

---

---

Reliability of Hybrid Technique in Word-Level Machine Translation for Pidgin to English  
Language: An Evaluation

By

**Nwaekwu C. Kingsley and Nwanze D.**

Computer Engineering School,  
Nigerian Naval Engineering College,  
Ogorode New Port, Sapele, Delta State, Nigeria  
Email: [enkaycee@yahoo.com](mailto:enkaycee@yahoo.com)

---

---

**ABSTRACT**

*An application for word-level translation of the Nigerian Pidgin English (NPE) to Standard English has been developed, in which the hybrid approach of literal translation technique was adopted, combining word-level and phrasal conjoins with the aim of improving quality of direct translation. The aim of this paper is to evaluate the reliability of the developed system in terms of accuracy of translation and ease of MT systems (MTs). Opinions were sampled from 100 randomly selected respondents within Benin City, amongst linguistic experts and natural speakers of both English and pidgin languages. The output of the system (translations) of selected texts were presented to the respondents to elucidate unbiased evaluation of the stated objectives. The study shows that, though not very accurate, hybrid approach of literal MTs offers appreciable level of reliability in translation of pidgin expressions to Standard English compared to other approaches.*

**Keywords:** Machine Translation, Nigerian Pidgin English, Computational Linguistic, bitext alignment, direct translation

**INTRODUCTION**

Machine Translation (MT), is a subfield of Computational Linguistics which deals with the use of a special computer application (software) to translate texts from one natural language to another. Often referred to as MT systems (MTs) or Computer-Aided Translation systems (CATs), it basically uses software programs that have been specifically designed to translate verbal or written texts from one language to another (Ager, 2018). In written texts, MTs literally translates texts by substitution or replacement of words in one natural language for words in another. Abdolmedhi (2002) described three techniques or approaches in translation, including:

1. translation at the word level
2. translation at the level of sentence, and
3. conceptual translation

Word-level translation, which is also referred to as *direct translation*, or literal translation (LT) is the core element of MTs (Padmamala, 2014). In LT, texts are rendered from one language (source) to another language (target) *word-for-word*, normally with or without conveying the sense of the original text (Wikipedia; Grassilli, 2016). Hence, in most cases, LT does not give any sense or useful order of the original expression, always resulting in mistranslations of idioms in the target language.

In LT, priority is given to lexical correspondences and results in ungrammatical sentences (Palumbo, 2009). Hence, LT has been often associated with scientific, technical, technological, or legal texts. According to Chen (2009), every natural language has its own idioms, which Webster's New World Dictionary of American English defines as "the syntactical, grammatical or structural form peculiar to a language".

Literal MTs simply employ a database of words and their translations, and probably produces translations full of errors since the translator makes no efforts to convey correctness of grammar or shades of meaning of the source language. However, it might be useful in seeing how words are used to convey a meaning in the source language. For instance, the German word *kindergarten* would be literally translated into English as *children garden*, whereas the real expression refers to the school year between pre-school and first grade.

The LT adopts simple strategy of translating by taking word-to-word from the original (source) phrases or sentences (Abdul *et al*) and replacing it with equivalent texts in the target language. The translation normally follows closely the form of the source language. Palumbo (2009) explained that LT could though be as close as possible to the original but still does not ensure its naturalness in target language. Grassilli (2016) stated that the concept of literal translation can be regarded as an *oxymoron* (contradiction in terms), for the fact that literal denotes *something existing without interpretation*, whereas a translation, by its very nature, is an interpretation (an interpretation of the meaning of words from one language into another).

A major challenge of MTs in word-level translation technique is that some phrases cannot be translated literally (Koehn and Knight, 2008). Another challenge is also apparent where a word in one language may have more than one meaning in another language (Padmamala, 2014). To address this challenge, this study has taken a single *main* or popular meaning for source-target texts, while ignoring other unpopular meanings. Also, conjoining of texts has also been considered to reduce the frequency of mis-translatable phrases, while isolated texts are left unchanged.

MTs however can be useful in aiding translation tasks with some sorts of human tweaking. Ager (2018) had cited the benefits of MTs to include the following:

1. *Time saving* – whole lot of time that could be spent poring over bilingual dictionaries to translate words can be saved by MTs as the application can provide content translation quickly
2. *MTs is comparatively cheap* – the cost of MTs in a long-run is comparatively cheaper than the cost of hiring professional translators
3. *Confidentiality* – sometime, sensitive data need not be given to hired translators if confidentiality is not to be compromised.
4. *Versatility and reliability* – often, MTs can be adapted to translate texts in any language with same result.

Some known and identifiable drawbacks MTs include:

1. *Lack of accuracy* – MTs does not render accuracy in translation. It only provides contextual ideas. This often refers to the mistranslation of idioms that affects the meaning of texts, making it unintelligible.
2. *Lack of order* – systematic and formal rules are flouted by MTs. It does not make use of experience or mental agility like human translator does. It is often considered a bad practice of conveying word by word translation in non-technical texts.

Literal translation is very much applicable with languages that are really very close in cultural terms. It is acceptable only if the translated texts retain the same syntax, the same meaning and the same style as the original text (Mathieu, 2016). The technique can be used in some languages and not others, depending on the sentence structure. LT renders translation that represents the precise meaning of the original text but does not attempt to convey its style, beauty, or poetry. Sometimes it works satisfactorily and sometimes it does not.

Early machine translations were famous for word-level translation by simply creating a database of words and their equivalent translations. Attempts in further studies in translation have utilized common phrases which resulted in better grammatical structure and capture of idioms but with many words left in the original language (Grassilli, 2016). Ungureanu (2008) also described word-for-word translation as not being capable to produce a satisfying target language texts, and in conclusion stated that MT, although imperfect, is not only a possibility, but also an actuality.

Nowadays, the most systems in use are based on a combination of techniques and algorithms that can improve on correct natural sound of the translation. However, professional translators can use machine translation to create a rough translation first, which they can then tweak (Grassilli, 2016).

The objectives of the study presented in this paper include steps to:

1. Evaluate how contextual meanings can be obtained from literal translation
2. Evaluate how reliable the word-level translation can provide insights in language translation
3. Determine how the translation technique can be improved upon in terms of efficiency in non-orthographic language translation such as the Nigerian Pidgin English.

### **Texts Conjoining**

Conjoining of words or texts was adopted to form phrase(s) that have unit translations either as a source or target language. This approach was aimed at achieving further improvement in the translation output. The texts combination helps in reducing the chances of occurrence of multiple meanings and minimise mistranslations.

There are several instances where group or words have a meaning separate from literal meaning of individual words, eg. *He kicked the bucket* (he is dead) cannot be translated word-for-word but as a unit because the translation would remain unchanged. The need for combining texts has helped in improving the quality of our translation. For instance, in the following expressions:

*dey* meaning *am* (which is an adverbial for “to be”. Other conjugations include are, is, were, was). For instance:

*dey come* => (meaning the adverb phrase “be coming”) translates into “*am come*” if *am* is taken as root meaning while leaving out others.

This could be used in the expressions such as:

1. *I dey come* (I **am** [or **was**] coming) would be translated as “*I am come*”
2. *I dey sleep* (I **am** [or **was**] sleeping) is translated as “*I am sleep*”
3. *e dey come* (it [or he, she] **is** [or **was**] coming – where *e* is third person pronoun) is translated as “*it am come*”
4. *dem dey come* (they **are** [or **were**] coming) is translated as “*they am come*”

5. *we dey for house* (“we are at home” or “we are in the house”) is directly translated as “we am for house”
6. *I dey hammer* (a slang, meaning “I am [or was] making it”, especially fortune) is translated as “I am hammer”

There are other numerous text occurrences in pidgin that can produce mistranslations. In order to improve in the quality of machine translation results, however, words (or texts) are conjoined in this study, such as:

1. I dey => I am
2. dem dey => they are
3. e dey => he is
4. I don => I have
5. e don => he has
6. don hammer => have made it
7. we dey for => we are in

In the examples such as:

- *I dey come => I am come*
- *dem dey sleep => they are sleep*
- *e dey come => he is come*
- *we dey for house => we are in house*

This technique is a further improvement in the alignment. Thus, texts conjoining thus has an effect in improving the translation. However, tenses are being ignored. In pidgin English, there are numerous texts (or words) that can be combined as phrases or sentences and mapped into their direct meanings in an effort to make linear translation better off. However, the herculean challenge is the compilation time for such texts that can be joined together, as well as the growing database. However, the design of our system is made to enrich the bitext database (corpus) by being open to accept new text-entries.

## METHODOLOGY

Opinion survey was adopted in the study. A total of one hundred (100) linguistic experts were randomly selected from a pidgin English-speaking community in Benin City in Edo State, southern Nigeria. The study was based on simple bitext alignment approach. A list of 20 pairs of pidgin-to-English translations produced by the developed MTs was presented to the respondents for their unbiased evaluation. Analyses and evaluation of their opinions and responses are presented in the subsequent sections. The analyses obtained are also presented in the subsequent illustrations. In the design of the system, the following were considered during the bitext mapping:

1. Separate/individual words separated by white spaces
2. Words separated by white spaces but conjoined into a single unit of translation as phrases
3. Full text (sentence) whereby a whole expression is taken as single unit of translation.

In the translation, certain words were kept unchanged. These words either do not have pidgin or English equivalent, or their meanings are not affected by the translation. Two arrays of indexed texts to be processed during the translation were created, namely:

1. Array of defined texts – containing an array of texts entries found in the bitext database, and
2. Array of undefined texts – containing an array of texts not found in the database.

The two arrays were incrementally indexed. Undefined texts can likewise be defined as new entry in the database at any instance or used in its undefined form. Items in the two arrays are sorted and merged to produce final translation output.

This technique requires an exhaustive search (or brute-force) algorithm because it performs searches exhaustively for all possible alignment options within the time complexity of  $\Omega(2^n)$ .

The following algorithm or pseudocode was thus developed for the study:

```

// Pre-process
Procedure define_bitext (text, meaning, corpusIndex) Store_database (tableBitext)
Procedure tokenize(token, tokenIndex)
Store_database(tableExpression)
Procedure Lookup(token,tableBitext)
Token: = text
Loop While not End
Procedure translate(tokens)
Tokenize()
Lookup()
If token found: = true
Load(token, meaning, tokenIndex) to Array(defined)
Else:
Load(token, tokenIndex) to Array(undefined)
End
// Process
Gettext
Preprocess text()
Text cleanup //remove left and right non-text strings
Text count := 0 //initialize text count
Input = Source text, S
Output = target meaning, T
Process text;
n = text count
Split texts into elements n //tokenize input strings
For i = 1 to n;
Assign index k to each token, Si=1..n
Fulltext = merge(string(token(k), “ ”))
K++
Until k = n
Search [Fulltext]
If exists;
Replace(fulltext, T)
Display{Output}
Else:

```

```

For i = 1; j = i +1; z = 1
compare(merge(token(i), token(j)), source)
i++
z++
until (i = n-1)
if compare true
    replace(tokens(i,j), target)
i=i+2
z=z+2
display{Output}
else:
replace(token(i),target)
i=i+1
z=z+1
display{output}
End
Translation:= Merge((Array(defined)+Array(undefined))
SORT By tokenIndex ASC)
Display Translation

```

### **Specification of the system**

The flowchart in Figure 1 shows the working of the system, which is further explained in the sequence outline below:

- capture input (*source*) texts in pidgin and clean it up.
- search exhaustively for the entire sentence in the database, if it exists as a unit translation. If found, return the target text in the array of defined texts.
- If the full text doesn't exist in the database, split the sentence into tokens.
- After tokenization, perform combination of neighbouring tokens into phrases and search for phrasal words in the source field until end of token. If found, mark the entries in the array of defined texts.
- If or where phrasal words do not exist, return the individual word-for-word translation.

### **IMPLEMENTATION**

The implementation was based on hybrid linear alignment approach for source-to-target tokens-and-phrasal mappings. A database of bitexts was created, containing over 500 words/texts/phrases mapped from source to target. Here, pidgin texts are referred to as *source* and English meanings as *target*.

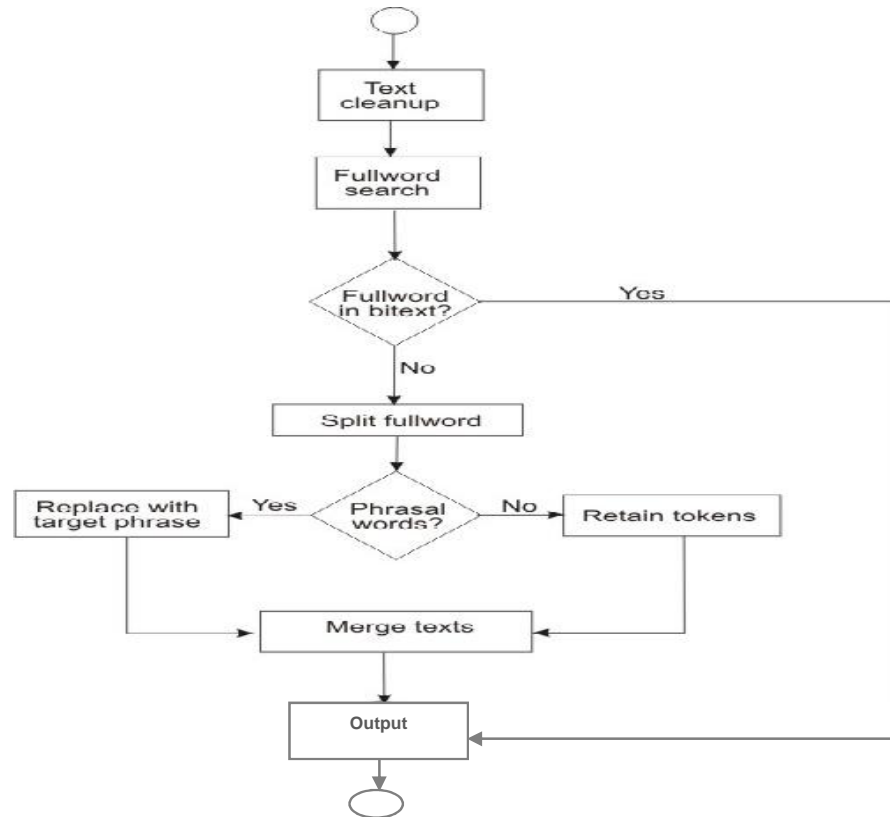


Figure 1: Process flowchart

The implementation was carried out on an object-oriented platform, using MySQL as the database backend. The front-end was a graphical user-interface (GUI) designed with PHP. Pidgin words/texts can be inputted via the input GUI.

Based on the parameters P<sub>1</sub> to P<sub>4</sub> in Table 1, the study considers the following question for evaluation:

1. Does the translation convey any reliability of MT system in terms of correctness as seen in the translation?

Table 1: Parameters of study

	Parameter	SA	A	I	D	SD
1	P <sub>1</sub> : Accuracy in translating Pidgin to English					
2	P <sub>2</sub> : Clarity of information					
3	P <sub>3</sub> : Provide insight into translation problems					
4	P <sub>4</sub> : Ease of use					

Where SA=Strongly Agree; A=Agree; I=Indifferent; D=Disagree; SD=Strongly Disagree  
 The implementation flowchart is presented in Figure 2.

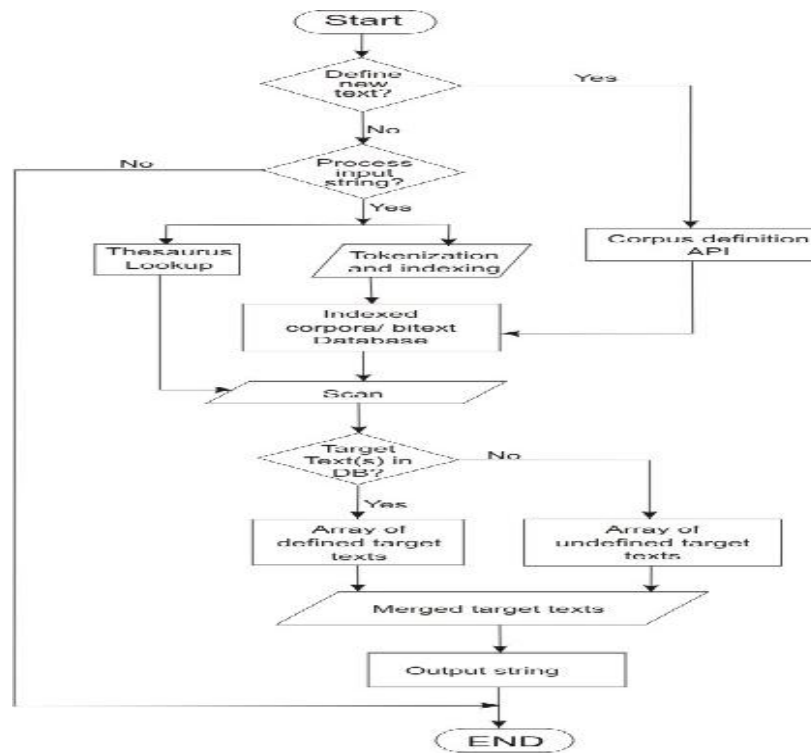


Figure 2: Implementation flowchart

Twenty source-to-target (S<sub>1</sub>-T<sub>1</sub>, ..., S<sub>20</sub>-T<sub>20</sub>) translations were sampled for the study as shown in table 2.

Table 2: Translation samples

	<b>Pidgin Expression (Source)</b>	<b>System Translation (English) (Target)</b>
1	<i>wetin be dat</i>	what is that
2	<i>where you dey go</i>	where are you going
3	<i>i don tire for your wahala</i>	I have tire for your trouble
4	<i>the two women get bad belle for dem dem</i>	The two women get jealousy for themselves
5	<i>the tin wey dey pain me be say you no dey hear word</i>	The thing that is worry me is that you do not listen
6	<i>e don carry belle</i>	it has pregnant
7	<i>Oyibo na real mumu</i>	White man is real foolish
8	<i>i see am korokoro wen im take am</i>	I see him vividly when he take him
9	<i>na so so dem say dem say dem dey hear from your mouth</i>	is all about rumour them is hear from your mouth
10	<i>im way na magomago na im make am dey fall for wahala</i>	his way is something dubious that is why him is fall in trouble
11	<i>i don hammer</i>	I have made it
12	<i>dem see me finish for this place</i>	them undermine me for here
13	<i>make dem waka comot</i>	Let them walk out
14	<i>na wuruwuru tin e dey do</i>	is unclear thing he is do



	Pidgin Expression (Source)	System Translation (English) (Target)
15	<i>i wan reach Lagos</i>	I want to reach Lagos
16	<i>e dey yab im papa</i>	he is insult his father
17	<i>martin don go jand</i>	Martin have gone abroad
18	<i>wahala full naija</i>	trouble full over Nigeria
19	<i>who be the guy</i>	who is the boy
20	<i>make we chop fast</i>	let us eat fast

The pairs of bitexts are being tagged  $S_1 - T_1, S_2 - T_2, \dots, S_{20} - T_{20}$ , for source to target, where S = Source, T = Target. The translations for the  $S_i-T_i$  pairs were evaluated and represented in Figure 3.

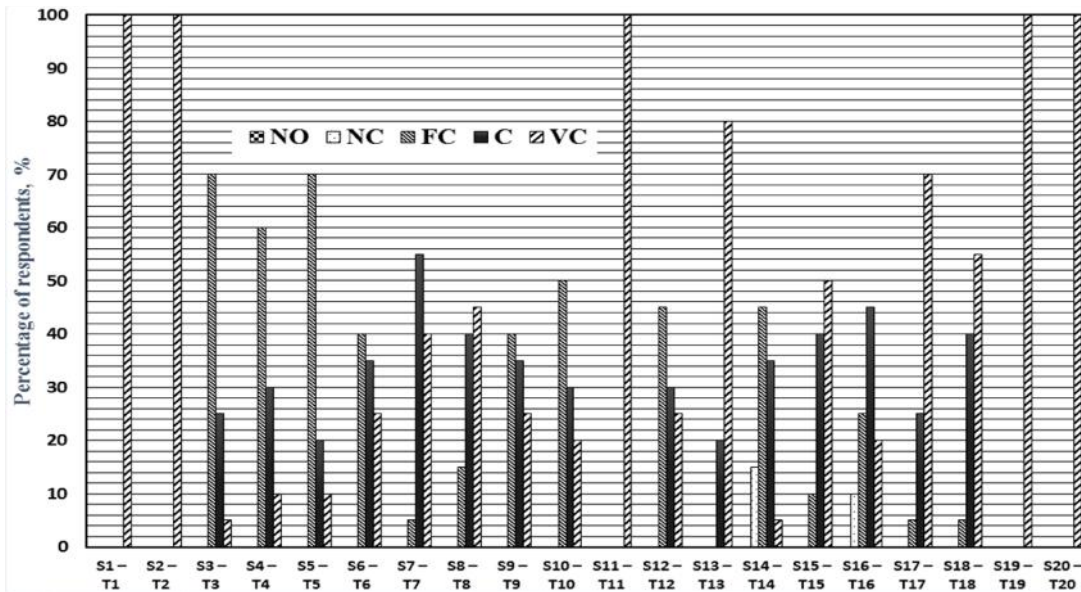


Figure 3: Respondents' opinions on the correctness of the system  
 VC: Very Correct; C=Correct; FC=Fairly Correct; NC=Not Correct; NO=No opinion

Based on the opinions of the respondents, the outlined parameters  $P_1 - P_4$  were evaluated and the results presented in Figures 4 - 7.

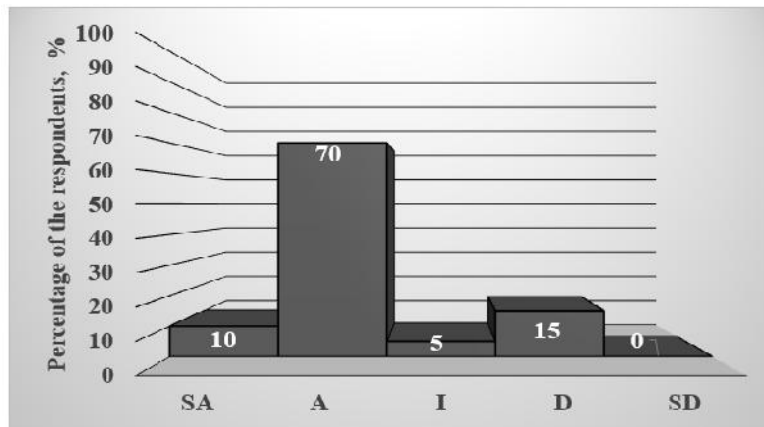


Figure 4: Respondents' opinions on accuracy of the system, P1

On the accuracy of the translations, P<sub>1</sub>, Figure 4 shows that 80% of the respondents (SA and A) collectively *agreed* that the translations were accurate, while 15% of them *disagreed* and 5% were indifferent.

Figure 5 represents the opinions of the respondents on the clarity of the information conveyed in the translation, P<sub>2</sub>.

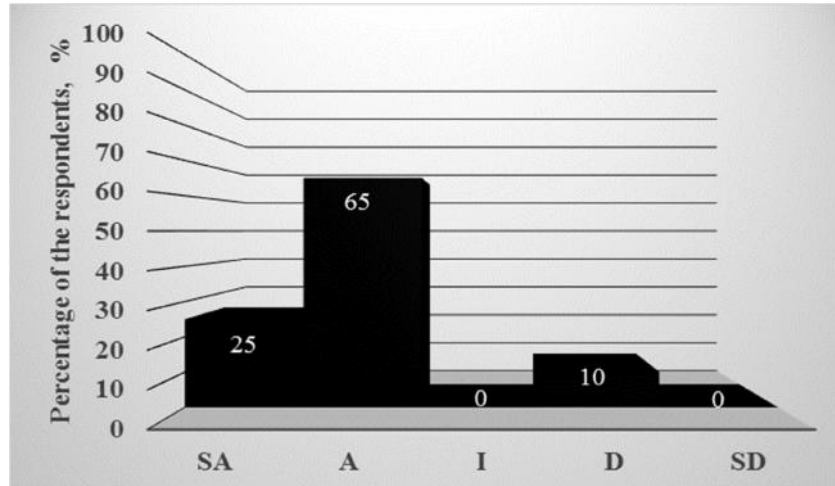


Figure 5: Respondents' opinions on clarity of the system, P<sub>2</sub>

While 90% of the respondents generally agreed that the system provides sufficient clarity for the translation to pass information from the source, only 10% of them disagreed.

Figure 6 is the presentation of the respondents' opinions on the insightfulness of the system i.e ability to comprehend the context of the translation.

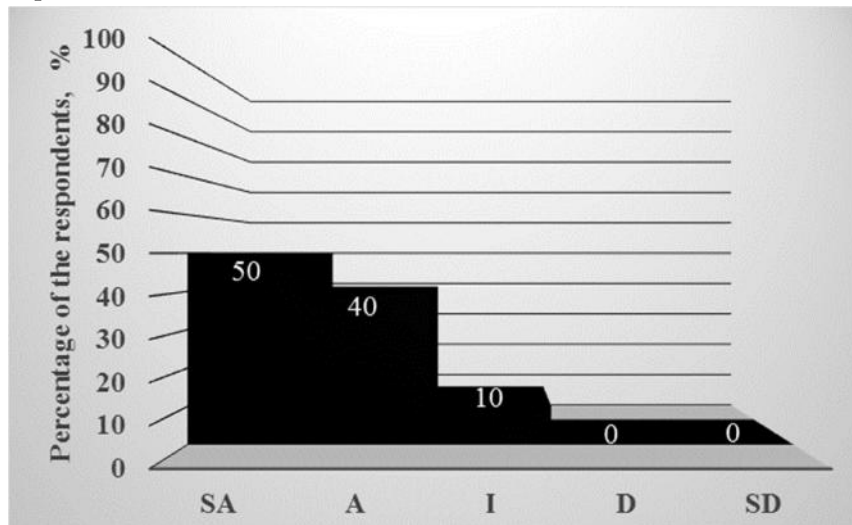


Figure 6: Respondents; opinions on the insightfulness of the system, P<sub>3</sub>

Altogether, 90% of the respondents agreed that the translation is insightful, however, 10% of them were indifferent.

Figure 7 is the presentation of the respondents' opinions on the ease of use (i.e simplicity) of the system. In a unanimous manner, all the respondents agreed so strongly that the system is simple to use. This is as a result of the good human-computer interaction technique that has been adopted in the design of the system.

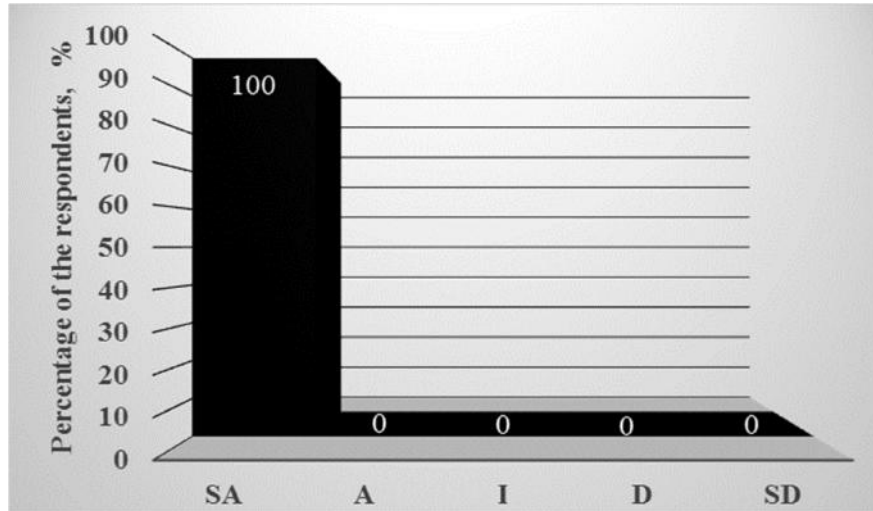


Figure 7: Respondents' opinions on the simplicity of the system

Finally, as shown in Figure 8, stacking of the entire study parameters shows that for P<sub>1</sub> to P<sub>4</sub>, the simple mean respondents are shown as follows:

- SA: 25%
- A: 57.5%
- I: 17.5%
- D: 0%
- SD: 0%

Thus, since the mean respondents for SA and A are above 20%, then the study is significantly relevant.

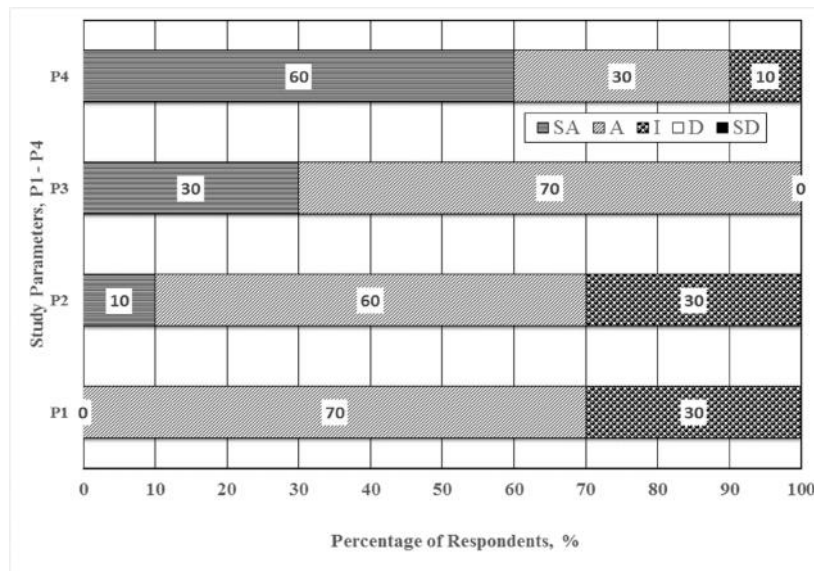


Figure 8: Stacked chart of P<sub>1</sub>-P<sub>4</sub>.

The one-way ANOVA was also evaluated as tabulated in Table 3 to further test the relevance of the study. The analysis shows that since P-value = 0.000763 < 0.05, the study is thus significant.

**Table 3: Single Factor ANOVA**

## SUMMARY

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
SA	4	100	25	700
A	4	230	57.5	358.3333
I	4	70	17.5	225
D	4	0	0	0
SD	4	0	0	0

## ANOVA

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	8950	4	2237.5	8.717532	0.000763	3.055568276
Within Groups	3850	15	256.6667			
<b>Total</b>	<b>12800</b>	<b>19</b>				

## CONCLUSION

The study has shown, with considerable satisfaction and significant reliability, that hybrid literal translation can produce better output in MTs for an unstructured language than normal word-to-word approach. Evaluations of the study show that MTs developed for Pidgin to English language translation using hybrid bitext alignment approach is acceptably accurate, clear in context, easy to comprehend and simple to use.

Further development could focus on algorithms and linguistic models on improved translation system for unstructured language like the Nigerian pidgin English and other native languages.

## REFERENCES

- Abdolmedhi, R., 2002. The Invisible in Translation: The Role of Text Structure. Bahrain University. <http://www.translationdirectory.com/article12.htm>
- Abdul, H. H.; Abdul, A.A.Y.; Mamat, R.; Abdul, R. N., 2013. Literal Translation from English and Malay in the written communication among Malay learners of French. In: Richards and Schmidh, 2002. *Asian Social Science*; Vol. 9, No. 9; 2013. Published by Canadian Center of Science and Education. [www.ccsenet.org/ass](http://www.ccsenet.org/ass)
- Ager, S., 2018. The Advantages and Disadvantages of Machine Translation. Omniglot.
- Grassilli, C., 2016. Translation. <http://translatorthoughts.com/2016/02/word-for-word-translation/>
- Koehn, P.; Knight, K., nd. Knowledge Sources for Word Translation Models. Information Sciences Institute, *University of Southern California*. Available at <https://www.aclweb.org/anthology/W/W01/W01-0504.pdf>
- Linli, C., 2009. On Literal Translation of English Idioms. *Journal of English Language Teaching*. Vol. 2 No. 2. [www.ccsenet.org/journal.html](http://www.ccsenet.org/journal.html)

- Mathieu, 2016. 7 Translation techniques to facilitate your work. <https://culturesconnection.com/7-translation-techniques/>
- Padmamala, R., 2014. Word Level Translation (Tamil-English) with word sense disambiguation in Tamil using Ontnet. *International Journal of Engineering Sciences*, Vol. 15. Vidya Publications, 2014. <http://www.ijoes.vidyapublications.com>.
- Palumbo, G., 2009. "Key terms in translation studies". London: Continuum.
- Ungureanu, C., 2008. Advantages and Disadvantages of Machine Translation. Universitatea Pitesti. Published by the University of Alba Iulia, Romania. Available at [http://www.uab.ro/reviste\\_recunoscute/philologica/philologica\\_2003\\_tom3/56.doc](http://www.uab.ro/reviste_recunoscute/philologica/philologica_2003_tom3/56.doc)
- Wikipedia. Literal translation. <https://en.wikipedia.org/wiki/Literaltranslation>