Nigerian Journal of Engineering Science Research (NIJESR).
Vol. 6, Issue 03, pp.01-07, September, 2023

ISSN: 2636-7114
Journal Homepage: https://www.iuokada.edu.ng/journals/nijesr/

Evaluating English to Nupe Machine Translation Model Using BLEU

[1]* Sayuti Musa Shafi'i, [2] Hassan Umar Suru and [3] Danlami Gabi

[1] Department of Information Technology, Continental Transfert Technique Limited, Abuja, Nigeria, sayutimusa87@gmail.com

[2] Department of Computer Science, Kebbi State University of Science and Technology Aliero, Nigeria, suruhassan@yahoo.com

[3] Department of Computer Science, Kebbi State University of Science and Technology Aliero, Nigeria.

**Corresponding Author: Sayuti Musa Shafi'i; sayutimusa87@gmail.com (08034663899)*

Manuscript History
Received: 25/06/2023
Revised: 20/09/2023
Accepted: 27/09/2023
Published: 30/09/2023

Abstract: *In Nigeria, the dominance of the English language as a medium of communication is alarming, as the minor indigenous languages are gradually moving towards extinction. Access to information written in another language is of great interest because it helps express our feelings, desires, queries to the world around us and the means of sharing information across languages is translation, therefore creating tools for translating from one language to another is a very crucial contribution to human development .The study aims to evaluate the effectiveness of machine translation system used to translate English phrases to Nupe relative to the effectiveness of English to Nupe human translation. There are many automatic methods used to evaluate different machine translators, one of these methods; Bilingual Evaluation*

Keywords: *Effectiveness, Machine Translation Model, English, Nupe, Bilingual Evaluation Understudy.*

INTRODUCTION

Machine learning is the process of using Artificial Intelligence to automatically translate content from one language (source) to another (target) without any human intervention. To deal with any translation, human or automated, the text of the original language (the source) must be completely restored in the target language, i.e., the translation. While it may seem simple at first glance, it is actually much more complex. Translation is more than just word-to-word conversion. The translator must interpret and analyze all the elements of the word and see how each word affects the other. This requires extensive knowledge of grammar, syntax (sentence structure), meaning (interpretation), etc. of source and target languages, as well as understanding of local region. The accuracy of any machine translation system is always evaluated by comparing the results to human judgment. The most popular automatic evaluation method which is Bilingual Evaluation Understudy (BLEU) proposed by IBM (Ken and Björn, 2018) was adopted for evaluating the output of the developed machine translation system. The method is language independent and highly correlated with human evaluation.

BLEU is based on the idea that the closer the output of a machine translation is to a reference (professional human) translation, the better. BLEU scores range from zero to one. Ken and Björn (2018) stated that BLEU scores above 0.30 generally reflect understandable translations and BLEU scores above 0.50 reflect good and fluent translations. This score is calculated through a series of steps. BLEU's baseline metric is a 'modified n-gram precision' that calculates precision using the maximum number of words in the TL found in any of the reference sentences, divided by the total number of words in the TL sentence. Nupe is a minority language spoken by the Nupe ethnic group in Nigeria. There are probably about 3.5 million Nupes, principally in Niger State. The Nupe language is also spoken in Kwara, Kogi and Federal Capital Territory. They are primarily Muslims, with some Christians and followers of African Traditional Religion (Sayuti *et al.,* 2022). Nupe is one of the 12 languages approved by the Federal Government of Nigeria for its mother tongue education policy (Jacob, 1999). It is still considered not to have been adequately studied compared to the three major Nigerian languages of Hausa, Yoruba and Igbo (Garba, 2012), this is because the available works on the Nupe language are still very few.

Nupe has a variety of words and word schemes that allow you to pronounce each sentence differently. In addition, the Nupe dialect has different features. Moreover, the existence of many Nupe dialects and the fact that words are not the same for source and target language, this often leads to the possibility of having more than one meaning of the same sentence. Machine translation of written English text to sign language. They study the existing systems and issues in order to propose an implantation of a statistical machine translation from written English text to American Sign Language (English/ASL) taking care of several features of sign language. The work proposes a novel approach to build artificial corpus using grammatical dependencies rules owing to the lack of resources for sign language. The parallel corpus was the input of the statistical machine translation, which was used for creating statistical memory translation based on IBM alignment algorithms. These algorithms were enhanced and optimized by integrating the Jaro–Winkler distances in order to decrease training process. Subsequently, based on the constructed translation memory, a decoder was implemented for translating English text to the ASL using a novel proposed transcription system based on gloss annotation. The results were evaluated using the BLEU evaluation metric (Achraf and Mohamed, 2019). It is usual to have more than one perfect translation of a given source sentence. According to this fact (Ken and Björn, 2018) casted BLEU in 2002 as an automatic metric that uses one or more reference human translation beside a candidate translation of an MT system. The increase in the number of reference translations leads to increase the value of this metric. BLEU metric aims to measure the closeness of a machine-translated (candidate) text to a professional human (reference) translation. BLEU uses a modified precision for *n*-grams at a sentence level and then averages the score over the whole corpus by taking the geometric mean, with *n* from 1 to 4. The BLEU metric ranges from 0 to 1 (or between 1 and 100). BLEU is insensitive to the variations of the order of *n*-grams in reference translations.

According to (Gamal *et al.,* 2022), Automatic machine translation becomes an important source of translation nowadays. It is a software system that translates a text from one natural language to one (many) natural language. On the web, there are many machine translation systems that give the reasonable translation, although the systems are not very good. Medical records contain complex information that must be translated correctly according to its medical meaning not its English meaning only. So, the quality of a machine translation in this domain is very important. In this paper, we present using matching stage from Example-Based Machine Translation techniques to translate a medical text from English as source language to Arabic as the target language. We have used 259 medical sentences that are extracted from internal medicine publications for our system. Experimental results on BLUE metrics showed a decreased performance 0.486 comparing to google translation which has an accuracy result about 0.536. English to Yorùbá Statistical Machine Translation System was developed by (Fásakin, 2017). The motivation for the work was due to the observation of Nigeria languages going towards total extinction.

Phrase-based machine translation system was formulated to translate English to Yorùbá sentences and vice-versa. The system was implemented with Moses toolkits and evaluated with BLEU score. Sani *et al.,* (2014) said we live in a multilingual society where large volumes of documents are produced in different languages. Translation is the means by which information generated in one language can be accessed by someone in a different language. Igala is one of the languages spoken in Nigeria. Igala is the ninth largest ethnic group in Nigeria and the language is spoken by about 2.5 million people. The main objective of this research is to model a language processor that can accept as input Noun Phrases in English language and translate same to Igala language. The two core technologies, corpus based and rule based technologies for building machine translation systems were carefully studied. Due to the structural differences between English and Igala, noun phrases coupled with the non- availability of large amount of parallel aligned corpus for English and Igala language, the rule based technology was adopted to develop the model. The model was implemented using VB.net programming language as front end and Microsoft Access as back end. The application was tested on 120 randomly selected English noun phrases using the Bilingual Evaluation Understudy (BLEU) method for evaluating Machine Translation systems. An accuracy of 90.9% was obtained. Abdulmusawir *et al.,* (2021) research work is aimed at bridging the knowledge gap between the most popular knowledge rich English language and the minority Ebira language spoken by the Ebira people, a minority ethnic group in part of Nigeria. Across the globe and on the internet, English language has become the most widely used language for knowledge dissemination. And presently, the majority of the indigenous people of Ebiral and also known as "Anebira" are still not proficient in their use of English language which as a result prevents them from gaining full knowledge disseminated in English language. Hence, the need to develop an automated Machine Translation System capable of translating English text to Ebira text which will help the people to tap from the abundant knowledge conveyed in English language for effective and fast development in their social, political, scientific, philosophical and economic areas of life. The system was designed to consolidate on human translators' effort and not to replace them. A comprehensive study and analysis of the two languages was carried out with the help of Ebira native speakers in Ebiraland Kogi central and some professional English language tutors at FCE Okene. The knowledge gathered provided the basis for the design and testing of the rule base, inference engine, bilingual dictionary which are important components for the proposed automated system for translation of English text to Ebira text using PHP. Making use of the word in the bilingual dictionary, the system will successfully translate your English text to Ebira. The system was evaluated using one of the popular automatic method of evaluating MT systems BLEU (Bilingual Evaluation Understudy). And an accuracy of 81.5% in translation was achieved. An improved system in the future is recommended to accommodate more complex sentences for the more benefit of the good people of Enebira.

However, Bilingual Evaluation Understudy (BLEU) correlates highly with human evaluation, it is language independent and easy to understand. Therefore, based on the above, the study used Bilingual Evaluation Understudy (BLEU) to evaluate English to Nupe machine translation system in relation to human translation. The research aims to evaluate the impact of machine translation system of English to Nupe language relative to the effectiveness of English to Nupe human translation. One of the necessary resources to accomplish this study is dataset of English phrases with Nupe reference translation. Therefore, this study involved creating data containing 50 Nupe sentences.

MATERIALS AND METHODS

BLEU is based on a core idea to determine the quality of any machine translation system which is summarized by the closeness of the candidate output of the machine translation system to reference (professional human) translation of the same text. The closeness of the candidate translation to the reference translation is determined by a modified n-gram precision which was proposed by (Papineni et al., 2002).

The modified n-gram precision is the main metric adopted by BLEU to distinguish between good and bad candidate translations, where this metric is based on counting the number of common words in the candidate translation and the reference translation, and then divides the number of common words by the total number of words in the candidate translation. The modified n-gram precision penalizes candidate sentences found shorter than their reference counter parts, also it penalizes candidate sentences which have over generated correct word forms. Bilingual Evaluation Understudy (BLEU) method is adopted in the study to evaluate rule-based machine translation system and human translation. The effectiveness of translation from English to Nupe tested using BLEU method. An n-gram can be defined as a sub-sequence of n items, from a given sequence of words (text or sentence). These items can be characters, words or sentences according to the application. Using this method to calculate modified precision, the proceed to calculate the modified precision for all n-grams up to length N. Then calculate the average logarithm with uniform weights, w_n, of all of the ρ_n. BLEU introduces a multiplicative brevity penalty, BP, but in order to compensate for harsh penalties of short sentences' deviations in length, the brevity penalty is only calculated over the whole corpus 'to allow some freedom at the sentence level' (Ken and Björn, 2018). In order to compute the precision score for each of the four *n*-gram sizes, we have to count first the number of common words in every candidate and reference sentence, and then we have to divide this sum over the total number of *n*-grams in the candidate sentence. To combine the previous precision values in a single overall score (called BLEU-score), we start by computing the Brevity Penalty (BP) by choosing the effective reference (i.e. the reference that has more common *n*-grams) length which is denoted by *r*. Then we compute the total length of the candidate translation denoted by *c*. Now we need to select Brevity Penalty to be a reduced exponential in (*r* / *c*) as shown in equation 1. To calculate the brevity penalty, let c be the length of the candidate translation and r be the effective reference corpus length:

$$\mathrm{BP} = \begin{cases} 1 & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c \leq r \end{cases} \tag{1}$$

The computation of the final BLEU score is shown in formula (2) and it is based on Brevity Penalty (*BP*) shown in formula (1).

$$\mathrm{BLEU} = \mathrm{BP}.\exp\left(\sum_{n-1}^{N} w_n \ \log p_n\right. \tag{2}$$

Where $N = 4$ and uniform weights $wn = (1/N)$. This indicates that higher BLEU score for any machine translator means that it's better than its counterparts with lower BLEU scores.

A. Dividing the Text into different N-Grams Sizes

An n-gram can be of any number of words and each of which has a name, when the sizes of the n-grams are equal to one, two, three, or four words, they are called unigram, bigram, trigram, and tetra-gram respectively. This study deals with these types. The n-gram extraction technique to extract any size of word(s) is described in Fig. 1.

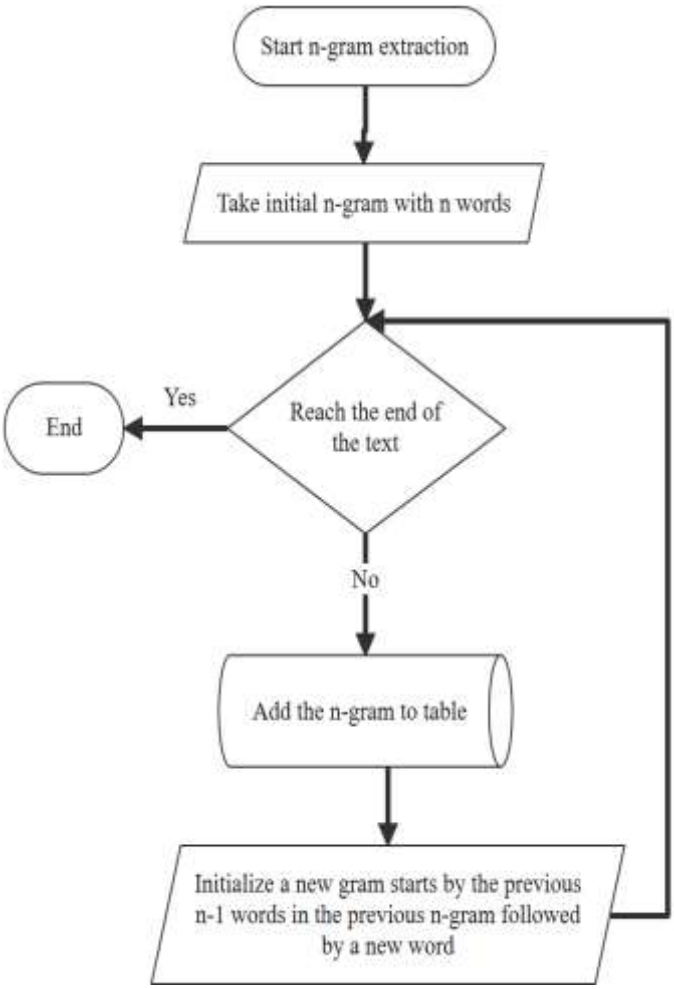

Fig. 1 N-grams Extraction Flowchart (Ken and Björn, 2018)

RESULTS AND DISCUSSION

A corpus of 50 English phrases was created and given to professional English Expert for translation from English to Nupe. This was used as the reference translation. The reference translation was stored in a table in the database called Evaluation Table which has the following fields: Translation ID, English Sentence, Reference Translation, Candidate translation and BleuScore. The same set of English phrases were given to the developed system for translation. When a phrase is entered into the control on the template, it is translated into Nupe, its ID is retrieved from the table, the translated text is stored in the Candidate translation field based on the Translation ID. This Operation was repeated until all the phrases were translated and stored. A module called BLEU evaluation module is then executed to compute the BLEU score for each of the translated phrase. The sample output is shown in Table-1.

Table-1 Sample Output

Translation ID	English Phrases	Reference Translation	Candidate Translation	BleuScore
1	The beautiful house	emi sa	Emi sa	1.00
2	My horse	Doko me	Doko	0.71
3	All their houses	Abugi kpata	Abugi kpata	1.00
4	Two doors	kpako guba	Kpako guba	1.00
5	My second Industry	Bajintun baci nyami	E baci nyami	0.45
6	The market is close	Dzuko tsoba	Dzuko tsoba	1.00
7	The food is ready	Nyagici Sajini	Nyagici giani	0.63
8	I want to sleep	Mi wa lele	Mia lele	0.88
9	The three doors	kpako guta	Kpako guta	1.00
10	Aisha swept the room	Aisha kata fin	Aisha kata fin	1.00

The result of the 50 test phrases was analyzed. The developed model was validated with two existing systems where Bilingual Evaluation Understudy (BLEU) and Experimental Subject Respondents (ESRs) datasets were used and the result is shown in Table-2.

Table-2 Model Performance Validation

Authors	Method	Accuracy
Abdulmusawir *et al.* (2021)	Bilingual Evaluation Understudy (BLEU)	81.5%
Safiriyu and Odetunji (2016)	Experimental Subject Respondents (ESRs)	82.3%
Proposed Research	Bilingual Evaluation Understudy (BLEU)	92%

The results show that the proposed model out-performed existing models in terms of accuracy.

CONTRIBUTION TO KNOWLEDGE

BLEU measures the closeness of the machine translation to human reference translation taking translation length, word choice, and word order into consideration. After applying the developed Nupe BLEU System on the collected data, the results of the study have showed an accuracy of 92%.

CONCLUSION

The accuracy of any machine translator is usually evaluated by comparing the results to human judgments. There is no standard Nupe corpus that can be used for such evaluations, for this we had to collect our data from English Nupe dictionary, Nupe novels and Nupe Teachers representing two types of data; a set of well-known English sayings and a set of sentences that were translated manually by two human translators for judgment purposes. Although the collected data was relatively small in size, as part of the future work more data will be collected and perform tests using the new data as well as any available standard data that could be found.

CONFLICT OF INTEREST

There is no conflict of interest for this research work.

REFERENCES

Abdulmusawir, T. M., Sani, F. A., Yahaya, M. K., & Eze, C. C. (2021). A System for Machine Translation from English to Ebira using the Rule-based Approach. *Journal of Scientific Research & Reports*, 27(11): 137-148

Achraf, O., & Mohamed, J. (2019). Designing High Accuracy Statistical Machine Translation for Sign Language Using Parallel Corpus: Case Study English and American Sign Language. *Journal of Information Technology Research* (JITR), 12(2)

Ayogu, I. I., Adetunmbi, A. O., Ojokoh, B. A. and Akandwanaho, S. M. (2018). Developing Statistical Machine Translation System for English and Nigerian Languages. *Asian Journal of Research in Computer Science, 1(4)*, 1–8. Retrieved from https://doi.org/10.9734/AJRCOS/2018

Fásakin, T. G. (2017). *An English to Yoruba Statistical Machine Translation System.* M. Tech Thesis, Federal University of Technology, Akure.

Gamal, D, Alfonse, M., Jimenez-Zafra, S., and Aref, M. (2022). Survey of Arabic Machine Translation, Methodologies, Progress, and Challenges. *2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*. 10.1109/MIUCC55081.2022.9781776. 978-1-6654-6677-6. (378-383).

Garba, I. A. (2012). A Morphological Analysis of Nouns and Adjectives in The Nupe Language. M. A. Thesis Ahmadu Bello University, Zaria Nigeria

Jacob, H. J. (1999) Translators and language engineering in Nigeria: A case study of Nupe language translators. *Meta: Translators Journal 44.* Retrieved April, 15th 2010, from http://www.erudit.Org/revue/meta/1999 /v4/n2/00323 ar. Pdf

Ken, P. T., & Björn, N. (2018). Design science research genres: introduction to the special issue on exemplars and criteria for applicable design science research. European. *Journal of Information Systems*, 129-139

Papineni, K., Roukos, S., Ward, T. and Zhu, W.J. 2002. "BLEU: a method for automatic evaluation of machine translation". In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL '02). Stroudsburg, PA, USA, pp. 311-318

Sani, F. A., Osuagwu, O. E., and Njoku, D. O. (2014). Machine Translation of Noun Phrases from English to Igala using the Rule-Based Approach. *West African Journal of Industrial & Academic Research, 1*(11): 17-26.

Sayuti M. S., Hassan, S. and Abdulrahman, M. S. (2022). English to Nupe Machine Translation System Using Rule-Based Approach. *International Journal of Advanced Research in Computer and Communication Engineering*. 11(1): 161-170