An empirical framework for parsing Bangla imperative, optative and exclamatory sentences

Conference Paper · February 2017

DOI: 10.1109/ECACE.2017.7912899

2 authors:

Tanzina Alamgir

2 PUBLICATIONS **3** CITATIONS

SEE PROFILE

Mohammed Safayet Arefin

Mississippi State University

13 PUBLICATIONS **24** CITATIONS

SEE PROFILE

An Empirical Framework for Parsing Bangla Imperative, Optative and Exclamatory Sentences

Tanzina Alamgir, Mohammed Safayet Arefin
Dept. of Computer Science & Engineering
Chittagong University of Engineering & Technology,
Chittagong, Bangladesh.
Email:tutul.cuet10@gmail.com; safayet_arefin@yahoo.com

Abstract— **Parsing is one of the most challenging task in the field of natural language processing and it plays an important role in order to analyze any natural language. To determine a legal structure for a sentence, we need to expose the rules of how sentences of a language are embodied and have a parsing algorithm to analyze sentences using those rules. This paper proposes a new technique to parse the Bangla sentences including imperative, optative and exclamatory sentences using a set of context sensitive grammars (CSG's) rules. This paper considers Bangla sentences based on the intonation or mood of the sentences rather than the structure of the sentences for parsing. The proposed framework can parse Bangla sentences with over 81% accuracy which is quite satisfactory.**

Keywords—Natural language processing, parsing, context-sensitive grammars. Lexicon, and parse tree.

I. INTRODUCTION

Natural Language Processing (NLP) which is one of the most precursory area of AI research allows people to communicate with computers in a human language such as, English, Bengali, Hindi etc. as easily as it is to communicate with other people. In other words, NLP should make it possible for people to use computers in much the same way that they would use a human assistant to get their work done. But developing programs that understands a natural language is a difficult problem. Natural language is the most versatile communication medium of mankind. They contain infinity of different sentences. No matter how many sentences a person has heard or seen, new ones can always be produced. We speak to each other in a human language from our childhood and take most of its complex characteristics. Even, sometimes when we speak natural language incorrectly i.e., not strictly in accordance with rules of grammar and syntax, we can still make sense out of it. To explore language, we must gather good ideas about the structures of the sentences. Parsing allows us to find the structure of any sentence. To do these, we need to know the rules of how a sentence is structured and developed an algorithm by those rules to explore the sentence. By analyzing the structure of the sentence, we can find words in a sentence related to each Usually, we get a parse tree or structural representation after the parsing. Syntactic information about the sentence can also be gathered from parsing. In most of the cases, CFG is used for parsing Bangla sentences. But CFG can't parse all types of Bangla sentences. CFG fails to indicate discontinuous constituent, structural similarity and unable to solve the problem of ambiguity [7]. In order to generate different kinds of Bangla sentences, we have to use CSG's due to its capabilities to handle agreement between subject-verb and person class [3,4,5]

According to structure, there are three kinds of Bangla sentences including simple, compound and complex sentences. Depending on the mood or intonation, Bangla sentences can also be divided into five categories namely assertive, interrogative, imperative, optative and exclamatory sentences. The main contribution of this paper is to implement a function of the user based parsing framework that can parse the three types of Bangla sentences such as, imperative, optative and exclamatory sentences by using a set of CSG rules. The system represents the output of the parser in a list. The parser is also tested with different types of examples and the performance of the testing is satisfactory.

II. PREVIOUS WORK

Parsing of Bangla sentences is one of the challenging task in Bangla language processing. Many researchers worked on parsing technique of Bangla sentences of different ways. A way to translate Bangla imperative, optative and exclamatory sentences using a set of context-sensitive grammars (CSG's) is introduced in [1]. A technique to parse Bangla sentences in a new approach using context-free grammar rules that accepts all types of Bangla sentences including complex, compound, exclamatory and optative sentences is implemented in [2]. A technique to translate Bangla sentences including assertive, interrogative and imperative sentences into corresponding English sentences using context sensitive grammar rules is introduced in [3]. To parse the Bangla sentences including assertive, interrogative and imperative by a set of context-sensitive grammars (CSG's) is shown in [4]. A set of Context-sensitive phrase structure rules for structural representation of Bangla natural language sentences is implemented in [5]. Again, in [7], a method to analyze syntactically Bangla sentence using context-sensitive grammar rules which accept almost all types of Bangla sentences including simple, complex and compound sentences is implemented. A machine translation (MT) system which translates an English sentence into a Bangla sentence of equivalent meaning is in [8]. They

978-1-5090-5627-9/17/$31.00 ©2017 IEEE

have focused on context-free grammars (CFGs) for English simple assertive sentences to discover the syntactic structures of both English and Bangla correct and complete sentences. Some research work has been done in English to Bangla MT in [9]. They developed a way that represents a new solution for building a MT system for English to Bangla translation by modifying the rule based transfer approach of MT system. A Bangla to English machine translation system using parts of speech tagging is developed in [10]. A set of transformational generative grammar is introduced which in conjunction with phrase structure grammar in [11]. It is used to generate or recognize all types of Bengali sentences. After analyzing the related work, we have found that most of the previous parsing systems considered the Bangla sentences into English structurally rather than depending on mood or intonation.

III. PROPOSED PARSER MODULE

The structural representation of proposed Bangla natural language parser module is represented in fig. 1. Details explanation of this module is given in the following subsections.

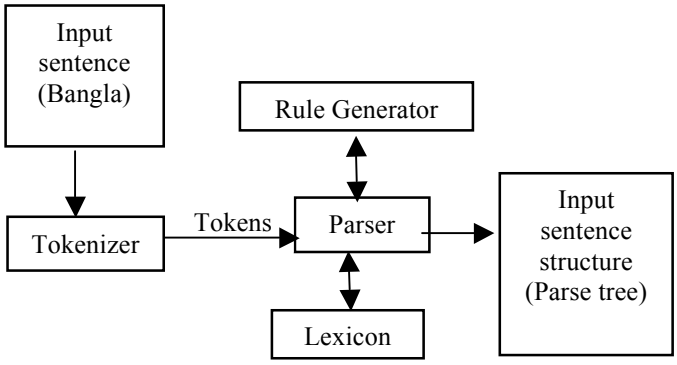

Fig. 1. Proposed Bangla natural language parser

A. Input sentence

The input of the parsing framework are Bangla sentences. In this system, only Bangla imperative, optative & exclamatory sentences are considered.

B. Tokenizer

The system initiates with the Tokenizer and in order to doing the task of parsing the input sentence enters into the Tokenizer. Then the sentence breaks into individual words which are called Tokens. For the further access the tokens are stored.

Input sentence: "কি সুন্দর পাখিটি !"

Output of tokenizer: ("কি", "সুন্দর", "পাখি", "টি", "!")

Then the tokens go to the parser to get the source language structure. With the help of lexicon tokens are taken for checking validity.

C. Lexicon

The typical entries in the lexicon which we have used in our system are shown in the Table 1.

TABLE I. TYPICAL ENTRIES IN LEXICON

Bangla	Features
বই	[N]
রহিম	[N, Per 3]
পড়	[V]
সাবাস	[E.W]
মঙ্গল	[V]
টি	[Det]
আল্লাহ	[N.C]
আস	[V]
!	[E. M]
আমার	[PN]
সুন্দর	[Adj]

[Abbreviations: PN: Pronoun, N: Noun, E.W: Exclamatory word, E.M: Exclamatory Marker, N.C: Noun.creator, ind: Indecllinable, V: Verb, Adj: Adjective]

D. Rule Generator

Bangla CSG for different kinds of Bangla sentences for parsing is described in the following subsections. List of used Bangla CSG to parse the sentences is given in the table 2. During parsing, in every cases when a token in the right side of the rule is not needed to parse the sentence then "Null" is used in that place.

TABLE II. CSG RULES OF BANGLA SENTENCES

Rule No	Bangla CSG's rule
1	S→OS|IS|ES
2	OS->NP VP
3	IS->NP VP
4	ES->AP NP|NP VP
5	NP→(N.creator)N|PN(BIV)
6	NP→N|PN(BIV)
7	NP→N(Det)E.M
8	NP→Null
9	NP→E.W E.M N|PN
10	VP (Adj)V OW
11	VP(N)(Det)V(Con)IW
12	VP→E.W E.M V(Con)V
13	VP->(Adv)(N)(BIV)V(NEG)
14	AP->WH Adj(Qtfr)
15	WH-> কি
16	V→ মঙ্গল |সহায়| সাহায্য
17	PN-> তোমার । আমার । তারা । আমাকে।আমরা

978-1-5090-5627-9/17/$31.00 ©2017 IEEE

18	Adj-> দীর্ঘজীবী । সুন্দর ।মজার ।বাজে ।সুখী
19	Det-> টি
20	V->হ । কর । পড় । যে । দেখ । খেল । গাই । নাচ । জিত
21	BIV-> এর ।কে ।য়
22	E.W-> কি । সাবাস। কি মজা । হায়
23	N.creator ->আল্লাহ
24	Adv-> আর
25	NEG->নেই
26	E.W->!
27	Con-> তে
28	Qtfr->একটি
29	N-> কাজ| বই।রহিম|মা ।ছবি ।বাড়ি
30	OW-> করুন| হউক
31	IW-> দাও

[Abbreviations: S: Sentence, OS: Optative sentence, ES: Exclamatory sentence, IS: Imperative sentence, NP: Noun phrase, N: Noun, PN: Pronoun, VP: Verb phrase, VF: Verb form, V: Verb, Qtfr: Quantifier, Biv: Bivokti (inflection), Adj: Adjective, Con: Concord, NEG: Negation, EW: Exclamatory word, EM: Exclamatory marker, OW: Optative word, N.creator: Noun creator, Det: Determiner, IW: Imperative word]

1) CSG for Imperative Sentences:

In Bengali imperative sentences, an imperative word "দাও" is used. As an example, we can consider the Bangla sentence "রহিমকে কাজটি করতে দাও".

In order to get the parse tree, we have used CSGs rules from Table 2.

S→IS [Rule no: 1]
→NP VP [[Rule no: 3]
→ N(BIV) VP [Rule no: 6]
→ রহিম (BIV) VP
→ রহিম কে VP
→ রহিম কে (N) (Det) V (Con) IW [Rule no: 11]
→ রহিম কে কাজ (Det) V (Con) IW
→রহিম কে কাজ টি V (Con) IW
→রহিম কে কাজ টি কর (Con) IW
→ রহিম কে কাজ টি কর তে IWs
→ রহিম কে কাজ টি কর তে দাও

Structural representation (SR) of this imperative sentence is:

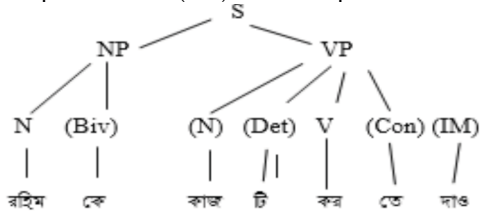

Fig. 2. Structural representation of "রহিমকে কাজটি করতে দাও"

2) CSG for Optative Sentences:

In Bangla optative sentence, we always find an optative word (OW) at the end of the sentence. As an example, we can consider the Bangla sentence "আল্লাহ তোমার মঙ্গল করুন" Here we find optative word "করুন", in the Bangla sentence.

In order to get the parse tree, CSGs rules from Table 2 is used.

S →OS [Rule no: 1]
→NP VP [Rule no: 2]
→(N.creator)PN(BIV)VP [Rule no: 5]
→ আল্লাহ PN(BIV)VP
→ আল্লাহ তোমার (BIV)VP
→ আল্লাহ তোমার null VP
→ আল্লাহ তোমার null (Adj)V OW[Rule no: 10]
→ আল্লাহ তোমার null null V OW
→ আল্লাহ তোমার null null মঙ্গল OW
→ আল্লাহ তোমার null null মঙ্গল করুন

SR of this optative sentence is:

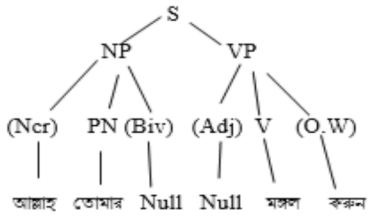

Fig. 3. Structural representation of "আল্লাহ তোমার মঙ্গল করুন"

3) CSG for Exclamatory Sentences:

In exclamatory sentences, we always find exclamatory marker (EM) at the end of the Bangla sentences. Let's consider a exclamatory Bangla sentence "কি সুন্দর বাড়িটি!. Here "!" is the exclamatory marker (EM).

In order to get the parse tree, we have used CSGs rules from Table 2.

S→ES [Rule no: 1]
→AP NP [Rule no: 4]
→ WH Adj (Qtfr) NP [Rule no: 14]
→ WH Adj (Qtfr) NP
→ কি Adj (Qtfr) NP
→ কি সুন্দর (Qtfr) NP
→ কি সুন্দর Null NP
→ কি সুন্দর Null N (Det) E.M [Rule no: 7]
→ কি সুন্দর Null বাড়ি (Det) E.M
→ কি সুন্দর Null বাড়ি টি E.M
→ কি সুন্দর Null বাড়ি টি !

SR of this exclamatory sentence is:

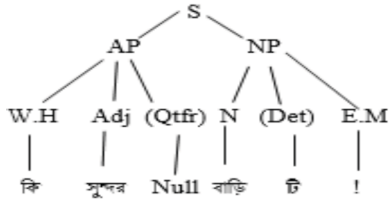

Fig. 4. Structural representation of "কি সুন্দর বাড়িটি!"

978-1-5090-5627-9/17/$31.00 ©2017 IEEE

E. Parser

Basic task of a parser to take tokens as the form of input and with the help of CSG rules it produces parse tree as output. In our system, list form is chosen for the representation of parse tree. The standard representation of structural representation (SR) can be considered as list that is one of the data structures that can be implemented and manipulated very easily with a computer [3].

1) Parsing Algorithm:

The input bangla sentence follows these steps for parsing. The steps are given below:

Step 1: Tokens those are produced from the tokenizer enter to the parser as input. For example, if we consider the input sentence as "আমাকে কাজটি করতে দাও" then the tokens will be ("আমা" ,"কে" , "কাজ" , "টি" , "কর" , "তে" ,"দাও"). The system will store the tokens in a stack for further accessing.

Step 2: In order to testing the validity, the tokes will enter into the lexicon in the next step. For example, the tokens ("আমা","কে", "কাজ", "টি", "কর", "তে", "দাও")will be gone to the lexicon to find its validity.

Step 3: By matching the tokens, grammar rules of Bangla appropriate parts of speech will be assigned in this step. Here, a token must be matched with a rule's right hand side. For example: PN-> আমা, Biv-> কে, N->কাজ,Det->টি V->কর ,Con->তে,IW->দাও will create a partial structure in this step.

Step 4: Starting from the left to right hand side of token list verify every rule whose right hand side will match one or more parts of speech. If a right hand side of a rule matches with appropriate parts of speech, then we have to select that rule [1].

Step 5 Repeat step 4, until no more words remains to generate.

Step 6: After generating all words, the system will generate SR of the sentence in a list format. For example: the parser output of the sentence**:** "আমাকে কাজটি করতে দাও" is given below-

S[NP[PN আমা][BIV কে][VP[N কাজ][DET টি][V কর][CON তে][IW দাও]

2) Parser Output:

The output of the parser is represented in fig. 5

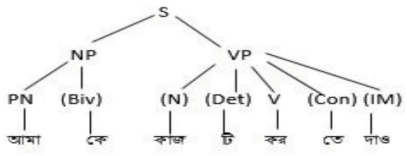

Fig. 5. Parse tree of "আমাকে কাজটি করতে দাও".

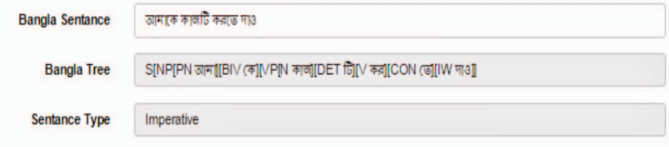

Fig. 6. Parser output of "আমাকে কাজটি করতে দাও"

Input: আমাকে কাজটি করতে দাও
Sentence Type: Imperative sentence
Bangla parse tree: S[NP[PN আমা][BIV কে][VP[N কাজ][DET টি][V কর][CON তে][IW দাও]

3) Comparison of Paring Technique

There are two ways of parsing. First one is top down parsing approach which begins with the start symbol and apply the grammar rules forward until the symbols at the terminals of the tree corresponds to the components of the sentence being parsed. Again, in the case of bottom up approach begins with the sentence to be parsed and apply grammar rules back word until a single tree whose terminals are the words of the sentence and whose top node is the start symbol has been produced. In this paper, the methodology of parsing is based on top down approach. In some cases, bottom up approach becomes so complex. As an example, we can consider the Bangla sentence "কি সুন্দর পাখিটি !". According to our parser, applying grammar rules for back word is tough cause some rules contains null. So, in every cases checking the "null" increases complexity.

According to top down approach:

S→ES [Rule no: 1]
→AP NP [Rule no: 4]
→ WH Adj (Qtfr) NP [Rule no: 14]
→ WH Adj (Qtfr) NP
→ কি Adj (Qtfr) NP
→ কি সুন্দর (Qtfr) NP
→ কি সুন্দর Null NP
→ কি সুন্দর Null N (Det) E.M [Rule no: 7]
→ কি সুন্দর Null পাখি (Det) E.M
→ কি সুন্দর Null পাখি টি E.M
→ কি সুন্দর Null পাখি টি !

According to bottom approach:

→ কি সুন্দর পাখি টি !
→ কি সুন্দর পাখি টি E.M
→ কি সুন্দর পাখি (Det) E.M
→ কি সুন্দর N (Det) E.M [Rule no: 7]
→ কি সুন্দর NP

According to our created rules and developed system, from this step the parser cannot generate the rule number 14 because it cannot identify "null" initially.

IV. Implementation

Snapshots of our parser for exclamatory, optative and imperative sentences are shown in Figs. 7, 8, and 9 respectively

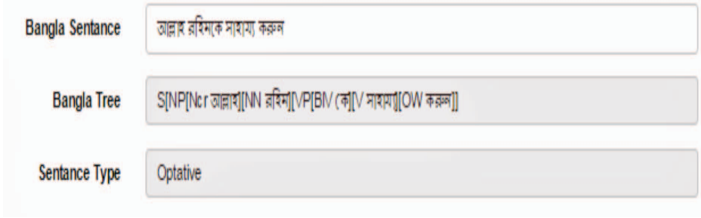

Fig. 7. Parser output of "আল্লাহ রহিমকে সাহায্য করুন"

978-1-5090-5627-9/17/$31.00 ©2017 IEEE

Input: আল্লাহ রহিমকে সাহায্য করুন
Sentence Type: optative sentence
Bangla parse tree: S[NP[Ncr আল্লাহ][NN রহিম][VP[BIV কে][V সাহায্য][OW করুন]]

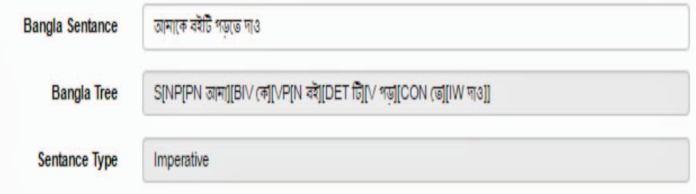

Fig. 8. Parser output of "আমাকে বইটি পড়তে দাও"

Input: আমাকে বইটি পড়তে দাও
Sentence Type: Imperative sentence
Bangla parse tree: S[NP[PN আমা][BIV কে][VP[N বই][DET টি][V পড়][CON তে][IW দাও]]

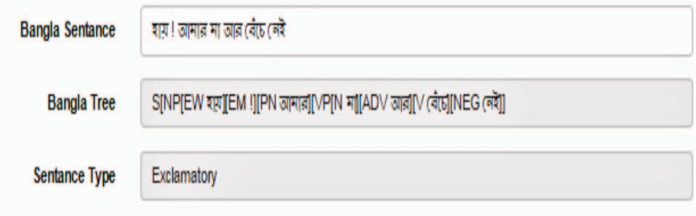

Fig. 9. Parser output of "হায়! আমার মা আর বেঁচে নেই"

Input: হায় ! আমার মা আর বেঁচে নেই
Sentence Type: Exclamatory sentence
Bangla parse tree: S[NP[EW হায়][EM !][PN আমার][VP[N মা][ADV আর][V বেঁচে][NEG নেই]]

V. Experimental Result

In order to analysis the effectiveness of our proposed parsing system, we have tested our system for 400 different sentences and found that total 326 sentences parsed correctly. Most of the sentences are taken from newspaper, story books and grammar books. From our total sentences, we have given 30% of the sentences from different sources and 70% is given from the artificially generated sentences.

TABLE III. SUCCESS RATE FOR DIFFERENT TYPES OF SENTENCES

Sentence types	Sentence length	No. of input sentences	No. of correctly parsed sentences	Overall success rate (%)
Imperative	3	70	59	80.00
	4	54	44	
	5	70	52	
Exclamatory	2	55	48	80.0
	3	30	25	
	4	55	45	
	5	40	32	
Optative	3	11	9	80.7
	4	15	12	
Total	-	400	326	81.5

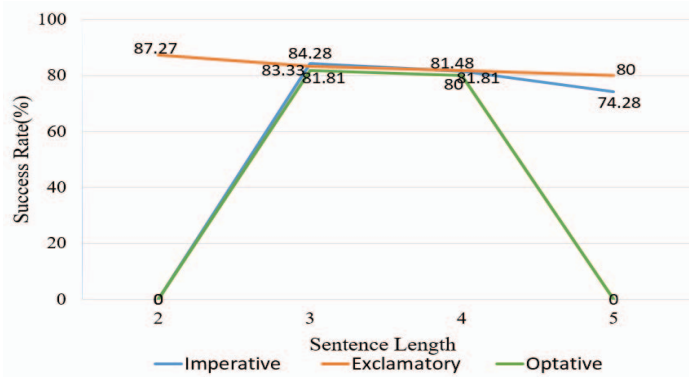

Fig. 10. Success rate vs. sentence length for different types of sentences

Table 3 illustrates the success rate of different types of sentence and Fig. 10 represents the success rate versus sentence length of the system. From the graph, we have found that when sentence length increases then the success rate decreases.

VI. Conclusion

The aim of this paper is to introduce and implement CSG rules to handle Bangla imperative, optative and exclamatory sentences for parsing. The motive of the entire work is to design a parsing technique that is capable for parsing Bangla sentences according to mood or intonation. The proposed parsing algorithm can detect the sentence type and generate the corresponding parse tree efficiently by using CSG's. Experimental result shows that the proposed parsing framework can parse the sentence with about 81% accuracy. For some sentences our system cannot get the accurate parse tree. We have considered very short and limited Bangla sentence prototype to evaluate our proposed parsing system. Our system cannot work for bottom up approach. Further modify the CSG rules, a more powerful parser can be designed for Bangla sentence by combining the bottom-up approach and top down approach together. CSG rules can be develop for semantic features and the thoughts of voice, narration, infinitive, composition of words can be considered for further research.

References

[1] T. Alamgir, M. S. Arefin and M. M. Hoque, "An Empirical Machine Translation Framework for Translating Bangla Imperative, Optative and Exclamatory Sentences into English ", *In Proc. 5th International Conference on Informatic, Electronics and Vision*, ICIEV'16, Dhaka University, Dhaka, Bangladesh, 2016

[2] L. Mehedy, S. M. Arefin and M. Kaykobad, "Bangla Syntax Analysis: A Comprehensive Approach", *In Proc. International Conference on Computer and Information Technology*, ICCIT'03, Dhaka, Bangladesh, pp. 287-293, vol. 5, 2003

[3] M. S. Arefin, M. M. Hoque, M. O. Rahman and M. S. Arefin, "Machine Translation Framework for Translating Bangla Assertive, Interrogative and Imperative Sentences into English ", *In Proc. 2nd International Conference on Electrical Engineering and Information Communication Technology*, ICEEICT'15, Jahangirnagar University Dhaka, Bangladesh, 2015.

[4] M.S . Arefin, L. Alam, S. Sharmin and M. M. Hoque, "An Empirical Framework for Parsing Bangla Assertive, Interrogative and Imperative Sentences" *In Proc. 1st International Confernce on Computer and Information Engineering*, ICCIE'15, RUET, Bangladesh, 2015.

978-1-5090-5627-9/17/$31.00 ©2017 IEEE

[5] M. M. Hoque and M. M. Ali, “Context-sensitive phrase structure rules for structural representation of Bangla natural language sentences ”, *In Proc. ICCIT’04*, Bangladesh, pp. 615-620, 2004.

[6] Md. M. Anwar, M. Z. Anwar and Md. A. Bhuiyan, “Syntax Analysis and Machine Translation of Bangla Sentences”, *IJCSNS International Journal of Computer Science and Network Security*, VOL.9 No.8, August 2009.

[7] S. S. Ashrafi, Md. H. Kabir, Md. M. Anwar and A. K. M. Noman, “English to Bangla machine translation system using Context-Free Grammar ”, *IJCSI International Journal of Computer Science Issues*, Vol. 10, Issue 3, No 2, pp. 287-293,May 2013.

[8] M. G. R. Alam, M. M. Islam, N. Islam ,“ A New Approach To Develop An English To Bangla Machine Translation System ”, *Journal of Modern Science and Technology*,Vol 6, No 1,2011.

[9] M. S. A. Chowdhury, “Developing a Bangla to English Machine Translation System Using Parts Of Speech Tagging”, *Journal of Modern Science and Technology*.

[10] M. R. Selim and M. Z. Iqbal, “Transformational Generative Grammar for Various Types of Bengali Sentences”, *SUST Studies*, Vol. 12, No.1, pp. 276-280, 2010.

[11] M. Rabbani, K.M.R. Alam, M. Islam, "A new verb based approach for English to Bangla machine translation*," ICIEV’14*, pp.1-6, 2014.

[12] H. Azad, *Bakyatattya(বাকযতত্ত্ব)*, The University of Dhaka, Dhaka, 1994.

[13] H. Mamud, *Vasha-shikha (ভাষা-শিক্ষা)*, 2006.

[14] M. Z. Iqbal, *Dipu Number two (দিপু নাম্বার টু),* 1996.

[15] *Bangla Bakoron(বাংলা ব্যাকরণ),* Class six(ষষ্ঠ শ্রেণী), 2014.

[16] S.M Zakir Hossain , *A Passage to the English Language,* 2005.

[17] D. Arnold, *Machine translation an introductory guide*, Ncc, Blackwell Ltd., London, 1994.

978-1-5090-5627-9/17/$31.00 ©2017 IEEE