The Gregorian Calendar Dates in Yoruba Language: A Machine Translation Approach

F.O. Durodola[1*], S.I. Eludiora[2], S.O. Owoeye[3]
[1,3]Department of Mechatronics Engineering, Federal University of Agriculture, Abeokuta, Nigeria
[2] *Department of Computer Science and Engineering, Obafemi Awolowo University, Ile Ife,* Nigeria
Email: durodola*fo@funaab.edu.ng*

Abstract
This study discussed the development of a system for Gregorian calendar date referencing in the Yoruba language. This is with a view to solving the problem of translating Gregorian calendar date expressions into their equivalent Yoruba form with the aid of Machine Translation (MT). The study analyzed the Gregorian calendar and its various pattern of date expression, the translation into Yoruba based on context and also developed a computational-based system that translates the Gregorian calendar date phrase to the corresponding Yoruba text. The Yoruba date translation process was achieved by identifying and extracting a dated phrase in the Gregorian calendar using regular expressions, the extracted date phrase is then converted to the British pattern of date referencing which is the DD/MM/YYYY. It is this pattern of date referencing that is then translated to Yoruba form. A Context-Free Grammars (CFGs) that capture the various patterns of date referencing according to the Gregorian calendar was specified. The CFG to capture the Gregorian calendar date phrase was simulated using the Push-Down Automata (PDA) and the software for the translation of the Gregorian calendar date phrase to the corresponding Yoruba phrase was implemented with the Python programming language and the PyQt5 Graphical User Interface (GUI) module. The evaluation of the system was carried out using the Mean Opinion Score (MOS). The MOS was done using a survey method, with a sample size of 45 study area of a broadcasting station in southwest Nigeria, with linguistics students and newscasters as respondents, to elicit their knowledge of the translation of selected dates in the Gregorian calendar. The responses were compared with the output of the system. A desktop application for the translation process was deployed for the windows operating system. The result of the evaluation showed that the application was able to translate the English date phrases to Yoruba text and produce 100 percent accuracy relative to a Yoruba expert translation. The result of the respondents to the questionnaires shows that most of the respondents are not familiar with the orthography of the Yoruba language, which means that a large percentage of Yoruba native speakers cannot tone mark their responses correctly. Another observation is that most of the respondents could not translate the year numbers to Yoruba as they are not familiar with the translation of large numbers. In conclusion, this study has shown that the date referencing translation process can be computationally analyzed using the modern knowledge of computing.

Keywords: Calendar date, Gregorian calendar, translation Processes, Yoruba language.

1.0 INTRODUCTION

The Gregorian calendar is the calendar used in most parts of the world. It has officially been adopted as the calendar for commercial and administrative purposes in many parts of the world where it runs alongside other calendars used for religious purposes (Cohen, 2000). A Gregorian calendar year is divided into 12 calendar months of irregular length with no relationship among the variations. The Gregorian calendar is based on a 365-days common year divided into 12 months, and on 366-days in leap count year in every year which is divisible by 4, except for centuries not divisible by 400, which are not leap year (Cohen, 2000). The calendar year runs from January 1 to December 31. Just as most of the world has come to adopt the Gregorian calendar, Nigeria as a nation is not left behind. The Gregorian calendar is the official calendar used for civil activities in Nigeria. This is evident in all official documents released by the government. Also, the education sector, the banking sector and all of our transactions or documentation that requires the referencing of dates make use of the Gregorian calendar. Other calendars in use are the Islamic calendar used by Muslims for ceremonial purposes and the Yoruba calendar popularly called "**Kọ́jọ́dá"**(Longstaff, 2005).

This study, therefore, aims at exploring the numbers and words used in Gregorian calendar date referencing and developing a system to capture how these numbers and words are correctly translated into the Yoruba language

2.0 METHODOLOGY

2.1 Study Area

The Yoruba language is predominantly spoken by the Yoruba people. Geographically, the Yoruba people are dominant in the immediate west of the river Niger, below the confluence and south of the Quorra, i.e the west branch of the same river above the confluence. There are bordered on the west by Dahomey. The land boundaries are between latitude 60 and 90 north and longitude 20 and 60 east (Johnson, 1966). Before the advent of the colonial masters, the Yoruba people of old have a way of counting days. Literature reveals that the Yoruba runs four days a week pattern, and a new month starts with the arrival of the moon. Hence, the month is usually 28 days or 29 days. The year's record is usually in relation to an event that has happened or the reign of a particular king (Jemiriye and Awosusi, 2007; Wande, 1976; Zaslavsky 1999).

2.2 Collection of Gregorian calendar date samples

Calendar dates can be written in a wide variety of ways, and often depend on formal or informal writing, personal style and whether you are writing British or American English. Whatever the format, elements of dates referencing are usually separated using a separator which could be a slash, a hyphen, a comma, or a full stop sign (Cambridge, 2017).

2.2.1 Sample 1. DD/MM/YYYY format

Referencing the day first, followed by the month and lastly by the year. This is the pattern that is predominantly used in European countries.

(a) When all the elements are digit, either of the separator (/ -.) is introduced.

Examples are (i) 2/2/2002. (ii) 2-12-1987. (iii) 15.06.2100.

(b) When the name of the month is referenced rather than expressed in numbers, a coma that serves as a separator is introduced after the name of the month. Also, the digit representing the day carries a superscript of the order of numerals.

Examples are (i) 2nd May 2013. (ii) 23rd February 2010. (iii) 14th July,1986

(c) The numbered day can be represented in a textual form just as the name of the month. Examples are (i) Seventh of July 2017, (ii) Third of May 2010. (iii) Fifteenth of Aug. 1987.

2.2.2 Sample 2. MM/DD/YYYY

Another way is to reference the month first, followed by the day and lastly by the year. This pattern is predominant in American countries (Cambridge, 2017).

(a) The three elements can be expressed using a digit. When this is done, a separator is introduced. Examples are (1) 2/30/1987 (ii) 12/28/2010 (iii) 11-30-1999

Now, this pattern is a bit confusing especially when the second item referenced is less than 12. Example 6/8/2012 can be translated as the 6th day in the month of August, it could also be translated as the 8th day in June.

(b) A clearer way of presenting this pattern is to give the name of the month, rather than representing it with a digit. When this is followed, the digit representing the number of the day is written as an ordinal numeral. Examples are(i). May 2nd, 2018 (ii). July 4th, 2001. (iii) August 8th, 1999

2.2.3 Sample 3. YYYY/MM/DD

The third pattern of writing dates from the Gregorian calendar is to reference the year first, followed by the month and lastly by the day (Cambridge, 2017). Examples are i. 2018-05-23. ii. 2001- 23rd May. All the above seven patterns of writing out dates from the Gregorian calendar were harvested for this research.

2.3 System library

Having collected the data, a data dictionary structure was created for the date number, the long form of the month and the short form of the month. The date number data dictionary structure maps the date number 1, 2, 3, up to 31 to its Yoruba number text. Some examples of the mappings are shown in table 1. The short month and the long month dictionary map the short form and the long forms of writing the months to a digit corresponding to the position of the month in the Gregorian calendar. This is further mapped to its Yoruba number text as shown in table 2. When a dated phrase is inputted into the system, the system library breaks it down into day, month, and year. Where appropriate, the input is stripped of all suffixes. Table 3 shows how samples of date are broken down and stored in the System

Table 1: date number dictionary structure

Date number	Cardinal Yoruba numbers
1	Kinni
2	Keji
5	Karuun
9	kesaan
10	Kewaa
25	keedogun

Table 2: Short months and long months

Short month	Long month	Digit representation
Jan	January	1
Feb	February	2
Mar	March	3
Apr	April	4
May	May	5
Jun	June	6
Jul	July	7
Aug	August	8
Sept	September	9
Oct	October	10
Nov	November	11
Dec	December	12

Table 3: System Library

Input	Day	Month	Year
2099-06-12	12	06	2099
23/05/1980	23	05	1980
06/30/1989	30	06	1989
15th Sept.,2011	15	09	2011
Oct. 17th,2200	17	10	2200
1980-05-23	23	05	1980
2009/01/02	02	01	2009

2.4 Model Formulation and Design

2.4.1 System model

The Date translation model is shown in Fig 1. This model engages several techniques which include the identification and extraction of date phrases from the Gregorian calendar. The model takes in different inputs as explained in section 2.2. The input goes through the first stage which is conversion. At the stage of conversion, any input is converted into its 'DD/MM/YYYY' equivalent pattern; date numbers with orders are stripped of their order, leaving only the numbers, names of months either in short form or full are converted to their numeric representation. For example, May 1st, 2012, becomes 01/05/2012.

The output of the conversion stage is fed as an input into the translator. At this stage, the input goes through number generation, morphological rules are applied, and an output is generated in standard Yoruba text.

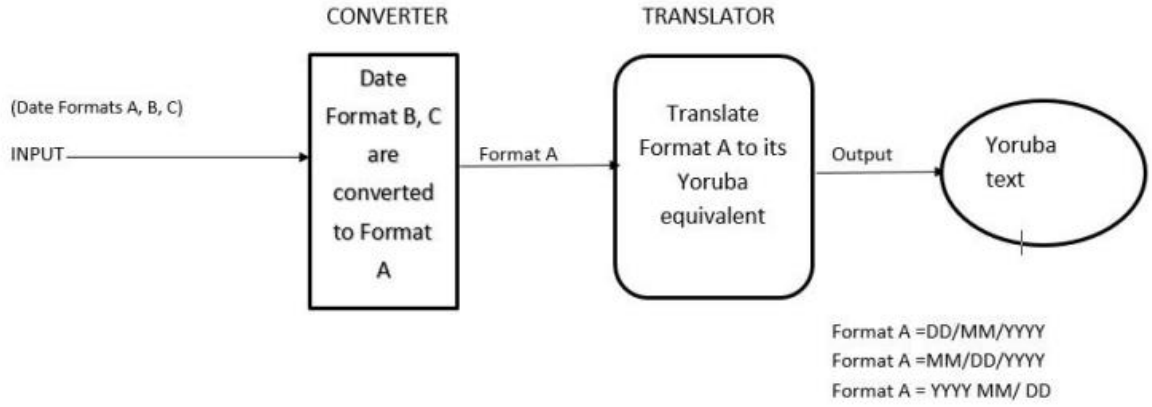


Figure 1: Date translation model

2.4.2 Model formulation

Hornby (2006) describes Grammar as the rules in a language for changing the form of words and joining them into sentences. These rules are useful to determine the category of words syntax in the two languages. Chomsky (2000) explained four different types of grammar rules, the most appropriate one for this research is the Push Down Automata. The Push Down Automata is a finite state Automata with a stack memory Context-Free Grammar (CFG).

The grammar production rule for Gregorian calendar dates is as follows:

G ::=< D, T, P, S > (1)

G implies the grammar,

D implies the non-terminal symbols,

T implies the terminal symbols and

P implies the production rule.

S is the start symbol.

D = Date, DP, MP, YP D,S, SF, MW_S, MW_I

T = list of D, List of S, List of SF, List of MW_S, List of MW_I

S = Date

Date → < DP >< O >< MP >< O >< Y P > (2)

Date → < MP >< O >< DP >< O >< Y P > (3)

Date → | < Y P >< O >< MP >< O >< DP > | (4)

DP → < D >< D > | < D >< D >< SF > (5)

MP → < D >< D > | < MW_S> | < MW_I> (6)

Y P → < D >< D >< D >< D > (7)
D → < 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 > (8)
O → < ., −, / > (9)
SF → < st, nd, rd, th > (10)
MW_s → < JAN, FEB, MAR, APR, MAY, ...DEC > (11)
MW_l → < JANUARY, F EBUARY, MARCH, APRIL, MAY, ...DECEMBER > (12)

The grammar production rule For Yoruba Date grammar, is as follows:
G ::=< D, T, P, S >
D = Date, S, S1 , NP, V P, N, Adj
T = OJO , OSU, ODUN
S = date

Date → SS (13)
S → S 1S 1 |NP V P (14)
S 1 → NP V P (15)
NP → NAdj (16)
V P → ϕ (17)
N → OJO,OSU,ODUN (18)
Adj → daynumber|monthnumber|yearnumber (19)

2.4.3 Pushdown Automata Model

the PDA machine execution table and the PDA model are shown in Table 4 and Figure 2 respectively

Table 4: Machine execution table to capture 21st May, 1967 - [ddsm, YYYY]

Input	Symbol	Current state	Translation	Next state	stack
ddsm,YYYY	-	q	-	q_1	Z
ddsm,YYYY	d	q_0	(q_0,d, -, push(DD), q_1)	q_1	DDZ
dsm,YYYY	ds	q_1	(q_1,ds, -, push(/), q_1)	q_1	/DDZ
m,YYYY	m	q_1	(q_1,m, -, push(MM), q_2)	q_2	MM/DDZ
,YYYY	,	q_2	(q_2,, , -, push(/), q_2)	q_2	/MM/DDZ
YYYY	YYYY	q_2	(q_1,YYYY, -, push(YYYY), q_3)	q_3	YYYY/MM/DDZ
-	-	q_3	(q_3, -, pop, -, q_4)	q_4	/MM/DDZ
-	-	q_4	(q_4, -, pop, -, q_4)	q_4	MM/DDZ
		q_4	(q_4, -, pop, -, q_4)	q_4	/DDZ
		q_4	(q_4, -, pop, -, q_4)	q_4	DDZ
		q_4	(q_4, -, pop, -, q_4)	q_4	Z

d → 21
ds → st
m → May
YYYY → 1967

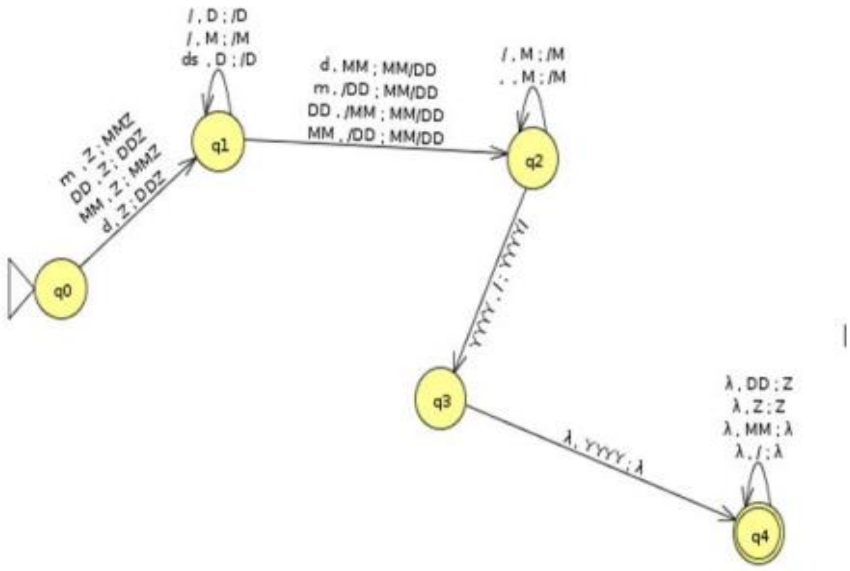


Figure 2: PDA model for the Gregorian calendar date phrase grammar

2.4.4. System Implementation

In this section, the software implementation is reported. Natural language toolkit (NLTK), PyQt4, file-based custom database and Python are the tools used for the software implementation. The NLTK (is a version of the python programming language) was used for sentence parsing. The NLTK was used to determine the correctness of the rewritten rules designed. The phrase structure grammar model was used for the rewrite rules. The PyQt4 (a version of Python) was used for the implementation of the graphical user interface (GUI). The GUI has two planes. The first plane is where the Gregorian calendar date text can be typed. The second plane displayed the translated Yoruba language date.

2.4.5 System design

The user can type any date phrase of any of the patterns from the Gregorian calendar into the first plane. The user clicks the translate button. If some of the words are not in the library, the system generates errors. Though only the administrator can add the new words to the various lexemes' databases. Figures 3 and 4 are the sample outputs of the date referencing translation system. Figure 3 is the date phrase with the pattern DD/MM/YYYY. Figure 4 is a dated phrase with the pattern **mm, dd (order), yyyy.**

3.0 RESULTS AND DISCUSSION

The implemented system was evaluated using the qualitative Mean Opinion Score Test. The Mean Opinion test is a test that has been used for decades in telephony networks to obtain the human user's view of the quality of the network. It is also used in the field of Artificial Intelligence (AI) to get an expert opinion on AI systems. The Mean Opinion Score has been used to evaluate this system. This evaluation is necessary so as to compare the translations from the system (machine translator) with those obtained from literate Yoruba speakers in terms of translation quality and accuracy. The translation quality and accuracy evaluation were determined by semantic accuracy, sentence orthography and syntax correctness. The machine and Experimental Subject Respondents were rated on how close their translations were to those of the human expert (speaker of Yoruba and professional translator). In essence, we were able to determine which one performed better, the machine or the human expert.

3.1 Questionnaire Design

The MOS evaluation was carried out to capture how well people can translate Gregorian calendar dates as it is represented in a different format to the corresponding Yoruba form. The questionnaire was developed to carry out the syntax, semantic and orthographic analysis of the responses received. In order to perform the evaluation, 50 questionnaires were administered and 45 were returned. The questionnaire contained 14 date phrases: 3 date phrases to cover the DD/MM/YYYY format, 3 date phrases to cover the MM/DD/YYYY format, another 3 phrases to cover the YYYY/MM/DD format, 2 date phrases to cover the DD (st, nd, th, rd) month, YYYY format, and the last 2 date phrase to cover the month dd(st, nd, rd, th), YYYY format.

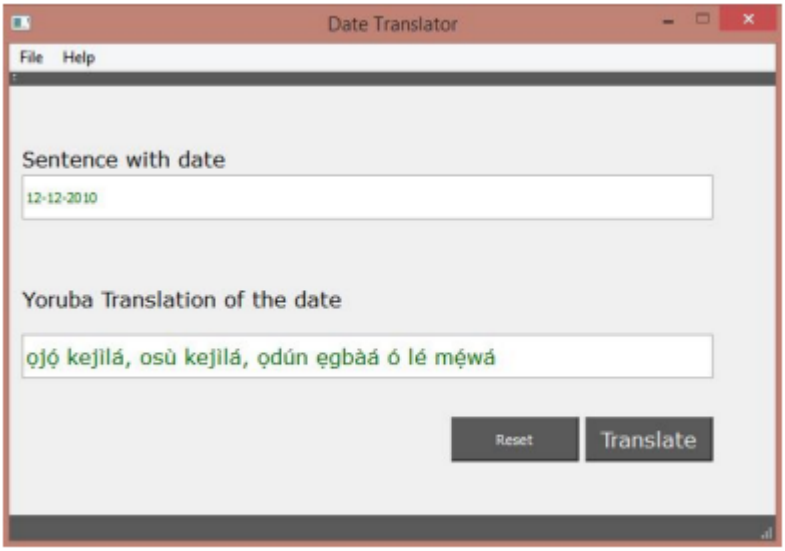


Figure 3: Translation for 12-12-2010

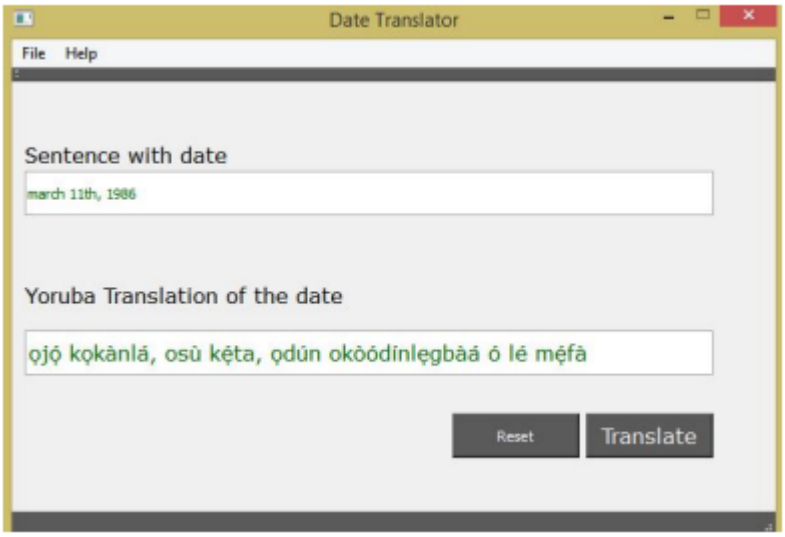


Figure 4: Translation for March 11th, 1986

3.2 System Evaluation Results

We evaluated the system by considering some different patterns/formats of date referencing in the Gregorian calendar. We considered the ESRs and the machine-translated sentences. The expert translations are used to control the choice of words between the ESRs and the machine-

translated date phrase. The three criteria used for the evaluation are word semantic accuracy, word orthography, and sentence syntax correctness. Word orthography accuracy means that a word in a sentence is properly tone-marked and under-dotted. For example, the word “Ọsu” has two syllables, “O” has one syllable with mid-tone and “su” has one syllable with a low tone on u. We evaluated each word using this approach and we scored the ESRs and the machine based on their ability to identify these syllables. The word semantic accuracy was evaluated by considering the meaning of each word within the sentence. All these and some other Yoruba writing styles were considered during the evaluation. The ESRs and the machine-translated phrases were scored. The phrase syntax correctness was evaluated based on the phrase structure of the target language (TL). We considered the position of each word in a phrase. A well-positioned word in a phrase attracts a good score.

3.2.1 Word Semantic Accuracy

The semantic analysis carried out is to determine the number of respondents who understands the correct meaning of the date phrase. This is done in order to validate their familiarity with the numbers and words used in the Gregorian calendar and how they are used in the context of a date. The scores of the machine translator are equal to that of the Expert except for one instance as is shown in Figure 5. The respondents' average score is 64.9 percent, the machine translator's average is 100 percent.

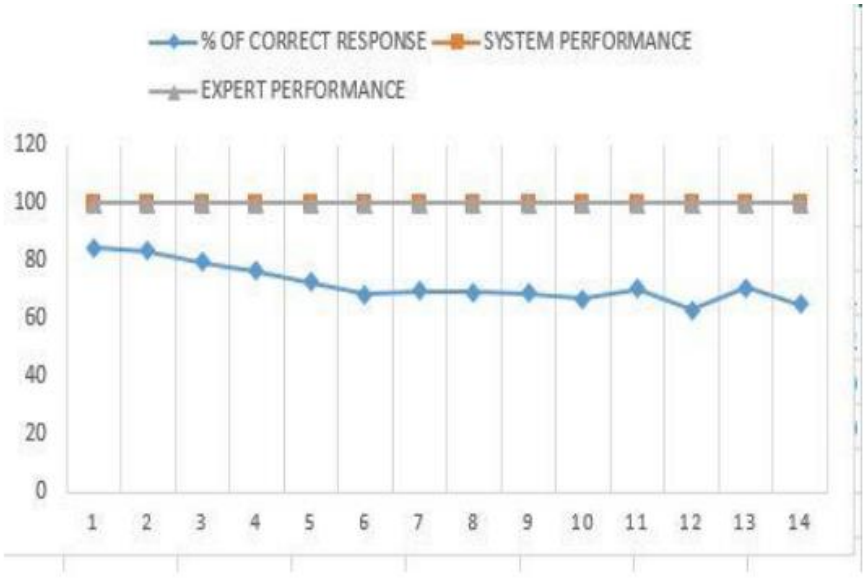


Figure 5: Word Semantic Analysis

3.2.2 Word Orthography Accuracy

The results present the extent of familiarity of the respondents with the orthography of the Yoruba language. Fourteen date phrases were extracted from the Gregorian calendar. It was observed that the experimental subject respondents (ESRs) scores are lower than the score of the developed system. The scores of the developed system are almost equal to that of the Expert except for one instance, where the score was about 90 percent as it is shown in Figure 6. Also, there are no instances where the ERSs' average score is equal to that of the developed system. The respondents' average score is 35 percent, while the developed system's average score is 99.1 percent. Another point to note is that only about two-thirds of the respondent correctly tone-marked their responses, leading to a reduction in the ESRs' average score. This was responsible for the performance recorded by these ESRs.

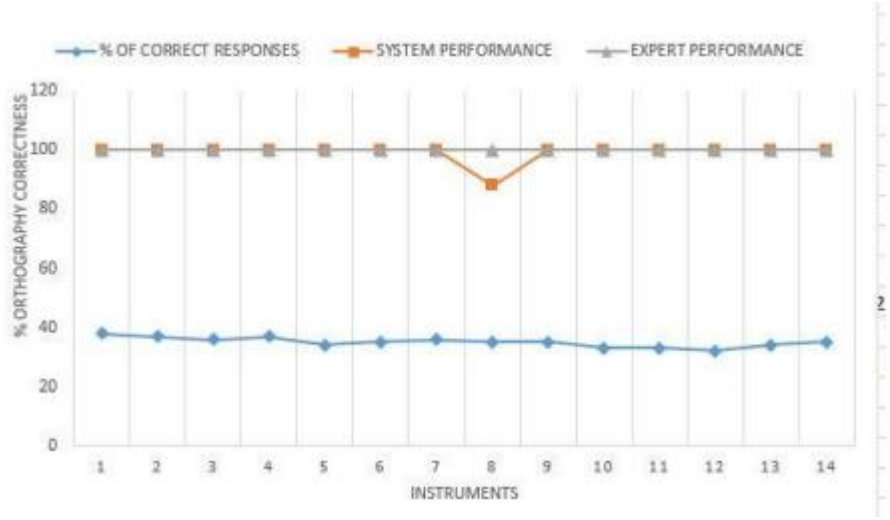


Figure 6: Word orthography accuracy

3.2.3 Word Syntax Accuracy

Word syntax accuracy is the analysis carried out to determine the familiarity of the respondents with the syntax of date phrases in Yoruba language. The result of the analysis indicates that about half of the respondents were not familiar with the syntax of date phrases in the Yoruba language. Figure 8 shows the results of syntax accuracy of 14 different date phrases as extracted from the Gregorian calendar. It was observed from the result that the developed system scored higher in terms of word syntax than the experimental subject respondents (ESRs). Also, the developed system's score is almost equal to that of the Expert except for a few instances where the scores are below that of the Expert. There are no instances where the ESRs scores are equal to that of the Expert as is shown in Figure 7. The respondents' average score is 56.9 percent, the developed system average score is 97 percent and the Expert average score is 100 percent. The analysis is done in Figure 5, 6 and 7 clearly show the comparison between the respondents and the developed system. This validates that the developed system performs relatively well compared to the respondents in terms of accuracy.

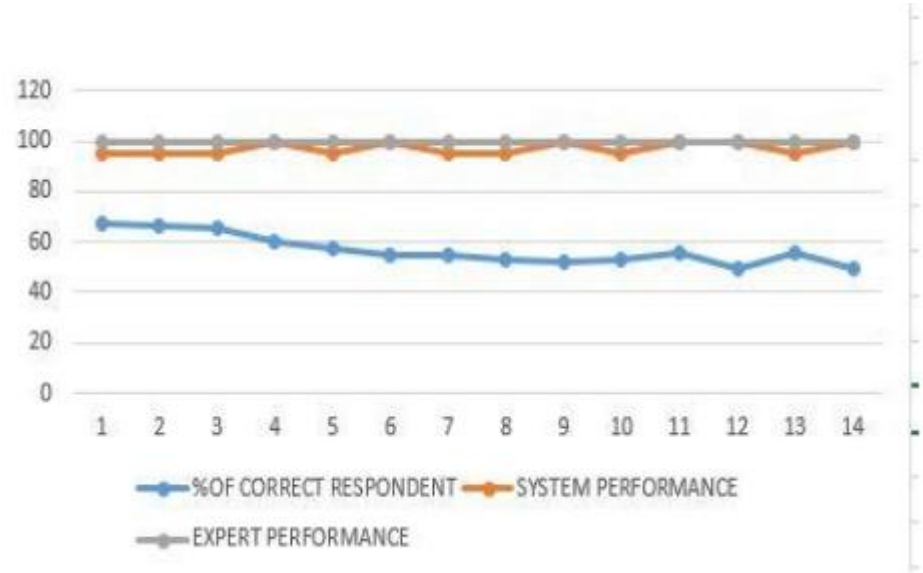


Figure 7: Word syntax accuracy

4. CONCLUSION

The Authors studied and analyzed date phrases from the Gregorian calendar with a focus on extracting the knowledge needed and the conditions required for its translation into the standard Yoruba language. It has been pointed out and discussed that there are various patterns of the Gregorian calendar date referencing. The authors discussed the phrasal context of the dates in the English language and how they are translated to the corresponding Yoruba text. Also, presented a grammar capturing the English date phrase and a model for the recognition. The processes involved in the translation have been discussed. The results of the study have been presented and discussed with a view to providing a key to the proper Gregorian calendar date referencing in the standard Yoruba language. In this study, the authors found out that: 1. There are various patterns of Gregorian calendar date referencing with peculiarities in the date compositions and arrangements, how they are translated from English phrasal context and the linguistic rules involved in their pronunciation. 2. The software developed was able to give 100% accuracy on tested data.

ACKNOWLEDGEMENT

The authors acknowledge the staff of Computer Intelligent Systems Research Group (CISRG) Unit of the Computer Science & Engineering Department, Obafemi Awolowo University for providing the needed computational knowledge.

REFERENCES

Akinade, O. O. and Odejobi D. A. (2014). Computational modelling of Yoruba numerals in a number-to-text conversion system. Journal of Language Modelling, 2(1):167-211.

Babarinde, O. (2013). The Linguistic Analysis of the Structure of the Yoruba Numerals, Journal of Literature, Languages and Linguistics, 45(1):78-88.

Cambridge, D. (2017). English Grammar Today, https://dictionary.cambridge.org/grammar/british-grammar/writing/dates, (Accessed on 04-01-2020).

Chomsky, N. (2000). The architecture of language, Oxford University Press

Cohen, E. L. (2000). Adoption and reform of the Gregorian calendar. Maths Horizon, 7(3): 5-11.

Coyne, G. V., Hoskin, M. A., and Pedersen, O. (1983). Gregorian Reform of the Calendar: In Proceedings of the Vatican conference to commemorate its 400thanniversary. edited by G.V. Coyne, M.A. Hoskin and O. Pedersen, Specola Vaticana.

Ekundayo. S. (1977). Vigesimal numeral derivational morphology: Yoruba grammatical competence epitomized. Anthropological Linguistics, 19(9):436-453.

Dershowitz, N. and Reingold, E. M. (2008). Calendrical calculations. Cambridge University Press.

Eludiora, S. I. (2014). Development of an English to Yoruba Machine Translation System, Unpublished Ph.D. Thesis, Obafemi Awolowo University.

Eludiora, S. I., Agbeyangi, A. O., and Ojediran, D. (2015). Word Sense Disambiguation in English to Yoruba Machine Translation System. Journal of Multidisciplinary Engineering Science and Technology, 2(7):1814-1819.

Eludiora. S.I, Agboola. A.F, and Abijo, T. V. (2013). Indigenous calendar system for time and date reckoning: Computational Analysis and Software Development, In Adegbite, W., Ogunsiji, A. and Taiwo, O. (Eds.) *Linguistics and the globalization of African languages for sustainable development: a festschrift in honour of Prof. Kola Owolabi*. Ibadan: Universal Akada Books Nigeria Limited. Pp 148-160)

Handelman, C. (1999). Calendar: Humanity's Epic Struggle to Determine a True and Accurate Year, A Journal of Medieval and Renaissance Studies, 30(1),205-208.

Harper, D. (1998). A brief history of the calendar. (www.obliquity.com. Accessed on 10-04-2020).

Jemiriye, T. and Awosusi, A. (2007). The Concept of Planning in Yoruba Traditional Religion. Soc. Sci, 2(3):351-356.

Johnson, S. (1966). The history of the Yorubas: From the earliest times to the beginning of the British protectorate. Cambridge University Press.

Longe, O. (2009). A Yoruba Decimal Number System. Bookbuilders.

Longstaff, A. (2005). Calendars from Around the World. Greenwich: National Maritime Museum.

Reingold, E. M. and Dershowitz, N. (2002). Calendrical Tabulations, 1900-2200, Cambridge University Press.

Snell-Hornby, M. (2006). *The Turns of Translation Studies: New paradigms or shifting viewpoints?* (Vol. 66). John#

Benjamins Publishing
Theodossiou, E., Manimanis, V. N., and Dimitrijevic, M. S. (2011). Six calendar systems in European history from the 18th to the 20th Century. Bulgarian Astronomical Journal, 16:109-129.
Thesaurus.com (2018). Dictionary.com Unabridged, http://www.dictionary.com/browse/day, (accessed on 10-03-2020).
Tripathi, S. and Sarkhel, J. K. (2010). Approaches to machine translation, Annals of Library and Information Studies, 57 (388-393).
Wande, A. (1976). Ifa: An Exposition of Ifa Literary Corpus, Oxford University Press.
Zaslavsky, C. (1999). Africa counts: Number and Patterns in African cultures, Chicago Review Press.