Computing, Information Systems, Development Informatics & Allied Research Journal
Vol. 10 No. 4, December, 2019 - www.cisdijournal.net

A Computational Modelling of Yorùbá Numerals in A Text-To-Text Format

***Agbeyangi, A.O., **Eludiora, S.I. & ***Olorunlomerue, A.B.**
[*]Department of Computer Science, Chrisland University, Abeokuta, Ogun State, Nigeria.
[**]Department of Computer Science & Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria.
[***]Department of Computer Science, Federal Polytechnic, Ede, Osun State, Nigeria
E-mail: aagbeyangi@chrislanduniversity.edu.ng[*], olorunlomerue.adam@federalpolyede.edu.ng[***]

ABSTRACT

Yorùbá numeral system has been seen to have the most unusual and complicated of any of the world's natural language numeral systems due to the complexity involved in its derivation. In this paper, we present the development of an English to Yorùbá text-to-text number translation system. In our approach, we examined English and Yorùbá numerals to understand the concepts and techniques underlining their translation process. The theories, rules and computational processes underlying the numeral system were used to formulate the model for the work. The design, requirement and the specification for the system were done in Unified Modeling Language (UML) and Automata theory. The designed system was implemented using Python programming language and other modules (PyQt4 and NLTK). The experimental results showed that the system scored 98% on all the output considered.

Key words: *Numeral system, Machine translation, Yorùbá numerals, Number, English numerals.*

CISDI Journal Reference Format
Agbeyangi, A.O., Eludiora, S.I. & Olorunlomerue, A.B. (2019): A Computational Modelling of Yorùbá Numerals In A Text-To-Text Format. Computing, Information Systems, Development Informatics & Allied Research Journal. Vol 10 No 4, Pp 1-. Available online at www.cisdijournal.org. DOI Affix - https://doi.org/ 10.22624/AIMS/CISDI/V10N4P1

1. INTRODUCTION

Counting and numbering is an important aspect and an inseparable part of the grammar of the most natural language. It has been noted that there is hardly any known linguistic discourse in a language that will likely not refer to quantity, time, size, weight, and distance using numbers (Omachonu, 2011; Agbeyangi et al., 2016). The use of numbers and their uniqueness in capturing relevant concepts makes them indispensable in effective communication (Choudhury et al., 2018). Numbers are believed to be an indispensable tool in our day-to-day activities because we can hardly communication without referring to number to quantify things. It was noted in Zaslavsky (1994) that key advancement in modern civilization can be traced to the conception, representation, invention, and manipulation of numbers to facilitate accurate rendering of measurable objects. They are mostly used in trade, mathematics, cosmology, music, divination, medicine, and in many other areas.

Human linguistic knowledge will remain insufficient if we cannot count fluently and analyze how this counting is done using our various languages (Klamer et al., 2017). According to Babarinde (2013), numerals in whatever language and or in whichever forms it takes can be seen as a clear manifestation of the ability of human to manipulate his innate knowledge of counting to cater for his communicative needs. But it appears that numeral is an aspect of human language in which researchers, scholars and linguists seem to pay less attention to. This neglect could be connected to the impression that there is little or nothing else to say about it (Akinade and Odejobi, 2014). The notion of neglect was stressed in Babarinde (2013) that “Little did we know that there is much to say about this little?”.

Knowing that part of human linguistic knowledge is contained in the ability to express ideas in our local dialect. Odejobi et al., (2015) also buttressed the notion that "aside from the power of language, numeracy and arithmetic are perhaps the next most important skills in problem-solving". Thus, man's ability to conceptualized and make abstractions numerically and also use arithmetic skills to manipulate the labelled abstraction is believed to be the most effective tool in computing (Vaidya and Joshi, 2015; Muhammad, 2016; Aina, 2016).

Numeral system is a mathematical notation to represent numbers of given set with digits or other symbols consistently (Schapper & Klamer, 2014). Boyer (1944) opined that the concept of number and the ability to count doesn't seem to follow language patterns, or a fortiori, of any written representation. Schapper & Klamer (2014) believed that numerals and numeral systems have a typological and historical interest to linguists. Thus, the concept of number notations was developed only when there were attempts to transmit, preserve records and ideas concerning number, and also when aids to mental calculations are required. Some of the world's recognised numerals system are Roman numerals (Schlimm and Neth, 2008), Arabic numerals (Smith & Karpinski, 2013), Mayan numerals (Keller, 1955), Bangla numerals (Obaidullah et al., 2015), Chinese numeral (Her, 2017), Yorùbá numerals (Babarinde, 2016), etc. In Figure 1, a sample of Mayan numeral system is shown.

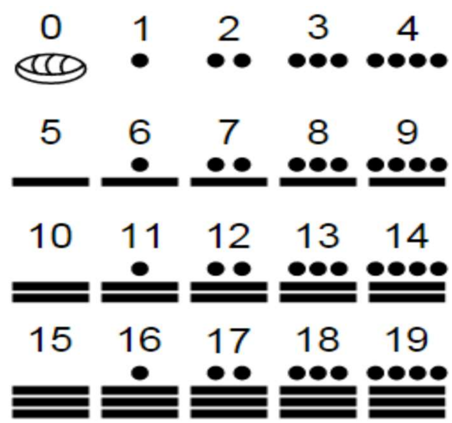

Figure 1: Maya Numerals
Source: Derksen (2019)

In this study, we seek to develop an English textual number to Yoruba textual number system. In the translation, the system will accept English numbers in a textual form and produce the equivalent number in Yorùbá language. The remaining part of the paper is structured as follows: Section 2 explains the concept of the Yorùbá numeral system; section 3 explains the system design and implementation, while Section 4 discusses the results. Section 5 concludes the paper.

2. YORÙBÁ NUMERAL SYSTEM

In an early study of the Yorùbá numeral system by Mann (1887), it was explained how large numbers could be represented as an arithmetic combination of the basic number units and reveals that the subtraction operation plays an important role in number naming. The Yorùbá numeral system is a form of a vigesimal numeral system like the Mayan numerals because its numbers are mostly represented in base 20 (Ekundayo, 1977; Akinade and Odejobi, 2014). Although the Yorùbá numeral system is not fully vigesimal since there are elements of decimal and quinary (base 5) used in its representation.

This notion was also supported by Lounge (2009) that "Yorùbá numeral has a rather elaborate vigesimal (base-20) numeral system that involves addition, subtraction and multiplication". More so, it has been argued that where the European decimal system is based upon units of ten and the functions of additions and multiplication, Yorùbá numeral system is based upon the units of both ten and twenty and the function of addition, multiplication and subtraction (Hall, 1973; Babarinde, 2016).

The Yorùbá numeral system uses a complex number system that combines vigesimal (base-20), decimal (base-10) and quinary (base-5) in a unique systematic way with heavy use of subtraction and over-counting. In Oyetade (1996) the pattern of the traditional Yorùbá vigesimal number system was given as follows:

- One (1) to ten (10) are basic words and eleven (11) to fourteen (14) are expressed as 1 + 10, 2 + 10, 3 + 10 and 4 + 10 respectively.
- Fifteen (15) to nineteen (19) are expressed as 20 - 5, 20 - 4, 20 - 3, 20 - 2, 20 - 1 and twenty (20) is a basic word.
- Twenty-one (21) to twenty-four (24) are expressed as 20 + 1, 20 + 2, 20 + 3 and 20 + 4.
- Twenty-five (25) to twenty-nine (29) are expressed as 30 - 5, 30 - 4, 30 - 3, 30 - 2 and 30 - 1.
- Thirty (30) is another basic word.
- A pattern similar to the one used for twenty-one (21) to twenty-nine (29) is followed for thirty-one (31) to thirty-four (34) and thirty-five (35) to thirty-nine (39).
- Forty (40) is expressed as 20 x 2.
- The pattern of the addition for forty-one (41) to forty-four (44) and subtraction for forty-five (45) to forty-nine (49) is followed for numbers after fifty (50), sixty (60), seventy (70), eighty (80), etc.
- Fifty (50) is 60 - 10. This same pattern is followed for seventy (80 - 10), Ninety (100 - 10), One hundred and ten (120 - 10), One hundred and thirty (140 - 10), One hundred and fifty (160 - 10), One hundred and seventy (180 - 10), and One hundred and ninety which is expressed as (200 - 10).
- The same pattern of multiplication used for forty is followed for sixty (20 x 3), eighty (20 x 4), One hundred (20 x 5), One hundred and twenty (20 x 6), One hundred and forty (20 x 7), One hundred and sixty (20 x 8), One hundred and eighty (20 x 9).

Thus, the basic Yorùbá numerals are one to ten (óókan sí ẹ□ẹ□wá). Twenty (ogún), thirty (ọgbọ□n), two-hundred (igba), three-hundred (ọ□dúnrún), four-hundred (irínwó) and twenty-thousand (ọ□kẹ□ kan) (Akinade and Odejobi, 2014). All other numerals apart from the basic numerals are derived using rules. These patterns seem not to follow any meaningful or easy way of representation. Hence, it poses a great challenge in its computational modelling.

3. METHODOLOGY

The developmental processes involve in this research work starts with the design of a model to describe each stage in the development of the system. There are five important stages in this model (Figure 2). First, is the part that converts the textual number to Arabic number, second is the number decomposition process: where numbers are expressed as a sum of smaller numbers in harmony with their sub-grouping. The output of number decomposition stage forms the magnitude stack. Next is the stage that generates the possible forms of a single number. This is done by careful combinations of the neighbouring elements of the magnitude stack and parsing them with the designed numeral grammar. The third stage is where tokens of the number forms are converted to their equivalent lexical forms, and the final stage is where the morphophonological rules employed in Yorùbá numbers naming are applied. These processes were modelled to organize the processes required for the translation.

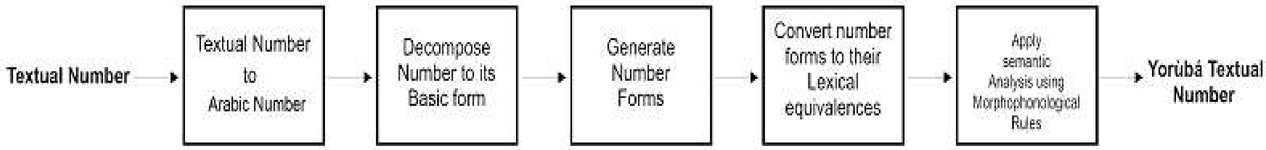

Figure 2: The Yoruba Numeral System Developmental Stages

In the design process, Context Free Grammar (CFG) was built for the translation process. The set of terminal symbols used are: DIGIT, M, V, VV, and REDUCE.

- **DIGIT** are the set of lexemes known as the basic numbers in Yorùbá numeral system. They are *ọ̀kan* (1), *èjì* (2), *èta* (3), *ẹ̀rin* (4), *àrún-ún* (5), *ẹ̀fa* (6), *èje* (7), *ẹ̀jọ* (8), *ẹ̀sán-án* (9), *ẹ̀wá* (10), *ọgbọ̀n* (30), *ọ̀dúnrún* (300) and *irínwó* (400).
- **M** are the set of multiplicative bases. M = {*ogún* (20), *igba* (200), *ọ̀kẹ́* (20000)}.
- **V** are set of operators that occur *within* number phrases. V = {*lé ní* (+), *dín ní* (-)}.
- **VV** are set of operators that occur *between* number phrases. VV = {*ó lé* (++), *ó dín* (--)}.
- **REDUCE** are set of implied subtraction operators represented by the prefixes *aadín* (reduction by 10) and *ẹẹdín* (reduction by 5, 100 and 1000). REDUCE = {*aadín*, *ẹẹdín*}.

Thus, the rewrite rules (which was a modification to what was used in Akinade and Odejobi (2014)) for the text-to-text Yorùbá numeral system are given below:

S $\rightarrow$ NUM
NUM $\rightarrow$ NP | NUM SN
SN $\rightarrow$ VV NP
NP $\rightarrow$ DIGIT | MP | VP NP
NP $\rightarrow$ REDUCE MP | NP M
MP $\rightarrow$ M | MP NP
VP $\rightarrow$ DIGIT V

In the rewrite rule, NP represents Noun Phrase or simply Number Phrase, MP represent Multiplicative-Base Phrase, NUM is the generated number which form the number in a sentence form. An outputs from NLTK parser where the re-write rules were tested in Figure 3 and 4. In Figure 3, how to generate number fifty-five (55) and One hundred and eighty-two (182) are shown. In Figure 4, same numbers were parsed in their textual form.

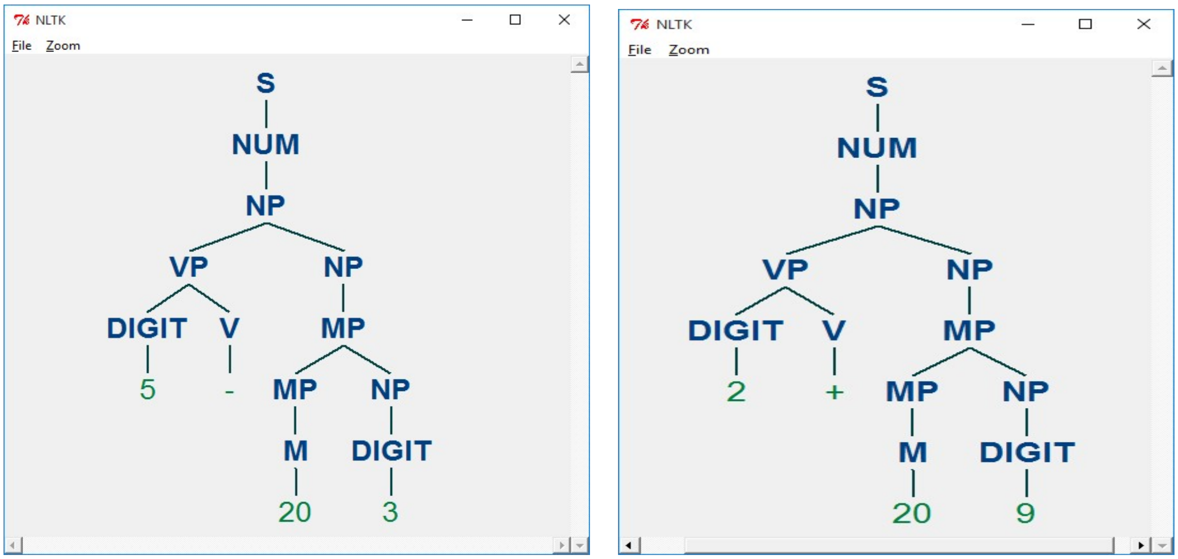

Figure 3: NLTK Parsing output for number Fifty-five (55) and One hundred and eighty-two (182)

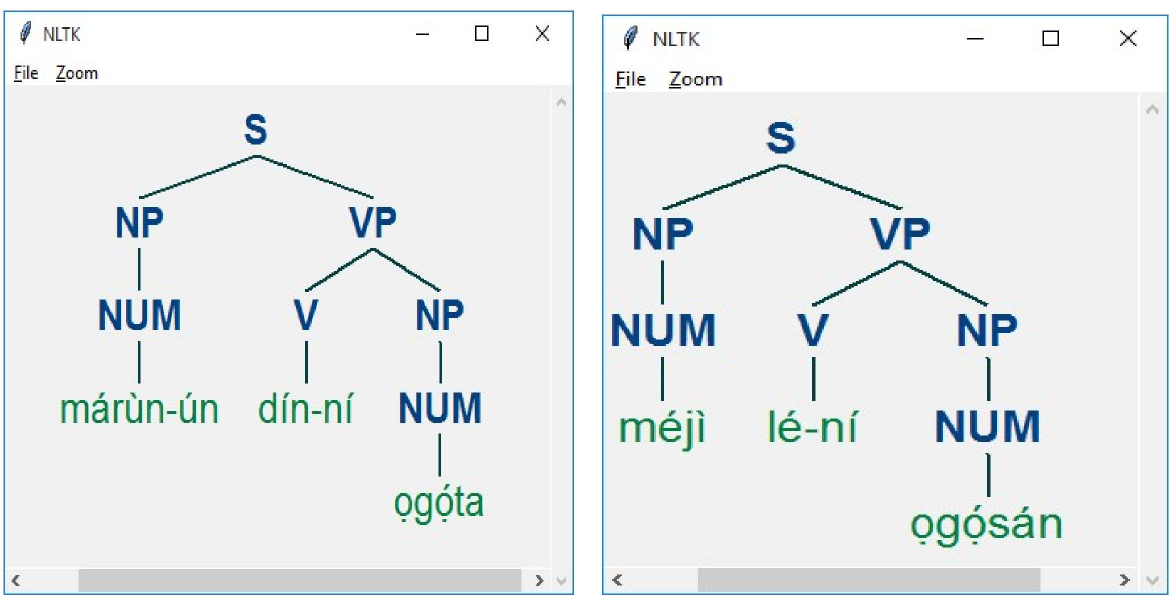

Figure 4: NLTK Textual output for number Fifty-five (55) and One hundred and eighty-two (182)

The behaviour of the translation process was also modelled using Automata. The behaviour was expressed as an abstract machine exhibiting the characteristic of outputting the standard output required. In JFLAP (a Java-based Automata simulator), the rewrite rule was simulated as illustrated in Figure 5. The Finite State Automaton (FSA) has six states (q_0 to q_5). The initial state is q_0 while the final state where the state transitions are accepted or rejected is q_3. In Figure 6, the re-write rule was simulated to show accepted and rejected inputs.

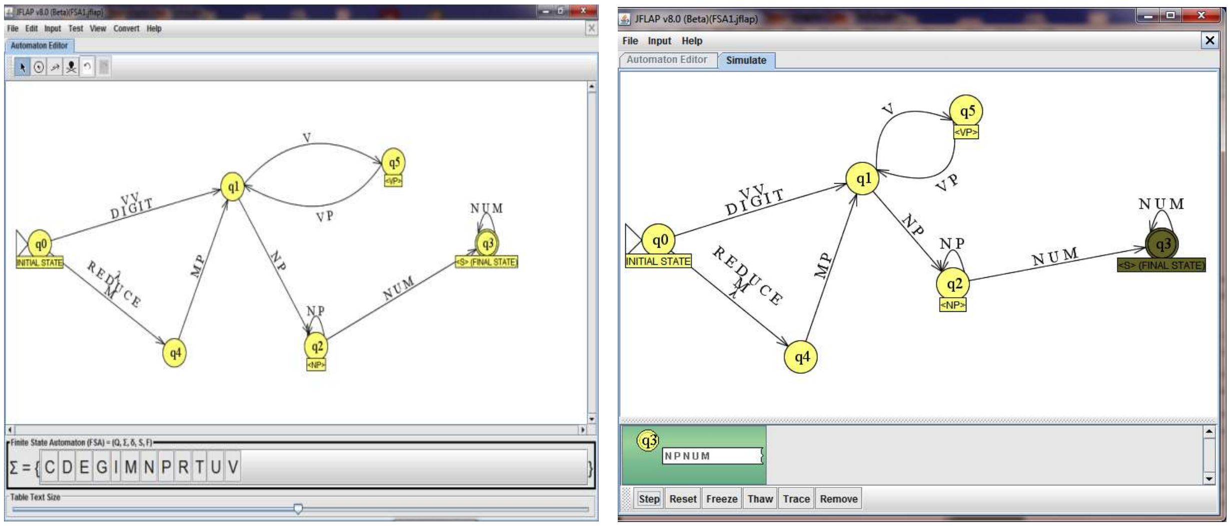

Figure 5: The Rewrite Rule Automaton Simulation

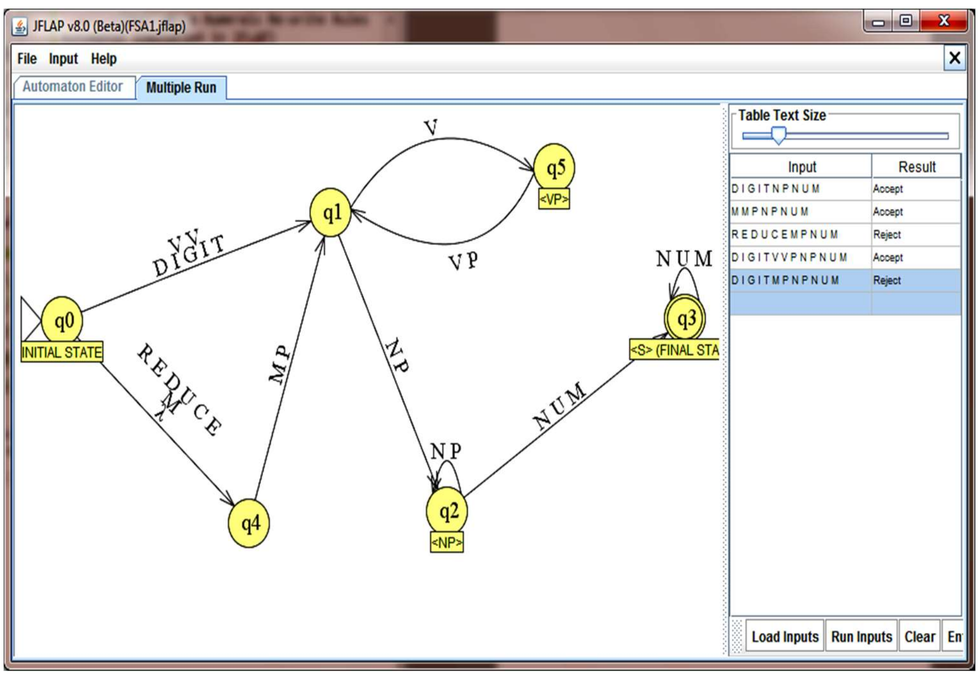

Figure 6: The Rewrite Rule Automaton with accepted and rejected inputs

The sequence diagram of the translation process is shown in Figure 7. It depicts identified objects of the translation process as lifelines running down the stages with their interactions over time which are represented as messages drawn as arrows from the source lifeline to the target lifeline.

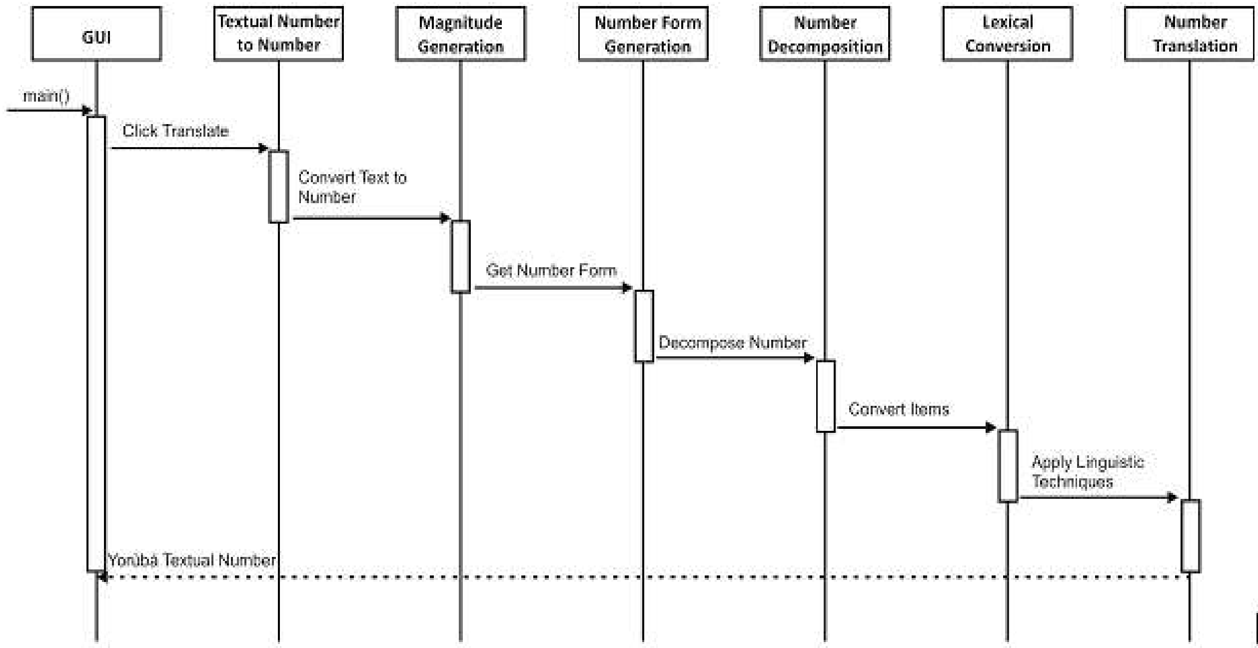

Figure 7: The Translation Process Sequence Diagram

3.1. Data Collection

The data collected and used for the system are the basic numbers in Yorùbá numerals, the multiplicative bases, the within/between phrases operator and the implied subtraction operator. The other entries in the database (stack) were generated using the rewrite rules. The sample of this is shown in Figure 8.

3.2. Development Tools

The system was developed as a Windows desktop application using PyQt (the software used for the application GUI), Python (the core programming environment) and NLTK (the support tool used for parsing). The developed application will work in Windows XP, 7 and 10 (both 32-bit and 64-bit).

4. RESULTS AND DISCUSSIONS

The system output in Figure 9 shows how seventy-eight (78) and four hundred and thirty-two (432) are translated to standard Yorùbá number (méjìdínlọgọ□rin) from English numeral in textual form. In Figure 10, the another translation is shown for three thousand five hundred and eighty (3580) and one billion (1000000000). The capacity of the system to translate higher numbers was tested in Figure 10 for the translation of one billion. The application was able to translate one billion to its Yorùbá equivalent as ọ□kẹ□ ọ□nà ọ□kẹ□ méjì ó lé ẹgbàrún ọ□kẹ□. The translation capacity of the system is enormous owing to the robust algorithm employed.

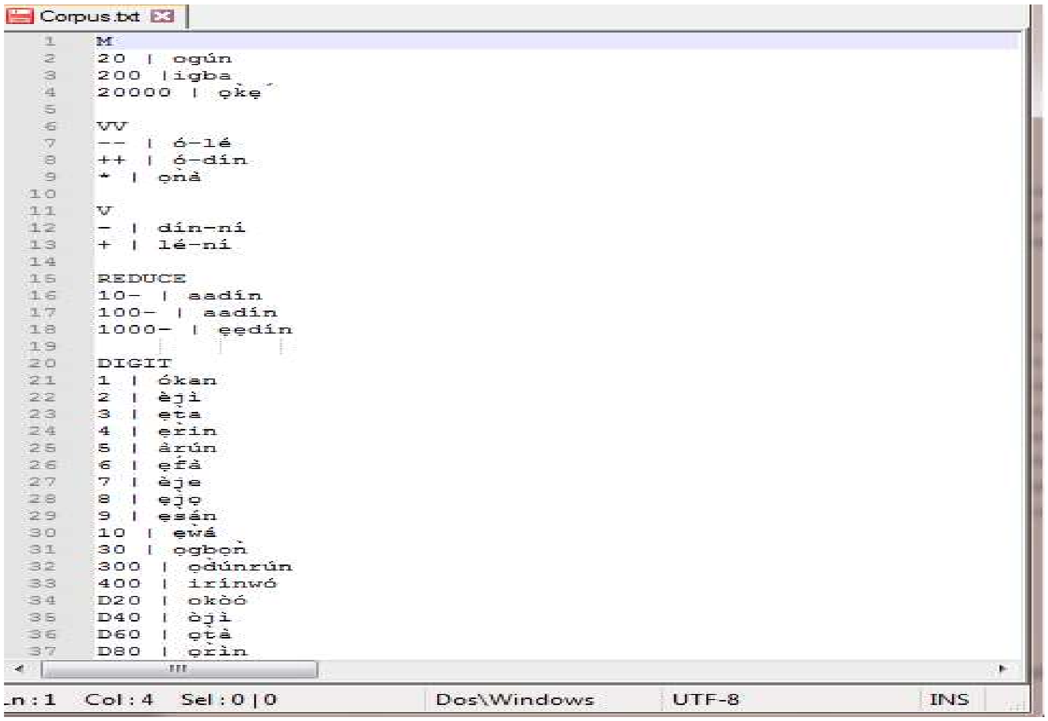

Figure 8: The Basic Numbers Corpus

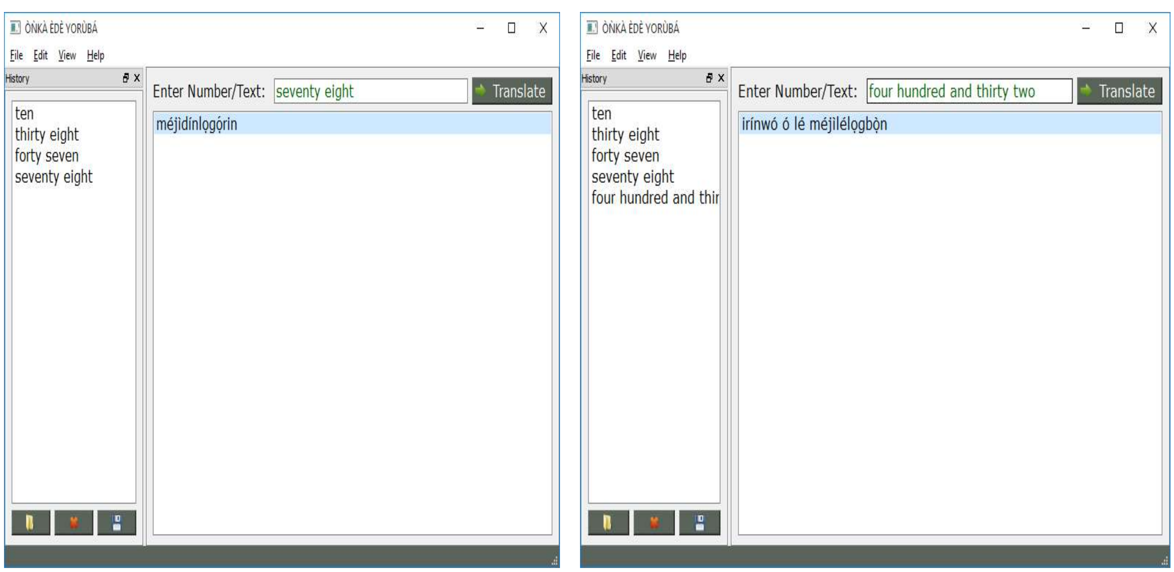

Figure 9: The System Sample Output I

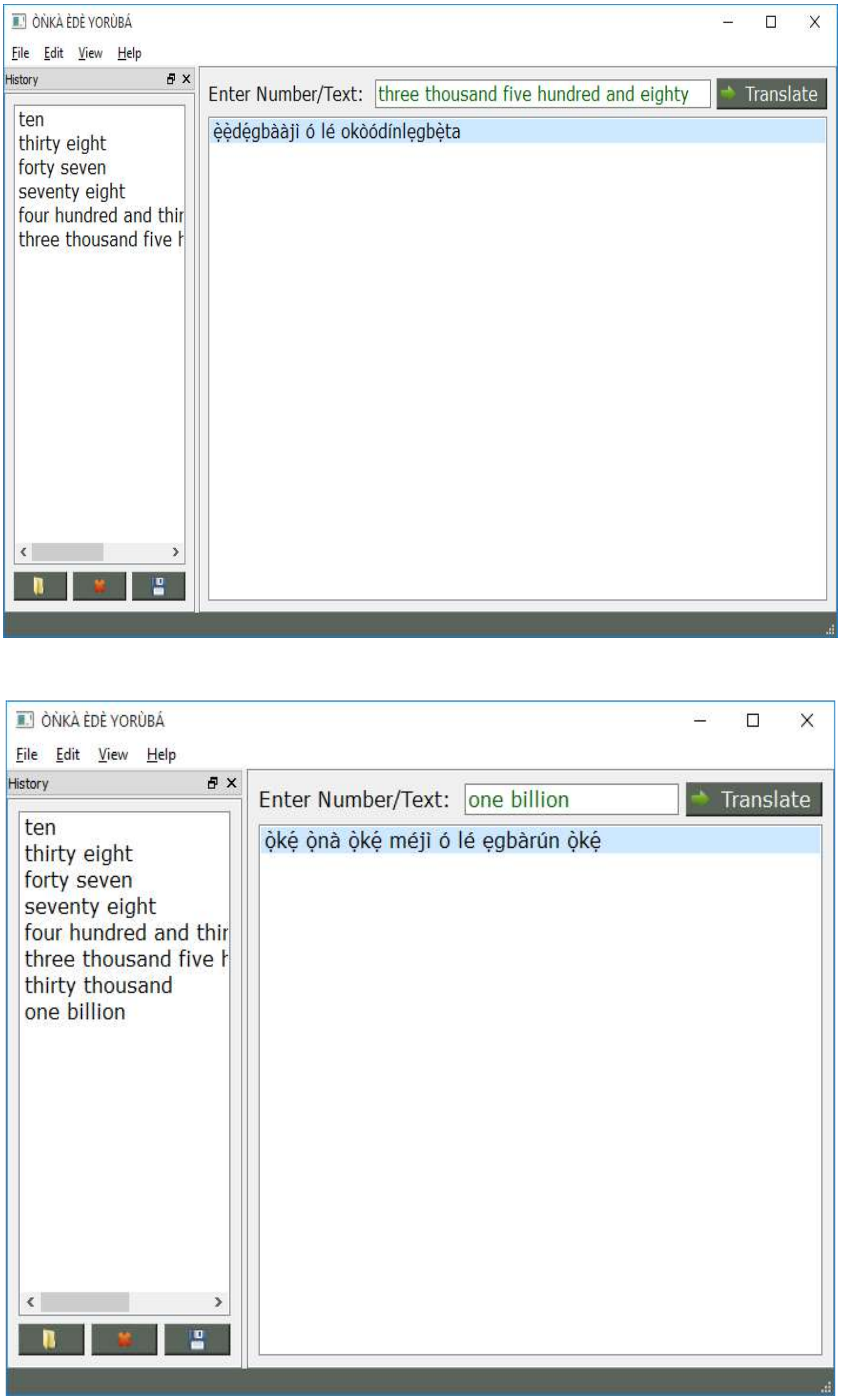

Figure 10: The System Sample Output II

4.1. System Evaluation

The system was evaluated using the mean opinion score test. This was done to get feedback from expert (Yorùbá language lecturers) and others general users to give an accurate representation of some numbers to compare with the system output to ascertain the level of accuracy of our system.

The numbers in the questionnaire distributed are:

- Frequently used numbers: Number 8, 15, 25, 43, 67, 82, 100, 132, 150 and 197.
- Numbers with multiple representations: Number 635, 753, 4500, 2400 and 967.
- Higher numbers: Number 20000, 50000, 100000, 1000000 and 20000000.

Although numbers with multiple representations were addressed in Akinade and Odejobi (2014), our system only generates one standard output. From the results gathered and analyzed, it was observed that all the respondents are familiar with frequently used numbers by providing near accurate response to these numbers. This shows that a large percentage of Yorùbá native speakers understand these numbers and mostly use them. However, the results gathered on higher number usage showed that many of the respondents never use or hear about these higher numbers as they are barely used nowadays. The graph of the results of the comparison between our system, general respondents and the expert is shown in Figure 10. This validate that our system performed relatively well if compared to the general respondents.

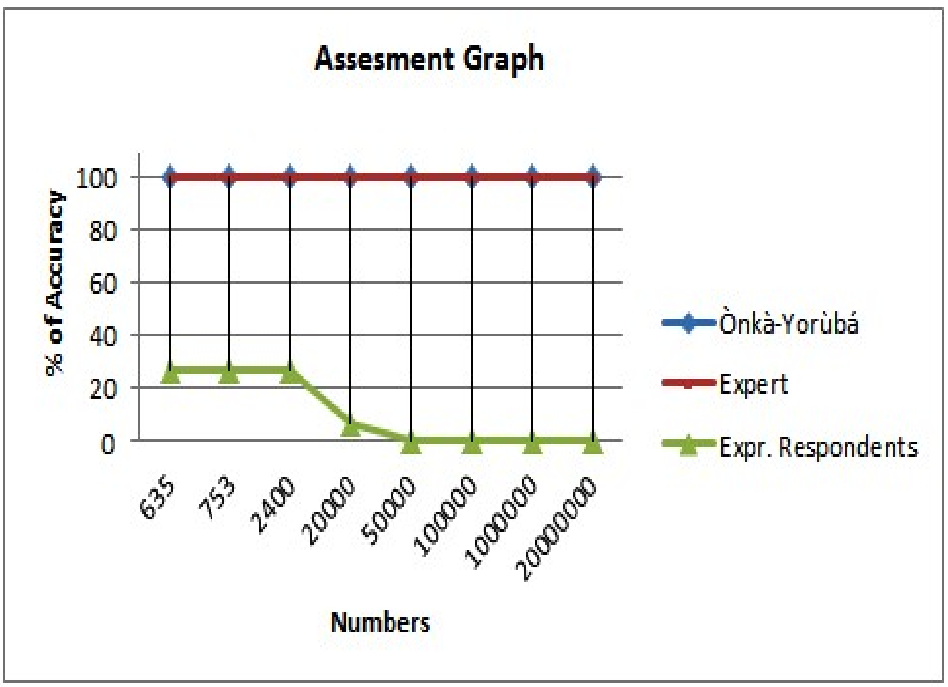

Fig. 7: Assessment Graph
(Expert, Experimental respondent and the System)

5. CONCLUSION

In this study, we have been able to explore the rich techniques to generate standard Yorùbá numbers and its numeral system. The developed system after passing through rigorous testing and evaluation is believed to have fulfilled the core aim of the research in rendering accurate translation in standard Yorùbá numeral. The research also avails us the need for other areas of study to further enhance this work. Some of these are the need to study the use of Yorùbá numbers for dates, the use of ordinal and adverbial numerals in Yorùbá language.

REFERENCES

1. Akinadé, O. O. and Ọdẹ̀jobí, D. A. (2014). Computational modelling of Yoruba numerals in a number-to-text conversion system. Journal of Language Modelling, 2(1):167-211.
2. Aina, A. A. (2016). Compounding in the Yoruba Numeral System. The Numeral Systems of Nigerian Languages, (2), 303.
3. Agbeyangi, A. O., & Eludiora, S. I. (2016). The Yorùbá Numeral System: A Review in Relation to other Numeral Systems. The Numeral Systems of Nigerian Languages, (2), 361.
4. Babárìndé, O. (2013). The Linguistic Analysis of the Structure of the Yorùbá Numerals. Journal of Literature, Languages and Linguistics, 1:78-88.
5. Babarinde, O. (2016). A Linguistic Analysis of the Structure of the Yoruba Counting System. The Numeral Systems of Nigerian Languages, (2), 225.
6. Boyer, C. B. (1944). Fundamental steps in the development of numeration. ISIS, JSTOR, pp. 153-168.
7. Choudhury, A., Rana, H. S., & Bhowmik, T. (2018). Handwritten Bengali Numeral Recognition using HOG Based Feature Extraction Algorithm. In 2018 5th International Conference on Signal Processing and Integrated Networks (SPIN) (pp. 687-690). IEEE.
8. Ẹkúndayọ̀, S. (1977). Vigesimal numeral derivational morphology: Yorùbá grammatical competence epitomized. Anthropological Linguistics, pp. 436-453.
9. Derksen, B. (2019). Maya. Retrived from https://upload.wikimedia.org/wikipedia/commons /1/1b/Maya.svg on 6th December, 2019.
10. Hall, G. (1973). Yorùbá Numbers. In Folklore Forum, pp 1.
11. Klamer, M., Schapper, A., Corbett, G., Holton, G., Kratochvíl, F., & Robinson, L. C. (2017). Numeral words and arithmetic operations in the Alor-Pantar languages. The Alor-Pantar languages, 337.
12. Keller, K. C. (1955). The Chontal (Mayan) numeral system. International Journal of American Linguistics, 21(3), 258-275.
13. Lounge, O. (2009). A vigesimal number system. Bookbuilders.
14. Mann, A. (1887). Notes on the Numeral System of the Yorùbá Nation. Journal of the Anthropological Institute of Great Britain and Ireland, pp. 59-64.
15. Odejobi, O. O., Ajayi, A. O., Eludiora, S. I., Akanbi, L. A., Iyanda, I. R., & Akinade, O. A. (2015). A web-based system for supporting teaching and learning of Nigerian indigenous languages. OAU TekCONF 2015 proceedings, Nigeria, 350-360.
16. Omachonu, G. S. (2011). Derivational Processes in Igala Numeral System. Journal of universal language, 12(2):81-101.
17. Oyetade, B. A. (1996). Comparative Number Words of Five Àkókò Dialects and Ẹ̀dó with Standard Yorùbá. Journal of Nigerian Languages and Literatures, 2:17-46.

18. Schlimm, D., & Neth, H. (2008). Modeling ancient and modern arithmetic practices: Addition and multiplication with Arabic and Roman numerals. In CogSci (pp. 2097-2102).
19. Schapper, A., & Klamer, M. (2014). Numeral systems in the Alor-Pantar languages. The Alor-Pantar languages, 285.
20. Smith, D. E., & Karpinski, L. C. (2013). The Hindu-Arabic Numerals. Courier Corporation.
21. Vaidya, M. V., & Joshi, Y. V. (2015). Marathi numeral recognition using statistical distribution features. In 2015 International Conference on Information Processing (ICIP) (pp. 586-591). IEEE.
22. Zaslavsky, C. (1994). " Africa Counts" and ethnomathematics. For the Learning of Mathematics, pp. 3-8.