Journal of West African Languages

Volume 43.1 (2016)

HUMAN EVALUATION OF YORÙBÁ-ENGLISH GOOGLE TRANSLATION

By

Clement Odoje
lekeclement2@gmail.com
Department of Linguistics and African Languages
University of Ibadan
Nigeria

Abstract

The task of Machine Translation is not just about translating the text of a language to another but also its evaluation so as to monitor its improvement particularly in fluency, accuracy and efficiency. However, the only available free machine translation on Yoruba-English is "Google Translate" which has been observed to be grossly inadequate. This paper therefore examines translations done by Google Translate as against human translation in order to investigate why machine translation applications make some errors while translating human natural language. There are many matrix evaluators to do this. This paper adopts human evaluation also known as manual evaluation which is considered to be more efficient, but costly. The paper adopts Ibadan and Akungba Structured Sentence Paradigm to evaluate the translators (Google Translate and human). The translations were sent to twenty human evaluators out of which only eleven responded. The responses were subjected to statistical analysis. Findings show that human translation fares better in terms of accuracy and fluency which are informed by the quality and the quantity of training data. This paper suggests that more data, especially literary texts, should be acquired to train the translator for general efficiency and fluency.

Keywords: Machine Translation, Statistical Machine Translation, Google Translator, Human/Manual Evaluation

Introduction

The quality of a Machine Translation[1] (MT) is measured through its evaluation. The purposes of the evaluation are: (i) to help state which system (Machine) translates better; (ii) to help measure improvements on a system as a result of changes made on it. Evaluation is also carried out to enhance the quality of translation made by a machine. As Koehn (2010:24) explains, evaluation is a hotly debated topic in Machine Translation. He explains further that because there are many valid translations for an input sentence, it becomes a challenge to rate a machine better than the other if the machines use different correct sentences to translate an input sentence. From Koehn's explanation, one could conclude that as much as evaluation assists to compare the translation of a machine to another, it becomes a challenge to do that in a situation where the machines involved

[1] MT is a branch of computational Linguistics (an Applied Linguistics) is an automated translation. It is the process by which computer software is used to translate a text from one natural language to another.

translate a sentence correctly though using diferent syntactic stuctures. It also becomes a challenge to state whether there is any improvement in a machine after alteration/correction must have been made on either the pipeline or the training corpus. The only free and available MT in Yoruba-English language pair is Google Translate. This study investigates the extent towhich Google Translate translates Yoruba-English language. It also aims at a systematic description of 'Google Translate' output so as to reveal the pattern(s) of errors and how such systematic description can inform the inventors' necessary information on how to (re)train or re-invent the translator for better performance.

Google Translate adopts statistical approach, a data-driven appraoch to MT. This is a deviation from the exsisting approaches in the language pair. The approaches in the language pair are Rule-Based approach (Awofolu 2002, Odoje 2010; Eludiora et al 2011) and Hybrid approach (Abiola et al, 2015). The reason for the adoption of these approaches is the lack of electronic corpus necessary for other approaches of MT. This is the situation of many African languages because they are considered resource scarce languages[2] from a technological point of view (Kamssu, Siekpe and Elizy 2004). Oflate, some researchers are adopting the Hybrid Approach which is a combination of Rule-Based and Statistical-Based approaches. Whichever approach an MT adopts, there is a need to evaluate its output, at least to ensure the quality of its translation as well as measure its improvement regarding data pruning, tuning and other procedural processes.
There are two broad ways of evaluating MT systems: human evaluation/judgment (also known as manual evaluation) which has been adjudged extensive but expensive and time consuming, as it involves human labour which is un-reuseable (Papineni, et al 2001; 2002; Koehn 2000, 2004, 2010; Callison-Burch et al 2009; and Padó et al 2010). Human judges assess the accuracy (preservation of meaning) and fluency of the output, or rank different possible translations of sentence. Automatic evaluation is another means of MT evaluation. According to Koehn (2010:25) Automatic Evaluation metrics typically compare the output of machine translation systems with human translations. While common metrics measure the overlap in words and word sequences, and word order, advanced metrics take synonyms, morphological variation, and/or preservation of syntactic relations into account. They are however, evaluated by their correlation to human judgment. There are several automatic evaluation metric systems of which the prominent ones are: Bilingual Evaluation Understudy (BLEU), and Metric for Evaluation of Translation with Explicit Ordering **(**METEOR) (see Agarwal and Lavie 2008; Lopez 2008).

Motivation for Manual Evaluation

Lavie et.al (2007) report that automatic metrics for MT evaluation have been receiving increasing attention over the past five years. These scholars reiterate that such metrics are critical tools for current and future MT research, as they allow research teams to guide the development of their system based on frequent concrete performance evaluations. They emphasize that the models used by MT systems today and probably in the future contain a variety of parameters that need to be tuned for optimal performance. They opine that as translation quality improves,

[2] Resource scarce languages are languages that have small users advantage in relation to technology which could be ignored by the commercial world (see Chan and Rosenfeld 2012)

attention should be given to small but sensitive differences at the sentence level to further achieve better translation qualities. The following excerpt from Lavie (2007:10) puts this in perspective:

> As MT systems improve and achieve high level of translation quality, it becomes ever more important to have evaluation metrics that are sensitive to small differences between translations at the sentence-level, so that minor improvements can still be detected, concrete translation errors can be isolated and identified and system parameters can be optimized to truly achieve the best translation performance...

This implies that MT and its evaluation metrics still do not take the so-called minor but complex and complicated words into account while translating or evaluating. This is the possition of MT critics. For example, Bar-Hellie (1954) maintains that MT, especially Statistical Machine Translation (SMT)[3], limits human natural language to counting and merging. Each word's features/properties are checked before merging takes place in human natural language. SMT on the contrary, does not reckon with the inherent properties of words in the lexicon before they enter computation. This makes it easy for critics to assume that no matter the success recorded by SMT, there would still be some intricacies such as nuances, contextual meaning, semantic extension, etc, yet to be covered.

The blame is not entirely SMT practitioners'; linguists and translators also have a huge part in the blame in that they left the discourse entirely to SMT practitioners (Way 2012:9). Way (2012) provides two suggestions to reduce the impending challenge in the discourse of SMT as shown in the extract below:

> ... as SMT became the principal way of doing MT, this conciliatory tone soon changed, to the point today where many people who want to understand have been left so far behind that they feel that it is impossible to ever catch up. We expressed that the view that linguists and translators have to share the blame in allowing the field to move almost entirely in the statistical direction, especially when the seminal IBM papers very much left the door open for collaboration with the linguistic community. However, in our view SMT research will soon have to alter their position, if the use of syntax (and later, once a further ceiling has been reached, semantics) is to become mainstream in today's model. These syntactic improvement have largely come about from those practitioners with wider background than is the norm in SMT. Those without a linguistic background, then, appear to have two choices: (i) to attempt to include the linguists, so that they may be

[3] Statistical machine translation (SMT) is a machine translation paradigm where translations are generated on the basis of statistical models whose parameters are derived from the analysis of bilingual text corpora.

> of help; (ii) to continue to exclude linguists, while at the same time trying to make sense out of their writings...

To corroborate Way's point of view, one wonders the volume of parallel corpora needed for a MT to translate the Yorùbá verb *pa* with its different meanings as contexts of usage demand. The examples (1) below show how *pa* can have different senses/meanings depending on context and usage which the available corpora may not accommodate.

1 a Adé pa ejò
Adé kill snake
'Ade killed a snake'

b Adé pa irọ́
Ade tell lie
'Ade lied'

c Adé pa ilẹ̀
Ade clear bush
'Ade cleared the bush'

d Adé pa ilé ní aró
Adé paint house prep colour
'Ade painted the house'

e Adé pa èkùrọ́
Ade crack palmnut
'Ade cracked the palmnut'

f Adé pa ojú dé[4]
Ade v eyes v
'Ade closed his eyes'

g Ojò pa Adé
Rain beat Ade
'Rain beat Ade'

h Adé pa itu ọdẹ
Ade perform wonder hunter
'Ade perfomed wonders'

[4] padé is a split verb meaning to close

i Adé pa owó rẹpẹtẹ
Ade make money plenty
'Ade made lots of money'

j Adé fi ìja pa ẹ́ẹ́ta pẹ̀lú Ṣadé[5]
Ade use fight perform three prep Sade
'Ade fought Sade'

k Adé fi òróró pa orí
Ade use oil rub head
'Ade creamed/rubbed/ anointed himself with oil'

l Adé pa àsẹ fún wa
Ade issue command prep us
'Ade commanded us'

m Adé pa òṣé sí ọ̀rọ̀ rẹ̀
Ade made hiss prep word pro
'Ade hissed at his word'

n Adé pa kuuru sí wọn
Ade perform rush prep they
'Ade rushed at them'

o Adé a máa pa ariwo
Ade aspect make noise
Ade makes noise

p Ọ̀rọ̀ pa èsì jẹ[6]
Word kill response eat
'No comment'

r Adé pa àtẹ́wọ́
Adé clap palm
'Ade clapped'

s Adé pa ilẹ̀ mọ́
Ade clear ground clean
'Adé prepared/ cleaned up the surroundings'

[5] pa in example 1j is used in its idomatic and figurative sense as such its meaning is different from that of everyday usage.
[6] Example 1p is an idiomatic expression.

Examples like those in (1) above show how difficult it is to achieve translation via machine, whether rule or statisticalbased. While rule-based MT limits translation to structures of a language and is faced with structural ambiguity issues among other challenges, statistical-based machine translation is devoid of usage in context because no corpus of any language is big enough to have all the possible words of the language and its contextual usage in a print whether literary or in other forms. For example *pa* is used 304 times[7] in *Igbó Olódùmarè* which has 1744 sentences. Fágúnwà's *Igbó Olódùmarè* uses *pa* in all contexts above except example (1 b,c,e,and i). This means that SMT that uses the novel as its training corpus will not be able to translate the said examples in the context of usage as mentioned above. This is even better compared to Folajimi and Omonayi (2012) who use Genesis (the first book of the Bible) as their training corpus whereby, the whole of Yorùbá Bible has 7 out of 17 contextual usages of (1) (i.e. a, b, f, g, k, l, and o). By and large, as much as we agree with Jurafsky and Martin (2009) that it may be difficult to acquire the legal right to fiction as well as translate literal sentences which may require compromise, yet we are of the opinion that these literary texts could serve as bases for resource-scarce languages like Yoruba to start the building of SMT pending when there would be enough resources. We also believe that neutral/non-religious literary materials will yield better results than the Bible that is restricted to religious vocabulary, although, this suggestion is just one of the numerous ways of overcoming language modelling in MT.

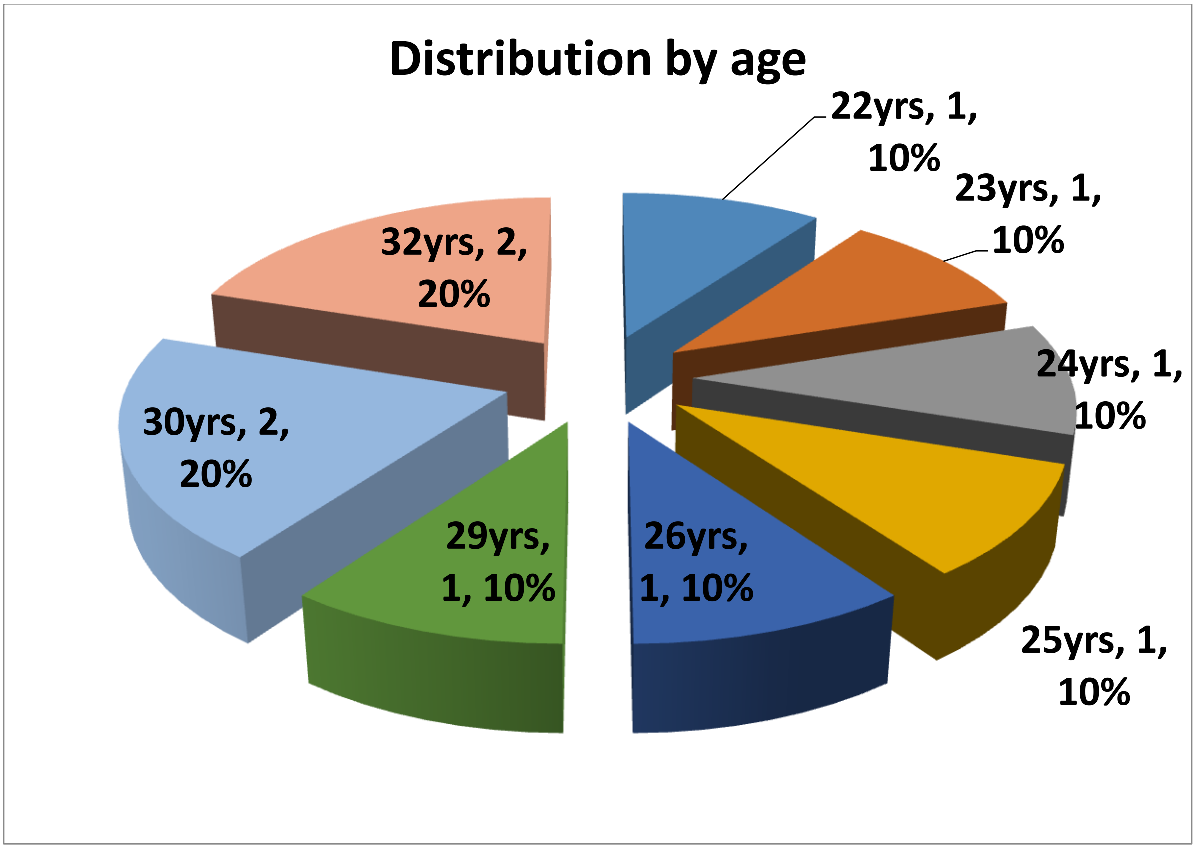
Koehn (2010:236) explains that all automatic evaluation metrics use the same trick. Each system translation is compared against one or more human translations of the same sentence. The human translations are called reference translations. An automatic evaluation metric matches the output of an MT with Reference Translation whereby an output closer to Reference Translation is preferred by giving it a high score. This could be used to compare two or more MTs or evaluate improvement in an MT based on more training or alterations made on it (also see Lin and Och 2004). With this in mind, we believe that human/manual evaluation should be conducted on Google Translate's output, measuring the extent to which it translates Yoruba-English language as well as describes the patterns of error in its output. It is noteworthy to state that there are other free online automatic translators: Facebook, Bing Translator, etc. but only Google Translate has Yoruba on the list of languages within its components.

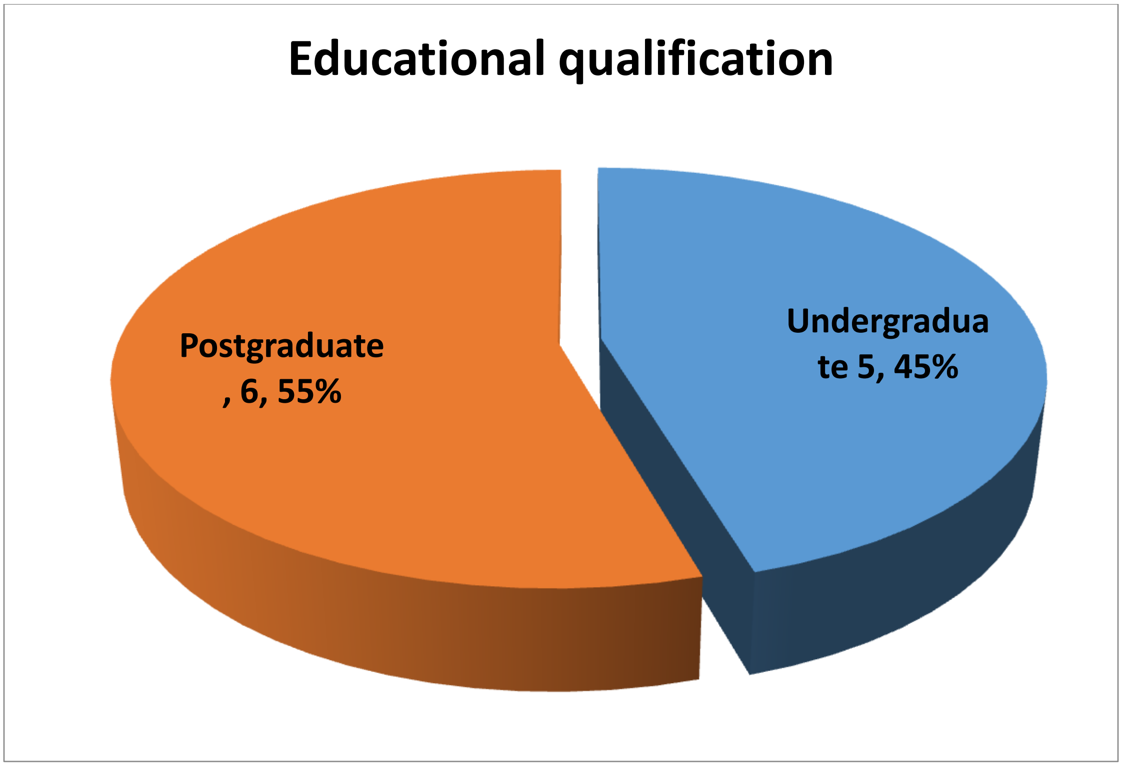
Evaluation of Google Translate

As of today, Google is the only available free Machine Translation application for Yoruba-English translation task. It translates bi-directionally. This study subjects its translations to human evaluation using Ibadan and Akungba structured sentence paradigm (SSP) to test the extent at which it translates the language pair and its bi-directional claim of translation. While Ibadan SSP is drafted in English, Akungba SSP is in Yoruba. The SSPs were translated by human and Google Translate and their translations were then subjected to human evaluation.

[7] Going by Awobuluyi (2013:88) explanation of the verb form in Yorùbá, we consider words like *pàdé, padà, parí* and *papọ̀* as compound verbs derived from *pa* and other verbs in the language which make up 304 occurences of the verb.

Twenty (20) translation practitioners in the Department of Linguistics and African Languages, University of Ibadan were contacted to evaluate the translations but only eleven (11) responded. The charts below reflect the age and educational qualifications of the respondents





According to Koehn (2010:232), one of the common approaches in human evaluation is the use of graded scale. The graded scales are based on fluency and adequacy. Fluency indicates that the output is a fluent target language involving both grammatical correctness and idiomatic word choices. Adequacy is concerned with the output conveying the same meaning as the input sentence without losing, adding or distorting any part of the message. The graded scale is given in Table (1) below:

Table 1

Adequacy	
5	All meaning
4	Most meaning
3	Much meaning
2	Little meaning
1	None

Fluency	
5	Flawless English
4	Good English
3	None-native English
2	Disfluent English
1	Incomprehensible

Table 1 informs the scale which the evaluators used in evaluating both machine and human translations. Table 2 shows each evaluator's rating of both human and machine translations and their average scores (the evaluator's rating divided by the number of items rated e.g 247/160 = 1.544).

Table 2

Score of fluency of interpretation for Google Translate	**Average Score for fluency of interpretation for Google Translate**	**Score of Accuracy of translation for Google Translate**	**Average Score of accuracy of translation for Google Translate**	**Score of fluency of interpretatio n for human translation**	**Average Score for fluency of interpretation for human translation**	**Score of Accuracy of translation for human translation**	**Average Score of accuracy of translation for human translation**
247	1.544	228	1.425	735	4.594	705	4.406
323	2.019	257	1.606	776	4.85	756	4.725
218	1.363	216	1.35	731	4.569	724	4.525
201	1.256	198	1.238	620	3.875	612	3.825
245	1.531	199	1.244	692	4.325	685	4.281
219	1.369	217	1.356	629	3.931	629	3.931
213	1.331	209	1.306	620	3.875	622	3.888
234	1.463	230	1.438	575	3.594	579	3.619
195	1.219	181	1.131	633	3.956	635	3.869
264	1.65	270	1.688	705	4.406	709	4.431
190	1.188	192	1.2	624	3.9	623	3.894
2549	**15.933**	**2205**	**14.982**	**7340**	**45.875**	**7279**	**45.394**
	Mean Score **15.933/11=** **1.449**		**Mean Score** **14.982/11 =**		**Mean Score** **45.875/11=** **4.171**		**Mean Score** **45.394/11** **4.127**

			1.362				

Mean is an arithmetic average of scores, calculated by adding all the scores, divided by total number of scores. This helps to decide which group has higher performance in cases where means are compared. Table (2) shows that the mean score of the accuracy and fluency of both Google Translate as well as human translation. The mean score of Google Translate accuracy is 1.362 which is approximately equal to 1. This implies that using the ranking in Table (1), the accuracy of Google Translate falls into category 1 which is very poor. This then means that the accuracy of Google Translate translating Yoruba to English is very poor. The mean score of human accuracy is 4.127, which approximately equals 4. From the rank of our scoring, 4 falls into category of very good, meaning that human translation is more accurate than Google Translate.

The result of the fluency shows that the mean score of MT is 1.449 which approximately equals to 2. From the ranking of our scoring about the fluency of translation in Table (1) above, 2 falls into the category of poor, implying that the fluency of a computer translation from Yoruba to English is poor. However, the mean score of the human translation is 4.171, which approximately equals to 4. Since 4 above falls into the category of very good in the ranking of our scoring in table (1), it means that the fluency of human translation is better in comparison to that of MT.

We also carried out a paired t-test[8] at 0.05 confidence interval to check if there is significant difference between the means of MT and human translation. The result of the analysis is significant with p-value at 0.000. The result of the paired t-test is significant; in other words, there is a significant difference between human translation and Google Translate's translations. Also, the mean of human translation (1329.00) is higher than that of Google Translate (449.73), showing that human translation is more efficient and effective than Google Translate as the tables below show:

Table 3
T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Human Translation	1329.00	11	119.174	35.932

[8] A T-test is a statistical examination of two population means. It examines whether two samples are different and commonly used when the variances of two normal distributions are unknown and when an experiment uses a small sample size (www.investopedia.com)

Machine Translation	449.73	11	62.629	18.883

Table 4

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Human Translation & Machine Translation	11	.674	.023

Table 5

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Human Translation - Machine Translation	879.273	89.790	27.073	818.951	939.594	32.478	10	.000

We should not just conclude that human translation is much better in terms of adequacy and fluency of translation when compared with translation done by Google Translate. At this point, the concern should be on why there are errors. For example, Google Translate translates *olú* as *emperor, divine* and *capital.* We observe that the error is not entirely the machines rather the training corpus. We are not sure of Google Translate's source of training corpus so we assume that because the general translation of *capital city* is *olú-ìlú, Olú* was translated as *capital.* Also, *Divine/Divinity* could be interpreted as *Olú ọ̀run/Òrìsà/Olúwa* which may inform the machine to tanslate *Olú* as *Divine* sometimes. Also, *Olú* could mean *emperor* depending on the context but the general usage of Olú being a personal name does not need any translation. Therefore, there could be high frequency of Olú's translation as emperor, divine and capital based on the information provided in the training corpus as shown in example (2) below:

2	**Source sentence**	**Translation by Google Translate**	**Translation**
a.	Olú rí mi	My capital	Olú saw me
b.	Olú rí ẹ/ọ	Capital	Olú saw you
c.	Olú rí i	Capital found	Olú saw him/her
d.	Olú rí wọn	Their capital	Olú saw them
e.	A rí Olú	A capital	We saw Olú
f.	Ẹ rí Olú	Their capital	You(pl) saw Olú

g.	Ìwọ rí Olú	You see the Divine	You saw Olú
h.	Àwa pàápàá rí Olú	We even saw the Emperor	We too/even saw Olú
i.	Ẹ̀yin rí Olú	Emperor eggs	You(pl) saw Olú
j.	Ẹ̀yin gan-an rí Olú	Emperor eggs	You(pl) too saw Olú

The frequency of translating *Olú* as Capital, Divine and Emperor is not consistent. Its inconsistency needs further investigation. From the data above, whenever Olu is used with short pronouns especially in the object position, *Olu* is translated as *capital*. When used with pronominal, then, the choice is between *Emperor* and *Divine.*

It is observed that (2 a,b,c,d,e,f,i,and j) that translations of Google Translate substituted phrases for sentences, and the translation did not convey in any way the meaning of the source language. Only (2 g and h) show a reasonable level of translation except that *Olú* was mis-translated.

The translations also show that Google Translate failed to learn Yorùbá with its tones as well as its manipulations (see Owolabi 2013). (2 i and j) show that Google Translate did not notice the difference between egg (ẹyin) and the pronominal (ẹ̀yin); hence the translation of plural third person pronominal for egg.

However, we need to point out that some complex sentences like (3) below were translated appropriately.

3	**Source Sentence**	**Translation by Machine**
a.	Kò yé mi bí mo ṣe ṣe é mọ́	I did not understand how I did it
b.	Bóyá ó bọ́ ní àpò mi	Perhaps it was lost in my bag

Even though the emphasis in the source sentence is lost in the translation by Google Translate in (3a), it is still meaningful to a large extent. (3b) is a bit ambiguous but its translation could be one of the meanings or a necessary compromise as opined by Jurafsky and Martin (2009:875). They said that “true translation, which is both faithful to the source language and natural as an utterance in the target language is sometimes impossible and if you are going to go ahead and produce a translation anyway, you have to compromise”. It should be noted that the aim of MT has shifted from good, quality, direct and unedited translation to the production of the first draft meant to be edited by humans.

Conclusion

It is not enough to just say that MT is not translating African languages appropriately; hence we should limit ourselves to human translation. We suggest that literary texts could be used for training so as to expand the frontiers of the activities of MT in relation to African languages to at least produce a first draft for its users before necessary editing is done. Even though literary texts have their unique challenges, they could be used as a means to build corpora for this kind of exercise. Hence efforts should be channeled towards corpora resource building. When we have enough corpora, with improved technical know-how, the machine should be able to give close to reasonable translation. This will save time and cost. For example it took a computer an average of 0.7 seconds to translate a sentence while it took humans an average of 4 to 5 minutes to do the same. Human translators charge a substantial amount for the service of translation while Google translate charges nothing. It would also be an avenue to contribute to the ongoing discussion over the global space on the internet; that Africans with their language can showcase their potentials without any language limitations whatsoever. Scholars should therefore, come together with their expertise to achieve the necessary needed improvement in terms of resource building and technicalities. African governments need to ensure that projects in relation to technology that are language focused are sponsored no matter how the small financial resources in the region are.

REFERENCES

Abiola O.B, Adetumbi A.O. and Oguntimilehin A. 2015. Unising Hybrid Approach for English-Yoruba Text to Text Machine Translation System (Proposed). *International Journal of Computer Science and Mobile Computing.* Vol 4, Issue 8, August, pp 308-313

Agarwal, A. and Lavie A. 2008. METEOR, M-BLEU and M-TER: Evaluation Matrics for High-Correlation with Human Ranking of Machine Translation Output. *Proceeding of the Third Workshop on Statistical Machine Translation*. pp 115-118 Columbus, Ohio USA @ June, 2008 Association of Computational Linguistics.

Awofolu, O. 2002. *The Making of A Yoruba-English Machine Translator*. St Mary's City: St Mary College of Maryland

Awobuluyi, Ọladele. 2013. Ê*kö Gírámà Èdè Yorùbá*. Oṣogbo, ATAM Limited

Banerjee, S. and Lavie, A. 2005. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgment. *Proceeding of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measure for Machine Translation and/or Summarization*. Ann Arbor MI 65-72

Bar-Hillel Y. 1954. Some Linguistic Problems Connected with Machine Translation. *Philosophy of Science*, vol.20, pp 217-225

Bible Society of Nigeria. 2003. The Holy Bible in Yoruba. Amity Printing Co. Ltd.

Chomsky N, 2006. *Language and Mind*. Cambridge. Cambridge University Press

Callison-Burch C., Koehn P., Monz C., and Schroeder J. 2009. Findings of the 2009 Workshop on Statistical Machine Translation. In Proceedings of the Fourth Workshop on Statistical Machine Translation (WMT09), March.

Eludiora S.I, Salawu A.A, Odejobi and Agbeyangi's (2011) Ife MT: An English Yoruba Machine Translation System http://www.slideshare.net/aflat/ifemt-an-englishtoyorb-machine-translation-system. Retrieved on 08/08/2014

Fagunwa D.O 2005. *Igbó Olódùmarè*. 20th ed. Ibadan. Nelson Publishers Limited

Folajimi Y.O. and Omonayin I 2012. Unising Statistical Machine Translation (SMT) as a Language Translation Tool for Understanding Yoruba Language. *EIE's 2nd Intl' Conf.*

Com., Engergy.Net., Robotic and Telecom.| eie con2012

Hutchins, J. 2001. Machine Translation and Human Translation: in Competition or in Complementation? *International Journal of Translation* 13, 1-2 Jan-Dec 2001, pp.5-20.

__________ and Somers H.L 2009. *An Introduction to Machine Translation*. London: Academic Press Limited

Jurafky, D and Martin J 2009. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. New Jersey. Pearson Education, Inc

Kamssu, J.; Siekpe, J.S. and Elizy, J.A. 2004 Shortcomings to Globalization: Using Internet Technology and Electronic Commerce in Developing Countries. *The Journal of Developing Areas*, Vol. 38, No. 1 pp. 151-169. Retrieved Jan 25, 2012 from http://www.jstor.org/stable/20066700

Koehn, P 2004. Statistical Significance Test for Machine Translation Evaluation. Proceeding of EMNLP 2004. Lin D and Wu D Eds 1-8

_______ 2010. *Statistical Machine Translation*, Cambridge. Cambridge University Press

_______ 2012. Statistical Machine Translation System: User Manual and Code Guide. Retrieved Mar 12, 2012 from www.moses.org.

Lavie A. and Agarwal A. 2007. METEOR: An Automatic Metric for MT Evaluation with High Levels of Correlation with Human Judgments. In Proceedings of the Second ACL Workshop on Statistical Machine Translation, pages 228–231, Prague, Czech Republic, June.

Lavie, A. 2014. Automated Metrics for MT Evaluation. Pp 11-731; Feb 20, 2014

Lopez, A 2008. Statistical Machine Translation. *ACM Computing Survey*, Vol 40, No 3, Article 8, August

Mearne, M. and Way, A. 2011. Statistical Machine Translation: A Guide for Linguists and Translators. *Language and Linguistics Compass* pp 1-21

Odoje C.O 2010. The Role of Syntax in the Yoruba-English Rule-Based Machine Translation. A Master Project Submitted to the Department of Linguistics and African Languages, University of Ibadan

Owolabi k. 2013. *Ìjìnlẹ̀ Ìtúpalẹ̀ Èdè Yorùbá: Fònẹ́tîìkì àti Fonọ́lọ́jì*, Ibadan. Universal Akada Books Nigeria Limited.

Pado, M. Galley, D. Jurafsky, and C. D. Manning. 2010. Robust machine translation evaluation with entailment features. In Proceedings of ACL.

Papineni, K., Roukos, S., Ward, T., and Zhu W. 2002. BLEU: A Method for Automatic Evaluation of Machine Translation. *Proceeding of the 40th Annual Meeting of the Association for Computational Linguistics* (ACL) Philadelphia, July 2002 pp 311-318.

Snover, M., Dorr, B., Schwaetz, R., Micciulla, L., and Makhoul, J. 2006. A study of Translation Edit Rate with Targeted Human Annotation. *Proceeding of the Conference of the Association for Machine Translation in America* (AMTA) 223-231