Context-Sensitive Phrase Structure Rule for Structural Representation of Bangla Natural Language Sentences

Article

2 authors:

Moshiul Hoque
Chittagong University of Engineering & Techn…

34 PUBLICATIONS   102 CITATIONS

SEE PROFILE

Muhammad Masroor Ali
Bangladesh University of Engineering and Te…

16 PUBLICATIONS   39 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   Wikipedia Entry Augmentation View project

Context-Sensitive Phrase Structure Rule for Structural Representation of Bangla Natural Language Sentences.

Mohammed Moshiul Hoque

Dept. of Computer Science and Engineering
Chittagong University of Engineering and Technology,
Chittagong-4349, Bangladesh
Email: moshiulh@yahoo.com

Muhammad Masroor Ali

Institute of Information and Communication Technology
Bangladesh University of Engineering and Technology,
Dhaka-1000, Bangladesh
Email: mmasroorali@iict.buet.ac.bd

ABSTRACT:

Phrase-structure rules are used in context-sensitive grammars to describe a given language /'s syntax. This is accomplished by attempting to break language down into its constituent parts. Phrasal categories include the noun phrase, verb phrase, and adjective phrase; lexical categories include noun, verb, adjective, adverb and many others. This paper proposes a set of context-sensitive phrase structure rules to parse the all kinds of Bangla natural language sentences including simple, complex and compound sentences. The proposed rules can parse the all types of Bangla sentences. This paper also describes a technique to decompose a complex sentence into a dependent and independent clause and a compound sentence into a simple sentence respectively. The inflection of Bangla verb phrase called auxiliary can have different forms depending on the tense, the person and the class of the subject of the verb. In this paper, we also presented a decomposition procedure for Bangla verb phrase into several subparts and then extracting necessary information from the auxiliary part.

Keywords: Phrase structure, SR, Parsing, Decomposition, Clause.

INTRODUCTION:

Natural language processing (NLP) are developed both to explore general theories of human language processing tasks such as providing natural language interfaces or front ends to application system. The subject of NLP is an immense field with many potential applications, including translation from one language into another, retrieval of information from databases, human/computer interaction, and automatic dictation. A language understanding program must have considerable knowledge about structure of the language including what words are and how they combine into phrases and sentences. It must also know the meaning of the words and how they contribute to the meaning of the sentence and to the context within which they are being used. In this paper, we represent a set of context-sensitive grammar (CSG) rules to parse the all kinds of Bangla sentences. To interpret any language requires determining the sentence structure of that language. Phrase structure (PS) grammars define how the basic components of symbol strings, the symbols themselves, can be aggregated into phrases,

and how these phrases can themselves be aggregated finally into sentences. A grammar captures the legal structure in a language and thus allows a sentence to be analyzed [2]. By using the Context-free grammar (CFG) rules, it cannot be possible to generate all the sentence patterns in Bangla languages. To generate all the sentences of Bangla, we have to use CSG's. The CSG PS rules says that how parts of speech can put together to make grammatical sentences.

A parser module can used the CSG's phrase-structure rules and also can find out how words in sentence related to each other. Therefore, parsing a sentence by using a set of PS rules produces underlying structure for sentence. The result is usually a tree or structural representation (SR) that can be used as input for a Machine Translation Engine. Machine translation (MT) [1] [2] is an emerging paradigm for processing natural languages. Now a day, MT and NLP are being used by language industries for translation one language to another and allow people to interact with computers in a natural human language that will save a lot of time and cost of money.

1.1 Previous work and Contribution of this Paper:

Parsing of Bangla sentences by using CSG's rules is in rudimentary stage still now. Very few research work have been conducted regarding parsing of Bangla sentences but a significant number of research activities have been conducted on the recognition of Bangla characters. Among them [4] describe the parsing of Bangla sentences for simple sentence, while [5] describes detail analysis of Bangla phrases and different types of sentences by using the Transformational Generative Grammar's (TGG's). Syntax analysis for Bangla sentences is presented in [7]. Besides these, [6] focuses on designing a Bangla conversion processor using natural language processing. Parsing methodology for Bangla simple sentences by using context-free grammars (CFG's) is presented in [8]. This paper is an extension to scheme presented in [8]. The extensions and refinements presented in this paper over previous work are as follows:

- A set of context-sensitive PS rules for simple sentence.
- A sets of context-sensitive PS rules for complex sentences.

- A sets of context-sensitive PS rules for compound sentences.
- Decomposition procedures for auxiliary of verb phrase.
- Decomposition technique for complex and compound sentences into simple sentence and parse these sentences as simple sentences.
- SR has been used to represent the complex and compound sentences respectively.

2. CONTEXT-SENSITIVE GRAMMAR (CSG's) FOR BANGLA:

Contex – sensitive grammars (CSGs) are the collection of a set of PS rules in which, at least one PS rule is context- sensitive [3]. Applications of PS rules are restrictive by the context that is if certain conditions regarding the context are fulfilled then they will be activated. In CSGs, auxiliaries are depended on subjects. The category of person + class for both the auxiliary and subject should be the same. Now, according to the above discussion, we can represent a set of context – sensitive phrase grammar rules for simple sentence in table 1. Form the table 1, we find the structure of the verb phrase (VP) and verb form (VF).

3. STRUCURE OF BANGLA VERB PHRASE (VP):

Every sentence in Bangla must have a verb phrase. The compulsory part of a VP is the verb Form (VF).
The original part may contain a NP or an AP or a NP and AP. As a rule VP can be expressed as follows-

VP ⟶ (NP) (AP) VF

3.1. Verb Form (VF):

Verb Form is the most important word in a sentence. It is the word used for stating something about a person or thing. The Bangla verb form can be segmented into two parts: These are- i) Verb Root (VR) ii) Auxiliary (AUX)

4. AGREEMENT:

In every language, there is an agreement between the subject and verb form. Similarly, there is also an agreement of person + class between verb form and subject in Bangla. The auxiliary changes for tense, aspect and person + class. For example-
 "Tini Boi Porchen"(¢a¢e hC fs-Re). Here, --------------
VP⟶ Boi Porchen (eB co‡Qb)
VF ⟶ Porchen (co‡Qb)
VR⟶ Por (co)
AUX⟶ Chen (‡Qb)
AUX ⟶ Aspect + Tense + Concord
Che = Ch + Φ+ en
 = Continuous + Present + Concord of Third Person Honorific for Present Tense
That means, in Bangla, the person + class of AUX follows the person + class of subject NP. For example, if subject is "Apani"(Bf¢e) then person + class of

Table 1. CSG's PS rules for Simple Sentence of Bangla Natural language.

Rule No.	CSG's Rule
01.	SS ⟶ NP VP
02.	VP ⟶ (NP) (AP) VF
03.	NP ⟶ N
04.	NP ⟶ N (DET)
05.	NP ⟶ NP (Biv) (NP)
06.	NP ⟶ (DEMO) (SPR) (AP) N
07.	NP ⟶ N PM
08.	N ⟶ PER
09.	AP ⟶ AD/AD'S
10.	DEMO ⟶ (DD) (DO)
11.	SPR ⟶ QFR (PP)
12.	VF ⟶ VR AUX
13.	PER ⟶ FP / SP/ TP
14.	FP ⟶ ami / amara (Avwg/Avgiv)
15.	SP ⟶ SPH / SPNH / SPP
16.	SPH ⟶ apani / apanara (Avcwb/Avcbviv)
17.	SPNH ⟶ tumi/tomara (Zzwg /†Zvgiv)
18.	SPP ⟶ tui/tora (ZzB/†Zviv)
19.	TP ⟶ TPH / TPNH
20.	TPH ⟶ tini/tara/shikak (wZwb/Zviv/wk¶K)
21.	TPNH ⟶ se/tara/pakhi/boi (‡m/Zviv/cvwL)
22.	DET ⟶ ti (wU)
23.	QFR ⟶ ek/dui/bahu/anek (GK/`yB../eû/A‡bK)
24.	PP ⟶ ti/ta/khana/jon (wU/Uv /Lvbv/ Rb)
25.	AD ⟶ khub/bhalo/sundor (Lye/fvj/ my›`i)
26.	PM ⟶ era/ra/der/samuha (Giv/iv/†`i/mg~n)
27.	DD ⟶ ai/sei/oi (GB/†mB/H)
28.	DO ⟶ prothom/ ditia (cÖ_g/wØZxq.)
29.	Biv ⟶ ke/der/re/kar (‡K/†`i/†i/Kvi)
30.	VR ⟶ likh/por/chal/ja (wjL/co/Pj/ hv)
31.	AUX ⟶ e/che/bhe/bo/chhen (G/†Q/†e/e/†Qb)

auxiliary should be "en"(He) or for "Tumi"(a¤¢j), it should be "o"(A) or "e"(H) are accepted and so on. Such as- "Apani Boi Porchen" (Bf¢e hC fs-Re) [Apani (Bf¢e)+ Boi (hC) + Por (fs) + chh (R) + en (He)]. But it is not possible to expressed – "Apani Boi Porcho"(Bf¢e hC fs-R¡) [Apani (Bf¢e) + Boi (hC) + Por (fs) + chh (R) + o (A)], Because "o"(A) does not agree with the subject "Apani"(Bf¢e). Now, we will try to shown in table 2, 3, 4, 5 how the auxiliary part of verb form (VF) can relate or agree with the various person and classes of the subject and with the various tenses of the verb.

4.1. Subcategorization and updating the Grammar Rules:

Subcategorization expresses the constraints that a predicate (verb for now) places on the number and type of the argument it wants to take. Therefore, from the above analysis, we can concluded that the inflection of Bangla verb called AUX can have different forms and hence subcategorized depending on the tense, the person and the class of the verb. The forms are described in table 2, 3, 4 and 5. After a thorough observation of table 2, 3, 4 and 5, we proposed a set of updated Context-sensitive grammar rules to further decompose and extract information from the verb of a sentence, with the following PS rules of table 6.

Table 2: Agreement of Person + class with the verb auxiliary and subject for Past Tense:

		Person and Class					
		FP	SPH	SPNH	SPP	TPH	TPNH
	Subject	ami/amara (Avwg/Avgiv)	apani/apanara (Avcwb/Avcbviv)	tumi/tomara (Zzwg/‡Zvgiv)	tui/tora (ZzB/‡Zviv)	tini/tara (wZwb/Zviv)	se/tara (‡m/Zviv)
Aspect	Indefinite	Iam =l o+am (jvg=j+Avg) kor+lam = korlam (Kijvg)	len = lo + en (†jb=j+Gb) kor + len = korlen (Ki‡jb)	le = lo + e (†jb=j+G) kor + le = korle (Ki‡j)	li = lo + i (wj=j+B) kor + li + = korli (Kiwj)	len = 1o + en (†jb=j+Gb) kor + len = korlen (Ki‡jb)	lo (j) kor + lo = korlo (Ki‡jv)
	Continuous	chholam = chh + i + lo + am (wQjvg = Q+B+j+Avg) kor + hhi - alm = korchhialm (KiwQjvg)	chhilen = chh + i + lo + en (wQ‡jb = Q+B+j+Gb) kor + chhilen = korchhilen (KiwQ‡jb)	chhile = chh + i + lo + e wQ‡j = Q+B=j+G kor + chhile = korchhile (KiwQ‡j)	chhili = chh + i + lo + i (wQwj=Q+B+j +B) kor + chhili = korchhili (KiwQwj)	chhilen = chh + i + lo + en (wQ‡jb=Q+B +j+Gb) kor + chhilen = korchhilen (KiwQ‡jb)	chilo = chh + i + lo (wQj=Q+B+j +A) kor + chilo = korchilo (KiwQj)
	Perfect	echhilam =e + chh + i + lo + am (GwQjvg= G+Q+B+j + Avg) kor + chhilam = korechhilam (K‡iwQjvg)	echhilen = e + chh + i + lo + en (GwQ‡jb=G+Q+ B+j+Gb) kor + echhilen = korechhilen (K‡iwQ‡jb)	echhile = e + chh + i + lo + e (GwQ‡j) kor + echhile = korechhile (K‡iwQ‡j)	echhili= e + echh+i+lo+i (GwQwj) =kor + echhili = korchhili (K‡iwQwj)	echhilen=e+ chh+i+lo+en = (GwQ‡jb) =kor + echhilen = korechhilen (K‡iwQ‡jb)	echilo=e+chh +i+lo (GwQj = G+Q+B+j+A) kor+echilo = korechilo (K‡iwQj)
	Imperative	Not applicable					

Table 3: Agreement of Person + class with the verb auxiliary and subject for Future Tense:

		Person and Class					
		FP	SPH	SPNH	SPP	TPH	TPNH
	Subject	ami/amara (Avwg/Avgiv)	apani/apanara (Avcwb/Avcbviv)	tumi/tomara (Zzwg/‡Zvgiv)	tui/tora (ZzB/‡Zviv)	tini/tara (wZwb/Zviv)	se/tara (‡m/Zviv)
Aspect	Indefinite	bo (e) kor + bo = (Ki+e = korbo (Ki‡ev)	ben (†eb) kor + ben = (Ki+†eb=Ki‡eb) korben	be (†e) kor + be = (Ki+†e=Ki‡e) korbe	bi (we) kor + bi = (Ki+we=Kiwe) korbi	ben (†eb) kor + ben (Ki+†eb= Ki‡eb) = korben	be (†e) kor + be (Ki+†e = korbe Ki‡e)
	Continuous	te thakbo (†Z _vKe)	te thakben (†Z _vK‡eb)	te thakbe (†Z _vK‡e)	te thakbi (†Z _vKwe)	te thakben (†Z _vK‡eb)	te thakbe (†Z _vK‡e)
	Perfect	e thakbo (G _vK‡e)	e thakben (G _vK‡eb)	e thakbe (G _vK‡e)	e thakbi (G _vKwe)	e thakben (G _vK‡eb)	e thakbe G _vK‡e
	Imperative	Not applicable	ben (†eb) kor + ben= (Ki+‡eb= Ki‡eb) korben	o (A) kor + o = (Ki+A=K‡iv) koro	ish (Bm) kor + ish = (Ki+Bm= Kwim) korish	ben (†eb) kor + ben = (Ki+eb= Ki‡eb) korben	be (†e) kor + be = (Ki+†e= Ki‡e) korbe

Table 4: Agreement of Person + class with the verb auxiliary and subject for Habitual Past:

		Person and Class					
		FP	SPH	SPNH	SPP	TPH	TPNH
	Subject	ami/amara (Avwg/Avgiv)	apani/apanara (Avcwb/Avcbviv)	tumi/tomara (Zzwg/‡Zvgiv)	tui/tora (ZzB/‡Zviv)	tini/tara (wZwb/Zviv)	se/tara (‡m/Zviv)
Aspect	Indefinite	tam=to+am (Zvg = Z+Avg) kor + tam = (Ki+Zvg=KiZvg) kortam	ten = to + en (†Zb = Z+Gb) kor + ten = (Ki+†Zb = Ki‡Zb) korten	te = to + e (†Z= Z+G) kor + te = (Ki+†Z=Ki‡Z) korte	ti = to + i (wZ=Z+B) kor + ti = (Ki+wZ=KiwZ) korti	ten = to + en (‡Zb=Z+Gb) kor + ten= korten (Ki‡Zb)	to (Z) kor + to = (Ki+Z=KiZ) korto
	Continuous	te thaktam (†Z _vKZvg)	te thakten (†Z _vK‡Zb)	te thakte (†Z _vK‡Z)	te thakti (†Z _vKwZ)	te thakten (†Z _vK‡Zb)	te tha kto (†Z _vK‡Zv)
	Perfect	e thaktam (G _vKZvg)	e thakten (G _vK‡Zb)	e thakte (G _vK‡Z)	e thakti (G _vKwZ)	e thakten (G _vK‡Zb)	e thakto (G _vKZ)
	Imperative	Not applicable					

Table 5: Agreement of Person + class with the verb auxiliary and subject for Present Tense:

		Person and Class					
		FP	SPH	SPNH	SPP	TPH	TPNH
Subject		ami/amara (Avwg/Avgiv)	apani/apanara (Avcwb/Avcbviv)	tumi/tomara (Zzwg/‡Zvgiv)	tui/tora (ZzB/‡Zviv)	tini/tara (wZwb/Zviv)	se/tara (‡m/Zviv)
Aspect	Indefinite	i (B) kor + i =kori (Ki + B =Kwi)	en (Gb) kor + en = koren (Ki+Gb)	o (A) kor + o = koro (Ki+A)	ish (Bm) kor + ish = korish (Kwim)	en (Gb) kor + en = koren (K‡ib)	e (G) kor + e = kore (K‡i)
	Continuous	chhi/chchhi = wQ/w"Q kor + chhi (Ki + wQ korchhi (KiwQ)	chhen/ (†Qb) chehhen (†"Qb) kor + chhen = korchhen (Ki‡Qb)	cho/ (†Qv) chchho (†"Q kor + chho = korchho (Ki‡Qv)	chhish/ (wQm) chchhish kor + chhish = korchhish (KiwQm)	chhen/ (†Qb) chchhen kor + chhen = korchhen (Ki‡Qb)	chhe/ (†Q) chchhe (†"Q kor + chhe = korchhe (Ki‡Q)
	Perfect	echhi/ (BwQ/Bw"Q) echchhi kor + echhi = korechhi (Ki+BwQ= K‡iwQ)	echhen/(G‡Qb) echchhen kor+echhen = korechhen (K‡i‡Qb)	echho/ (G‡Qv) echchho kor+echho = korecho (K‡i‡Qv)	echhish/ (GwQm) echchhish kor+echhish = korechish (K‡iwQm)	echhen/ (G‡Qb) echchhen kor+echhen = korechhen (K‡i‡Qb)	echhe/ (G‡Q echchhe kor+echhe = korechhe (K‡i‡Q)
	Imperative	Not applicable	un (Db) kor + un = korun (Ki"b)	o (A) kor + o = koro (K‡iv)	Null kor + Null = kor (Ki)	un (Db) kor + un = korun (Ki"b)	uk (DK) kor + uk = koruk (Ki"K)

5. TYPES OF SENTENCES:

Structurally, there are three (3) types of sentence in Bangla [3]. So, a PS rule for different kinds of sentences of Bangla can be expressed as table 7:

Table 7: PS rules for Bangla Sentence.

Rule No.	PS Rule for Simple Sentence
01.	S ⟶ SS / CS / COMS
02.	SS ⟶ Simple Sentences
03.	CS ⟶ Complex Sentences
04.	COMS ⟶ Compound Sentences

5.1. Structure of the Simple Sentence:

A simple sentence is formed by an independent clause or principle clause. A principle clause in Bangla can be rewritten as (or decomposes into, or consist of) a NP followed by VP. For example - "se kall ashbe" (®p L¡m Bp-h) or "Rahim eshkole jae" (l¢qj ú¤-m k¡u). Therefore, a simple sentence can be represented by the following PS rule-

SS ⟶ NP VP
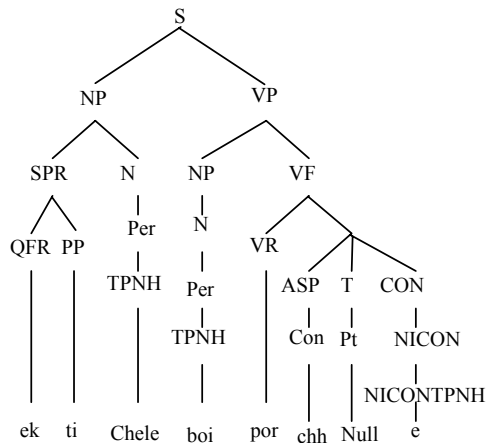
The structure of sentences that are allowed by grammar
Fig. 1: SR for the simple sentences " Ekti Chele boi Porche" (HL¢V ®R-m hC fs-R).

trees or an SR. By using the context-sensitive PS rule of table 1, 6, and 7, we can parse the simple sentence shown in figure 1.

Table 6: Context – Sensitive PS rules for Bangla with Subject -Verb and Person + Class agreement.

Rule No.	CSG's Rule
01.	VP ⟶ (NP) (AP) VF
02.	VF ⟶ VR AUX
03.	VR ⟶ kor (Ki) /por (co) /other verb roots
04.	AUX ⟶ ASP T CON
05.	T ⟶ Pr/Pt/Fu
06.	Pr ⟶ Null
07.	Pt ⟶ lo (j)/ il (Bj)
08.	Fu ⟶ b (e)/bo (†ev)
09.	ASP ⟶ I/Con/P/IMP/Hab
10.	I ⟶ Null
11.	Con ⟶ chh (Q)/ chchh (")Q)
12.	P ⟶ echh (GQ)
13.	IMP ⟶ Null
14.	Hab ⟶ to (Z)
16.	NICON ⟶ NICONFP/ NICONSPH/NICONSPNH
17.	NCON ⟶ NICONSPP/NICONTPH/ NICONTPNH/
18.	ICON ⟶ ICONSPH/ICONSPNH/ ICONSPP/ICONTPH/ICONTPNH
19.	NICONFP ⟶ i (B)/am (Avg)/Null/o (A)
20.	NICONSPH ⟶ en (Gb)
21.	NICONSPNH ⟶ Null/e (G)
22.	NICONSPP ⟶ ish (Bm)/I (B)
23.	NICONTPH ⟶ en (Gb)
24.	NICONTPNH ⟶ e (G)/Null
25.	ICONSPH ⟶un (Db)/en (Gb)/ben (†eb)
26.	ICONSPNH ⟶ o (A)
27.	ICONSPP ⟶ ish (Bm)/Null/i (B)
28.	ICONTPH ⟶ en (Gb)/un (Db)
29.	ICONTPNH ⟶ e (G)/uk (DK)/be (†e)

5.2. Structure of the Complex Sentence:

A complex sentence is formed by one principle clause and one or more subordinate clause. For example - "ami jani je tumi ashbe" (B¢j S¡¢e ®k a¤¢j Bp-h). In this sentence, Principle Clause: "ami jani" (B¢j S¡¢e), and Subordinate Clause: "je tumi ashbe" (®k a¤¢j Bp-h). This two-sentence "ami jani" (B¢j S¡¢e) and "je tumi ashbe" (®k a¤¢j Bp-h) are simple form but second part is the subordinate of the first part. Clauses are combined with subordinator and/ or corresponding subordinator complement.

Some subordinator and their respective subordinator complement used in Bangla are given in table 8.

Table 8: Some Subordinators and their respective Subordinator Complements.

Subordinator (SUBORD)	Subordinator Complement (SUBCOM)
jadi (hw`)	tahole (Zvn‡j)
jadi (hw`)	tobe (Z‡e)
jadi-na (hw`-bv)	tahole (Zvn‡j)
jekhane (‡hLv‡b)	sehkane (‡m¤nZz)
je (‡h)	se (‡m)
jemon (‡hgb)	temon (‡Zgb)
jehetu (‡h‡nZz)	sehetu (‡m¤nZz)
jara (hviv)	tara (Zviv)

From the above analysis, proposed a set of PS rules for parsing the complex Sentences of Bangla. Apply the rules of table 1, 6, 7, and 8 for parse the complex sentence is shown in fig. 2.

Table 9: A set of Rules for parsing Complex Sentence.

Rule No.	Rule for Complex Sentence
01.	C ⟶ DC IC
02.	CS ⟶ IC DC
03.	DC ⟶ NP (SUBORD) VP
04.	DC ⟶ (SUBORD) SS
05.	IC ⟶ NP (SUBCOM) VP
06.	IC ⟶ (SUBCOM) SS
07.	SS ⟶ All Simple Sentenc
08.	SUBORD ⟶ Jadio, Jadi, Jekhane, Jehetu
09.	SUBCOM ⟶ Tahole, Tobe, Sekhane, Sehetu

5.3. Structure of the Compound Sentence:

A compound sentence is formed by two or more principle clause joined by an indeclinable or connectives.

5.3.1 Connectives or Indeclinable:

Indeclinable means no changes i.e. word in the sentence is unchanged. The part of speech, which is always, unchanged form and which makes the sentence more meaningful and also used, as a connective for more than one part of speech, clauses or sentences is called indeclinable. As a PS rule Connectives can be represented as follows:

Conj ⟶ O (J), Ebong (Hhw), Fale (g-m), Sutorang (Bl p¤al¡w), Kingtu (¢L¿°).

5.3.2 Decomposition Technique of Compound Sentence:

To analyze the compound sentence, breaks the sentence into several independent clauses then analyzed the independent clauses separately. If independent clause is same as simple sentence then analyze that clause as simple sentence or if the independent clause is same as complex sentence then analyze that clause as complex sentence. For example - "Tomora manush noo, ar jader chalaoo taraoo manush noo"(®a¡jl¡ j¡e¤o eJ, Bl k¡-cl Q¡m¡J a¡l¡J j¡e¤o eu). It is a compound sentence and it consists of two independent clause- (i) "Tomara manush noo" (®a¡jl¡ j¡e¤o eJ) and (ii) "Jader chalaoo taraoo manush noo" (k¡-cl Q¡m¡J a¡l¡J j¡e¤o eu). These two independent clauses are combined a connective ar (Bl). First independent clause is a simple sentence and second independent clause is a complex sentence. In the second independent clause is the combination of two clauses: **(i)** Principle clause: "Tomara manush noo" (®a¡jl¡ j¡e¤o eJ) and **(ii)** Subordinate Clause: "jader chalaoo" (k¡-cl Q¡m¡J). These two clauses are simple sentence. Therefore, first two independent clause are separated and first one is analyze as simple sentence and second one is analyze as complex sentence and then the two clause under the complex sentence are analyze as simple sentence separately. Compound sentence can be constructed in several ways [8]. Such as-
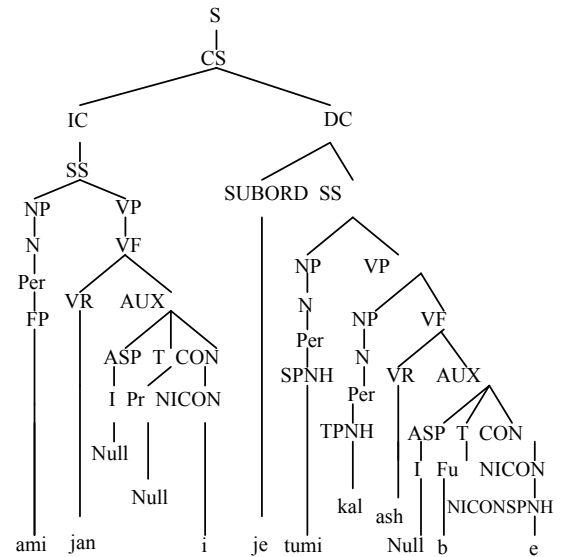

Fig. 2: SR for the Complex Sentence "Ami Jani Je Tumi kal ashbe" (B¢j S¡¢e ®k a¤¢j Bp-h).

(i) Simple + Simple:
Consider an example- "se soth kingtu tar bhai osoth" (®p pv ¢L¿° a¡l i¡C Apv). In this sentence, "se soth" (®p pv) and "tar bhai osoth" (a¡l i¡C Apv) are principal clause or simple sentence and kingtu (¢L¿°) is a connective.

(ii) Simple + Complex:
Consider an example- "se soth kingtu je bonduti tar sange ashechilo se osoth" (®p pv ¢L¿° ®k hå¤¢V a¡l p-‰ H-p¢Rm ®p Apv). In this sentence, "se soth" (®p pv) is a simple sentence and "je bonduti tar sange ashechilo se osoth" (®k hå¤¢V a¡l p-‰ H-p¢Rm ®p Apv) is a complex sentence.

(iii) Complex + Complex:

Consider a long sentence as an example -"jadi jante chaoo se keno asheni ta hole bolbo ami jani Na, ar jadi jante chaoo ami keno jaini ta hole bolbo amar echhe holona Tai" (k¢c S¡e-a Q¡J ®p ®Le B-p¢e a¡ q-m hm-h¡ B¢j S¡¢e e¡, Bl k¢c S¡e-a Q¡J B¢j ®Le S¡¢e a¡q-m hm-h¡ Bj¡l C-µR q-m¡ e¡ a¡C). In this example, two complex sentences are connected by a connective ar (Bl). In this similar manner, compound sentence can be constructed as simple + simple, simple + complex, complex + simple, Complex + Complex etc. Now, we represent a set of rules to express the Compound sentence in table 10.

Table 10: A set of rules for Compound Sentence.

RULE NO.	RULE for COMPOUND SENTENCE
01.	COMS ⟶ SS Conj SS
02.	COMS ⟶ SS Conj CS
03.	COMS ⟶ CS Conj CS
04.	COMS ⟶ CS Conj SS
05.	SS ⟶ All Simple Sentence
06.	CS ⟶ All Complex Sentence
07.	Conj ⟶ O, Ebong, Noile, Kintu, Fale, Notuba, Sutorang,...

Now, we apply the rule of table 1,6, 7, 8 and 10 to parse the compound sentence of "Karim Danio lok sutorang Tini Jakat Diben" in the following fig. 3.
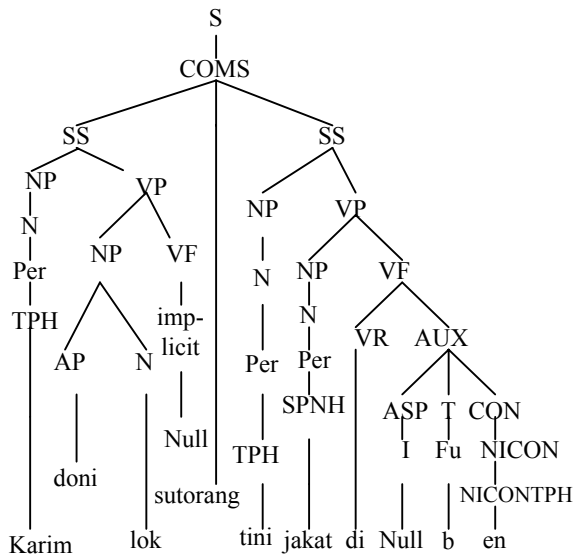

Fig. 3: SR for the Compound sentence "Karim Doni lok sutorang Tini Jakat Diben" (L¢lj de£ ®m¡L p¤al¡w, ¢a¢e k¡L¡a ¢c-he).

6. FUTURE IMPROVMENTS:
Our proposed CSG's PS rules can be parsed the all kinds of Bangla sentences in syntactic patterns. There are lots of researches scopes on the filed of Bangla NLP. The possible future development may mention in the following:

- CSG's for semantic analysis of Bangla sentences.
- CSG's for morphological analysis of Bangla sentences.
- Efficient algorithm for Bangla NL parser to solve the problem of ambiguity.

- CSG's rules to handle various types of Bangla punctuation symbols.
- The concepts of Voice, narration, composition, of words and inflection of noun or pronoun should be further analyzed.
- Pragmatic analysis on complex and compound sentences may be carried on.

7. CONCLUSION:
Parser is one of the most important parts of an NLP and an MT systems. Parsing a sentence then involves finding a possible legal structure for sentence and finally gets an SR. Our proposed context-sensitive PS rules can be assigned all types of Bangla sentences into an SR. To represents a complex and a compound sentence into an SR, we uses the decomposition technique. An SR generally graphic object and this representation cannot be deals with computer. The standard representation of an SR is a list that are one of the data structure that can be implemented and manipulated very easily within a computer. Further modify and extending the CSG rules (with the addition of sentences consisting of idioms and phrases, double word, change of voice, and narration), we can able to represents the all kinds of Bangla sentences which are used as an input of an MT engine to produced other equivalent sentences.

8. REFERENCES:
[1] A. Trujillo, *Translation Engines: Techniques for Machine Translation*, Springer-verlag, London, (1992).
[2] D. Arnold, L. Balkan, S. Meijer, R. L. Humphreys, and L. Sadler, *Machine Translation An Introductory Guide*, Ncc, Blackwell Ltd., London, (1994).
[3] A. Humayan, *Bakyattava,* 2nd ed., The University of Dhaka, 1994.
[4] M. M. Murshed, *Parsing of Bengali Natural Language Sentences*, in Proc., *International Conference on Computer and Information Technology, ICCIT' 98*, (Dhaka, Bangladesh), PP. 185-189, (1998).
[5] M. R. Selim and M. Z. Ikbal, *Syntax analysis of Phrases and Different types of Sentences in Bangla*, in Proc., *International Conference on Computer and Information Technology, ICCIT' 99"*, (Sylhet, Bangladesh), pp. 175-186, (1999).
[6] M. I. A. Khan, A. K. M. A. Hossian and R. C. Debnath, *"A Bangla Conversion Processor using Natural Language Processing"*, In Proc. *International Conference on Computer and Information Technology, ICCIT'02*, (Dhaka, Bangladesh), PP. 262-266, (2002).
[7] 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, (2003).
[8] M. M. Hoque and M. M. Ali, *A Parsing Methodology for Bangla Natural Language Sentences,* In Proc. *International Conference on Computer and Information Technology, ICCIT'03*, (Dhaka, Bangladesh), PP. 277-282, (2003).