© 2016 IJSRST | Volume 2 | Issue 3 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology

Design and Implementation of a System for Automatic Translation of IGALA to English Language

Sani Felix Ayegba[1], Musa Ugbedeojo[2], Negedu Philip[3]

[1,2] Department of Computer Science, Federal Polytechnic Idah, Kogi State, Nigeria
[3]Department of Electrical/Electronic Engineering, Federal Polytechnic Idah, Kogi State, Nigeria

ABSTRACT

Languages apart from being a tool of communication are vehicles of value systems and cultural expression and are essential component of the living heritage of humanity. Some languages are currently in danger because its speakers have almost ceased to use it and also ceased to pass it on from one generation to the next. The shift to language of wider communication which is English language in Nigeria has resulted in most children of the Igala race being unable to speak Igala language, a situation which greatly hampers intergenerational transmission that is the means by which a language and culture is preserved. This situation has put Igala language in danger. The aim of this research is develop a language processor that can accept as input text in Igala language and automatically translate same to English language. The system was implemented with PHP running on Apache Webserver connected to MySQL database as a backend. The output of the system was tested on 250 randomly selected Igala texts using human method for evaluating Machine Translation system. An accuracy of 81.2% was obtained.

Keywords: Language, Machine Translation System, Language of Wider Communication

I. INTRODUCTION

The babel myth documented in Genesis 11:6-9 indicates that there was a time when all human beings spoke one language. Men later developed an inordinate ambition of building a city and a tower contrary to The Creator's plan and purpose. As a result God gave people different languages. This resulted in movement of different groups of people to occupy different parts of the Earth. The resources of nature are not evenly distributed, what is found in one part may not be found in another. This made people to travel from one geographical location to another in search of their needs. A need for a means to communicate arose which gave birth to translation. [1]. Translation is necessary for communication for ordinary human interaction, and for gathering the information one needs to play a full part in society [2]. Translation is a social as well as political necessity in any multilingual society. Translation is essential for international and intercultural activities, for it facilitates mutual understanding among different and conflicting racial, ethnic, religious and cultural groups.

Human translators have **practical world knowledge** which gives them the ability to determine the proper connection between the words and between the sentences throughout the document. This way the translator creates a legible document that is also logical and contains the correct grammar and accurate connections. Despite the fact human translators produce accurate translation, only a limited number of human translators are available. According to market studies, the demand for translation outweighs its supply. Apart from being in short supply human translators are expensive and much time is spent carrying out translation. To meet up with the ever increasing demand for translation, translation technologies were evolved.

Translation technology is a general term used to describe the technologies or computerized tools available to translators to help them do their job [5]. Two of the most common technologies are probably Translation Memory (TM) and Machine Translation (MT). Machine Translation is the use of computers to automate some or all of the processes of translating from one language to another [2].It is worth noting that the emergence of translation technology does not mean the superannuation of human translators. Machine translation has not attained and may not attain the interpretive skill of the professional human translator. Machine translation can never be a substitute for human translation because no matter how fast a computer functions, the real problem in translation is not electronic but one of linguistics. It is hard to adequately represent through models the vast array of interconnections between the words of a single language and also the target language and correctly and completely map all the interconnections between them. It is impossible to make the correct choice between competing terms in a great majority of cases. [3] described the attempt to invent a method for fully automatic high quality translation as an exercise in futility and a dream that will not come true in the foreseeable future. Although there have been significant improvement in the quality of MT outputs over the years, factors such as infinite diversity of human language, complexity and impreciseness in human language which makes high quality translation difficult for MT has not changed. This means that human translators will remain indispensable in the field of translation. MT should therefore be seen as a translation support tool. Machine translation technology has both direct and indirect benefits [6]. The direct benefits include: reduced translation cost, improved delivery time, availability, consistency and throughput. The indirect benefits includes: reduced support cost, improved documentation, faster time to market and increased product/brand loyalty.

1.1 Rationale for the study

Languages apart from being a tool of communication are vehicles of value systems and cultural expression and are essential component of the living heritage of humanity. According to [7] language is a strong instrument of both ethnic and socio-cultural identity and the development of any group of people is directly a function of the extent to which their language is studied and developed. [4] noted that language embodies the unique cultural wisdom of a people and the loss of any language is a loss to humanity. Annelia Norris Hillman described language as the essence of our culture without which we do not really exist. These statements underscore the importance of the language of the people and the need for its preservation. [9] described a language on the path to extinction as an endangered language. A language is in danger when its speakers cease to use it, use it in an increasingly reduced number of communicative domains, and cease to pass it on from one generation to the next.

The shift to language of wider communication which is English language in Nigeria has resulted in most children of the Igala race being unable to speak Igala language, a situation which greatly hampers intergenerational transmission that is the means by which a language and culture is preserved. Interaction with families of Igala people in urban areas of Nigeria and diaspora show that a reasonable percentage have ceased to use the language for communication. Most of the parents communicate with their children using English language. This situation if not checked will lead to endangerment, then to moribund and finally total extinction, the consequence of which is the loss of the national identity of the people.

Machine translation can play a very vital role in the process of language documentation [8]. Safeguarding and enhancing the identity and rich cultural heritage of the Igala people using machine translation technology is the main thrust of this research. Automatic translation of Igala to English will greatly facilitate and enhance the teaching and learning of both languages. Through this medium the treasures and values conveyed and carried by the Igala language will be preserved.

1.2 Objective of study

The research is aimed at developing Igala to English Machine Translation System that would be capable accepting an Igala sentence and translating same from Igala to English.

II. METHODS AND MATERIAL

2.0 Methodology

There are two main approaches for developing machine translation applications namely: Rule based and corpus based or empirical or data centric technologies. Development of Machine Translation systems using the corpus based or data centric approach requires the availability of linguistic resources such as large parallel corpora. There is currently no English - Igala parallel corpus. Developing this resource is both expensive and time consuming. Rule based Machine Translation (RBMT) technology is less demanding on these resources than its corpus based or data centric counterpart and therefore more feasible to implement in respect of Igala language. The rule based approach was therefore adopted.

2.1 Igala-to-English MT Conceptualization

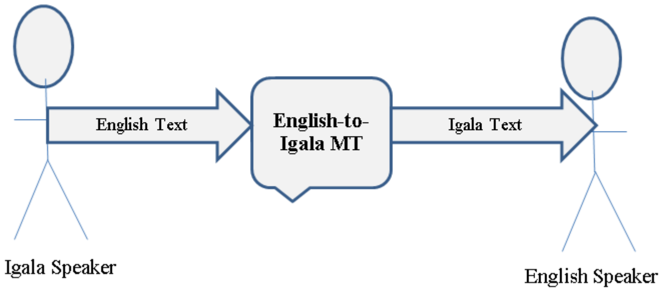


Figure 1. Igala-to-English MT Conceptual diagram

Figure 1 Shows an Igala speaker submitting Igala text as input to the Igala-to-English MT system and an English speaker receiving the English equivalent of the input Igala text.

3.0 System Architecture and Modules

The architecture of the Igala to English machine translation system is shown in figure 2.

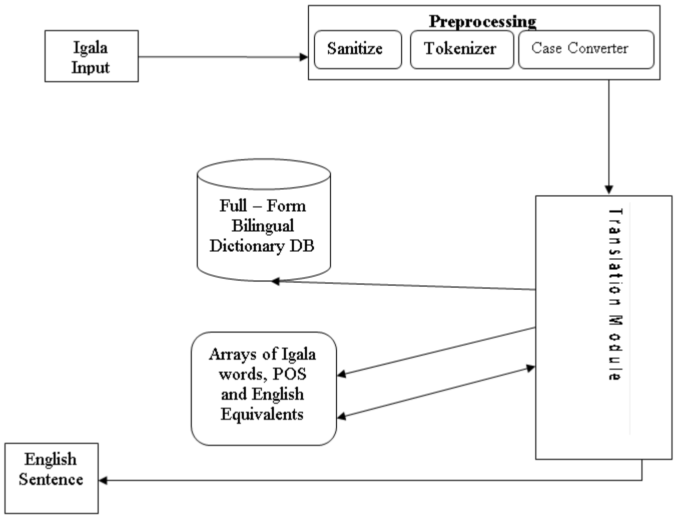


Figure 2. Architecture of Igala to English Translation System

The structure of the system above can be broken down into the following phases:

Phase (1) - Igala language text: The input to the system consists of text in Igala language.

Phase (2) – Preprocessing: This phase consists of three operations:

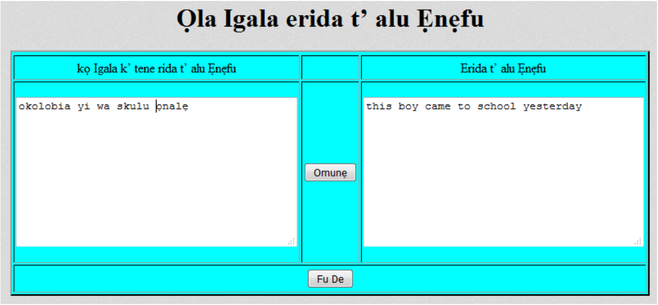
1. Sanitization: This is the process of formatting the sentence. It removes all punctuation marks such as coma, colon, semicolon, quotation marks etc. This is necessary because if a punctuation mark accompanies a word, the word will not be found in the dictionary. This function is performed by the procedure called sanitizer.
2. Tokenization: A basic text processing operation is tokenization which is the breaking up of raw text or sentence into words. This function is performed by a function called tokenizer. The input sentence is broken up at this point into words. It recognizes a word whenever a space is encountered which signifies the end of the word. The tokens are stored in an array.
3. Case Conversion: Tokens are converted to lowercases during this phase.

Phase (3) – Translation: This is the actual translation phase. The operations in phase 1-2 are preliminary operations. It consists of the following tasks:

1. English Equivalent and Pos Mapping: Each token or group of tokens (ngrams) is searched for in database tables, if found their English meaning, part of speech and other relevant information are retrieved and stored in arrays. A procedure called English Equivalent and Pos Mapper perform this function.
2. Translation Rule Application: A set of rules is applied on the tokens in accordance with Igala grammar formation rules. A function called ruleEngine carries out this operation. An array of English tokens is generated.
3. Synthesis: During this phase, translated tokens are joined together to form English sentence. The synthesizer function performs this operation. The generated sentence is then displayed on the screen as output and then saved in the database if necessary.

3.1 System implementation

The system uses full-form lexicon approach for morphological analysis (omar etal 2010). The full form bilingual lexicon which contains the Igala words and English meaning together with the parts of speech was developed using MySQL platform. The rule engine which applies a collection of lexical and syntactic transfer rules to generate the English sentences was developed using PHP. The translation interface is shown in figure 3.



Ọla Igala erida t' alu Ẹnẹfu

kọ Igala k' tene rida t' alu Ẹnẹfu

okolobia yi wa skulu ọnalẹ

Erida t' alu Ẹnẹfu

this boy came to school yesterday

Figure 3 : Translation Interface

III. RESULT AND DISCUSSION

Test and Evaluation

Machine Translation Evaluation is the examination of the quality of text which has been machine translated from one natural language to another. Quality here means the degree of correspondence between a machine's output and that of a human: "the closer a machine translation is to a professional human translation, the better it is". The aim of MT evaluation is to determine effectiveness and usefulness of the translation and the capability of the system to improve its quality.

Translation quality is judged along two key dimensions namely; adequacy and fluency. Adequacy is the extent to which the meaning conveyed by the reference translation is also conveyed by the machine translation being evaluated. It is the measure of the quantity of information existent in the original text that the translated text contains; it indicates whether the output is a correct translation of the original sentence in a sense that the meaning is transferred. Fluency on the other hand is the degree to which the translation is well-formed according to the grammar of the target language; fluency measures the extent of readability and understands ability. The scale used for evaluating adequacy developed by Linguistics Data Consortium is the following: 5 All, 4 Most, 3 Much, 2 Little, 1 None and that used for fluency is: 5 Flawless, 4 Good, 3 Non-native, 2 Disfluent, 1 Incomprehensible [10]. 300 Igala sentences were given to the Igala to English Machine translation system one after another for translation. For each translation, the input (Igala sentence) and the output (English sentence) were saved in MYSQL database table. The content of the database table was printed and submitted for evaluation with respect to adequacy and fluency. Two independent evaluators were used for the evaluation. The first evaluator is a has National Certificate of Education in Igala/English language from Kogi State College of Education Ankpa while the second a staff of Radio Kogi, Ochaja, responsible for translating English documents to Igala before broadcasting in Igala language. Table 1 is a sample from the table of scores.

Sn	Source	Target	Ascore1	Fscore1	Ascore2	Fscore2
1	ẹnẹkẹlẹ lẹ lekwu	the man died	5	5	5	5
2	iye mi chẹ gbo	my mother is aged	5	5	5	5
3	Ali chẹ onukwu mi	ALI is my friend	5	5	5	5
4	u na gbaa ọtakada-inabali ẹgba i wa	i was reading newspaper when he came	5	5	5	5
5	okolobia yi wa skulu ọnalẹ	this boy came to school yesterday	5	5	5	5
6	Ali jẹ ochikapa odudu yi	ALI ate rice this morning	5	5	5	5
7	ADA gwugwu oji ọgedegbe lẹ	ADA sit on the chair	4	4	5	4
8	Ali neke gwẹ ochibu lẹ	ALI can wash the plate	5	5	5	5
9	ẹnẹkẹlẹ lẹ che ukọlọ lẹ nyọnyọ	the man did the work very well	5	5	5	5
10	I du ujẹwn mi ku jẹ	he gave my food to ate	3	3	3	4

Ascore1 (Adequacy score by first evaluator) Fscore1 (Fluency score by first evaluator) Ascore2 (Adequacy score by second evaluator) Fscore2 (Fluency score by second evaluator) Percentage accuracy in translation obtained = 81.2%

IV. CONCLUSION

This paper has concentrated on the design and implementation of a machine translation system which translates Igala sentences to English. A rule based approach that satisfies the following requirement for translation of sentences: fast, correct, easy to edit was proposed and successfully implemented. Evaluation was carried out using human method. The result of the evaluation indicated that 81.2% accuracy in translation was achieved. This system will empower more Igala people to communicate online in their native tongue. Furthermore it will enhance the global visibility of the Igala race. We therefore recommend its adoption and use among Igala people.

Directions for future work

The work presented above can be continued in any of the following directions:

1. Continue to improve the rules and the lexicon to achieve better translation.
2. Ambiguity which is a situation that occurs when a word can have two or more different meanings is very common in Igala language. This means that in Igala language there are many words that can be interpreted in multiple ways depending on the context. The development and integration of a model that can computationally determine the sense of an ambiguous word that is activated by its use in a particular context in a given Igala sentence will greatly improve accuracy in translation.
3. Develop a large parallel corpus for Igala-English so that Statistical or Example based machine translation technologies can be used to effectively develop Igala to English automatic translation system.

V. REFERENCES

[1] Abdessalam Benabdelali, (2006). Fi Attarjama In translation], (first edition). Casablanca: Dar Toubkal, p. 13.

[2] Arnold, D, etal (1994) Machine Translation: An Introductory Guide, NCC Blackwell, London, ISBN: 1855542-17x.

[3] Bar-Hillel Y. (1964) Languages and information Reading, Mass: Addison-Wesley. Pp 74-79.

[4] Eric W. Wamalwa, Stephen B. J. Oluoch Language Endangerment and Language Maintenance: Can Endangered Indigenous Languages of Kenya Be Electronically Preserved? International Journal of Humanities and Social Science Vol. 3 No. 7; April 2013 pp. 258-266

[5] Marian Flanagan (2009) Recycling Texts: Human Evaluation ofExample-Based Machine Translation Subtitles for DVD. PhD Thesis. Dublin City University.

[6] Mike, Dillinger, Arle Lommel (2004) LISA Best Practice Guide: Implementing Machine Translation, available at www.translationoptimization.com/papers/DillingerLommel_MT_BPG.pdf, retrived on 14th April, 2014.

[7] Omachonu G.S, Igala Language Studies and Development: Progress, Issues and Challenges, Text of a paper presented at the 12th Igala Education Summit held at Kogi State University, Anyigba Kogi State, Nigeria. 28th -29th Dec. 2012.

[8] Steven Bird, David Chiang Machine translation for language preservation Proceedings of COLING 2012: pp 125–134, COLING 2012, Mumbai, December 2012.

[9] UNESCO (2003) Language Vitality and Endangerment, Ad Hoc Expert Group on Endangered Languages, UNESCO, Paris http://www.unesco.org/culture/ich/doc/src/00120-EN.pdf.

[10] Xavier G. Survey of Machine Translation Evaluation, Universitat des. Saarlandes, Computer linguistik, Germany; 2007