American Journal of Engineering Research (AJER)
e-ISSN: 2320-0847 p-ISSN : 2320-0936
Volume-5, Issue-6, pp-08-12
www.ajer.org
Research Paper Open Access

Morphological Analysis of Standard Yorùbá Nouns

Agbeyangi, A. O.[1], Odiete, J. O.[1], Lawal, T. A.[2], &Olorunlomerue, A. B.[2]
[1]*Department of Computer Engineering, MoshoodAbiola Polytechnic, Abeokuta, Ogun State-Nigeria*
[2]*Department of Computer Science and Engineering, ObafemiAwolowo University, Ile-Ife, Osun State-Nigeria*

ABSTRACT: *Computational morphological analysis is an important step in the natural language processing and it has always been a challenge for computational linguistics because of its richness and complexity. This paper presents the development a morphological analyser that generates standard Yorùbá nouns by affixing Yorùbá vowels with monosyllabic verbs. Data were collected in conformity with the rules. Computational model was designed for the system based on the knowledge obtained using Finite State Automata (FSA) technique and the system was implemented using python programming language.The results gathered from expert opinion shows that the system performs as expected owing to the size of the corpus used. Other areas noted for further study includes; morphology analyzer for other Yorùbá language part of speech (verb, preposition, etc).*

Keywords: *morphological analysis, nouns, Yorùbá language*

I. INTRODUCTION

The word 'morphology' means the study of forms or structures. It can be applied to different subjects. For example, in Biology, it is the study of forms and structures in organisms. Also, in Geology it is the study of land forms and their configuration and evolution [1]. In linguistics, morphology *is the study of the form and structure of words*. It is achieved by breaking down the word into the smallest constituents possible, these 'pieces' are then compared to other constituents from other words and their relationship is evaluated. The constituents that the words are broken down to, is known as 'morphemes'. Each morpheme has a grammatical function.

There are numerous rules involved in morphological analysis. Thus, creating a system which can successfully perform morphological analysis involve some sort of artificial intelligence. There are many instances in morphological analysis which a system find ambiguous, but human are able to distinguish easily. Most of these issues involve the words' affixes, and rules which are needed to be programmed into the system so that it will break up and analyze the word(s) correctly.

There are two basic approaches to morphological analysis: the analytical and synthetic approach. Analytical approach starts with the whole word being broken down into morphemes. This approach is used by the linguistics when examining a foreign language they are not familiar with. The Synthetic approach starts with the morphemes and attempt to put them together (usually through trial and error) to form the word [2].

This research work is based on synthetic approach, because it extends the principle of the first approach by saying that instances of the same morpheme may not necessarily be spelt the same, or pronounced the same. For example, *igbá* (calabash) and *ìgbá* (garden egg) are not pronounced with the same tone.

Morphology system is the backbone of a natural language processing system and no application in the field can survive without a good morphology system to support it [3]. Yorùbá language has its own features that are not found in other languages; reason why many researchers working on the language has to critically study this. Yorùbá language has being classified among the unwritten African languages comprises several dialects spoken by over 40 million people, mainly in West Africa [4]. In Nigerian, the language is been spoken in Lagos, Osun, Ogun, Ondo, Oyo, Ekiti and Kwara, as well as some part of Kogi state. Yorùbá is also one of the major languages spoken in Benin Republic and it is also spoken in some parts of the Republic of Togo [4].

The objective of the study is basically to design and implement a model for the generation of standard Yorùbá Nouns. In achieving this, there is need to analyze the structure of standard Yorùbá monosyllabic verbs affixed with some Yorùbá vowels, design a computational Finite State Automata (FSA) model and implement the model designed. The work is also restricted to Yorùbá noun generated by affixing with monosyllabic verbs.

II. THE BASIC MORPHOLOGY IN YORÙBÁ LANGUAGE

Morphology is the branch of linguistics that studies the internal structure of words and how words are formed in a language [5]. It accounts for word formation in languages. The basic unit of analysis in morphology is called the 'morpheme'. A morpheme is defined as the minimal meaningful unit of grammatical analysis, that is, a meaningful sequence of sounds which is not divisible into smaller meaningful unit [5]. E.g., the word *àdá* 'cutlass' can be broken into two morphemes i.e {*à*} and {*dá*}, morpheme {*à*} indicates an agent, one who does a thing while morpheme {*dá*} means 'break'. The main word or the *root morpheme* is '*dá'*, it is called a free morpheme because it can stand on its own and meaningful. The other parts of the word '*à'* that is attached as the prefix of the root word '*dá'* cannot stand on its own and do not have meaning. It is referred to as *bound morphemes;* they are meaningful only when affixed to the main word or the root morphemes [2].

2.1 Nouns in Yorùbá Language

Nouns in Yorùbá language generally in their simplest form are formed by prefixing a vowel to monosyllabic verbs.E.g. *dá* (to break); *àdá* (cutlass); *ṣè (*to offend); *ẹ̀ṣè* (sin); *ké* (to cut); *aké* (axe).Also, the verbals*àlọ* (going); *àbọ̀* (coming); from *lọ* (to go); and *bọ̀* (to come). These prefixes have certain peculiarities of their own i.e. *a* prefixed indicates an agent who does a thing.

The vowel '*o'* and '*ọ'* plays the same role as '*a'*, but limited in their use e.g. *dẹ* (to hunt); *ọ* + *dẹ→ọdẹ (*hunter). Vowel '*ẹ'* also denotes a noun in the concrete as '*rù'* (to carry); *ẹ* + *rù→ẹrù* (load) and '*i*' prefixed denotes a noun in the abstract as '*rí'* (to see); *ì* + *rírí→ìrírí* (experience). According to [3], vowels '*e'* and '*u'* are rarely used.

III. MATERIAL AND METHOD

This section provides a description of the software system architecture and techniques used to achieve the objectives of this paper. It also focuses on more technical details of the technologies involved in building the modules of the software.

3.1 Requirement Analysis and Specification

In order to address the problem definition of this research work, it is necessary to identify the requirements of the system and the system specification.

3.1.1 Requirement Analysis

The requirements analyses of the system are as follow:

i. to present a friendly local interface to the user;
ii. the system is built to work on a windows operating system;
iii. to give the user the power to enter Yorùbá vowel as a prefix;
iv. generate the nouns that can combine with the vowel entered;
v. allow user to quit or reset the system at any time.

The system will run on a Personal Computer (PC). No specific amount of free hard disk space or available memory is set, but it is recommended that the PC have a free hard disk space and available memory required to run an application.

3.2 Detailed Design

Corpus was generated to be used for our analysis. This is done by affixing the Yorùbá vowels with monosyllabic verbs to generate nouns. Some of the samples are shown in table 1 below;

Table 1: Examples of nouns derived from monosyllabic verbs

Vowel (Affix)	Verb (CV)	Noun (VCV)
a	bà bọ́ bẹ dé	abà abọ́ abẹ adé
e	wé yín fín	ewé eyín efín
ẹ	bí bọ dá	ẹbí ẹbọ ẹdá

ẹ̀	bẹ̀ dá fọ́	ẹ̀bẹ̀ ẹ̀dá ẹ̀fọ́
ì	bò dè dí	ìbò ìdè ìdí
ò	fì fò gbè	òfì òfò ògbè
o	bì gbó rí	obì ogbó orí
ọ	bẹ̀ dẹ fà	ọbẹ̀ ọdẹ ọfà
ọ̀	bẹ dẹ̀ fẹ́	ọ̀bẹ ọ̀dẹ̀ ọ̀fẹ́

FSA (Finite State Automaton)was used to model how a string will be accepted or rejectedin the language (Yorùbá). It uses regular expressions. For example, the FSA drawn in Figure1 below accept the word *àdá*(cutlass).

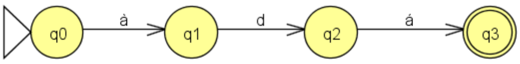

Figure 1: FSA for the word "àdá" (Cutlass)

It has 5-tuple. M = (Q, Σ, δ, S_0, S_f) where:

- Q = (q_0, q_1, q_2, q_3, q_4,............q_n) is a set of states that the FSA can be in;
- Σ = Σ_c ᴜ Σv ᴜ Σn is the set of SY alphabets comprising $\Sigma_c$andΣv, the sets of the consonants and vowels, respectively as well as Σn is the set of syllabic nasal;
- δ : Q x Σ →Q is a mapping of states onto input alphabet to produce another state;
- q_0 ϵ Q is the initial state of the FSA;
- q_fϵQ is the set of acceptance or final states.

Figure 2 shows the FSA architectural model of Yorùbá Nouns, themodel explains how each state transit to the other. Figure3 also depicts the generation of different nouns from a single vowel "a" following the model developed.
In the model in Figure 2, "c" represent consonant alphabet while "v" represent a vowel. For example the word ọba (king) which is a noun generated from the vowel letter "ọ" combining with the monosyllabic verb "ba", has a VCV pattern (V = ọ, C = b, V = a).

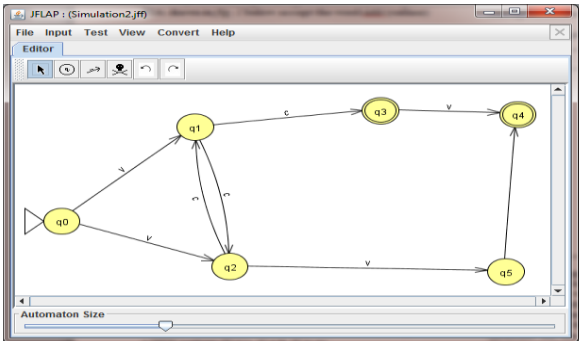

Figure 2: Automaton Transition Diagram for Noun

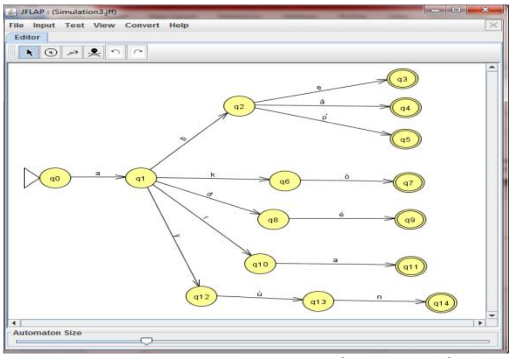

Figure 3:FSA transition diagram for nouns *(abẹ́, abà, abọ́, akò, adé, ara, ayùn)*

In Figure 4 below the generated noun "*ayùn"* was tested to see if it was accepted by the automaton designed.

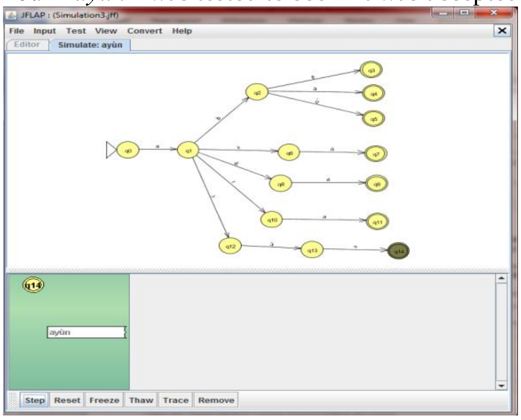

Figure 4: FSA showing the simulation of the noun "*ayùn"*

3.3 System Design

The system or application design layout is shown in Figure 5 below. In the design layout, the user is the person who will use the system, the window interface (application GUI) is the application interface; the middleware is the programming tool for the system (python) that perform the analysis while the backend is the application library that serves as the database for the corpus. Also in Figure5, the system sequence diagram is shown. The sequence diagram explains the interaction between the user and the system i.e. the flow of operations.

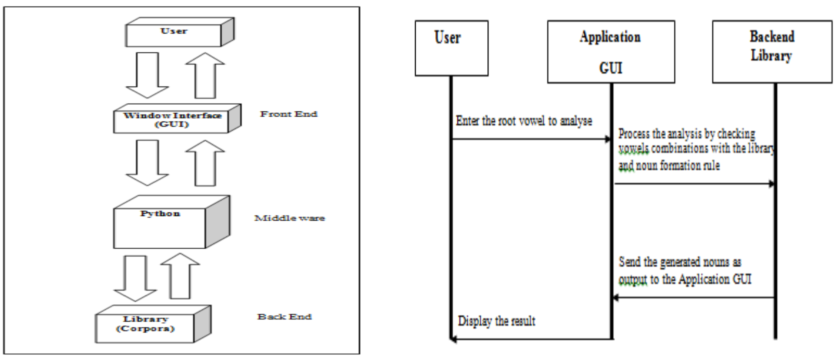

Figure 5: System design layout and the sequence diagram

IV. RESULTS AND DISCUSSION

Users are expected to enter Yorùbá vowels letters into the system. Any input other than "i e *a e ẹ i o ọ"* is invalid and no output will be display. It can either be uppercase or lowercase, but once the system processes the letter, output will be in lowercase.

The user interface is user friendly. The software has been aimed at linguists with a substantial level of computer literacy. The system designed could be used by linguists and non-linguists with basic computer skills. The GUI of the system is shown in Figure 7, and some operations of the system are shown in Figure8.

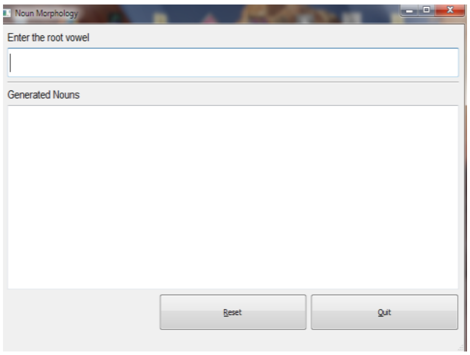

Figure 7: Graphical User Interface (GUI) of the System

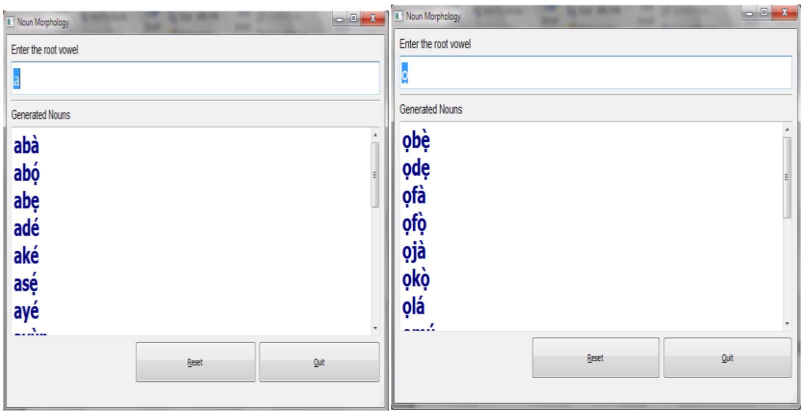

Figure 8: System operation example

V. CONCLUSION

Morphological analyzer are essential for any type of natural language processing works and it can be integrated to the language processing systems for a variety of applications in the Natural Language Processing field. The morphological analyzer developed has a variety of use, particularly to help linguists to discard the obsolete method of carrying out morphological analysis (i.e. pen and paper).Other areas noted for further study includes; morphology analyzer for other Yorùbá language part of speech (verb, preposition, etc).

REFERENCES

[1]. Aronoff, M. and Fudeman, K. (2004)What is Morphology?Blackwell Publishing: Cornwall.

[2]. Awobuluyi, O., (2008) Mofoloji Ede Yorùbá. Ibadan: Oxford University Press.

[3]. Martin, J. and Jurafsky, D., (2002) Speech and Language Processing: An Introduction to Natural Language processing, Computational Linguistics and Speech Recognition.

[4]. Adegbola, T., Owolabi, K., &Odejobi, T. (2011). Localising for Yorùbá: Experience, challenges and future direction. In *Proceedings of conference on human language technology for development* (pp. 7-10).

[5]. Lyons, J., (1981) LanguageandLinguistics*:* AnIntroduction. United Kingdom. Cambridge University Press.

[6]. Akinlabi, A., (2001) The Facts about Yorùbá Language. Ibadan: Oxford University Press.

[7]. Church, K. (2005) *The DDI* Approach to Methodology. Inquiries into Words, Constraints and Contexts. CSLI Publications: Stanford.

[8]. Igboanusi, H., and Lothar P., (2005) LanguagesinCompetition. Frankfurt and Main: Peter Lang.

[9]. Larman, C., (2003) "Applying UML and Patterns", 2nd Edition John Wiley & Sons, Inc.

[10]. Yona, S. and Wintner, S. (2005). A Finite State morphological grammar for Hebrew. Proceedings of theACL Workshop on Computational Approaches to Semitic Languages. P 9-16.

[11]. Agbeyangi A. O., Adenekan O. A., &Lawal T. A. (2015). Development of a computational system for morphological analysis of standard Yorùbá Nouns: *1st International conference of Environmental Studies, MoshoodAbiola Polytechnic, Ilaro, Ogun State, January 20-22, 2015.*