbo and Hausa are tonal languages and possibly some other African languages. Secondly, can other world languages be learnt using this approach? Morphological analysis is the first step in many natural language processing tasks. Tasks such as parsing, machine translation, information retrieval and part of speech tagging, among others [1].

Morphology is the study of the internal structure of the word. Morphological analysis can be used to retrieve the grammatical features and properties of a morphologically inflected word [2]. It is the process of segmenting words into morphemes and analysing the word formation. It is a primary step for various types of text analysis of any language [2].

As noted in [3], morphology studies the internal structure of words. The building blocks are called morphemes. One distinguishes between free and bound morphemes. Free morphemes are those which can stand alone as word. Bound morphemes are those that always have to attach to other morphemes (Söhn, 2008). Morph or Morpheme is a minimal distinctive unit of grammar. E.g., a word like unselfish (*ànífétaraenikan*) has three morphemes in the English language. They are: un-, self and -ish. In Yorùbá, it has nine morphemes: *à + ì + ní + ifé + tí + ara + eni + nì + kan* which can be segmented [4].

An allomorph is a possible way of realizing a morpheme. For example, the plural morpheme in English is realised by the allomorphs –es, -ies, -s. E.g., box (boxes), fly (flies) and book (books) [3]. The morpheme which expresses plurality in English, for instance, appears in several variants: cap/caps, log/logs, force/forces, mouse/mice, sheep/sheep, etc. Two of these variants form the voiceless [s] of caps, the voiced [z] of logs. The irregular shape of mice could be said to be an allomorph of the plural morpheme and the phenomenon is called allomorphs. The plural morphemes in Yorùbá are not expressed in this manner. For example, pupil and pupils; pupil means *omọ ilé-èkọ* and pupils means *àwọn omọ ilé-èkọ*.

Word formation in the two languages follows a simple process. In English, vowel can attach with a consonant to form a word, for example, in, on, of (mostly prepositions), etc. In Yorùbá, consonant can attach with a vowel to form a word or words (according to the variation of the tone marks usually verb), this word is usually a root word of a whole class of words, for example, *dé* cover, *gé* cut, *kà* read etc. Pre-fixing a vowel to these words, a noun is formed from the same root word, for example, *dé* → cover, *a + dé* → crown, *lù* → beat, *ì-lù* → drum [4].

In [5], extensive work on Yorùbá word syllabication had been done. The software developed can be used for tone marking and under-dotting of Yorùbá words. Our concern in this paper is to develop a tool that can simplify the word formation and how the syllable can be identified in a word. The tool can be used for other related languages within Africa and other parts of the world.

The remaining part of the paper is organized as follows: section 2 examines related works; section 3 gives the theoretical framework; section 3 section 4 discusses the results, while section 5 concludes the paper.

2. RELATED WORKS

In [5], is of opinion that if urgent steps are not taken by the stakeholders that Yorùbá Language will be endangered. He examined low-usage of the language in some states in South West of Nigeria. The findings show that the students find it difficult to communicate with the people using the Yorùbá language.

In [6], examined the factors or variables that are responsible for Yorùbá language endangerment. The results of the study show lack of commitment to indigenous language; habitat displacement, colonial legacy, and defective language planning are responsible to the gradual extinction. The study concluded that the Yorùbá language should be used at home, and it should be a criteria for post primary school admission.

In [7], laid down a basic way of learning the Yorùbá language. In the book the author explain different types of Yorùbá vowels and their features. The author discussed phonology and morphology of Yorùbá language. The author explained how words are formed. The book is informative and it will help learners.

In [8], examined the contributions of mother tongue in pre-Nursery or primary early childhood education. The study raised six issues: language background, situation in Nigeria, policy documentation on language in Nigeria, language theory and development, and problem statement and rationale. Success and failures of mother tongue usage in the country needs to be reviewed in order to address issues mentioned.

In [9], was of opinion that proper implementation of educational policies in Nigeria will increase the learning of indigenous languages in particular Yorùbá language.

In [10], presented that one of Nigerian educational policies stipulated it that “the first three years of primary education should be taught in learners’ mother tongue. The study conducted show that primary school teachers used were able to teach the subjects (like social studies), but with challenges.

In [11], examined which of the Yorùbá language (mother tongue learners) as a medium of instructions and English language as a medium of instructions to teach social studies in the nursery school would make the pupils perform excellently well. The experiment was carried out and the results show that the pupils taught with their mother tongue (Yorùbá language) perform better than those that were taught with English language.

In [12], learning is defined as a change in knowledge attributable to experience. Learning involves a change in the learner, learner’s knowledge and cause of the change is the learner’s experience. Learning is not measured through one operational definition. Rather, learning is a blend of comprehension, transfer of new material, and the retention of material. In fact, most transfer studies focus purely on the similarities and differences between the contexts of initial learning and subsequent transfer [13]. Given the current study, learning has been evaluated using a multimedia device.

In [13], proposed the use of E-learning approach to teach and learn the Yorùbá language. He opined that the Information and communication technology (ICT) is a good tool to increase the people’s interest in learning the language.

In [14], developed a windows mobile application for learning Yorùbá language. The learner can learn how to read alphabets, numbers and common words in Yorùbá language.

3. THEORETICAL FRAMEWORK

Finite state automata (FSA) technique was used to analysed different ways of forming words from the two languages' alphabets. The FSA was used for two to five letter words, while some notable examples of single letter words were mentioned. In subsequent subsections, they will be discussed in details.

3.1. Single letter words

In English language, 'I' (pronoun) and 'a' (determiner) are the single letter words.. In Yorùbá language it is mostly pronouns. Example:

'Ó' (3rd person singular, s/he, and it) and 'A' (we).

3.2. Two letter words

Two letter words (KF) in Yorùbá language are verbs and pronouns, and single syllable. In English language the two letter words can be CV (he, me, we, to, etc), and VC (an, on, in, it, of, etc). Figure 1 shows how two letters words are formed using the finite state automata. Table 1 shows all possible Yorùbá words that can be generated from Figure 1. The possible combinations of KF or CV are not sensible semantically in some cases, but syntactically they are sensible. The focus here is to address issue of syllable that will give meanings to all these possible combinations. Also in Figure 1, three different Yorùbá language tone marks are shown; high tone (H: á), low tone (L: à) and mid-tone has no symbolic representation (a). The application of tone marks on vowels are used to resolve possible ambiguities in some words. If the tone marks cannot be used to distinguish the words, then such words can be distinguished contextually. A sensible two letter words can only form three different words maximally (*bá*, *bà* and *ba*), each word may have different meanings. For example, '*bà*' is an ambiguous word. It means touch down or ferments.

Pure syllabic vowels (*m* and *n*) in Yorùbá language can be tone marked within a sentence or a phrase. In most cases they take high tone, and they are single letter words. For example: *m̀ b̀* and *ǹ l̀*. Also, consonant *Ṣṣ* is the only one that has under dot. In most case it is used to stress word (e.g., *ṣṣifìṣṣàn* - television). The alphabet *GB gb* is a combination of two letters (also termed as digraph) is consonant not two consonants (*gbé*, *gbà*, *gbò*, etc).

Figure 1 explains possible ways of forming two letter words from both languages. The Yorùbá has one way of forming two letter words. That is, Konsonanti (K) + Faweli (F) => KF. The English language has two forms. That is, consonants (C) + vowels (V) => CV and VC. Examples are shown in Table 2.

Table 1. Yorùbá language two letter words

K + F	á é ì í ó ò ú	à è è ì ò ò ù	a e e i o o u (no tone marks)
Bb			
B + a	bá	Bà	ba
B + e	Bé	Bè	Be
B + e	Bé	Bè	Be
B + i	Bí	Bì	Bi
B + o	Bó	Bò	Bo
B + o	Bó	Bò	Bo
B + u	Bú	Bù	Bu

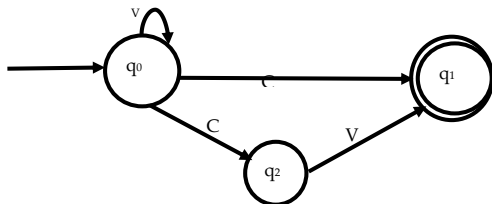


Figure 1. FSA state diagram for English and Yorùbá two letter words

Table 2. English Language two letter words

CV	POS	VC	POS
He	PRN	if	Conjunction
to	PRN	on	Preposition
be	Verb	an	Determiner
me	PRN	it	PRN
we	PRN	in	Preposition
go	Verb	of	Preposition
do	Verb	us	PRN

3.3. Three letter words

In Figure 2, the state diagram of three letter words for English and Yorùbá languages is shown. For the Yorùbá language there are possible three combinations.

The first scenario (FKF) is shown in Table 3 (possible combinations of F, K, and F). It means that seven vowels can be combined with consonants prefix and postfix. The tone marks and under-dots produce different words. This is A + b + other vowels.

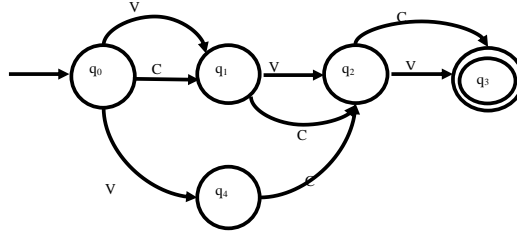


Figure 2. FSA State Diagram of English and Yorùbá Three Letter Words

Table 3. Scenario 1: F + KF

F + KF	Yorùbá word	Syllable	POS	English word	POS
A + ba	Àba	2	verb	to incubate	Verb
A + ba	Abà	2	noun	Hamlet	noun
A + ba	Àbà	2	noun	Ladder	noun
A + be	Àbe	2	noun	Village name	noun
A + be	Abẹ	2	noun	blade	noun
A + bi	Àbí	2		Isn't it	
A + bo	Abo	2	noun	Female	noun
A + bo	Àbo	2	noun		noun

In scenario 2 (KFF), KFF produces one word as shown in Table 4. Scenario 3 (KFK) as show in Table 5, gives the combination of KFK i.e. K + FK. The FK can represent the five nasal vowels mentioned in the previous section.

Table 4. Scenario 2: KF+F

KF + F	Yorùbá word	Syllable	POS	English word	POS
na + a	Náà	2	AIỌO	The	Det

Table 5. Scenario III: K+FK

KFK	Yorùbá word	Syllable	POS	English word	POS
K + an	kán	2	V	brake	V
K + an	kàn	2	V	knock	V
K + an	kan	2	V	sower	V
Y + ẹn	yẹn	2	article	that	article
G + un	gún	2	V	Stab	V
Y + an	yán	2	V	Yarn	V

Table 6 shows different combination of consonants and vowels forming three letter words in English language. The combinations are: CCC, CVC, CCV, VVC, CVV and VCV. There are different words formed.

Table 6. Three letter words formation in English language

Combinations	English word	POS
CCC	Fly	noun
CVC	Low	Adjective
VCC	Egg	noun
CCV	The	Determiner
VVC	Oil	noun
CVV	See	verb
VCV	Use	verb

3.4. Four letter words

Figure 3 depicts the English and Yorùbá language four letter words formation. KFKF and FKFK are two possible combinations for Yorùbá four letter words as shown in Table 7 and 8. These are sample words from possible words that can be generated.

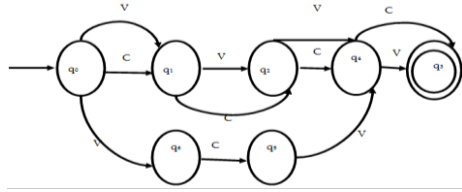


Figure 3. FSA state diagram for English and Yorùbá languages four letter words

Table 7. Scenario 1: KFKF

KF+KF	Yorùbá word	Syllable	POS	English word	POS
da + ra	dára	2	Adj	Good	Adj
pa + de	pádé	2	Verb	to close	Verb
pa + de	pàdé	2	Verb	to meet	Verb
pa + da	padà	2	Verb	to return	Verb
Ba + ba	Baba	2	Noun	Father	Noun

There are eight (8) combinations of English four letter words as shown in Table 9. The combinations are: CVCV, CVVC, CVCC, VCVV, VCVC, CCVC, VCCV, and VVCC.

Table 8. Scenario 2: F + KFK

F + KFK	Yorùbá word	Syllable	POS	English word	POS
E + gbon	Ègbón	2	Noun	brother	noun
E + gbin	Ègbin	2	Noun	animal	noun
E + dun	Èdun	2	Noun	animal	noun
a + òdín	àdín	2	Noun	Palm kernel oil	noun
i + yèn	iyen	2	Det	that	det
A + kan	Akàn	2	Noun	Crab	noun
ẹ + hìn	ẹhìn	2	Noun	back	noun
a + han	ahán	2	Noun	tongue	noun
ọ + kan	ọkàn	2	Noun	heart	noun
ẹ + òdún	ẹdun	2	Noun	animal	noun
i + òdún	ìdun	2	Noun	Bed bug	noun

Table 9. English four letter words

Combinations	English word	POS
CVCV	Zone	noun
	Base	noun
	Hake	verb
CVVC	Beak	noun
	quin	noun
	joey	noun
CVCC	tell	verb
VCVV	Aqua	noun
VCVC	Epic	verb
CCVC	Clap	verb
VCCV	Abba	noun
VVCC	Oink	noun

3.5. Five letter words

Five letter words pattern is shown in Figure 4 for the two languages. There are two possible ways of forming five letter words in Yorùbá language. It can be F + KF + KF as depicted in table 10 and F + F + KFK as depicted in Table 11.

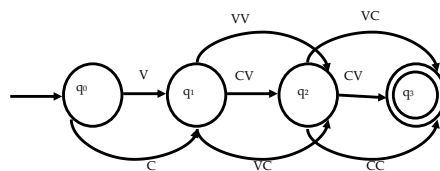


Figure 4. Pattern analysis of five English and Yorùbá letter words

Table 10. Scenario 1: F + KF + KF

Combinations	Yorùbá word	Syllable	POS	English Equivalent	POS
F + KF + KF					
Ì + sẹ + dá	Ìsẹdá	3	verb	create	verb
Ò + ò + dọ	Òòdọ	3	noun	flower	noun
À + kọ + kọ	Àkókọ	3	noun	time	noun

Table 11. Scenario 2: F + F + KFK

Combinations	Yorùbá word	Syllable	English Equivalent	POS
F + F + KFK				
O + ò + rùn	Oòrùn	3	sun	noun
Ò + ó + rùn	Òórùn	3	odour	noun

There are seven possible combinations for English language five letter word as shown in Table 12.

Table 12. English language five letter word

Combinations	English word	POS
CCVCV	Stone	noun
CVCVC	Widow	noun
CVVCC	Round	noun
CCVVC	Broom	noun
CCVCC	Chest	noun
CVCCV	Title	noun
VCVVC	Ocean	noun

4. CONCLUSION

4.1. System framework design

The system framework design covers the system database, and software design. Figure 5 depicted the system activity. In the theoretical framework section, two to five letter words were analysed for the two languages. The feature of each letter word was discussed in relation to the expected number of tones on each word. These tones determine the number of syllable formed. Morphemes are different from words as it was explained in the previous sections. According to Figure 5, the system determines whether a word typed is a Yorùbá or not. If it is a Yorùbá word, the system will check for the number alphabets to know if it is two, three, four or five letter words. The system will compare the word with the words in the database. If it matches, the system displays the possible number words that such a word can be represented. The user needs not to tone mark a word while typing the word. The system will analyse the word to determine how many possible words can be formed from that single word. It will provide different possible words that can be formed. For example; if *igba* is typed, after the analysis, the system will display five different words which are *igba* (200), *ìgbà* (season), *ìgbá* (garden egg), *ìgbà* (rope for climbing palm tree), *ìgbá* (calabash). The tone marks determine the syllables; in this case there are two syllables *i-gba*, it means the tone marks can be on the two vowels (i or a). These tone marks determine how words are pronounced (phonology). The Yorùbá orthography (writing styles) depends on this tone marks to make meaning out of a word. It might not be critical in speeches but critical in texts. It is problematic when reader is fixing meaning from the context. However, such a meaning may deviate from actual intended meaning. The system will displays the the total number syllables, the parts of speech (POS) and the English language equivalent. The system activities diagram showing the various actions been performed by the system is shown in Figure 5.

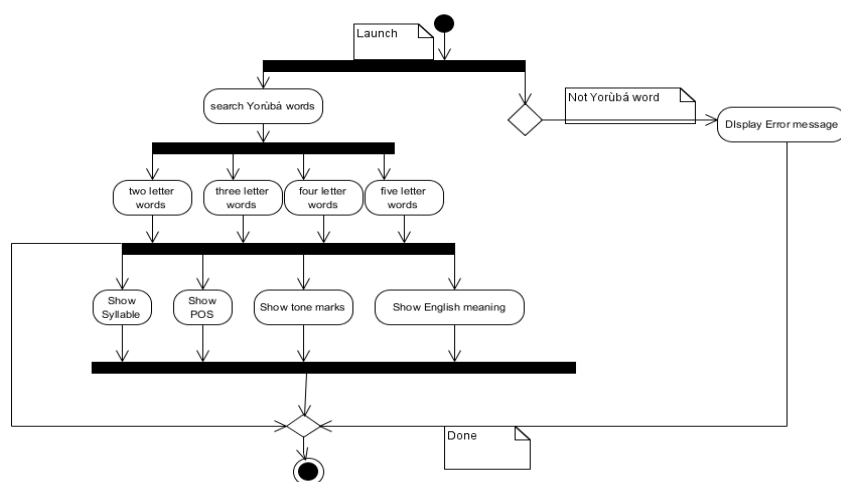


Figure 5. The System Activities Diagram

4.2. Database design

The database design is based on the theoretical analysis done in the previous section. The structure of the database is different from the tables presented in the previous section. The database consists of words from two letter words to five letter words. The words were manually tone marked. Three things were considered in the database: the Yorùbá word (tone marked), syllable, its equivalent in English language, and POS. Each letter word was separately designed for easy access. The system compare every word typed by the user with the word

in the database. The database was designed to accept new words, but must be vetted by the system administrator.

4.3. Software design and implementation

Figure 6 shows how different modules relate with each other. There are four modules: *startpage*, *selectview*, *analyse*, and *selectsearch*. The *startpage* coordinates other modules. The *selectview* displays the word’s attributes. It displays the tone marked word, syllable, and POS. The *selectserch* is the database that the startpage can access. SQLite was used to design the database. The code was implemented using python programming language. The system class diagram is shown in Figure 6.

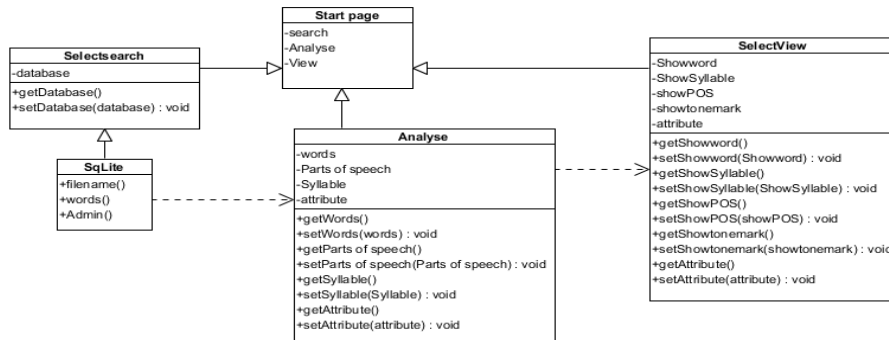


Figure 6. The System Class Diagram

5. RESULTS AND DISCUSSION

The system implementation considered all the modules and units to implement the whole system. Figure 7 depicts the Yorùbá word formation system. The system has user plane, where user can type a word. Below the user plane are: analyse, reset, and close buttons. The user can analyse the word by clicking the analyse button. The user can reset and type new word. Figures 8 show the system sample outputs.

The GUI displays the results of word analysed, it displays the number of syllables, Yorùbá language tone marked words, the equivalent words in English language, and the POS.

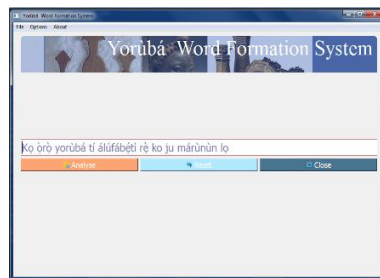


Figure 7. The System Graphical User Interface (GUI)

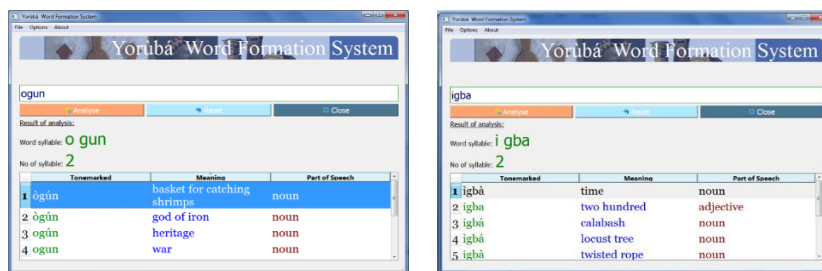


Figure 8. Sample Output of the System

6. CONCLUSION

There were many things considered in this study and there are many ways of using them. The alphabets combination can provide reasonable size of corpora for the English and Yorùbá languages which can be used for machine translation. The FSA state diagram can be used for other language pair to see whether it is suitable. The final application can be used by Yorùbá Teachers at any level of education. This is similar to (but detailed than) Nursery/Primary English word perfect. In future, we will make it a multimedia system. Text, pictures, sounds will be included.

REFERENCES

- [1] K. Chege, P.W., Wagacha, G. De Pauw, and J. Mutiga, “*Morphological Analysis of Gikuyu Using a Finite State Machine*” In Proceedings of Conference on Human Language Technology for Development, Egypt, pp. 112, 2011, URL: <http://www.hltd.org/pdf/HLTD201119.pdf>, Accessed: (23/11/2011).
- [2] M. A. Kumar, V. Dhanalakshmi, K.P Soman, and S. Rajendran, “A Sequence Labeling Approach to Morphological Analyzer for Tamil Language”, *International Journal on Computer Science and Engineering*, Vol. 02, No. 06, pp. 1944-1951, 2010.
- [3] J.P. Sohn, *Introduction to Computational Linguistics Morphological Analysis*, 2008, URL: <http://www.soehn.net/work/icl/morph.pdf>, Accessed: (23/11/2011).
- [4] O. Awobuluyi, *Èkó Ìsẹ̀dà Ọ̀rọ̀ Yorùbá*, Akure: Montem Paperbacks, 2008.
- [5] T.A. Balogun, “An Endanger Nigerian Indigenous Language: The case of Yoruba Language”, African Nebula issue 6, pp 70-82, 2013, URL: http://nobleworld.biz/images/6-Balogun_s_Paper.pdf, Accessed: (09/24/2016).
- [6] O.I. Aladesote, O.V. Johnson, and O. Agbelusi, “Selection of Factors Responsible for Yoruba Language Extinction using Feature Extraction Techniques”, *Pyrex Journals of Educational Research Reviews*, Vol 2(1), pp 001-005, 2016, URL: <http://www.pyrexjournals.org/pjerr/pdf/2016/january/aladesote-et-al.pdf>, Accessed: (09/24/2016).
- [7] O.O. Adébùsólá, (online) *Yorùbá Mọ̀nńkọ̀ Mọ̀nńka Mọ̀nńsọ̀ (Know how to write it, read it, and speak it)*, URL: <http://www.learnyoruba.com/Downloads/YorubaPrimer.pdf>, Accessed: (15/12/2016).
- [8] B.G. Abidogun, and O.I. Adebule, “*Contributions of Mother Tongue Education in early childhood Education*”, First Annual International Interdisciplinary Conference, Portugal Proceedings, pp 267-272, 2013, URL: <http://eujournal.org/files/journals/1/articles/1337/public/1337-4340-1-PB.pdf>, Accessed: (09/24/2016).
- [9] J.A. Abijo, “Emerging Trends in the use of Mother-Tongue as a Language of Instruction in Lower Nigerian Primary Schools”, *Journal of Emerging Trends in Educational Research and Policy studies*, Vol 5 (8) pp 124-127, 2014.
- [10] B.G. Abidogun, B.G. “Teachers Experiences of Using Yoruba as a Medium of Instruction in Primary Classes: Implications for Learning”, Unpublished Ph.D Thesis, University of Pretoria, South Africa, pp 1-229, 2012, URL: [http://repository.up.ac.za/bitstream/handle/2263/31910/Abidogun_Teachers\(2013\).pdf?sequence=1](http://repository.up.ac.za/bitstream/handle/2263/31910/Abidogun_Teachers(2013).pdf?sequence=1), Accessed: (09/24/26).
- [11] O.A. Oribabor, and A.D. Adesina, “Mother Tongue Instruction and Academic Achievement of Pupils in Nursery Schools”, *International Research Journals of Arts and Social Sciences*, Vol 2(5), pp131-133, 2013, URL: <http://www.interestjournals.org/full-articles/mother-tongue-instruction-and-academic-achievement-of-pupils-in-nursery-schools.pdf?view=inline>, Accessed: (09/24/2016).
- [12] R.E. Mayer, *Applying the science of learning*. Boston, MA: Pearson, 2011.
- [13] A.C. Butler, “Repeated Testing Produces Superior Transfer of Learning Relative to Repeat Studying”, *Journal of Experimental Psychology*, 36(5), 1118-1133, 2010.
- [14] A.A. Oyekan, “*E-learning Solutions for Language Teaching and Learning: Yoruba Language*”, International Conference “ICT for Language Learning”, 2013, URL: http://conference.pixel-online.net/ICT4LL2013/common/download/Abstract_pdf/150-ELE12-ABS-Abiodun-ICT2013.pdf, Accessed: (09/24/2016).
- [15] I.A. Adeyanju, O.I. Omotosho, and B.M. Fatunde, “Development of a Yoruba Language E-Tutor for Windows Phone”, *African Journal of Computing and ICT* Vol 8 No. 3, pp 135-140, 2015, URL: http://www.ajocict.net/uploads/V8N3P15-2015_AJOCICT.pdf, Accessed: (09/24/2016).
- [16] A. Adegoju, “Language Teaching, Learning and Utility: A Triadic Paradigm for Revitalising Indigenous Nigerian Languages”, issue 27, pp 1-11, 2009, URL: https://www.researchgate.net/publication/237777896_Language_Teaching_Learning_and_Utility_A_Triadic_Paradigm_for_Revitalising_Indigenous_Nigerian_Languages, Accessed: (09/24/2016).
- [17] F.O.Kumolalo, E.R. Adagunodo, and O.A. Odejobi, “Development of a Syllabicator for Yoruba Language”, 2010, URL: http://ifecisrg.org/sites/default/files/articles/SYoruba_%20syllabicator_4.pdf, Accessed: (12/15/2016).