IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 8, August 2015.
www.ijiset.com
ISSN 2348 – 7968

Digital Yorùbá Corpus

Olutola Fagbolu[1], Akinwale Ojoawo[1], Kayode Ajibade[2] and Boniface Alese[3]

[1] Department of Computer Science, The Polytechnic Ibadan, Oyo State, Nigeria
[2] Department of Art and Design, The Polytechnic Ibadan, Oyo State, Nigeria
[3] Department of Computer Science, Federal University of Technology Akure, Akure, Ondo State, Nigeria

Abstract

With the introduction of the Yorùbá corpus, translation of words or phrases from either English language or Yorùbá language to its target language would become easier, It is a large and structured set of texts usually stored and processed in electronic form. They are used to do statistical analysis and hypothesis testing, checking occurrences or validating linguistic rules within a specific language. The corpus aid in Natural Language Processing and Machine Translation, programming tools employed for mobile platform are JDK 6, Apache Ant 1.8 or later, Android Software Development Kit, Eclipse Integrated Development Environment, Android Developer and Android Studio while latest technologies such as PHP, Mysql, .net, Mssql 2005, 2008, Ajax techies, C# and Apache are used to develop its web platform. It brings the usefulness of Information Technology and graphic design to the doorstep of non- Yorùbá people who may wish to visit Yorùbá nations or learn how to converse, make friends with Yorùbá indigenes or transact business with Yorùbá people that are not literate.

Keywords: *Yorùbá corpus, English language, hypothesis testing, Natural Language Processing, Machine Translation, Information Technology, Graphic design.*

1. Introduction

Yorùbá is a dialect of West Africa with over 50 million speakers. It is a member of Niger-Congo family of language and it is spoken among other languages in Nigeria, Togo, Benin and partly in some communities in Brazil, Ghana, Sierra Leone (where it is called Oku) and Cuba (where it is called Nago) (Bamgbose, 1965). Yorùbá is one of the three major languages in Nigeria and language being the principal means used by human beings to communicate with one another (Ogunbiyi, 2003; Babalola, 2010); it is spoken and considered as the third most spoken native African language. Yorùbá language has ancestral speakers who according to their oral traditions is Oduduwa (son of Olúdùmarè), the supreme god of the Yorùbá (Biobaku, 1973). Yorùbá first appeared in writing during the 19th century and the first publications were a number of teaching booklets produced by John Raban in 1830 – 1832 and another major contributor to orthography of Yorùbá was Bishop Samuel Ajayi Crowther (1806 – 1891) who studied many of the languages of Nigeria (Oyenuga, 2007), he wrote and translated some of the Yorùbá phrases and words. Yorùbá orthography appeared in about 1850 although with many inherent changes since then. In the 17th century Yorùbá was written in the Ajami script (Ogunbiyi, 2003) and major development in the documentation of Yorùbá words and phrases were done by Anglican (CMS) missionaries that were working in places like Sierra Leone, Brazil, Cuba and they assembled the grammatical units in Yorùbá together which were published as short notes (Adetugbo, 2003), in 1875 Anglican communion organized a conference on Yorùbá orthography. Johnson (1921) remarked that several fruitless efforts had been made to either invent new characters or adapt the Arabic, which was already known to Moslem Yorùbá. Finally, Roman character-based alphabets that were acquainted with Anglican (CMS) missionaries were adopted (Johnson, 1921).

Yorùbá anthology can be traced to the publication of several Yorùbá newsprints in Lagos, Nigeria in 1920s such as Eko Akete in 1920 with Alaagba Isaac B Thomas as the editor, Akede Eko in 1922, Eletiofe in 1925 with E.A Akintan as the editor and many more which enhance the numerous usage of the language in the area of economic, political diplomatic and cultural relations. Yorùbá corpus will enhance intending visitors or learners of the language to tap into numerous advantages to be derived in its usage but nonexistence of English – Yorùbá corpus can inhibits.

Digital Yorùbá corpus may contain texts in a single language (*monolingual corpus*) or text data in multiple languages (*multilingual corpus*). Multilingual corpora that have been specially formatted for side-by-side comparison are called *aligned parallel corpora*. In order to make the corpora more useful for doing linguistic research, they are often subjected to a process known as annotation. An example of annotating a corpus is part-of-speech tagging, or *POS-tagging*, in which information about each word's part of speech (verb, noun, adjective, etc.) is added to the corpus in the form of *tags* as in fig.1. Another example is

ISSN 2348 – 7968

indicating the lemma (base) form of each word. When the language of the corpus is not a working language of the researchers who use it, interlinear glossing is used to make the annotation bilingual. Some corpora have further *structured* levels of analysis applied. In particular, a number of smaller corpora may be fully parsed. Such corpora are usually called Tree banks or Parsed Corpora. The difficulty of ensuring that the entire corpus is completely and consistently annotated means that these corpora are usually smaller, containing around one to three million words. Other levels of linguistic structured analysis are possible, including annotations for morphology, semantics and pragmatics.

# PoS	name	Total PoS
1	Definite Article	2
9	Noun Modifier	8
10	Adjective	1164
11	Noun	3632
12	Pronoun	40
13	Verb	1170
14	Preposition	36
15	Adverb	208
16	Conjunction	10
17	Interjection	4

Figure 1: Analysis of POS tagging

Yorùbá corpus will covers all must-know vocabulary and words for typical situations, from making reservations and getting around to shopping and obtaining any kind of help and includes quick access to the required topic or subject matter. It will provide easy-to-read transliterations to help travellers with pronunciation, as well as useful travel tips and includes "expressions you will hear" sections with minimum of 1,000-word dictionary. It will offer an appendix with abbreviations, national holidays, distances between cities, conversion tables, embassies and consulates, useful phone numbers, and common signs and notices in Yorùbá nations.

2.0 Research Objectives

There are series of difficulties in building Yorùbá corpus but because of its immense benefits to Machine Translation and Natural Language Processing, its objectives are:-

1. To design a graphical representation for electronic collection of Yorùbá words.
2. Implement the graphical interface created on any digital platform (Mobile phones and Web).
3. Evaluate the performance and usefulness of Yorùbá corpus.

3.0 Research Materials and Methods

A comprehensive review of manually composed corpora such as French, German, Spanish, Italian and other languages were appraised, including some corpora in digital format. Also, fundamental features of web and internet were employed in this research work which enhance the newly develop digital Yorùbá corpus that was developed in three modules viz:

1. The text processing module
2. Prosody prediction module
3. Concatenation signal processing module

The Yorùbá words are taken as input in the text processing section and are converted into a sequence of phonetic transcriptions (phonemes, dip hones, demi-syllables or syllables) with high level prosodic descriptions such as stress, focus, breaks etc. Then an appropriate set of prosodic contours such as fundamental frequency, duration and amplitude is calculated by prosody module. At last, a pitch and duration modification algorithm, such as PSOLA is applied to preset units to guarantee that the prosodic features of synthetic speech meet the predicted target values. These systems have the advantages of flexibility in controlling the prosody.
The Digital Yorùbá corpus were developed using the latest technologies such as PHP, MySQL, Microsoft.Net Framework, MsSQL 2005, 2008, Ajax techies, VB.net, C# and lots of cloud computing services all on both windows and Linux platform. This will allow the corpus to work on a wide range of platforms including internet, mobile phones and other local websites. All the aforementioned technologies will be integrated together in such a way that

ISSN 2348 – 7968

the front-end processor of the software will allow the user to enter or make a selection of one or more word as desire, determine in which language will the pronunciation be made, give its meaning and possibly translate if the need be. After these, the back-end processor will then analyze all the choices made by the user using the front-end processor to produce the desire output in the alternative language specified by the user and then give the user the choice to click and listen to the pronunciation of the output produced. The preferred approach in today's technology - Prototyping would be employed that is a working sample design method that are not as time consuming as paper and pencil design and less prone the considerable errors. It will encourage and require active participation of end users and would be friendly.

The first choice in choosing to implement a design is to decide on which platform the design was to be implemented and the language which the design will be implemented in. The following criteria were at the forefront of my mind in choosing the platform.

Productivity: It was apparent that the digital Yorùbá corpus would require an extensive amount of software engineering in a very short time scale. It would not be easy to build a reliable product which met the project specifications in time, and thus my most important criterion for choosing a platform was the productivity afforded to me by that platform.

Availability: I wanted a platform that was widely available so that a large number of people would be able to use my software if they so desired.

Speed: The development of this research work would require exponential time and memory. However there would be a constant factor involved which would depend upon the programming language and platform used and then once latest technologies would be employed in its development, factor of speed will be properly catered for.

3.1 Design Elements

The design of Digital Yorùbá Corpus can be categorized into the following elementary units:-

i. Inputs
ii. Outputs
iii. Database
iv. Procedures

A. INPUTS

It is a direct dependent of output, the design of the corpus translation, presentation of what various programs will do as part of the corpus, considerations are in data (created corpus), volume of what is supplied to the computer (input), type of input, available media and design layouts of input (interface).

B. OUTPUTS

It is the expected or what is required from the corpus before deciding on how to set about translating words or texts. Design of forms for query, reports for translation or any other report as the case may be is most evident in system's output.

C. DATABASE

It coordinates and controls the activities that bring about the required translation by linking the input and output. Considerations are on the volume of data, its storage media, its retrieval, method of access and organization of Database, its security and layouts.

D. PROCEDURES

In layman, it is the way of linking all the required steps together to produce the required translation, this will specify the responsibilities of the several modules making up the code. Modularization of code and specific action that are carried out by them are spelled out here.

3.2 System Construction

The transfer model is made up of three stages: analysis, transfer and generation as depicted in figure 2.

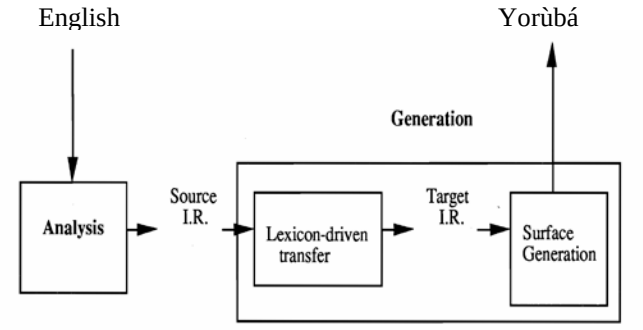


Figure 2: System Outline

In Example Based Machine Translation the search and matching process replaces the analysis stage, transfer is replaced by the extraction and retrieval of examples and recombination takes the place of the generation stage. Matching is the first task in translating Yorùbá words to English words or vice versa which takes the source language to be translated and find the example or set of examples which most closely match it. The procedure employed in searching depends on the way the examples are stored and matching process may be more or less linguistically motivated. In matching and retrieving phase, the input text is parsed into segments of certain granularity. Each segment of the input text is matched with the segments from the source section of corpora at the same level of granularity, the matching process may be syntactic or semantic or both depending on the domain. In

ISSN 2348 – 7968

syntactic, matching is done by the structural matching of the word while in semantic matching, semantic distance is found out between the words and the corresponding translated segments of target language are retrieved from the second section of the corpora. The mechanism for best match retrieval has the following tasks:-

i. Determine whether the search is for matches at sentence or sub sentence level that is determining the text unit
ii. The similarity between two text units.

Having matched and retrieved a set of examples with associated translations, the next step is to extract from the translations appropriate fragments (alignment or adaptation).

Recombination combines the aligned fragments so as to produce a grammatical target and output which is arguably the most difficult step in EBMT and can be done by either identifying which portion of the associated translation corresponds to the matched portions of the source text or recombining these portions in an appropriate manner.

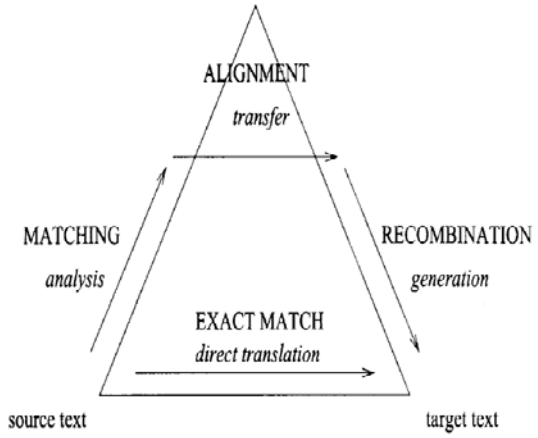


Figure 3: The Vanquois Triangle Modified for EBMT

Alignment model are currently used by almost all translation systems, the fundamental motivation of alignment modelling for machine translation is that, given a source sentence e and target sentence f and now wish to define a correspondence between the words of e and f. A link between two words can be viewed as an indication that they are translations of each other, each link can be represented as a pair (i, j) with $1 \leq i \leq I$, $1 \leq j \leq J$. We write A for the set of such links; each element of A can be viewed as an undirected arc in a graph with I + J nodes, one node for each word of each sentence.

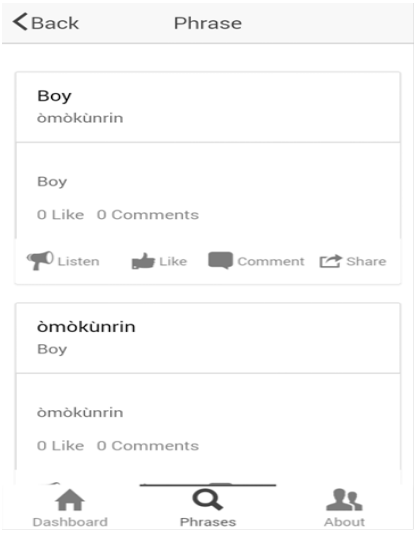
Every webpage developed for portable devices then had to be designed through XML (Sanborn and Lattig, 2000). This led to the improved language of the TML through the introduction of the Extensible Hypertext Markup Language (XHTML), which emerged as the best means of reducing the gap between HTML and WML to develop web content (Lee, 2001). XHTML was thus chosen as the most appropriate principal web development language for the software design process in this research work. Data used in the system was stored on a database. The interaction between the software application and the database was facilitated through PHP. PHP is known as a hypertext preprocessor. It is a scripting language that ensures the connection between a database and a website. It contains wording or coding.

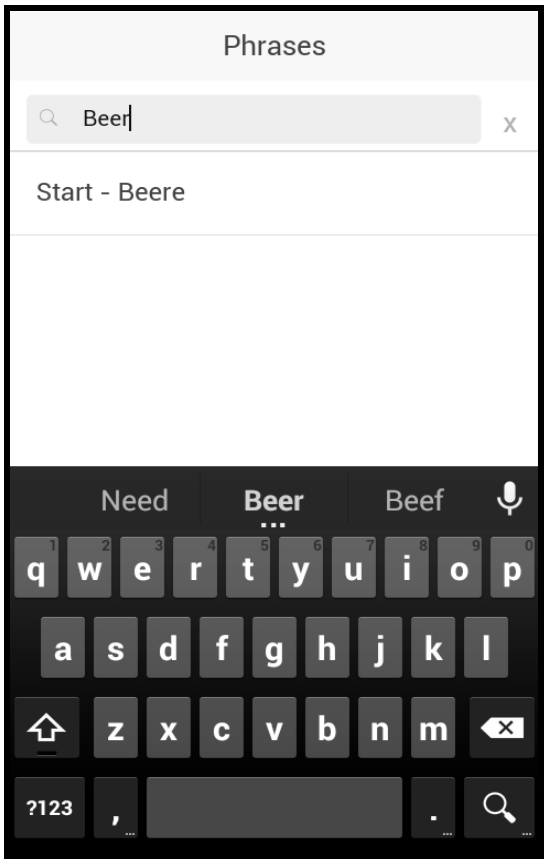
4.0 Implementation of Digital Yorùbá Corpus

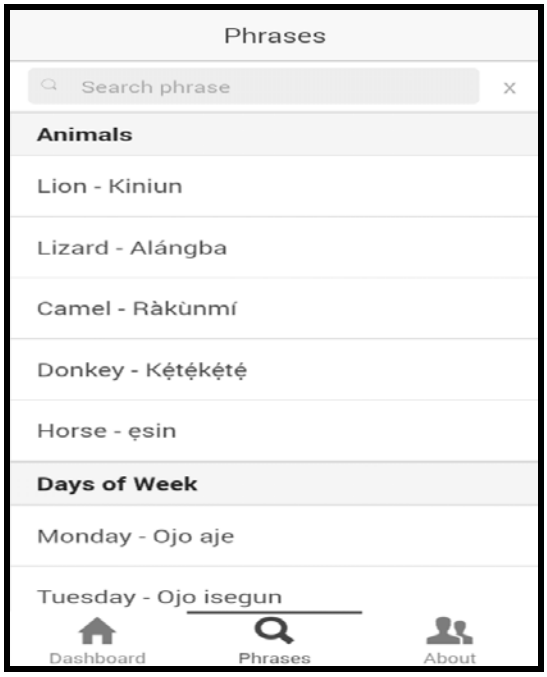
This research constructed an electronic database consists of collection, computerizing and checking of corpus data. Its activities include identification of data sources, defining criteria for sampling, computerizing the data and checking and revising both the corpus data and text headers regularly. Annotation of the corpus involves the steps of developing a parts of speech tagging software for Yorùbá. The next stage is text annotation and parts of speech tagging, preparing a lexicon of roots and affixes for parts of speech tagger; developing a parts of speech tagger that would tag at least 6000 lexemes; testing and upgrading the tagger. In developing the corpus interface, we will consider query criteria, visual design and cross-platform compatibility of the software that will operate the Corpus. Finally, a user friendly graphic interface that would ease access to the corpus is developed; defining query criteria and query fields that would appear in the interface in accordance with the fundamental design of the corpus; checking and upgrading compatibility of the interface (Windows, Linux, MacOS, SunOS, and Pardus) for different users and different operating systems.

Different samples of the use of constructed Yorùbá Corpus for searching numerous phrases or words as the case may be are shown below on both mobile platform and web platform.

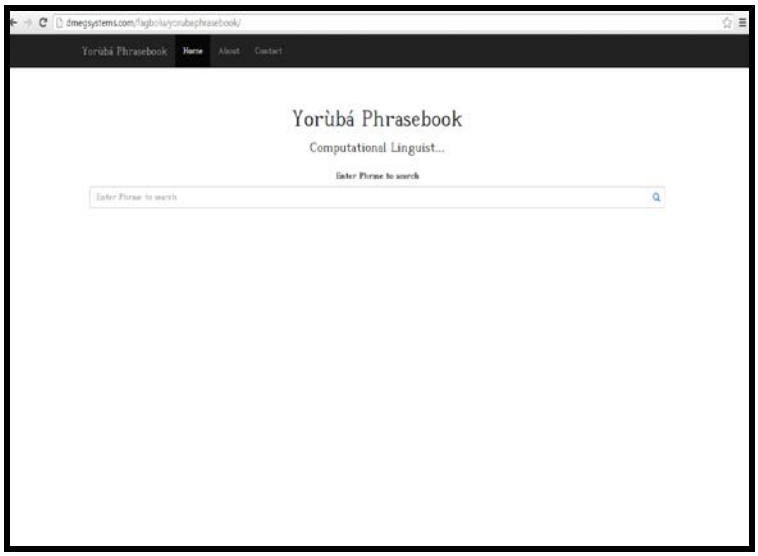
ISSN 2348 – 7968



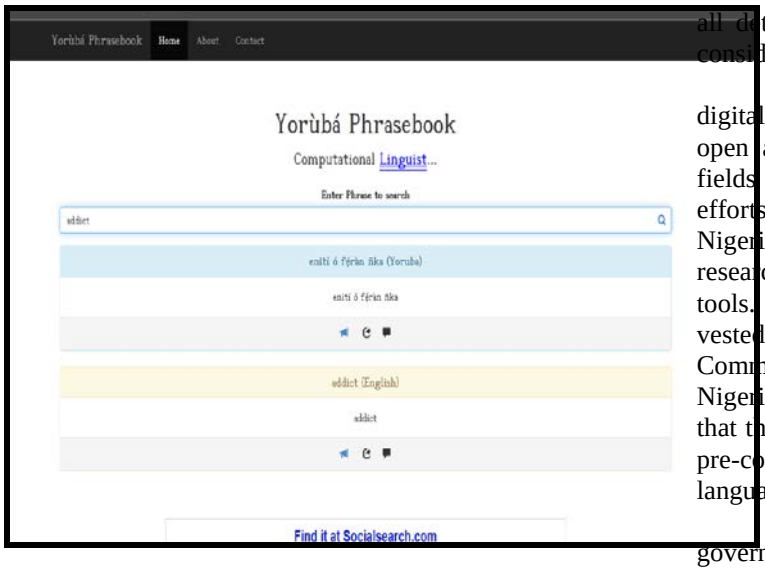




These are interfaces for mobile platform



ISSN 2348 – 7968



These are interfaces for web platform.

Any phrases queried from any of these platforms would make use of the constructed Yorùbá corpus to translate the phrase or word into its equivalent in Yorùbá language.

5.0 Conclusion and Recommendation

The aim and objectives were achieved, its design and implementation were carried out on web and mobile platforms. During the course of this research, Yorùbá language which is tending towards extinction were reawakened, this platforms promote indigenous African languages to native and non-native speakers, tourists and NYSC corps from another ethnic groups in Nigeria that may want to associate with Yorùbá people during their visit or their service year.

It serves as one of the collective efforts to expand words, phrases and expressions in Yorùbá language and make Yorùbá language normal and natural means of spoken and written communication for whoso ever desire, consequently the language will be more popular, gain value and prestige and no one will denigrate it.

New words and expressions that are suitable for situations, legislation, science, engineering, commerce, computing, mass communication and other sphere of life will be created in a large number via constructed corpus. In this research, several materials were consulted and it was realized that development of digital Yorùbá corpus one of the complex architecture similar to English – Yorùbá Translator and Digital Dictionary. Developing digital Yorùbá corpus proved to be a hard nut to crack if all details and expectations of machine translation were considered.

One of the challenges is the non-availability of digital text with diacritics and data gathering requires an open attitude and concerted efforts from researchers in fields related to study of Yorùbá. It is recommended that efforts at building corporal for national languages in Nigeria be encourage as this will reduce time spent on research and development of languages computational tools. There is also an urgent need for the authorities vested with development of Information and Communication Technologies and Languages in the Nigeria to present request to UNICODE consortium so that the characters used in the orthographies will come as pre-composed characters. This will aid the use of such languages using the computer platform.

Also research work like this need to be funded by government, institutions and corporate bodies because of their attendant benefit to the society at large, Ministry of Tourism and Ministry of Science and Technology would find its significance utmost for those who favourably disposed to the language. Multi domain of corpus created in this research work can be deployed and used by other researchers.

5.3 Future Work

More study are needed to improve the number of words and thereby achieving 100% intelligible and accurate Digital Yorùbá Corpus. This research work can further be extended to cater for other Nigerian local languages such as Igbo, Hausa, Ibibio and so on.

References

[1] Abiola, O.B et al. (2013). A computational model of English to Yorùbá Noun-phrases Translation System, FUTA journal of Research in sciences Vol 1, pp.34-43

[2] Abraham, R. C. (1958). Dictionary of Modern Yorùbá (Yorùbá-English). University of London Press, London.

[3] Ade Ajayi, J. F. (1960). How Yorùbá was reduced to Writing. ODUA Journal of Yorùbá, Edo and Related Studies, 8:49–58.

[4] Adeoye, O.B. (2012). ”A Web-Based English to Yorùbá Noun-Phrases Machine Translation System”, M.Tech Thesis, Federal University of Technology, Akure, Nigeria.

[5] Adetugbo, A. (2003). The Yorùbá Language in Yorùbá History.

[6] Ahmed, A. and Seong, D. (2006).”Sign Writing on Mobile phones for the Deaf ”, Proceedings of the 3rd International Conference on Mobile Technology, Applications and Systems-Mobility, Bangkok, 25-27 October, pp. 28.

[7] Alake, C. A. (2000). Early Descriptions of the Yorùbá Language: The Work of Samuel Ajayi Crowther. In P., D., L., J., P., S., and P., S., editors, The History of Linguistic and Grammatic Praxis. Proceedings of the XIth International Colloquium of the Studienkris ”Geschichte der

Sprachwissenschaft". Leuven, 2nd–4th July 1998, page 427–443. Peeters Publishers.

[8] Akshi, K. (2005). "Design and Development of Translator's Workbench for English to Indian Lang.", Translation J 9(3). Retrieved December, 2010. Source at: http://accurapid.com/journal/33TWB.htm. www.nlp.hive five.com/entity/profile/andrew-mccallum. Retrieved, 2010.

[9] Armentano-Oller, C. and Forcada, M. (2006). Open-Source Machine Translation between small languages: Catalan and Araneseoccitan, in strategies for developing machine translation for minority languages (5th SALTMIL Workshop on Minority Languages), pp. 51-54 (organized in conjunction with LREC 2006 (22-28.05.2006))

[10] Asahiah, F.O. (2014). Development of a Standard Yorùbá digital text automatic diacritic restoration system. Ph.D thesis, Obafemi Awolowo University Ile-Ife, Nigeria

[11] Awobuluyi, O. (1978). "Essentials of Yorùbá Grammar" Published by Oxford University Press Nigeria, Iddo Gate Ibadan.

[12] Awoyale, Y. (2008). The LDC Corpus Catalog (Global Yorùbá Lexical Database v.1.0). Web Available at http://www.language-archives.org/item/oai: www.ldc.upenn.edu:LDC2008L03.

[13] Babalola, A. (2010). Yorùbá Literature. In Andrzejewski, B. W., Pilaszewicz, S., and Tyloch, W., editors, Literatures in African Languages: Theoretical Issues and Sample Surveys. Cambridge University Press, reissue, reprint edition. ISBN 0521126258, 9780521126250.

[14] Bamgbose, A. (1965). Yorùbá Orthography: A Linguistic Appraisal with Suggestions for Reform. University Press, Ibadan.

[15] Bamisaye, O.T. (2000). "Essentials of English Syntax" Department of English, University of Ado-Ekiti, Nigeria. Published by Balfak Educational Publisher, Ado-Ekiti, Ekiti State

[16] Bar-Hillel ,Y. (1960). A demonstration of the non feasibility of fully automatic translation. Appendix III of 'The present status of automatic translation of languages', Reprinted in Bar- Hillel Y, 1964. Language and Information, Reading, Mass. Addison-Wesley, pp 174-179

[17] Batra, K. and Lehal, G. (2010). Rule based Machine Translation of Noun Phrases from Punjabi to English, *International Journal of Computer Science* Issues, Vol 7, Issue 5, pp.409-413

[18] Bennett, W. (1994)."Machine translation in North America". Encyclopedia of languages linguistics, ed. R.E. Asher and J.M.Y. Simpson, 5, 2332-2338. Oxford: Pergamon Press

[19] Bennett, W. and Slocum, J. (1985). The LRC Machine Translation System, Computational Linguistics 11, 111-121 reprinted in Slocum (1988) pp. 111-140

[20] Biobaku, S. O. (1973). Sources of Yorùbá History, London, Oxford Clarence Press

[21] Blank, D. (1998). "Definition of Machine Translation".Source at: http://www.macalester.edu/courses/ russ65/definiti.htm

[22] Bod, R. (1995). Enriching Linguistics with Statistics: Performance Models of Natural Language. PhD thesis, Universiteit van Amsterdam, The Netherlands.

[23] Boitet, C and Nedobejkine, N. (1981). Recent Developments in Russian-French Machine Translation at Grenoble. Linguistics 19, 199-271

[24] Booth, A.D et al. (1958). Mechanical Resolution of Linguistic Problems, New York, Academic Press.

[25] Booth, A.D. (1967). "The history and recent progress of machine translation". Aspects of translation, 88-104. London: Secker and Warburg

[26] Chowdbury, G. (2005). "Natural Language Processing". Department of Computer and Information Sciences, University of Strathclyde, Glasgow G1 1XH, UK. Source at: www.infotoday.com/books/assist/artist 37.shtml. Retrieved 2010.

[27] Corbi-Bellot, A., Forcada, M., Ortiz-Rojas, S., Prez-Ortiz, J., Ramirez-Sanchez, G., Sanchez-Martinez, F., Alegria, I., Mayor, A. and Sarasola, K. (2005). An open-source shallow-transfer machine translation engine for the Romance Languages of Spain. In Proceedings EAMT Conference, pp. 79-86.

[28] Costa-Jussa, M.R et al. (2012). Study and comparison of rule-based and statistical Catalan- Spanish Machine Translation System, *Computing and Informatics* Vol 31, pp. 245-270

[29] Dasgusta, S. (2006). Examining the Role of Linguistic Knowledge Sources in the Automatic Identification and Classification of Reviews, In Proceedings of COLING/ACL, Chinese Academy of Science.

[30] De Pauw, G., de Schryver, G.-M. and Wagacha, Peter Waignjo. (2009a.) A corpus-based survey of four electronic Swahili–English bilingual dictionaries. Lexikos, 19, p. 340–352.

[31] De Pauw, G., Wagacha, P.W. and de Schryver, G.-M. (2009b.) The SAWA corpus: a parallel corpus English - Swahili. In G. De Pauw, G.-M. de Schryver & L. Levin (Eds.), *Proceedings of the First Workshop on Language Technologies for African Languages (AfLaT 2009). Athens, Greece*: Association for Computational Linguistics, pp. 9–16

[32] Elliston, J. (1979). Computer-aided translation: a business viewpoint. In Barbara M. Snell (ed.) Translating and the Computer, Amsterdam, North-Holland, 149-158

[33] Eludiora, S.I. (2014). *Development of English to Yorùbá Machine Translation System*, Ph.D thesis, Obafemi Awolowo University, Ile-Ife, Nigeria

[34] Fagbolu, O.O, Alese, B.K. and Adewale, O.S (2014) Development of a Digital Yorùbá Phrasebook on a Mobile Platform, Nigerian Computer Society (NCS) 25th Annual Conference-Building a knowledge-based economy in Nigeria: The Role of Information Technology, Nike Lake Resort Enugu,(page 13 – 19 in the conference proceedings Vol. 25)

[35] Gary Schneider and James Perry. (2001). Electronic Commerce, Canada, Learning Inc.

[36] Howard, J. (1982). "Analyzing English an Introduction to Descriptive Linguistics" City of Birmingham Polytechnic, United Kingdom.www.bcu.ac.uk/pme/schoolof english/staff/howar d_jackson. Retrieved 2011.

[37] Hutchins, W. J. and Somers, H. L. (1992). An Introduction to machine translation, London, Academic Press

ISSN 2348 – 7968

[38] Johnson, D. (1985). Design of a robust, portable Natural Language Interface Grammar. IBM Research Reports RC 10867.
[39] Johnson, S. (1921). The History of the Yorùbá. C.M.S. Nigeria Bookshops, Lagos.
[40] Jurafsky, D. and Martin, J. H. (2000). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistic and Speech Recognition. Prentice-Hall.
[41] Krashen, S. and Terrell, T. (1983). The Natural Approach: Language acquisition in the classroom, Alemany Press, Great Britain.
[42] Kumolalo, F. O., Adagunodo, E. R., and Odejobi, O. A. (2010). Development of a Syllabicator for Yorùbá Language. *In Proceedings of OAU TekConf*, September 5-8, 2010, pages 47–51, OAU, Ile-Ife, Nigeria.
[43] Locke, W. and Booth, A. (1955) "Historical Introduction" Machine translation of languages: fourteen essays, ed 1-14, Cambridge, Mass., Technology Press of the Massachusetts Institute of Technology.
[44] Nagao, M. (1998). Language Engineering: The real bottle neck of natural language processing, Proceedings of COLING 86, 97-103
[45] Nirenburg, S., Raskin, V. and Tucker, A. (1987). The Structure of Interlingua in TRANSLATOR In: Machine Translation, Theorectical and Methodological Issues, Cambridge/London/New York: Cambridge University Press, pp. 90-103
[46] Nirenburg, S. (1989). Knowledge-based Machine Translation, Kluwer Academic Publishers 4, 5-24
[47] Ogunbiyi, I. A. (2003). The Search for a Yorùbá Orthography since the 1840s: Obstacles to the Choice of The Arabic Script. Sudanic Africa: A Journal of Historical Sources, 14:77–102.
[48] Oloruntoyin, S.F. (2014). Development of Yorùbá Language Text-to-Speech Learning System. *International Journal of Scholarly Research Gate 2 (1)*,pp 19-36 Source at: http:// onlinejournals.oscij.com
[49] Oyenuga, S. (2007). Learning Yorùbá Web Available at www.YorùbáForKidsAbroad.com
[50] Oyenuga, S. and Oyenuga, T. (2007). Learn Yorùbá in 27 days, Canada, Saskatoon, Gaptel Innovative Solutions Inc
[51] Papegaaij, B., Sadler, V. and Witkam, A. (1986). Word Expert Semantics: an Interlingual Knowledge-based Approach (Distributed Language Translation I), Dordrecht/Riverton:Foris
[52] Raskin, V. (1974). On the feasibility of fully automatic high quality machine translation, American Journal of Computational Linguistics 1, microfiche 9
[53] Sachnine, M. (1997). Dictionnaire Yorùbá-Francais, suivi d'un index Francais-Yorùbá, Paris: Karthala, Grammars and Sketches
[54] Salminen, T. (1999). UNESCO Red Book on Endangered Languages. UNESCO.http://www.tooyoo.l.u-tokyo.ac.jp/archive/RedBook/index.html.
[55] Sanchez-Martinez, F., Forcada, M. and Way, A (2009). Hybrid rule-based-example-based MT:Feeding Apertium with Sub-Sentential translation units. In Proceedings of the 3rd Workshop on Example-Based Machine Translation, pp. 11-18, November 12-13, 2009, Dublin, Ireland.
[56] Sanchez-Martinez, F. and Way, A. (2009). Marker-based filtering of bilingual phrase pairs for SMT. In Proceedings of the 13th Annual Conference of the European Association for Machine Translation, pp. 144-151, May 14-15, Barcelona, Spain.
[57] Shaheen, M. (1991). Theories of Translation and their Applications to the Teaching of English/Arabic-Arabic/English Translating, PhD thesis, University of Glasgow, U.K
[58] Slocum, J. (1985). "A survey of machine translation: its history, current status and future prospects".Computational Linguistics 11. 1-17.
[59] Somers, H. (1990). Current research in Machine Translation. Third International Conference in Theorectical and methodological Issues in Machine Translation of Natural Languages (Austin, TX), pp. 1-12
[60] Stroppa, N., Groves, D., Way, A. and Sarasola, K. (2006). Example-Based Machine Translation of Basque language. Proceedings of AMTA, Cambridge, MA, USA, pp232-241
[61] Tinsley, J. (2010). Resourcing Machine Translation with Parallel Treebanks Ph.D thesis, Dublin City University, Ireland.
[62] Toma, P. (1997). My first 30 years with MT. *MT Summit VI, Machine Translation: Past, Present, Future, Proceedings*, 29 October -1 November 1997, San Diego, California, USA; pp 33-34
[63] Tyers, F. M. (2010). "Rule-Based Breton to French Machine Translation". St. Raael, France. European Association for Machine Translation, EAMT. Source at: http://www.mt-archive.info/EAMT-2010-Tyers.pdf, (Accessed: 23/11/2011).
[64] Van den Bosch, A., Stroppa, N. and Way, A. (2007). A memory-based classification approach to marker-based EBMT, Proceedings of the METIS-II Workshop on New Approaches to Machine Translation , pp. 63–72.
[65] Vasconcellos, M. (1988). Technology as Translator Strategy, American Translators Association Scholarly Monograph series, II, viii 248pp.
[66] Veale, T. and Way, D. (1997). Gaijin: A Bootstrapping Approach to Example-Based Machine Translation. In International Conference, Recent Advances in Natural Language Processing. Tzigov Chark, Bulgaria, pp. 239-244
[67] Vauquios, B. (1968). A survey of formal grammars and algorithms for recognition and transformation in machine translation, IFIP Congress-68 (Edinburgh); pp 254- 260.
[68] Whitten, J., Bentley, L. and Dittman, K. (2001). System Analysis and Design Methods, North America, McGraw-Hill Companies, ISBN 0-07-231539-3
[69] Witkam, A. (1983). Distributed Language Translation: Feasibility study of a Multilingual Facility to Videotex Information Networks, Utrecht:BSO

First Author He holds B.Tech, M.Tech and Ph.D in Computer Science and has over 15 years professional and teaching experience with several scholarly acclaimed publications, Member Nigerian Computer Society. His research interests are Quantum computing, Software Engineering and Natural Language Processing.

Second Author He holds B.Tech and M.Tech in Computer Science and Engineering with several years of teaching experience and his special

ISSN 2348 – 7968

interest lies in hardware repairs and networking optimization techniques.

Third Author is a graphic savvy and a lecturer in Fine Art, Graphics and Design who had participated in several exhibitions both locally and internationally. His creativity were deployed in the design of the platform for the corpus.

Third Author is a reader in the Federal University of Technology Akure who had supervised several Ph.D research work and published numerous research papers in conferences all over the world. He is on the Professorial seat of First Bank of Nigeria Plc, member of the IEEE and the IEEE Computer Society, member Nigeria Computer Society.