International Journal of Trend in Research and Development, Volume 7(1), ISSN: 2394-9333
www.ijtrd.com

Example Based Machine Translation of English-To-Igala Language

[1]Attah Joshua, [2]Dr.Sani Ayegba Felix and [3]Orah Richard Ojochegbe,
[1,2]Department of Information Technology, Faculty of Science National Open University of Nigeria, Idah Community Study Centre, Kogi State, Nigeria
[1,3]Department of Information Technology and Computer Salem Univerity, Lokoja Kogi State, Nigeria

Abstract: Today's world is highly integrated and globalized and large quantities of information are generated in different languages that need to be accessed. Translation is the only means by which information rendered in one language can be accessed by people of another language. Translation has always been carried out by humans, but human translation is time consuming, expensive and always in short supply. This limitation of human translation has made the development of system for automatic translation to enable efficient communication across cultures very necessary. Developing a system for automatic translation of English text to Igala text will greatly enhance economic activities in the territory of Igala nation. Different technologies for building machine translation systems (corpus based and rule based) were carefully studied; Example based technology which is one of the corpus based approach was adopted. Our objective is to model a language processor which will accept as input a text in English language and produce as output its corresponding equivalent in Igala language using the Example based approach. The quality of the translation was evaluated using the Bilingual Evaluation Understudy (BLEU) method. The output of the system was tested on 300 English sentences. An accuracy of 95.4% was obtained.

Keyword*: Bilingual Evaluation Understudy (BLEU), Igala Language Target text, English Language Source Text and Parallel Corpus Based*

I. INTRODUCTION

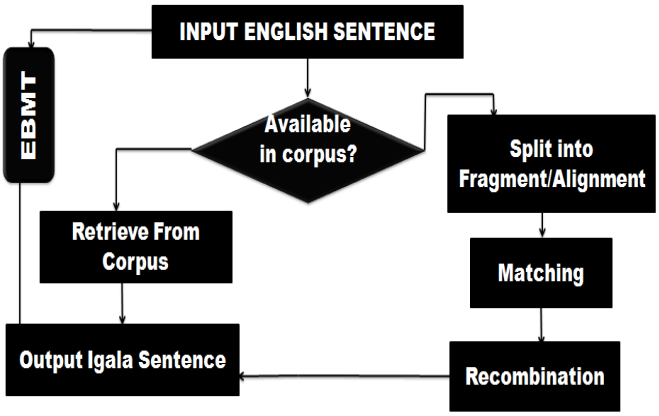
Language is an adequate and systematic medium of communication which explicitly represents the ideas, emotions, desires and other expressions of the human taught. According to (Masaru Tomita & Jaime G. Carbonell. et al., 1987). There are over 6,800 living languages exist in today's world. Thus, in order to access information written in another language we need to develop machine (computer) for translating information from one language to another. Translation is the transfer of the meaning of a text from one language to another for a new relationship. Translation is not always a straight forward case of substituting word(s) in the source language with the equivalent word(s) in the target language. In business community, language barrier is a factor stifling the growth of international business deal. It is difficult to gain foothold in international trading because it is more convenient for the global audiences to read and communicate in their native languages. Translations were first made by human translators but this human translation was unable to meet demands for translated content. These limitations of human translation led to the discovery of Machine Translation. According to (Abiola O.B, Adetunmbi A.O and Oguntimilehin A. et al., 2015).

Machine Translation (MT) is a subfield of computational linguistics that investigates the use of computer software to translate text or speech from one natural language to another. It is the area of information technology and applied linguistics dealing with the translation of human languages (Abiola O.B, Adetunmbi A.O and Oguntimilehin A. et al., 2015). Translation by or with the aid of machines can be faster than manual translation and can also reduce the cost per word of a translation. Example Based Machine Translation (EBMT) this is an approach based on recalling or finding analogous samples (of the language pair).

The system is given a set of sentences in the source language (the language from which one is translating) and corresponding translations of each sentence in the target language which point to point mapping.

These examples are used to translate similar types of sentences of source language to the target language, so if a previously translated sentence occurs again, the same translation is likely to be correct again. It also uses dictionaries to translate text. The three main steps in Example Based Machine Translation (EBMT) are Example acquisition (Alignment), Matching and recombination.

Example based machine translation will formalized from the sentence for the equivalent target language which were hand-coded in a computer program to handle the process of transfer and output generation. Here is the working flow of EBMT



The Igala race is closely related to the Yoruba and Itsekiri languages, but is recently being affected to a large extent by other tribal groups sharing border with them, including the Hausas. As a consequence, the divergence is more pronounced near the borders than within the central communities like Anyigba. A more general Igala language is however, used in this translation project.

II. RELATED WORKS

The translation of English-to-Igala using Example-based approach to machine translation is similar to a translation between any other pair of languages. According to (Sani Felix Ayegba, 2014), it is virtually understood that, the origin of

languages has become the basic need in everyday life and there is a need for translation. From the research work, there are some basic backgrounds of machine translation, a brief history of machine translation, the difference between machine translation and similar technologies, classification of machine translation and the different types of MT demand were illustrated.

Also the effort to is discussing the strengths as well as to leverage the different technologies to produce better MT systems through hybridization and samples of successful Machine Translation applications in the world. The various presented with the strengths and weakness of each technology of machine translation. The Machine Translator can translate, modify and non-modify simple sentences (Subject Verb Object (SVO)). For instance, if one were to build a translator that can translate English to Hausa, the same procedure is entailed as in the translation of English-to-Igala: the only difference lies in the morphology of each pair of languages. The features of Igala language and English language are comparatively examined here.

A. Conceptual Framework

Machine Translation (MT) is the changing of text which denotes a meaning in one language to text that denotes an equivalent meaning in another language by computer. It is the process of converting one natural language into another natural language by means of the computer (Peng, 2013). Machine translation can be a complete automation or merely a partial assistance of the computer system (computer-aided translation).

In the early 1920s, governments of some countries displayed enormous interests in the new effort to translate languages by means of the computer. These interests were demonstrated by granting patents to some experts to enhance their pursuit of the noble endeavor, such as George Arsrouni, who developed the mechanical brain (Hutchins, 2005). In 1949, Warren Weaver suggested that the problem of translation can be attacked with statistical methods, with some ideas from information theory (Peter et al., 1990).

Machine Translation comes in place when a computer translates an entire document automatically and then presents it to human. Human Translation has to do with process of composes a translation, perhaps calling on a computer for assistance in specific tasks such as looking up specialized words and expressions in a dictionary, regarded as traditional human translation (Sani Felix Ayegba, Osuagwu O.E. ,Njoku Dominic Okechukwu(2014)).

Machine translation is classified into: rule-based (RBMT) and corpus-based translations (Peng, 2011). Transfer approach to Machine Translation and others like direct translation and Interlingua belong to the rule-based category of machine translation, whereas, Statistical technique and Example-based method are examples of Corpus-based MT system.

Due to the limited expert of human efforts, the demand for translation is not satisfied because there are not enough human translators. Apart from the scarcity of human translator, human translation is also slow and expensive. The economic necessity of finding a chapter and faster alternative gave birth to the machine translation technology.

III. MATERIALS AND METHOD

Example-based Machine Translation tries to overcome the differences which exist between two languages by applying contrastive knowledge (i.e. knowledge about the differences between the source language and the target language). This strategy is referred to as direct transfer model. The model requires some representation of the structure of the source language, which results in the structure for the target language, followed by a generation sentence in order to create the output sentence.

The Example based machine translation is one of the approaches in machine translation. The concept uses the corpus of two languages and then translates the input text to desired target text by proper matching. Here is conceptual framework of the model.

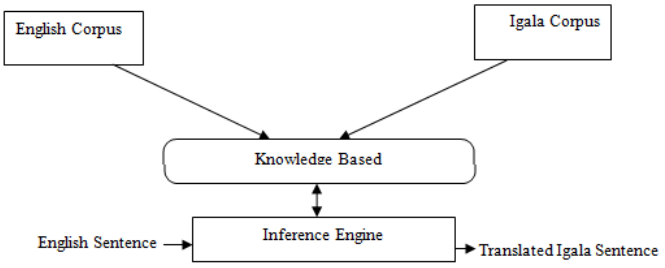


Figure 3.4.5: Example Based Machine Translation of English to Igala Language

The different languages have different language structure of the subject-object-verb (SOV) alignment. The matching is then arranged to give proper meaning in target text language and to form proper structure. In this research work, we describe the Example Based Machine Translation using Natural Language Processing. The proposed EBMT framework can be used for automatic translation of text by reusing the examples of previous translations. This framework comprises of three phases, matching, alignment and recombination.

A. Example Based Machine Translation

- English Corpus: We have used 300 English sentences for forming a corpus. The sentences are the news headlines from reputed newspaper.
- Igala Corpus: It consists of the translated sentences in Igala for each of the English sentences.
- Knowledge Base: It stores the patterns of how English sentences are translated into Igala form.
- Inference Engine: It is a collection of facts and rules. Inference Engine compares the given English sentence with the English sentences stored in the corpus. After finding the best match, it translates it into Igala according to the Igala translation present in the Igala corpus.

Automatic translation of text form one language into another is a Machine Translation.

The implementation comprises the Input being supplied and analyzed by the Java Programming language. The EBMT structure of the input text is then examined and then passed to the transfer module. The transfer module entails all the transfer that enables the process of transformation from the source structure to the target language structure. The next phase is the translation phase, which takes place by means of the translation engine. In the translation engine, the target EBMT structure is passed over to the translation module, where the translation rules and re-ordering transfer are combined and implemented to yield the output. The output is ultimately the results of the translation process. The diagram in Figure 4.4 illustrates the implementation details.

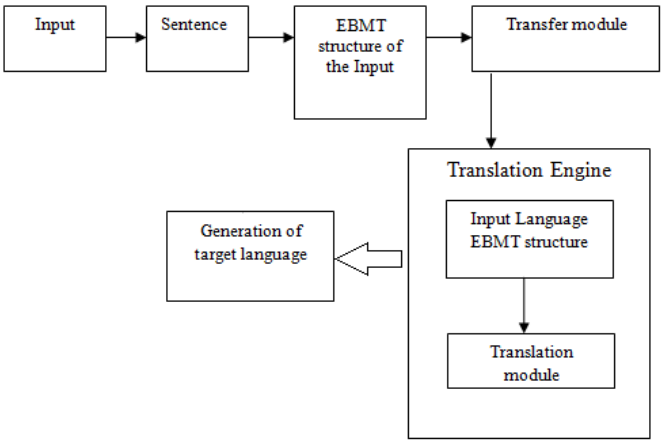


Figure 4.4: Implementation architecture

B. Translation

The translation sub-module of the system is made up of a combination of processes such as loading of the database from MySQL, a search for a match for each element of the input string among the MySQL entries, the sentence of the target set, and the appending of the target string to the output pane. The translation sub-module is where the translation resides, making the sub-module very important. It must be pointed out that the rich set of Java libraries and methods such as query language, record set, split, concatenate, append, Arrays. To String, substring, etc were prudently explored.

C. Bilingual English Lexicon

To make it easy for hash table operation, the database was built on a simple MySQL database platform. Only two columns of entries were used, and each of the elements on a row was separated by means of the equality sign. Elements found to the left side of the equality sign is the English words, while the element on the right side of the equality sign is the Igala sentence. Each of the Igala entries is concatenated to a substring. These tags are invaluable as a means for identifying the order in which the tokens of an input string are arranged and how they should be re-arranged in the output.

Program Interface Moodle-1

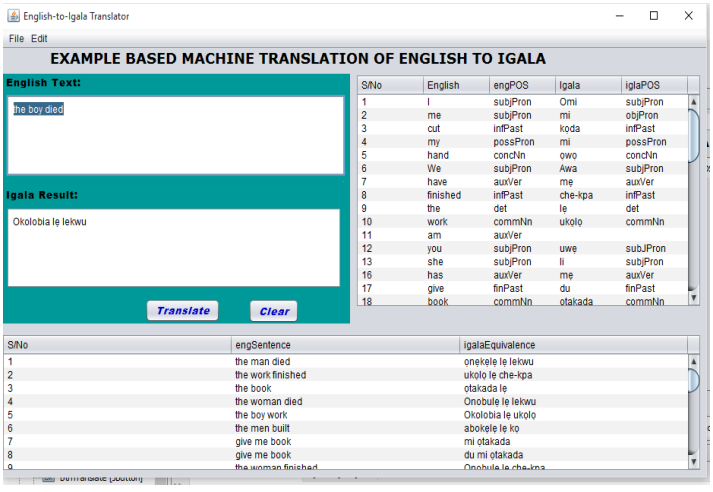
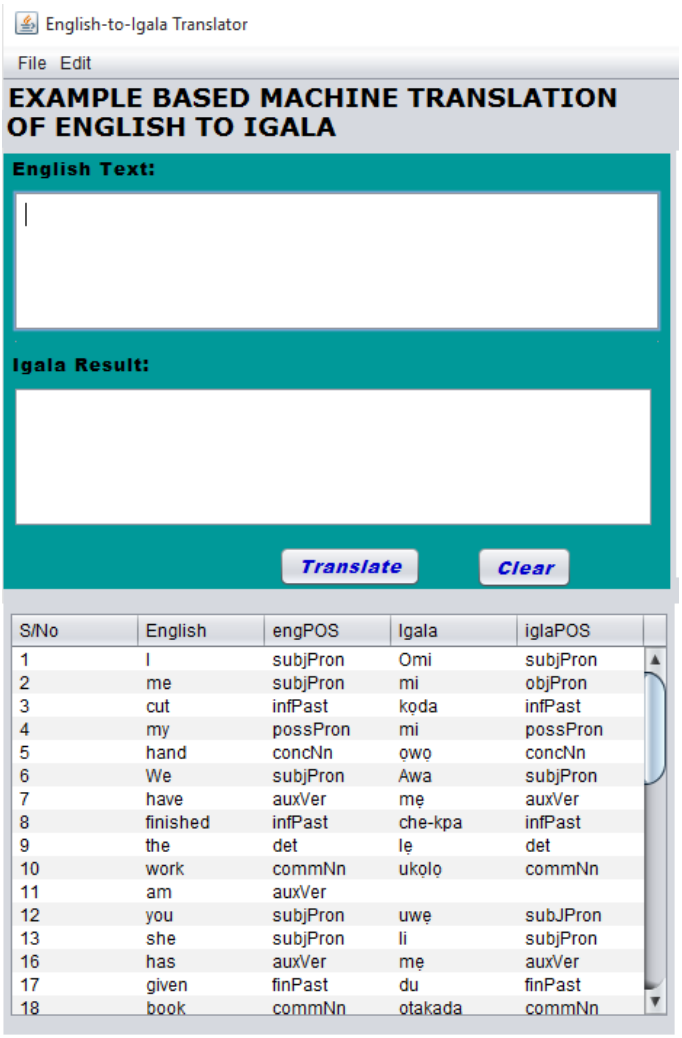


Figure 4.4.3a Program moodle to accept input from user

This will allow user to input English sentence from the moodle or source language and click on the translate to display equivalent language in Igala as show in figure 4.4.3 above

Program Interface Moodle-2



S/No	English	engPOS	Igala	iglaPOS
1	I	subjPron	Omi	subjPron
2	me	subjPron	mi	objPron
3	cut	infPast	koda	infPast
4	my	possPron	mi	possPron
5	hand	concNn	owo	concNn
6	We	subjPron	Awa	subjPron
7	have	auxVer	mẹ	auxVer
8	finished	infPast	che-kpa	infPast
9	the	det	lẹ	det
10	work	commNn	ukọlọ	commNn
11	am	auxVer		
12	you	subjPron	uwẹ	subJPron
13	she	subjPron	li	subjPron
16	has	auxVer	mẹ	auxVer
17	given	finPast	du	finPast
18	book	commNn	otakada	commNn

Figure 4.4.3b Program moodle to accept input from user

This will allow user to input English sentence from the moodle or source language and click on the translate to display equivalent language in Igala as show in figure 4.3.3b above

IV. EVALUATION AND RESULTS:

The evaluation of a MT system helps to ascertain the level of the system's performance. If that is the case, any translation system must be evaluated to determine performance, and ascertain areas that need future improvement.

The methods of evaluation that can be used include heuristic methods such as BLEU, NIST, TER, Precision and Recall, and METEOR (Jurafsky et al., 1999). These methods, most of non-human MT systems provide more effective unlike human rating method, of evaluation technique (Jurafsky et al., 1999).

In this research work, use BLEU technology to evaluate the translator's performance. BLEU technology like other evaluation methods measure system performance along the dimensions of fidelity and fluency (Miller et al., 1958). While fluency entails how intelligible, how clear, how readable and natural a translation output is, fidelity looks at how adequate and how informative the translation output is.

For this translation system, BLEU technologies was used to rate the quality of the translation output along five dimensions:

Raters were asked to rate how natural, informative and unambiguous the translation output is by awarding a score between the ranges of 0-5.

BLEU were also given some English sentences that were initially translated by the system, and were asked to give a translation of the in their own words. These were recorded and later compared to with the translation offered by the system; the number of similar words occurring between the human translation and the system's translation were recorded and used to rate the level of performance of the system.

On the merits of this work, we are constrained to judge our work as good. Here is our valuation table for BLEU

Evaluation of BLEU:

Our evaluation was carried below:

EvaluationTable				
TransID	**EnglishSentence**	**RefTranslation**	**CandTranslation**	**BLEUScore**
1	The little child died.	ọma kẹkẹ lẹ lekwu	ọma lekwu lẹ	0.72
2	I finished the work.	Omi fukọlọ che kpa	Omi che-kpa lẹ ukọlọ	0.86
3	they finished work	a fukọlọ che kpa mẹ	ukọlọ che-kpa	0.21
4	The man died.	ọnẹkẹlẹ lẹ lekwu	ọnẹkẹlẹ lẹ lekwu	1.00
5	He give him the book	Ii du ọtakada lẹ wnu	Ii du ọtakada lẹ	0.78
6	The woman died.	Onobulẹ lẹ lekwu	Onobulẹ lẹ lekwu	1.00
7	The man give him book.	ọnẹkẹlẹ lẹ du ọtakada nwu	ọnẹkẹlẹ du lẹ ọtakada	0.78

BLEU Score for translated sentence

BLEU Score	No. of phrases	Percentage Value
>=0.8 and <=1	233	77.7
>=0.5 and <=0.79	53	17.7
<=0.4	1	0.33

Analysis of Results

BLEU score above 0.50 reflect good and influent transformation. From the table percentage score above 0.50 is 77.7 + 17.7 = 95.4%

SUMMARY

This research successfully described the concept of Example based machine approach to MT system: it has been found to be most appropriate for a translation of English-to-Igala for the fact that rules for NLPs could be easily formalized, and also because Igala has no available body of text in 'soft form', that can be easily acquired and used as training corpus based. If this were the case, it would have been possible to use Statistical Machine Translation method.

It has been learned in the course of this work, that language pre-processing, which is a part of the analysis phase provides ample opportunity for understanding the grammar of the languages that are translated. Analysis also ensures that the 'grammaticality' of each English NLP and its equivalent in Igala is confirmed.

After establishing the validity of some of the structures of English NLPs, patterns were identified about how some components should be combined to form NLPs in both English and Igala language: these patterns constituted the transfer rules with which translation tasks were implemented. Syntactic parsing which yielded syntax structure is of importance because trees actually laid bare the patterns which were used to formalize the transfer rules.

References

[1] Alegria, I. et al. (2006): An Open Architecture for Transfer-based Machine
[2] Ayegba, S. F. et al. (2014): Machine Translation of Noun Phrases from English to Binghamton, NY.
[3] Bowen, D. and Bowen, M. (1990): Interpreting: Yesterday, Today, and Tomorrow.
[4] Brown, P.F. et al. (1990): A Statistical Approach to Machine Translation.
[5] Computational Linguistics, Vol. 16, No. 2. Yorktown Heights, NY.
[6] Gehlot, A. et al.(2015): Hindi to English Transfer Based Machine Translation
[7] Hutchins, W.J. (2005): Prospect in Computer-based Translation. Proceedings of
[8] Hutchins, W.J. and Somers, H.L. (1992): An Introduction to Machine Translation.
[9] Igala Using the Rule-based Approach. West African Journal of Industrial and India.
[10] Lavie, et al. (2003): Experiment with a Hindi-to-English Machine Translation under
[11] Lavie, et al. (2004): Rapid Prototyping of a Transfer-based Hebrew-to-English Machine Translation Summit, Singapore. Pages 30 – 34. Machine Translation System.
[12] Machine Translation. IEICE Trans. Inf & Syst., volume E80-D, No. 11. Nagoya-Matsubara, S. and Inagaki, Y. (1997): Incremental Transfer in English-Japanese
[13] Nagao, M. and Tsujji, J. (2005): Transfer Phase of the MU Machine Translation Processing (TALIP), 2(2); pages 143 – 163.
[14] Rich and Resource-poor Languages: the Case of Hebrew and Arabic. An MA shi, Japan.
[15] Shilon, R. (2011): Transfer-based Machine Translation Between Morphologically-System. International Journal of Advanced Computer Research, Vol. 5. Pune
[16] System. Kyoto, Japan.
[17] Thesis. Tel Aviv.
[18] Translation (AAMT), page 7-14. HAL archives.
[19] Translation between Spanish and Basque. Asia-Pacific Association for Machine