

**GRAMMAR FOR BANGLA VERB PHRASE (KRIYAGUCCHO)****Nazmin Akter***

Department of Computer Science and Engineering, Metropolitan University, Sylhet,
Bangladesh.

Article Received on 01/02/2017

Article Revised on 22/02/2017

Article Accepted on 14/03/2017

Corresponding Author*Nazmin Akter**

Department of Computer
Science and Engineering,
Metropolitan University,
Sylhet, Bangladesh.

ABSTRACT

Grammar for Bangla verb phrase is specifying some exact grammar for identifying bangla verb phrase from sentences. There are few rules we can find to identify verb phrase from sentence but that's not well enough to identify verb phrase. A verb phrase is the portion of a sentence that contains both the verb and either a direct or indirect

object (the verb's dependents). It is an important part of Natural Language Processing and it is often the first stage many applications of NLP. Grammar for Bangla verb phrase used in Text to Speech Applications, Information Retrieval, Linguistic Research, Machine Translation and many others. Now-a-days bangla language gaining popularity and it is one of the widely used languages all over the world. In this thesis I will develop grammar for bangla verb phrase, so my main focus will be on verb phrase. There are few strategies to identify verb phrase, Supervised and unsupervised are two main categories. I used supervised and also use unsupervised approaches for making verb phrase grammar.

KEYWORDS: Grammar for Bangla verb phrase.**INTRODUCTION**

Bangla language is rich with finite verb, but there are also many infinite verbs. These forms can be used other Part-of-Speech (verbs, adjectives or nouns) tagging. In bangla grammar there is a wide range of grammatical meanings of verb. This fact resulted in contradictions in the form's analysis not only in scientific researches, but also in Bengali text-books. My approach for Grammar for bangla verb phrase is to first use an annotated corpus with tagged Part-of-Speech and search this corpus for identify verb phrase from a sentence, then I will use them

to make chain rule form a large text set and finally use Hidden Markov Model, N-Gram, Brill-Tagger machine learning techniques to use this chain rules for finding Part-of-Speech for unknown words and then I separate the verb part only any try to make a structure and grammar to identify verb phrase from a sentence. It's very hard to find verb phrase from a sentence that's why in my approach I uses both supervised and unsupervised methods.

Though the computerization of Bangla is an inevitable need, only a few researches have been made to efficiently recognize natural Bangla sentences. This paper proposes a technique to parse Bangla sentences in a new approach using context-free grammar rules that accept all types of Bangla sentences.^[2]

Previous Work

In this section I will review some of the previous works on Bangla verb phrase tagging. Although some works has been done on verb phrase tagging in Bangla, but those are not satisfactory level due to lack of resources. Center for Research on Bangla Language Processing, Bangladesh University of Engineering and Technology, has done some research on Bangla Syntax analysis and Context-Sensitive Phrase Structure. One of their work is, Context-Sensitive Phrase structure Rule for Structural Representation of Bangla Natural Language Sentences 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.^[2] Another comparison work is Bangla Syntax Analysis: A Comprehensive. This paper proposes a technique to parse Bangla sentences in a new approach using context-free grammar rules that accept all types of Bangla sentences. Also describe the diagrammatic model of proposed parser and lists the first set of grammar rules for the parser.^[1] They also develop a parser, the parser should accept all Bangla sentences that are syntactically correct, parse each of them using the rules of the proposed grammar, or report an error otherwise.^[1] University of Asia Pacific and Jahangirnagar University, describe Semantic Analysis Approach improves correct method of reconversion

of UNL expression of Bangla language. A new technique has been proposed in this thesis work. The new technique used a constructive approach to determine the universal words of Bangla language. The novelty of this method is that, it used straightforward and simple technique to determine the ambiguity of Bangla word as well as the diversified usage of words in sentences for a given Bangla sentence.

Verb Phrase Structure in Bangla

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-

- i) $VP \rightarrow (NP) (AP) VF.$
- ii) $VP \rightarrow VF.$

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).^{[2][3]}

I found this above two structure of bangla verb phrase that are used to identify verb phrase from a sentences. At first we need to use tag set Noun, Pronoun, Verb, Adverb, Adjective, Finite verb and Infinite verb for bangla verb phrase tag. First major obstacle in bangla POS tag is lack of resource. There no huge corpus available for bangla. Another problem is one word may have different POS tag depending on its surrounding words. If I do not know POS tag for neighboring word it may be difficult to find exact POS tag for a word. Bangla word may also change its POS when suffix are added.

Example:

sey <u>vat</u> <u>khai</u>	tini <u>boi</u> <u>porchen</u>
Here,	Here,
Noun \rightarrow vat	Noun \rightarrow boi
Rootword \rightarrow kha	Rootword \rightarrow por
Suffix \rightarrow a	Suffix \rightarrow en

amar <u>kharap lagche</u>	Mahi <u>lal Jama poreche</u>
Adjective → kharap	Noun → lal
Rootword → lag	Adj → jama
Suffix → che	Rootword → por
	Suffix → eche

My Approach

There is already some work going on bangla language processing in Shah Jalal University of Science & Technology. And there few structures to identify verb phrase. It's not possible to find all verb phrases to use this current structure. That's why I implement some new structure to identify all verb phrases correctly. At first I will tag all the noun, pronoun, adjective from sentences and then I will identify all verbs then categorize the entire verb like finite verb, infinite verb, and root verb. I already apply my approach to 50 Bangla sentences and I can easily find verb phrase from those sentences. But still I face lots of problem because I don't identify all verb phrases using just this structure and I further I work on it.

Approached Verb Phrase Structure

i) VP → IV VF

Example:

Tumi oita khete parbe na

Infinite verb → Khete

Rootword → paar

Suffix → a

Auxilieires → na

Tini Khete jaben

Infinite Verb → khete

Rootword → ja

Suffix → en

ii) VP → IV (NP) (AP) VF

iii) VP → (NP) (AP) IV VF

Example:

dekhte dekhte din gelo

Infinite verb → dekhte

Noun → din

Rootword → gelo

Auxlieries → o

se porte porte gumiye gelo

Infinite verb → porte

Noun → Ghum

Rootword → gelo

Suffix → yea

Experiment & Result Analysis

I experiment with my current implementation but there's a lots of verb form. After stemming a word, I just check it that is it match to my structure and if it matches with my approach structure then I separate the verb phrase portion from sentences. But there are some problems with stemming. A word can be stemmed by several ways and so there may be more than one root word for a word. Sometimes it is also possible for a root word with more than one POS tag. Here are some examples:

aasbe =aa+seb/ADV

aaseb =aasob+কো/NN

proshonge = pro+shonge/PRP

proshongo = proshongo+কো/NN

There are some other problems also. কো suffix is a “kaal”,”Suffix” and “nirdeshok”. When i will make chain rule this problem may be solved by machine learning techniques.

Futere work

Grammar for Bangla verb Phrase identifying greatly depends on size of pos tagged and specially verb form. My plan is to enrich it, the more efficient the structure, the more bigger corpus, the better performenc it will be. I also have plan for feature base morphological parsing for বিশেষ্য পদ. My future plan is Bangla Verb Phrase identify, transform Verb Phrase in to specific grammar generate Parser and Reduce ambiguity form grammar. (Use CSG grammar)

REFERENCES

1. Lenin Mehedy, S. M. Niaz Arifin and M Kaykobad, “Bangla Syntax Analysis: A Comprehensive Approach”, BUET.
2. Mohammed Moshiul Hoque, Muhammad Masroor Ali, “Context-Sensitive Phrase Structure Rule for Structural Representation of Bangla Natural Language Sentences”, CUET, BUET.
3. M. F Mridha, Alope Kumar Saha, Jugal Krishna Das, “Solving Semantic Problem of Phrases in NLP Using Universal Networking Language (UNL)”, University of Asia Pacific, Jahangirnagar University.
4. Sibansu Mukhopadhyay, Tirthankar Dasgupta, Manjira Sinha, Anupam Basu “Automatic Extraction of Compound Verbs from Bangla”, Society for Natural Language Technology Research, Kolkata, Indian Institute of Technology Kharagpur, Kharagpur.
5. Dr. Mohammad Shahidullah, “Bangla Grammar”
6. Md. Daniul Haque, “Bhasabigganer kotha”
7. “Bangla bhasar Baykoron”
8. Professor Dr. Alauddin Al Azad “Uccotoro Bangla Basharity”
9. Humayun Azas, “Bakkototto”