Automated conversion of numeral to words in Hausa language

Rabi Mustapha*, Falalu Ibrahim Lawal and Muhammad Aminu Ahmad

Department of Computer Science, Faculty of Science, Kaduna State University, P.M.B 2339, Kaduna State, Nigeria.

*Correspondence: rabichubu@kasu.edu.ng

Abstract
Working with cardinal numbers is undoubtedly an integral part of our everyday life, which tells us about quantity. They are used for instance, in numeracy lessons, on receipts, slips, tellers in banks, to represent figures on documents, to tell time and are often represented inconsistently. However, the words form representations are often inconsistent with the standard Hausa. Standard Hausa as compared to other languages such as Arabic and British English, the Numeric System is characterized by a multitude of concatenations as well as dialect variations which causes misinterpretations and confusion. Also, initial findings show that no such automation published work has been done specifically in the numerical system of Standard Hausa language. In order to fill the gap, this study proposed an algorithm for converting numerals to words for the numeric system of Hausa language automatically in standard practice. The aim of the study is to examine the effect of iteration and mapping structures algorithm for converting numeral to cardinal form in standard Hausa language. The algorithm has been tested on approximately one hundred million sequences of numbers produced accurate results. The results showed that our method is successful in automating conversion of numerals to words with a high precision.

Keywords
Algorithm; Automation; Conversion; Hausa Language; Numerical System

License: CC BY 4.0*

Open Access Article

How to cite this paper: Mustapha, R., Lawal, F.I. and Ahmad, M.A. (2023). Automated conversion of numeral to words in Hausa language. *Gadau J Pure Alli Sci*, 2(2): 140-145. [https://doi.org/10.54117/gjpas.v2i2.100](https://doi.org/10.54117/gjpas.v2i2.100)

1.0 Introduction

It is estimated that there are about 7,139 spoken languages today in the World and 40% of which are endangered (Eberhard et al., 2021). Hausa language on the other hand is the widest spoken language in sub-Saharan Africa. Several factors such as; war, trade, search for livelihood, quest for knowledge and migration are attributed to the spread of Hausa language. However, the most indispensable aspect in the treasury of human language, the numeral counting system of Hausa language are also affected (Bunza, 2018). Hausa is thought to be one of the most unusual and sophisticated numerical systems of any natural language in the world (Bala, 2015; Garba, 2018). Introduction of the Arabic language, literary tradition extending back several centuries before the invasion of the colonial masters, has dispossessed many of the purely older names. Words such as ‘gomiya’ and ‘hauya’ (used for expressing multiples of ten, twenty respectively) for counting and many were borrowed from Arabic (Migeod, 1914). Doing with numbers is inherent in today’s digital world where information is ubiquitous. Numbers are inconsistently written or printed in books, newspapers, etc. and on cheques, tellers, and papers etc., due to the evolutionary dynamic of languages, influence of the Arabic language, and a number of geographical dialects. One can notice variations in the dialects in the area of Kano, Zaria, Bauchi, Daura, Sokoto, Gobir, and northwards into Niger. Standard Hausa is coined based on ‘Kananci’ the dialect of Kano and has been recognized as the norm for the written language as contained in books and mass communication (Africa: linguistik- und-sprachen/african-languages/hausa, 2015). Undoubtedly in human history, numbers and numerals share a common root of origin and be that it may for Hausa method originated from either of the following (Bunza, 2018):
1. **Fingers and toes:**
The system of counting using fingers has since been and still in use today in different parts of the world. People often use it to fill the communication gaps especially in the market places (Maika'nti, 2014). Traditionally, Hausa people use fingers to count, mostly from right hand to left one - as the case may be.

![Figure 1: Counting using fingers](image)

---

2. **Drawing/writing:**
Historically, ancient Hausa culture, counting is represented, see Table 1 by either drawing lines on the ground/sand or putting dots equal to the numbers

<table>
<thead>
<tr>
<th>Numerals</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dots</td>
<td>.</td>
<td>..</td>
<td>...</td>
<td>..</td>
<td>. .</td>
<td>..</td>
<td>...</td>
<td>..</td>
<td>. .</td>
<td>....</td>
</tr>
</tbody>
</table>

Following British colonization, which forcefully intruded Roman alphabet although, had to be learned from scratch into newly created education system in 1909 to slow down already established Ajami - the adapted Hausa language Arabic script (Abdalla, 2005). Today, this representation of the language has been superseded for most purposes by the Roman script (Africa: linguistik-und-sprachen/african-languages/hausa, 2015).

**Phonemic inventory of Hausa**
There are total of 34 consonants in Standard Hausa. The vowels comprising of 5 short vowels and 5 corresponding long vowels and 3 diphthongs making a total of 13. Some consonants are not found in English. Most common of these are usually the hooked letters i.e. ɓ, ɗ, ƙ and the semi vowel ‘y’, which are entirely different from the corresponding plain letters b, d, k and y. Table 2 represents Sakkwatanci dialect, standard Hausa and the English equivalence of counting system from 1-10.

![Table 2. SEQ Table ∵ ARABIC 2: Cardinal](image)

Table 2 represents Sakkwatanci dialect, standard Hausa and the English equivalence of counting system from 1-10.

The counting system from eleven changes with additional ‘sha’ before the next unit with optionally additional word ‘goma’. The ‘sha’ stands for ‘swallow’. In this view, ‘(goma) sha’ is to be repeated
in counting up to nineteen ‘19’ as represented in Table 3.

### Table 3. SEQ Table * ARABIC 3: Cardinal representation of numbers 11-19

<table>
<thead>
<tr>
<th>Num. ‘Sakkwatanci’</th>
<th>Phoneic Representation</th>
<th>Standard Hausa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dialect</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 (gōmā) sha ḍā́i</td>
<td>(gōmā) sha ḍā́i/dā́jā́/</td>
<td>(gōmā) sha ḍā́yā</td>
</tr>
<tr>
<td>12 (gōmā) sha biyū́</td>
<td>(gōmā) sha biyū́/</td>
<td>(gōmā) sha biyū́</td>
</tr>
<tr>
<td>13 (gōmā) sha ụkkù́</td>
<td>(gōmā) sha ụkkwù́/</td>
<td>(gōmā) sha ụkkù́</td>
</tr>
<tr>
<td>14 (gōmā) sha ụdhū́</td>
<td>(gōmā) sha ụdhū́/</td>
<td>(gōmā) sha ụdhū́</td>
</tr>
<tr>
<td>15 (gōmā) sha biyà́́ ́́</td>
<td>(gōmā) sha biyà́́́/</td>
<td>(gōmā) sha biyà́́́</td>
</tr>
<tr>
<td>16 (gōmā) sha šālī́́́́́́</td>
<td>(gōmā) sha šālī́́́́́́/</td>
<td>(gōmā) sha shālī́́́́́́</td>
</tr>
<tr>
<td>17 (gōmā) sha bākwá́́́́́́́</td>
<td>(gōmā) sha bākwá́́́́́́́/</td>
<td>(gōmā) sha bākwá́́́́́́́</td>
</tr>
<tr>
<td>18 (gōmā) sha tā́kwā́́́́́́́</td>
<td>(gōmā) sha tā́kwā́́́́́́́/</td>
<td>(gōmā) sha tā́kwā́́́́́́́</td>
</tr>
<tr>
<td>19 (gōmā) sha tā́rā́́́́́́́</td>
<td>(gōmā) sha tā́rā́́́́́́́/</td>
<td>(gōmā) sha tā́rā́́́́́́́</td>
</tr>
</tbody>
</table>

Compound numbers from twenty and greater, the numbers will then, be derived by adding ‘da’ morpheme in between the two numbers 20 and the primary numbers (1-9) till it gets to 29 before the number would change to a unique number which is 30 see Table 4 for the examples.

### Table 4. SEQ Table * ARABIC 4: Cardinal

<table>
<thead>
<tr>
<th>Num. ‘Sakkwatanci’</th>
<th>Phoneic Representation</th>
<th>Standard Hausa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dialect</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 šālrī́n dā́ dā́í</td>
<td>/ʃālrī́n dấ dấ/dā́jā́/</td>
<td>šālrī́n dấ dấyā</td>
</tr>
<tr>
<td>22 šālrī́n dā́ biyū́́</td>
<td>/ʃālrī́n dấ biyū́́/</td>
<td>šālrī́n dấ biyū́́</td>
</tr>
<tr>
<td>23 šālrī́n dā́ tā́kk̠̱</td>
<td>/ʃālrī́n dấ tā́kk̠̱*̱/</td>
<td>šālrī́n dấ tā́k̠̱</td>
</tr>
<tr>
<td>24 šālrī́n dā́ lū́df̠̱</td>
<td>/ʃālrī́n dấ lū́df̠̱/</td>
<td>šālrī́n dấ lū́df̠̱</td>
</tr>
<tr>
<td>25 šālrī́n dā́ biy̱̱</td>
<td>/ʃālrī́n dấ biy̱̱/</td>
<td>šālrī́n dấ biy̱̱</td>
</tr>
<tr>
<td>26 šālrī́n dā́ šālī́́́́́́</td>
<td>/ʃālrī́n dấ šālī́́́́́́/</td>
<td>šālrī́n dấ shālī́́́́́́</td>
</tr>
<tr>
<td>27 šālrī́n dā́ bākẃ̱̱</td>
<td>/ʃālrī́n dấ bākẃ̱̱/</td>
<td>šālrī́n dấ bākẃ̱̱</td>
</tr>
<tr>
<td>28 šālrī́n dā́ tā́kẃ̱̱̱</td>
<td>/ʃālrī́n dấ tā́kẃ̱̱̱/</td>
<td>šālrī́n dấ tā́kẃ̱̱̱</td>
</tr>
<tr>
<td>29 šālrī́n dā́ tā́ṟ̱̱́</td>
<td>/ʃālrī́n dấ tā́ṟ̱̱́/</td>
<td>šālrī́n dấ tā́ṟ̱̱́</td>
</tr>
<tr>
<td>30 tā́lḻ́ṯ̱̱̱</td>
<td>/tā́lḻ̱̱̱/</td>
<td>tā́lḻ̱̱̱</td>
</tr>
</tbody>
</table>

Hundreds are formed starting with the word for hundred ‘dari’, followed by the multiplier unit, for example ‘dari biyu’ i.e. 200, ‘dari uku’ i.e. 300, ‘dari hudu’ i.e. 400 and so on. Thousands on the other hand, are formed starting with the word for thousand ‘dubu’, followed by the multiplier unit. 2,000 is represented as ‘dubu biyu’, 3,000 as ‘dubu uku’, 4,000 as ‘dubu hudu’ and so on. The equivalent word for million is ‘miliyan’, for billion is biliyan, for trillion is ‘tiriliyan’. Automatic conversion of numbers has attracted much attention due to its broad applications in day to day activities. Several algorithms have been proposed to automatically convert a number to its textual equivalent. Object-oriented approach in C++ to convert numeric values into corresponding words in the Uzbek language (Utkir, Mukhriddin, Bakhtiyor, and Zarmasov, 2020). A web-based English to Yorùbá numeral translation system implemented using Google Web App Engine with support for python and a Python package for natural language processing support NLTK, translates English numbers both in figure and text to its standard Yorùbá form has been proposed in (Agbeyangi, Eludiora, and Popoola, 2016; Olúgbénga and Ode, 2014). Another algorithm has demonstrated conversion of numbers to words, however the range is limited from 1 to 1000 and displays the result onscreen. For example, 235 would be “Two Hundred and Thirty-five.” (Ben, 2017). It is undoubtedly that, unlike other languages, number forms in Tamil have recursive properties and details of their nuances in dealing with 90s, 900s, 9000s numbers, to word conversion using base-10 has been proposed in (Muthiah and Sathia, 2020). The algorithm works for parsing integral and floating point non-negative numbers; it parses text forms of numbers in both Indian and American standard (i.e. using millions, billions).

With all these developments, however, the methods were perfectly doing well when applied to a particular language, and do not apply to other languages, not even the dialects within a language. Although previous research showed successful outcomes, they for unknown reasons failed to discuss the feasibility of automating number conversion in Hausa language. With the complexity of the number system of Hausa language pointed out by Bala (2015), the researchers however, considered the feasibility of automating the conversion of numbers into words in Hausa language.

### 2.0 Materials and methods

This section focuses on the design of a rule based algorithm with lookup functions using pseudocode. It was later implemented using Python Programming language. With the complexity of the number system of Hausa language pointed out by Bala (2015), the researchers however, considered the feasibility of automating the conversion of numbers into words in Hausa language.

### 3.0 Results and discussion

The algorithm was tested on approximately ten billion numbers, Figure 3 - 4 and produced accurate results with a high precision, Figure 5. This shows how powerful algorithms can be in automating seemingly difficult tasks. The result has also shown a rule-based system can be powerful and consistent in its task.
Figure 2. SEQ Figure 5: preview translated_Hausa.csv with Pandas
The algorithm has been subjected to a sequence of integer numbers from 1 to 9,999,999 generated by Python: Figure 3:

**Figure 3: Python generated numbers**

Figure 4 shows the codes for:

```python
obj = kalma()
with open("translated_Hausa.csv", "w", encoding="utf-8") as myfile:
    for number in range(1,1000000):
        myfile.writelines(f'{number},{obj.fassara(number)}\n')
import pandas as pd
df = pd.read_csv(r'translated_Hausa.csv')
df.head
```

**Figure 3. SEQ Figure \* ARABIC 2: Pseudo code of the algorithm**

1. **Algorithm**
   The algorithm is divided into three different parts, the *Kalma*() method - a constructor that initiates the lookup arrays up on the creation of objects in the class. Another method *fassara*() takes a positive integer as an argument called to make the translation task. The last method *daruruwa*() serves as a helper to the *fassara*() method. The algorithm converts integral numbers received to generate corresponding words for that number. Rule based systems (algorithms) that automatically translate numeral representations into word representations are of great interest to us.

4.0 **Conclusion**

We have successfully designed, implemented and experimented an algorithm, *Kalma*, and consistently performs well.

In future work, we plan to expand the algorithm to also convert floating numbers.

**Competing interests**

All authors declare no competing interests.

**Funding**

There was no funding for the current report.

**References**


