ADAPTING RULE BASED MACHINE TRANSLATION FROM ENGLISH TO BANGLA

JUDITH FRANCISCA*

Computer Science & Engineering Department, United International University,
House # 80, Road # 8/A, Dhaka-1209, Bangladesh
judithfran@gmail.com
http://<www.uiu.ac.bd>

MD. MAMUN MIA

Computer Science & Engineering Department, United International University,
House # 80, Road # 8/A, Dhaka-1209, Bangladesh
md.mamunmia@gmail.com
http://<www.uiu.ac.bd>

DR. S. M. MONZURUR RAHMAN

Computer Science & Engineering Department, United International University,
House # 80, Road # 8/A, Dhaka-1209, Bangladesh
mrahman99@yahoo.com
http://<www.uiu.ac.bd>

Abstract

This paper presents the adapting rule based machine translation from English to Bangla. The proposed language translation model relies on rule based methodologies especially fuzzy rules. There are "If - Then" basis rules are applied for the English to Bangla language translation. In this language translation, we use rough set technique for Knowledge Representation System. This technique is used to classify each English sentence to a particular class using attributes of that English sentence and then translate them to the Bangla sentence using the rules that is produced earlier in the language translation system. Here English to Bangla bilingual dictionary has been formed for the purpose of language translation.

***Keywords*:** Machine Translation; Natural Language Processing; Language Translation; Artificial Intelligence; Knowledge Representation System.

1. Introduction

Bangla (can also be termed as Bengali) is one of the richest languages among all the languages exist in the world and one of the largely spoken languages in the world. More than 220 million people speak in Bangla as their native language. It is ranked sixth based on the number of speakers [1]. Bangla is the mother language of Bangladesh and also a large number of people in eastern area of India (West Bengal and Kolkata as its capital) speak and write in Bangla. The written Bangla was found approximately 650 A. D. But during 1500 B. C. to 1000 B. C. there was a language called "Shangskrit" (ms¯‹Z) which was used only by the upper level people of the society and was used only for literature. And the local people used "Aarjo" (Avh) of "Adim Prakrito" (AvW`g cvKZ). However, there were more than one type of "Adim Prakrito" language and one of them was "Pali" (cvWj). "Pali" was changed to "Praccho Prakrito" (cvP¨ cvKZ) and then to "Gouro Prakrito" (‡MŠo cvKZ). Then, the next stage is "Gouro Opovrongsho" (‡MŠo Acåsk). And from this "Gouro Opovrongsho" the language Bangla was created [2]. Bangla is a member of the Indo-Aryan languages. It is derived from Sanskrit. It is written left-to-right, top-to-bottom of page (same as English). Its vocabulary is akin to Sanskrit. Though the vocabularies are quite difficult at first, but to some extent there are similarities with Latin as exemplified by the following words in Table 1.

* Present address of the author.

Table 1: Comparison of the similarities between different languages.

Language	Word					
English	month	mother	new	night	nose	three
Latin	mensis	mater	novus	nox	nasus	tres
Sanskrit	mãs	matar	nava	nakt	nãs	trayas
Bangla	Mash	mata	nobo	ratri	naakh	tin

Natural Language Processing (NLP) is quite a difficult task. There were many researches in the field of language translation but there is no fully successful language translation machine so far. Since it is a Human Language Technology (HLT), there are lots of varieties and lots of opportunities for research. It is not possible to work on the whole language translation process together. Therefore, it is segmented into many parts. But, the fact is, most of them choose a part of the source language to be translated to the target language. For example, in our paper the source language is English and the target language is Bangle. There are varieties types of sentences both in English and Bangla language. In this paper, we have kept all types of sentences though the types of sentences are not the main issues of our thesis. We have mainly focused on the process of language translation and the effectiveness of the language translation. The efficiency is measured by different methods. In this paper, we can adapt a new system so that it will be 100% efficient at least theoretically.

The paper is organized as follows: Section II discusses the background of the research. Section III explains the problem statements. Solution Frameworks are analyzed in Section IV, while the methods for improving to the adaptive system are provided in Section V. Finally, Section VI draws some conclusion and provides some remarks on the future works.

2. Background

A very few work has been done on English to Bangla language translation both in Bangladesh and West Bengal of India. A hybrid scheme is shown in [3]. Only the present indefinite and present continuous forms of English sentences are concerned in [4]. They represent a simple algorithm for language translation. Using Artificial Intelligence (AI) a Natural Language Processing (NLP) algorithm is proposed in [5]. In [6], Cockey-Younger-Kasami (CYK) algorithm is used for language translation. The main change they brought was that they used normal parse tree than the Chomsky Normal Form (CNF) parse tree because they proved that some problems arose during transfer from English parse tree to Bangla parse tree. Fig. 1 shows pars tree for English and Bangla sentence. Morphological analysis is done in [7] where morphemes means minimal unit of meaning of grammatical analysis. A phrasal Example Based Machine Translation (EBMT) is represented in [8]. Another paper [9] concentrated specially on prepositions of English. In other languages, excluding Bangla it is interesting to study Machine Translation (MT) because it will show some other efficient techniques for language translation. This type of research is Spanish to English language translation [10]. Using five (5) components and three (3) algorithms they have done the language translation. Components are i) PENMAN Upper Model, ii) ONTOS Model, iii) Longman's Dictionary of Contemporary English (LDOCE), iv) WordNet, v) Harper Collins Spanish English Bilingual Dictionary.

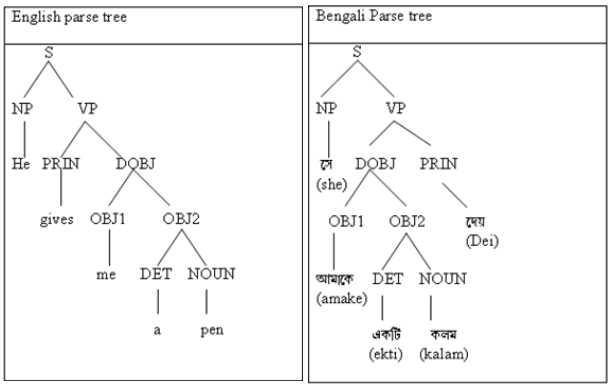

Figure 1: Parse tree for English and Bangla sentence.

3. Problem Statement

Before knowing that how the language translation works we need to understand both the structure of English and Bangla language. The methods of language translation are depended on the structure of sentence and grammar of the each language.

3.1 *Grammatical Analysis for English and Bangla Language*

Analysis of English grammar is a huge task. But we have analyzed only those things which are essential for our language translation. These are verbs, phrase, prepositions and Bivokti (Inflection).

Verbs are the most important term in grammar. Using the verb we can identify the tense of a sentence. In our language translation we omit auxiliary verb and joining with the subsequent verb we make a verb phrase. This is done because there is no direct use of auxiliary verb in Bangla sentence. For example the English sentence, "I am going to school" has two verbs, "am" is the auxiliary verb and "going" is the subsequent verb in this sentence. It will be considered only one verb like "am going". Here "am" is translated in Bangla as "nq/nB" (hoy/hoi) which is not used in Bangla sentence. So the dictionary is also developed in this way. This "am going" is considered as verb phrase. In the same way "to" and "school" also merged as "to school" and considered as noun phrase. Here "school" and "to school" are translated differently in Bangla sentence. In Bangla sentence "school" is translated as "¯‹j" (school) and "to school" is translated as "¯‹‡j" (schoole). This "‡" (e) is called Bivokti in Bangla grammar. Bivokti means a single letter or more than one letter which is added after the noun to make relation with the other words of the sentence. Since there is no existence of Bivokti in English grammar different phrases and also different prepositions are translated as Bivokti in Bangla sentence. For example "I am sleeping in my room" will be translated as "Awg Avgvi N‡i NgvwQ" (ami amar ghore ghumachchhi). Here "in" is translated as "‡" (e) and added after the noun "room" as "N‡i" (ghor + e = ghore). The issue is preposition. There is no use of preposition in Bangla grammar. Prepositions are translated as Bivokti in Bangla. Sometimes prepositions are considered as prepositional phrase and then translated. The dictionary has to be rich enough to handle prepositions.

3.2 *Comparative structure between English and Bangla Language*

For the sake of language translation from English to Bangla we need to discuss a comparative structure in between these two languages. It will help us to understand about the major problems for language translation. Here, we have described the sentence patterns of these two languages at first. And then we have described the grammar patterns of these two languages. Later, we have compared these patterns for both English and Bangla language.

Generally sentences are divided into three parts: i) Simple Sentence, ii) Complex Sentence and iii) Compound Sentence. Compound Sentence can also divide into Double and Multiple Sentence. Here we give an example of Simple Sentence. For English pattern: p1: Subject + Verb + Object (should have one subject and one finite verb). Example – He plays football (He + plays + football). For Bangla pattern: ~ p1: Subject + Object + Verb. Example – ‡m dUej †L‡j (she football khale) {‡m + dUej + †L‡j (she+football+ khale)}.

A grammatical analysis is needed to produce rules for language translation. Both English and Bangla have their own grammars and we need the proper mapping from English grammar to Bangla grammar. For instance, the English grammar and Bangla grammar for a particular sentence is illustrated here. English grammar for the sentence, "I washed my hands and face" is stated below.

$$n + v + (d + n' + c + n'')$$

n $\rightarrow$ Head Noun, v $\rightarrow$ Verb, d $\rightarrow$ Determiner, $n' \rightarrow$ Noun 1, c $\rightarrow$ Conjunction, $n'' \rightarrow$ Noun 2

Noun $\rightarrow$ I | Karim | Rahim, Verb $\rightarrow$ washed | wash, Determiner $\rightarrow$ my | all | some, Noun 1 $\rightarrow$ hands, Conjunction $\rightarrow$ and | or, Noun 2 $\rightarrow$ face

Bangla grammar for the corresponding Bangla sentence, "Awg Avgvi nvZ Ges gL aBjvg" (Ami amar hat ebong mukh dhuilam) is given below.

$$n + (d + n' + c + n'') + v$$

n $\rightarrow$ Head Noun, v $\rightarrow$ Verb, d $\rightarrow$ Determiner, $n' \rightarrow$ Noun 1, c $\rightarrow$ Conjunction, $n'' \rightarrow$ Noun 2

Noun $\rightarrow$ Avwg (ami) | Kwig (korim) | iwng (rohim), Verb $\rightarrow$ aBjvg (dhuilam) | ‡avqv (dhoa), Determiner $\rightarrow$ Avgvi (amar) | mevB (shobai) | wKQ (kichhu), Noun 1 $\rightarrow$ nvZ (hat), Conjunction $\rightarrow$ Ges (ebong) | A_ev (othoba), Noun 2 $\rightarrow$ gL (mukh)

A morphological analysis is done while mapping from English to Bangla grammar. It is essential to compare the two grammar structure for proper language translation. The dissimilarities between the two languages are the main problems of language translation. We have already discussed some of them. The auxiliary verb and prepositions of English grammar are not found in Bangla grammar. On the other hand, Bivokti does not exist in English grammar. This problem is solved by making auxiliary verb as verb phrase. Here, auxiliary verbs are considered as verb phrase that is added with the subsequent verb of the sentence. Prepositions are translated as prepositional phrase or Bivokti that is adding one or more than one letter at post position after the noun in Bangla sentence.

4. Solution Framework

In our research we have concentrated on rules those have been assigned by us for proper language translation from English to Bangla. We have taken a set of English sentences and translate them into Bangla sentences as training set. Afterwards, we have taken a testing set to measure the efficiency of our system.

4.1 *Flow Diagram of the Translation Architecture*

The flow diagram of the language translation architecture is shown in Fig. 2. There are different blocks which represents different steps of the language translation process. First of all, an English sentence is constructed with words and is ended commonly with a full stop (“.”). The first task of the translator is to split each word. But splitting doesn’t mean that all are distinct and have no connection with each other. Attributes are the information of that word. Using these attributes the translator will analyze that word lexically. Lexical analysis is done for understanding the grammatical and sentence structure of that sentence. Using the attributes from the dictionary the sentence is then classified. The result of the Lexical Analyzer is to generate the class of that sentence. Since the previous step finds the class of the English sentence in this step a searching method is done. The target of this method is to find out the matched fuzzy rule (*r*) for the English sentence. It might be happened that a rule is matched fully or partially. Both are in concern because those are partially matched will produce a partial translated Bangla sentence. It is better than totally void result. Moreover, we can enrich the Knowledge Based rules of the machine while facing the partial searching result.

In the next step the dictionary will be used for finding the corresponding Bangla words. Here morphological analysis is done. It is necessary to analyze that which word meaning is needed in which place. The meaning should be appropriate and correct. Accuracy is the big measurement of the efficiency. For taking the right decision and choosing the right word we have used the various attributes of that English word from the dictionary. The final task is to reconstruct the Bangla sentence. It is already half done in the second step. We have found the matched production rule (*r*) for the English sentence. Now, from the production rule table the corresponding rule (*r'*) for the Bangla sentence is found. According this rule (*r'*) the final result of the translator has been provided.

4.2 *Algorithm*

Input: $S_e \rightarrow$ English sentence, $D \rightarrow$ English to Bangla dictionary, $n \rightarrow$ Total number of rules

Output: $S_b \rightarrow$ Bangla sentence

Steps: $k := 0$;

$EnglishWordsArray[k] := SplitSentenceIntoWords(S_e)$;

for i:= 1 to n do

for j:= 1 to k do

$R_n := MatchedRule(EnglishWordsArray[j])$;

endfor

endfor

$l := 0$; //length of the dictionary. $l := LengthOfDictionary(D)$;

for i:= 1 to k **do**

for j:= 1 to l **do**

if($EnglishWordsArray[i] == DictionaryEnglishWords[j]$) **then**

$BanglaWordsArray[i] := DictionaryBanglaWords[j]$;

endif

endfor

endfor

$S_b := 0$; //Bangla sentence.

$S_b := ReconstructBanglaSentence(BanglaWordsArray[k], Rn)$;

Return: S_b

$ErrorEstimationOfBanglaSentence(S_b)$:

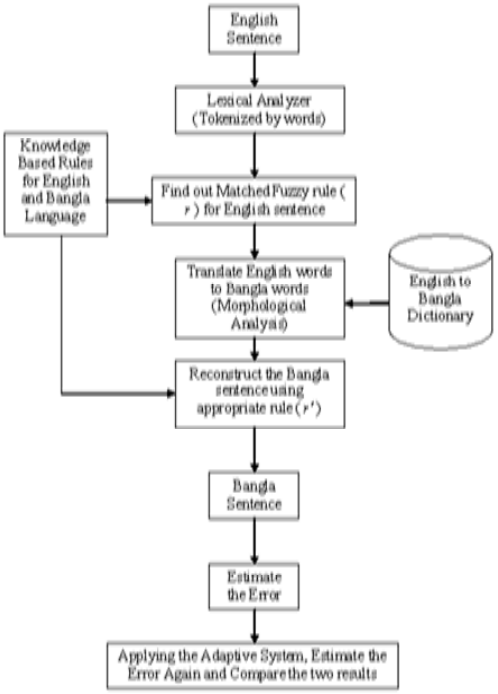

Figure 2: Flow Diagram of English to Bangla Translator.

4.3 *Production Rules*

While translating the training set we have produced a set of rules. These rules are stated in Table 2. These rules are extendable. The more the training set sentences the more the rules can be added as required. In the table, there are production rules represented. We have showed the production rules for both English and the Bangla sentences side by side. There are individual sentence patterns for English and Bangla sentences. We have been represented here all the English sentence patterns as *r* and all the Bangla sentence patterns as *r'* in common. These rules are in pair wise. Because a sentence pattern in English must have a corresponding sentence pattern in Bangla which is used for language translation. These rules are predefined and must be precisely given in the language translation system. For the language translation purpose, an English sentence pattern will change to a Bangla sentence pattern according to a particular rule. This rule is given in the production rule table. In this table there are very few rules represented to give the idea that how the production rule works. All the rules are numbered and each pair is in the same number. For example, if an English sentence pattern represents r_3, then the Bangla sentence pattern for particularly r_3 will be r'_3. So, r_n sentence pattern for English sentence will have the r'_n sentence pattern in Bangla in the production rule table.

4.4 *Database*

Since we needed a dictionary, it was mandatory to use a digital dictionary. A flat file has been used as dictionary. Here different attributes have been given for finding out the classification and each result has been in a definite class. These are the information about the words. This information has been used to classify the sentence. There are more than 200 words generated in the dictionary. It is expandable to as much as we need. The more the sentences are applied to the model the more the dictionary will be enriched. Both for training and testing sets, the dictionary is enhanced with the words along with the attributes.

Table 2: Generated rules for language translation.

PRODUCTION RULES			
English Pattern (r)		**Bangla Pattern (r')**	
r_1	$s \rightarrow n + v + n'$	r'_1	$s \rightarrow n + n' + v$
	Salma + was peeling + potatoes		mvjgv + Avjyi + †Lvmv Qvovw”Qj
r_2	$s \rightarrow n + v + n' + n''$	r'_2	$s \rightarrow n + n'' + n' + v$
	Knowledge + lights + the way + to heaven		Ávb + †e‡nk&‡Zi + c_wU + Av‡jvwKZ K‡i
r_3	*p + art + adj + n*	r'_3	*p + art + adj + n*
	It + is a + costly + pen		GUv + GKwU + `vgx + Kjg
r_4	*p + v*	r'_4	*p + v*
	We + were playing		Avgiv + †Ljv KiwQjvg
r_5	*n + v*	r'_5	*n + v*
	The moon + shines		Puv` + wKiY †`q
r_6	*p + d + v*	r'_6	*p + d + v*
	We + all + breathe		Avgiv + mevB + wbtk¦vm †bB
r_7	d + art + n	r'_7	*d + art + n*
	This + is a + picture		GUv + GKwU + Qwe
r_{8a}	$n + v + (p + n')$	r'_{8a}	$n + (p + n') + v$
	Karim + cut + (his + finger)		Kwig + (Zvi + Av½yj) + †K‡UwQj
r_{8b}	$n + v + (p + art + adj + n')$	r'_{8b}	$n + (p + art + adj + n') + v$
	Grandfather + told + (us + a + funny + story)		`v`vRvb + (Avgv‡`i + GKwU + gRvi + Mí) + e‡jwQj
r_{8c}	$n + v + (n' + adv)$	r'_{8c}	$n + (n' + adv) + v$
	Habib + goes + (to college + regularly)		nvwee + (K‡j‡R + wbqwgZ) + hvq
r_{8d}	$n + v + (d + n' + c + n'')$	r'_{8d}	$n + (d + n' + c + n'') + v$
	I + washed + (my + hands + and + face)		Avwg + (Avgvi + nvZ + Ges + gyL) + ayBjvg

r_{9a}	$p + v + n$	r'_{9a}	$P + n + v$
	I + eat + rice		Avwg + fvZ + LvB
r_{9b}	$p + v + (n + p')$	r'_{9b}	$p + (n + p') + v$
	He + told + the news + to everyone		‡m + (msev`wU + mevB‡K) + e‡jwQj
r_{9c}	$p + v + (n + d + n')$	r'_{9c}	$p + (n + d + n') + v$
	We + visited + (Moynamoti + last + year)		Avgiv + (gqbvgwZ + MZ + eQi) + †`‡LwQjvg
r_{9d}	$p + v + adv$	r'_{9d}	$p + adv + v$
	He + was sleeping + then		‡m + ZLb + Nygvw”Qj
r_{10}	$p + n + v + pre + d + p' + n'$	r'_{10}	$p + n + pre + d + p' + n' + v$
	Each + Muslim + is brother + of + every + other + Muslim		GK + gymjgvb + cÖ‡Z¨K + Aci + gymjgvb + Gi + fvB
r_{11}	$v + p + d + adj + n$	r'_{11}	$p + d + adj + n + v$
	Give + him + some + crisp + potato-chips		Zv‡K + wKQz + gP&g‡P + Avjy fvRv + `vI

4.5 *Programmatic Solution*

To implement the theory that we have stated earlier, we needed to create a program. To create a program we had to write a set of instructions. These instructions are called programmatic codes. These programmatic codes are usually written in some special languages such as C, Java and VB etc. We have used Java as our programming language. Initially it is required to tokenize the words from the English sentence. This means each word is a token in a sentence. We need to detach all the words. This process is called tokenizing. This tokenizing process is done by *split( )* method. This splits all the words separately. One part of coding is used for defining phrases of the English sentences. In grammar there are different types of phrases. For this we need to define all the phrases and find out which phrase is appropriate for the particular sentence. Then we need to find out the corresponding Bangla word from the dictionary. In this part, each word of the English sentence is searched in the dictionary for matching. When the matched word is found all the information of that word is collected for analysis. This information is termed as attributes. These attributes are used to find out the fuzzy rules of the particular sentence. There is also a calculation of probability of finding out the words in dictionary. A percentage of searching result is showed in this program. It shows the rate of successful searching result.

4.6 *Experiment*

In the experiment we use two data sets. One is training data set and the other is testing data set. The model is built with the sentences of training data set. Afterwards, the sentences of testing data set are used for calculating the efficiency of the model.

4.7 Error Estimation

To estimate the error different types of method can be used. We have used three methods for error estimation. These are: i) Hold Out, ii) Cross Validation (Leave one out) and iii) Bootstrap. In hold out method 2/3 of the data is used for training set and 1/3 data is used for testing set. We selected data for testing set randomly. In this case we have taken 79 sentences as training data set and 27 sentences as testing data set. Among the 27 sentences of testing data set 25 sentences are translated correctly without any error. Two sentences are found incorrectly translated. These two sentences are 75% and 33% correct respectively. In cross validation method the data set is divided into subsets of approximately equal sizes. Each time one subset is leave out for testing and all other subsets are used as training data set. In our experiment we have taken 8 fold data that is there are 8 subsets. So the model is executed 8 times. Each time the error rate is calculated in testing data set. The final error rate is found by calculating the average of all the error rates. The sampling size is 10. Total 79 sentences are taken. Since it is not divisible by 10, only the last fold contains 9 sentences. The bootstrap is based on the statistical procedure of sampling with replacement. The previous methods were sampled without replacement to built training or testing set. In the bootstrap method the testing set is selected randomly from subsets by replacement. The formula of this method is

$$e = \frac{1}{n}\sum_{i=1}^{n} 0.632 * E_1 + 0.368 * E_2 \quad \text{..... (1)}$$

Where, $e =$ Error rate, $n =$ Sampling size, $E_1 =$ Error on the testing data set, $E_2 =$ Error on the training data set and $i =$ Number of iteration.

Training set will contain about 63.2% with size n whereas the testing set will contain only 36.8%. Here Sampling size is 10 that is n = 10 and the samples are taken 10 times with replacement, that is i = 10. The error on the training data set is zero that is $E_2 = 0$.

4.8 *Comparison of the Methods*

We needed to compare into different learning methods on the same problem data set to see which one is better to use. To find out the best method we have to estimate the error using several methods and the method which results the smallest value of error estimation is the best method. Comparing among all the above three methods the bootstrap method is the best. It gives the smallest error estimation shown in Fig 3.

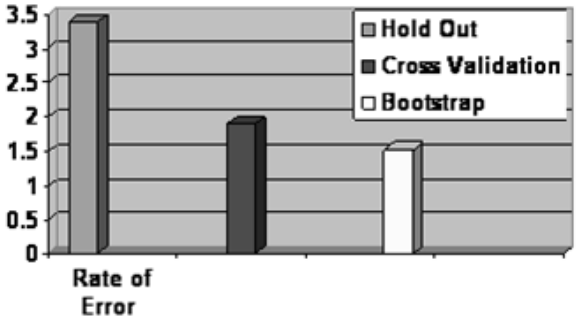

Figure 3: Comparing Chart among Three Methods.

5. Improving to the Adaptive System

It is necessary to improve the system by adapting some techniques so that it can be performed with higher efficiency. Here, we would like to improve the system to make an errorless translator. It is almost impossible to generate a cent percent correct translator. But we should give a try to improve the system.

5.1 *Adaptive System*

An adaptive system adapts to the new conditions of the previous system. Adaptive System refers to the change of the conditions that are created in preceding system. The adaptive system either adjusts its set of rules by example or that develops gradually into a new system which adjusts its set of rules by using any ways [11]. By [12] there are four components of adaptive systems.

5.2 *Different ways of Adaptation*

Adaptation can be taken in different ways to update the model like, i) Increasing the data set, ii) Increasing the number of rules, iii) Merging the previous unnecessary rules, iv) Increasing the number of attributes in dictionary and v) Applying the preferences.

5.3 *Error Estimation in Adaptive System*

It is very interesting to observe the difference between non adaptive and adaptive system. There is significant difference between the results of the two systems. The error estimation for both the non adaptive and the adaptive system are done by the same process, using the same methods as described in Section 4.7. The previous testing data set is merged with the previous training data set. New testing data set is taken for the comparison of the two systems. Comparison between non adaptive and the adaptive system are shown in Table 3.

Table 3: Comparing the Non Adaptive and the Adaptive System.

Systems / Methods	Previous Testing Data Set	New Testing Data Set	
	Non Adaptive (Errors in %)	**Non Adaptive** (Errors in %)	**Adaptive (Errors in %)**
Hold Out	3.3951	2.1826	1.5874
Cross Validation	1.895875	1.1787	0.9802
Bootstrap	1.5379088	0.8544	0.7666

6. Conclusion

In this paper, we have shown a totally new approach for language translation. In Bangladesh, there is very little work on English to Bangla language translation done. Among them this research is totally a different one. The language translation architecture that is represented here is not developed before. According to this architecture the algorithm has constructed. The task that we have done in this paper can be extended more. A lot research is possible in this field. We have tried to keep variation among the English sentences that we have translated into Bangla sentences. But we have not completed all the variety of sentences. Since it is Natural Language Processing (NLP) the number of variation is almost unlimited. It is because the language is changeable according the time. Many words are expired and not used nowadays. On the other hand, many new words are added in the language. This is a Human Language Technology (HLT) that is people are making new words of languages. So there is unlimited opportunity to upgrade the current research.

REFERENCES

[1] http://en.wikipedia.org/wiki/List_of_languages_by_total_speakers, Last accessed July 12, 2010.

[2] Ajit Kumar Guho and Dr. Anisuzzaman, “Notun Bangla Rochona”, Agaust 1997.

[3] Saha Goutam Kumar, “The EB-ANUBAD translator: A hybrid scheme,” in Journal of Zhehiang University SCIENCE, pp. 1047-1050, 2005.

[4] S. Ahmed, M. O. Rahman, S. R. Pir, M. A. Mottalib, and Md. S. Islam, “A New Approach towards the Development of English to Bangla Machine Translation System,” in *International Conference on Computer Information and Technology (ICCIT)* , pp. 360-364, Jahangirnagar University, Dhaka, Bangladesh, 2003.

[5] S. A. Rahman, K. S. Mahmud, B. Roy, and K. M. A. Hasan, “English to Bengali Translation Using A New Natural Language Processing Algorithm,” in *International Conference on Computer Information and Technology (ICCIT)* , pp. 294-298, Jahangirnagar University, Dhaka, Bangladesh, 2003.

[6] S. Dasgupta, Abu Wasif, and S. Azam, “An Optimal Way of Machine Translation from English to Bengali”.

[7] A. N. K. Zaman, Md. A. Razzaque, and A. K. M. K. Ahsan Talukder, “Morphological Analysis for English to Bangla Machine Aided Translation” in National Conference on Computer Processing of Bangla, Dhaka, Bangladesh, 2004.

[8] S. K. Naskar, and S. Bandyopadhyay, “A Phrasal EBMT for Translation English to Bengali,” in MT Summit X, Kolkata, India, 2005.

[9] S. K. Naskar, and S. Bandyopadhyay, “Handling of Prepositions in English to Bengali Machine Translation,” Kolkata, India.

[10] Kevin Knight, and Steve K. Luk, “Building a Large-Scale Knowledge Base for Machine Translation”.

[11] Dr. Kevin E. Voges and Dr. Nigel K. Ll. Pope, “An Overview of Data Mining Techniques from an Adaptive Systems Perspective” in ANZMAC 2000 Visionary Marketing for the 21 Century: Facing the Challenge, pp. 1323-1329, School of Marketing, Griffith University.

[12] Holland and John H., “Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence,” Cambridge, MA: MIT Press, 1992.