SciencePG
Science Publishing Group

Machine Learning Translation of English into Igbo Language: A Review

Orji Ifeoma Maryann[*], Sylvanus Okwudili Anigbogu, Ekwelaro Oluchukwu Uzoamaka, Asogwa Doris Chinedu

Department of Computer Science, Nnamdi Azikiwe University, Awka, Nigeria

Email address:
oma4goodness@gmail.com (O. I. Maryann), so.anigbogu@unizik.edu.ng (S. O. Anigbogu), ou.ekwealor@unizik.edu.ng (E. O. Uzoamaka),
dc.asogwa@unizik.edu.ng (A. D. Chinedu)
[*]Corresponding author

Abstract: Machine learning is the machine translation used in language. A lot of people have been using it and it's working extremely well for them. It works without delay and is getting better as day passes. Of course, this advancement will change the way the modern world operates because language barriers have been an essential impediment to international commerce since time immemorial. Language is tagged with cultural disposition of people's ways of life. Without language there will be no existence of the world and the people living within it. The necessity of translation comes into existence because of the migration of people from one country to another. Language translation is the key technology for the in-coming generation of IT and every digital device uses it and is dependence on it. The objective of the paper is to review a machine learning translation of English to Igbo language. The major concept reviewed was machine learning in term of dealing with language translation. Machine learning is a wide and interesting field in computer science. This has been tagged as an awesome professional area for research in demands for translation of languages. Many researchers preferred using supervised learning in doing language translation while few researchers used unsupervised learning in doing language translation.

Keywords: Machine Learning, Algorithm, Translation, English Language, Igbo Language

1. Introduction

Language is the medium of exchanging information. Communication should include the feeling that language has been generally practiced and approved as the most estimable human institution and is absolutely necessary in all spheres of life [6]. To communicate ideas, emotions, feelings, desires, to co-operate among social groups, to exhibit habits etc which can be translated along a variety of channels, Human Language plays a vital role [4]. Means of getting information from other language is of paramount importance but the ability to give out information across language is translation. Hence, designing of devices to have a translation into another language is an essential achievement to human development. In the absence of translation, communication will not take place apart from those who have usual language and large number of scholars will not be recognized in the absence of this crucial role.

Translation is very important towards accessing knowledge disproportion. It is one of the essentials in the business community. The work of translation was carried out at first by human translators. The supply of translation services could no longer meet with the demand for translated content at a time. Also, human translation is very expensive, takes a lot of time and inappropriate for accessing time duration businesses requirements which include serving multilingual prospects, partners and customers. The intrinsic prescriptions of human translation created the search for other means of translation paramount. And today, the search gives on to the finding of machine learning algorithm for translation. A sub section in artificial intelligence that is also known as the capability of IT systems to independently detect solutions to problems by acknowledging patterns in databases is called Machine

Learning. That is to say that Machine Learning assists IT systems to acknowledge patterns on the basis of existing algorithms and data sets and to create adequate solution concepts [11]. In order to enable the software to produce solutions independently, the preceding motion of people is mandatory. For example, the expected algorithms and data must be keyed into the systems beforehand and the specific rules of analysis for the acknowledgement of patterns in the data stock must be stated. For this reason, we decided to develop a machine learning algorithm for translation of English into Igbo language.

The transformation of information into knowledge is known as Machine learning. There have been an explosion of data in the past decade and this mass of data is unhelpful except been analyzed and the patterns hidden within discovered. To find the fundamental valuable technique Machine learning skills are used to perfunctory find the fundamental valuable techniques within complex data which we would eventually find difficult to come across. To forecast time ahead results and carry out different difficult analysis, secret techniques and knowledge pertaining to a problem can be applied. Many people are ignorant that they have already come across with Machine Learning each day. Whenever we search for one thing in the internet and pay attention to music or snap a picture, Machine Learning is one of the core factors supporting it, regularly accumulating knowledge and enhancing from each conversation. It also contributes to the world-changing advances like discovering cancer, producing fresh drugs and automotive vehicles. To this end, artificial knowledge is created on the basis of experience in Machine Learning.

2. Related Literature

The major engine of machine learning is called Machine learning algorithms, which may also be defined as the algorithms that change data set into a model. Goyal et al [8] suggested Hindi to Punjabi Machine Translation System. This system is on the basis of immediate word-word translation method. It comprises of floristic analysis, rephrasing of word sense, documentation and post processing.

Amharic to English Language Translator for iOS was created to utilize a Translation system which was based on Microsoft Interpreter Center point and utilized Microsoft Interpreter Programming interface and the application can be utilized to make a Translation of writings from Amharic to English [9].

An approach which would aid instruction and studying of Hausa, Igbo, and Yoruba languages was proposed by Odejobi et al [14]. The investigation took into account body parts of human ID, trees distinguishing proof, and creatures' lists. The English to Yoruba machine translation and Yoruba number tallying approaches are a piece of the fundamental approach. The model was developed in order to create a system for the learner of the Hausa, Igbo, and Yoruba languages.

The theorem-proven approach for English to Yoruba Machine Translation System was proposed by Agbeyanyi et al [2]. The three ways to deal with the Machine Translation process were identified. The creators investigated all the three methodologies and regarded principle found methodologies for the Translation Method. As indicated by creators, there exists constrained body which is accessible for Yoruba language, which illuminates the standard found methodology.

Chinenyeze et al [5] created a Natural Language Processing System for English to Igbo Language Translation in Android. The aim was to build a translator Android application which can be used by people who want to translate texts from English into Igbo. He used Statistical Machine Translation to utilize statistical translation models generated from the analysis of monolingual and bilingual training data. Essentially, this approach uses computing power to build sophisticated data models to translate one source language into another. Statistical Machine Translation consists of Language Model (LM), Translation Model (TM) and Decoder.

Ifeanyi et al [10] proposed an analysis and representation of Igbo text document for a text based system considering its compounding nature and described its representation with the Word-based N-gram model to properly prepare it for any text-based application. Their result showed that Bigram and Trigram n-gram text representation models provide more semantic information as well addressed issues of compounding, word ordering and collocations which are the major language peculiarities in Igbo.

Olufemi et al [15] developed a computational analysis of Igbo numerals in a number-to-text conversion system which was used to convert Arabic numbers between 1-1000 in Igbo language. The translation system is very crucial in Igbo language because it comprises of three major Nigerian languages feared to be among the endangered African language. Python programming language and pyQt was used to implement and designed system which contains sequence and activity diagram in it.

3. Machine Learning and Language Translation

Animikh [3] noted that Machine learning algorithms can be classified as:

Supervised learning: This machine learning algorithm provides the output of a new set of user's input if a model is trained to ascertain the results by using an old set of inputs and its relative known set of outputs. That is to say that the system makes use of the past instances used before. These models compare the new results to the old ones and update their data accordingly to improve the forecasting pattern after training. A supervised learning algorithm analyzes the past experiences and brings out an inferred results, which is called a classifier (as long as the result is discrete) or a regression function (as long as the output is continuous). Consequently,

the two main works performed with supervised learning are REGRESSION (trying to forecast a real value. For example, forecast the value of a good tomorrow given its past performance or forecast a student's score in an exam based on his/her homework scores) and CLASSIFICATION (trying to forecast the appropriate category of an object amongst a set of categories specified in training model). Examples of supervised learning algorithms are support vector machines, decision trees, random forest, naïve Bayes classifier, nearest neighbor algorithm, logistic regression etc.

Unsupervised learning: This type excludes training the model based on old data, i.e. there is no tutor to assist in giving the model former examples. The system is not trained to provide any set of inputs and relative outputs. Rather, the model by itself study and forecast the result based on its own experiences. CLUSTERING is one of the major example of works carried out under unsupervised learning which is the partitioning of data into a set of distinguished groups or clusters. Some of the common clustering algorithms are K-means, mean-shift, and expectation maximization etc. They share data points based on similar or shared characteristics. K-means algorithm a good popular algorithm used in unsupervised learning.

Semi Supervised Learning: Semi-supervised learning is a combination of both the features of supervised and unsupervised algorithms. Hither, both labelled and unlabeled instances exist, and in many cases of semi-supervised learning, the count of unlabeled instances dominates that of labelled ones.

Reinforcement Learning: In reinforcement learning, the device learns from the howl it has gotten from the environment. It often acquires and improves its existing knowledge by taking the howl round within its environment.

Essentials of Machine Learning Algorithms

Data: Data is an input to the computer. For instance, to design a computer that can forcast the weather over the next few days, then you should enter 'data' that will consist of maximum and minimum air temperatures, the speed of the wind, amount of rainfall, etc.

Algorithm: Data is seen as the 'food' to the machine, an algorithm can be regarded as its digestive system. An algorithm performs task on the unprocessed information. It crushes it; analyses it; permutates it; identifies the gaps and fills in the blanks. These methods used by machines to perform task on the data input is viewed as algorithms.

4. Language Translation

There are different morphology, syntax and semantic in Igbo and English language. In Igbo language, we have two main components of speech - Nouns and verbs, the rests are gotten from the two main components. In Igbo language tone plays a significant part in realizing parts of speech [12]. Tones or contrastive pitch are used to show words with similar means. For instance, the similar word in Igbo may have four derivative meanings relying on its pitch. Pitch is an attribute of words and very relevant in tone language when

dealing with relative pitch. High and low tone are two main tone in Igbo language. In high tone, "U" in "RULE" is pronounced with the tongue crooked towards the roof of the mouth while in low tone, "A" in "FATHER" is pronounced with the tongue flat and the mouth a bit wider than high tones. Looking at these examples, *akwa* can mean crying (high-high tone), *bed* (high-low), *egg* (low-high) or *bridge* (low-low).

Many words are structured from smaller words in language but only few words from English have been duplicated straightly. We have wide variety of dialects with resemblance from each other although, some have different vocabularies and pronunciations. Besides, word order and tone are consistent in every grammatical Igbo structure. English language is specific in her choice of words; Igbo language on the other hand is sentential. In most cases, a sentence is used to describe a concept. English language has definite and indefinite articles, whereas such articles do not exist in the Igbo language.

Example:

A man	Nwoke
A goat	Ewu
The house	Ulọ
The pot	Ite

Igbo language operates a centrifugal syntax while English language operates a centripetal. This means that the qualifiers and determiners come quite often after the noun. Noun Phrase performs the function of a noun in a sentence, a noun phrase can be introduced with articles such as; (a, an, the), adjectives (participles) and possessive pronouns. A noun phrase could be introduced by a noun (i.e a single word) and more than one word.

Example:

nwunye m	my wife
nwunye gï	Your wife
nwunye ya	her wife

Compound verbs are formed under some constraints in Igbo. There is a strong relationship between the conceptual properties of the compound verbs and their compound types; one which according to Gamerschlag [7] derives the argument structure from such conceptual.

Properties. Some of examples of compound verbs are:

kpó-chi:	VI-KPÓ = peg.	V2-CHI = cover
pị-chi:	V1-PÌ = press, squeeze	V2-CHI = cover, close

5. Machine Learning and Translation

The source language input first resolves and produces an internal representation in a machine learning system. This representation is operated and conveyed to make it appropriate for the target language. Hence, the output is produced from target language. Machine translation is concerns with change from one natural language to another. Natural language Interface is used for day to day communication and allows the people to convey with the

computer in a natural language. The key relevance of automatic machine translation in natural languages involves the used of computational machine. This is crucial as interaction among people from unusual language historys still pose a great difficulty [16]. Also, Machine translation systems are relevant tools used in creating culturally sensitive contents on the web, this gives room to environments for multilingual diversity [13]. Machine translation of natural languages is tasking. It can simply be viewed as a replacement of dialects from one natural language to another. Although, it is not so easy due to the complex nature of some natural languages. Different words are interpreted in different ways in some languages, so they have different translations. Ambiguous sentences also have various meanings. Varieties in grammatical relations depends on the languages, and interpreting sentence from languages having different relations means re-coining the sentence. Besides, difficulties may occur due to the associated world knowledge and these are not usually easy to solve [1]. Translate from English to Igbo using any of the approaches to machine translation, owing to the dissimilarities in their syntax and structures, it is important to have a very good knowledge of the mother language, being conversant with the grammatical features of the target language, having an idea the subject matter of the text being translated and a clear understanding of the etymological and idiomatic correlations between the two languages, it is also necessary to have a sense of when to metaphase ("translate literally") also when to paraphrase so as to ensure a true rather than spurious equivalents among the source and target-language texts and a good understanding of the syntax and grammatical features of the two languages and their vocabularies.

6. Methodologies of Machine Learning

Machine learning is a competence of artificial Intelligence. It categorises the source language input (English) and provides an internal representation. This representation aids in forming a target language (Igbo). It is usually concerned with the conversion from one natural language to another. Unsupervised learning was used for the translation. The unsupervised technique used to translate from English to Igbo is Clustering. It comprises of language and translating models.
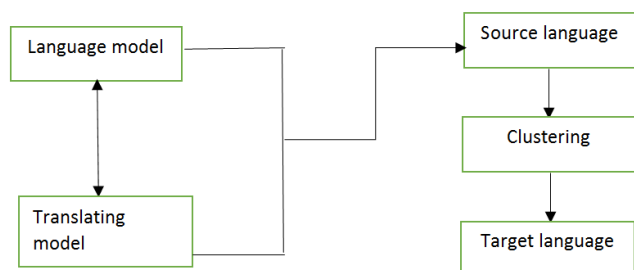

Figure 1. *Basic structure of machine learning Algorithm for translation of English into Igbo.*

Language Model: It is the core component of machine learning algorithm for translation. Language model tries to

examine patterns in the language of human. It is used to predict words or text in any language. It provides context to distinguish between words and phrases that sound similar.

Translation Model: It explains the mathematical relationship that exists among two or more languages. They can be regarded as models of translational equivalence because they all ascertain expressions in different languages having equivalent meaning.

7. Summary and Conclusion

This paper looks into review of literature related to the concept of machine learning and language translation. Machine learning translation of languages help mankind in learning multiple languages across the global. It is a key technology for the next generation IT, which will be much smarter and human-centered in its functionality. This technology has emanated to aid human life from the high cost of human translator and provide a better translation paradigm. Thus, learning a new language also opens up a new view of the world and gives one a different way to put reality into words.

References

[1] Abiola O. B, Adetunmbi A. O, Oguntimilehin A (2015) "Review of the Various Approaches to Text to Text Machine Translations" International Journal of Computer Applications. Vol 120 No 18, pp 7-12. ISSN: 0975-8887.

[2] Agbeyangi, A. O., Eludiora, S. I., and Adenekan, D. I. (2015). "English to Yorùbá MachineTranslation System using Rule-Based Approach". Journal of Multidisciplinary Engineering Science and Technology (JMEST), Vol. 2 Issue 8, August, Nigeria.

[3] Animikh. A. (2021). Machine Learning Algorithms: [With Essentials, Principles, Types and Examples covered]. https://www.knowledgehut.com/blog/data-science.

[4] Banjo, A. E. and Jibowo, A. V (2011) The Use of Principles and Theories of Translation in Languages: A Case Study of Yoruba. Journal of Communication and Culture; International Perspective Vol. 2 No 3.

[5] Chinenyeze C. E, Bennett E. O. and Taylor O. E. (2019) A Natural Language Processing System for English to Igbo Language Translation in Android. International Journal of Computer Science and Mathematical Theory ISSN 2545-5699 Vol. 5 No. 1. www.iiardpub.org

[6] Danladi, S. S. (2013). Language Policy: Nigeria and the Role of English Language in the 21st Century. European Scientific Journal, ESJ, 9 (17).

[7] Gammerschlaag, T. (2000) Deriving Argument Structure in Japanese V-V Compounds. Working Paper of the SFB Theorie Des Lexikons No. 282, University of Düsseldorf.

[8] Goyal, V., and Lehal, G. S. (2010). "Hindi to Punjabi Machine Translation System". Department of Computer Science, Punjabi University, Patiala, India.

[9]   Hana, B. D. (2016). "Amharic to English Language Translator for iOS". Department of Information Technology, Helsinki Metropolia University, Finland.

[10] Ifeanyi. R. N, Ugwu. C. and Adegbola. T. (2017) "Analysis and Representation of Igbo Text Document for a Text-Based System". International Journal of Data Mining Techniques and Applications Volume: 06, Issue: 01, Page No. 26-32 ISSN: 2278-2419.

[11] Lina. M. (2018) "Architecture Global Market and Information Technology in the Digital". Published in Ardet U.S.A.

[12] Mbah, E and Ayegba M.(2012) Tone in Igala Language: An Autosegmental Analysis. Journal of Igbo Language and Lingustics, No. 4 ISSN 05987518, pp 67-75.

[13] Montviloff. V. (2015) Meeting the challenges of language diversity in the information society. World Congress on Language Policies, Bacelon Available: www.linguapax.org/wpcontent.

[14] Odejobi. O. A., Ajayi. A. O, Lukman. A. and Safiriyu. I. E (2015) "A Web Based System for Supporting Teaching and Learning of Nigerian Indigenous Languaage". Faculty of Technology Conference at Obafemi Awolowo University, Ile-Ife, Nigerian.

[15] Olufemi. D. N., Abimbola. R. I., Isacc. O. E and Olamide. E. O. (2017) "Computational Analysis of Igbo Numeral in a Number-To-Text Conversion System". Journal of Computer and Education Research Decemeber. Volume 5. Issue 10.

[16] Sangeetha. J, S. and Jothilakshmi, R. N. (2014), "An Efficient Machine Translation System for English to Indian Languages Using Hybrid Mechanism" International Journal of Engineering and Technology (IJET), pp 1909-1919, Vol 6 No 4, ISSN: 0975-4024.