

Valency Analyzer of Verb Arguments for Bangla

Subhash Chandra

subhash.chandra@cdackolkata.in
NLP Group, Centre for Development of
Advanced Computing (CDAC), Kolkata

Pampa Bhattacharyya

pampa.bhattacharyya@cdackolkata.in
NLP Group, Centre for Development of
Advanced Computing (CDAC), Kolkata

Abstract

This paper presents a methodology for analysis of the requirements of Bangla verbs for its arguments. Literary it can be called the searching for expectation of Bangla verbs and compatibility of Bangla nouns. There are two phases in the total study. First one is to build up a method for generating the expectation of verbs and compatibility of nouns for Bangla. Bangla ontological tree and verb expectation system are the outcome of this method. Second is to establish the relationship between verb and noun. As a result an analyzer will be generated for valency checking. After completion of this work system can distinguish between different types of noun entities. The work is under development.

Keywords:

Verb Semantic, Verb Argument, Compatibility of Noun, Ontological Tree, Noun Ontology, Expectation for Verb, Verb Valency, Verb Categorization, Valency Analyzer

1. Introduction

Verbs play a key role in sentence production and comprehension because they specify the relations among words in a sentence. It is well accepted that some crucial aspects of sentence processing are determined by the semantic and syntactic attributes of the verbs that appear in the sentence [1]. It is a well known linguistic fact that a verb has some expectations for its arguments whereas noun has compatibility for going with specific verb. The study mainly aims at establishing these types of relations. Unfortunately, there is no significant achievement regarding verb valency for Bangla till date. In natural language, Verb compatibility determination is very useful in the field

of language analysis. The ultimate goal of this study is to develop a useful parser for Bangla. The ‘Valency Analyzer of verb arguments’ will act as a crucial input for the Bangla parser which is under development. Its output will be used as a part of the semantic disambiguation phase of the parser. Now this study is going on regarding verb categorization on broad level. The option is given here to increase the verb categories further when and where required. Along with this categorization, an ontological tree is prepared for subject and object information. This will help to develop an ontological dictionary with subject and object property information. Finally a rule file will be developed for mapping the arguments with respect to the specific verb valency.

2. Structure of Bangla Sentence

Bangla language descends from Sanskrit language belonging to an Indo-European Family. Bangla has generally Subject-Object-Verb (SOV) structure [2]. For example-

Sr.	Bangla	ITRANS	English
1	ছেলেটি (subject)	chheleTi	the boy
2	রুটি (object)	ruTi	bread
3	খাচ্ছে (verb)	khAchche	is eating

“Table-1: Structure of Bangla Sentence”

If the structures of Bangla sentences are observed very minutely, then the following sequence comes out -
1. Subject→ 2. Time Adverbial→ 3. Place Adverbial→ 4. Indirect Object→ 5. Direct Object→ 6. Verb [3].

Sr.	Bangla	ITRANS	English
1	আমি	Ami	I
2	প্রতিদিন	pratidina	everyday
3	অফিসে	aphise	in office
4	বিবেককে	bibekake	to Bibek
5	পেন	pena	pen
6	দিতাম	ditAma	gave

"Table-2: Structure of Bangla Sentence"

Although in most of the cases Bangla is free word order language [4] and nouns occur with zero morphemes, the main problem lies in identifying the subject and object in sentence. For example-

Sentence-1

Bangla: পাখি আম খায় ।

ITRANS: pAkhi Ama khAYa.

English: Birds eat mango.

Sentence-2

Bangla: আম পাখি খায় ।

ITRANS: Ama pAkhi khAYa.

English: Birds eat mango.

In the above two examples, bird occurs as subject and mango occurs as object in different word order and with zero morphemes. So the problem is to extract the subject and object information by some rules. So the main challenge is to identify the subject and object through the system.

3. Bangla Verb

Bangla is one of the famous languages that show agglutinative behavior [5]. The inflections are added in root verb with respect to tense, aspect, modality and person only, but not with gender and number as in the case of Hindi and other languages. It is identified that a single Bangla root verb generates 59 inflected forms along with its non-finite forms [6]. It is possible to categorize the Bangla verb in two corpulent levels.

3.1 One-part Verb

One-part verbs are generated with one word. These are classified in two categories by nature-

3.1.1 Verb based on object

This type of verbs are classified into two categories according to the expectation of the object of the specific verb as an argument-

3.1.1.1 Transitive verb

Verbs which can take an object called Transitive verb. For example- Bangla: -শোন, -কেন, ITRANS: *shona*, *-kena* etc, English: to listen, to buy etc.

Again Transitive verbs are classified into two categories based on the number of objects it can take-

(i) mono-transitive

(ii) di-transitive

3.1.1.2 Intransitive verb

Verbs which can't take an object called Intransitive verb. For example: Bangla: -ঘুমা, -দৌড়া, ITRANS: *-ghuma*, *-dau.DA* etc., English: to sleep, to run etc.

3.1.2 Verb based on sense

Based on completeness of sense (tense, aspect, modality and person), verbs are categorized into two-

3.1.2.1. Finite verb

The finite verbs show tense, aspect, modality and person information and carry a complete sense, for example

Bangla: আমি গতকাল বইটা পড়েছিলাম।

ITRANS: Ami gatakAla baiTA pa.DechhilAma

English: I read the book yesterday.

3.1.2.2. Non-finite verb

In other hand, non-finite verbs are those which don't carry tense, aspect, modality and person information. For example-

Bangla: আমি গতকাল বইটা পড়ে ফেরত দিয়েছিলাম ।

ITRANS: Ami gatakAla baiTA pa.De pherata diYechhilAma

English: I returned the book yesterday after reading.

3.2 Two-part verb

Two-part verbs are generated with more than one word. This type of verbs are classified in three categories

3.2.1 Conjunct verb

The combinations of Conjunct verb are 'noun+verb', 'adjective+verb' and 'verbal noun+verb'. For example: 'noun+verb'- Bangla: ব্যবহার কর, চড় মার, ITRANS: *byabahAra kara*, *cha.Da mArA*, English: to use, to slap, 'adjective+verb'- Bangla: গৃহীত হ, পরাজিত হ, ITRANS: *gRRihIta ha*, *parAjita ha*, English: to accept, to defeat, 'verbal noun+verb'- Bangla: খাওয়া হ,

করা হ, ITRANS: *khAoYA ha, karA ha etc.*, English: to feed, to do etc..

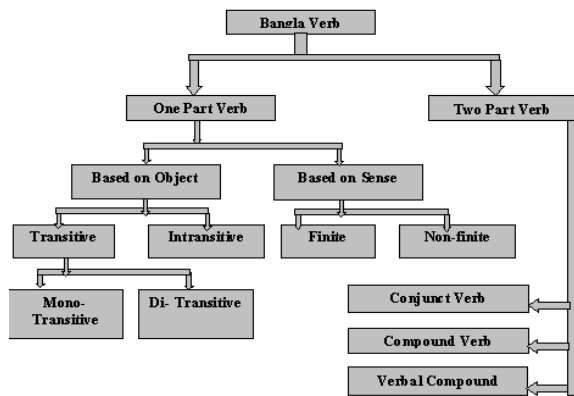
3.2.2 Compound verb

The combination of compound verb is ‘non-finite verb+finite verb’. For example- Bangla: উঠে পড়, কেঁদে ফেল, ITRANS: *uThe pa.Da, ke.Nde phela*, English: to get-up, to start crying.

3.2.3 Verbal compound

The combinations of verbal compound are ‘one verb+another verb’/ ‘two verbs with compounding’/ ‘consecutive occurrence of same verb’ [3]. For example- Bangla: যাওয়া-আসা, মারামারি, হেসে হেসে, ITRANS: *yAoYA-AsA, mArAmAri, hese hese* English: come and go, to fight etc.

Bangla verb can be understood by following hierarchy-



“Figure 1 Bangla Verb Classification”

The study revolves around on the analysis of one-part verb only.

4. Expectation and Compatibility

4.1 Expectation of Verb

Intransitive verbs expect one noun as an argument which is the subject of the sentence. Transitive verbs expect two nouns as an argument, first one is the subject and second one is the object of the sentence but there are some restrictions. All verbs cannot take every noun as an argument. Actually verb expects a valid combination driven by rule of subject and object. For example-

Expectation of Verb			
Bangla	রাম	লোহা	খাচ্ছে
ITRANS	rAma	lohA	khAchchhe
English	Ram	iron	is eating

“Table-3 Expectation of Verb”

The above sentence though syntactically correct is semantically wrong because Ram cannot eat an iron. The verb ‘eat’ expects an animate as subject, here Ram is a member of animate group so this rule is correct for this sentence, but ‘eat’ also expects an edible thing as an object, that rule is breaking here because iron is not an edible thing. In other word, if subject occurs as animate and object as edible thing in a sentence in the context of “eat” verb then sentence is acceptable otherwise sentence might not be acceptable. This is a depiction of the expectation of verb.

4.2 Compatibility of Noun

As expectation of verb, the compatibility of noun is another viewpoint. There are also some restrictions in noun because all nouns cannot go with all verbs. Example-1

Compatibility of noun			
Bangla	পাখি	আকাশে	উড়ছে
ITRANS	pAkhi	AkAshe	u.Dachhe
English	The bird	in the sky	is flying

“Table-4 Compatibility of Noun”

Example-2

Compatibility of noun			
Bangla	মানুষ	আকাশে	উড়ছে
ITRANS	mAnuSha	AkAshe	u.Dachhe
English	Man	in the sky	is flying

“Table-5 Compatibility of Noun”

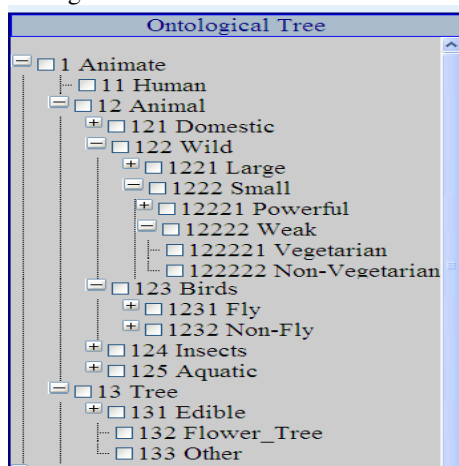
In the above examples, the first one (table-4) is acceptable and the second example (table-5) is not acceptable. The main difference between the sentences is compatibility of noun. Bird and man are both common noun and both are also animate but there are differences between two nouns regarding some basic attributes, because bird has the property for flying but man has no such property. This poses a problem also, because what are the attributes and how many of them should be considered for sense disambiguation is really a challenge that has to be addressed. There is always a tension between exploring deep in the attribute structure of the noun and its manageability.

5. Ontology

The structure of ontological class hierarchy has been build up with the help of Roget’s Thesaurus

ontology [7], Mikrokosmos Ontology [8], Buffalo ontology [9], Navya-Nyaya [10]. After a lot of study, the elements of the ontological tree have been defined according to Bangla corpora collected from Internet by NLP group of CDAC, Kolkata and MIT and self-knowledge. The ontological tree covers each noun which is found in Bangla texts for deciding property and developing ontological dictionary.

The ontological class will be complete through manual effort and to be modified time to time according to need. A sample of ontological class is shown in Figure 2-



“Figure 2 Ontological Tree”

This is not the final class hierarchy but only a sample for developing ontology. It can be changed through program according to the requirements.

6. Methodology

Firstly Bangla verbs are collected from “Samsad Banan Abhidhan” by Ashoke Mukhopadhyay [6] and MIT corpus. After scrutinizing, verbs are classified into two categories-Transitive and Intransitive. Then the compatibility of subject of intransitive verbs has been checked. After that transitive verbs are studied and classified with its expectation of the object. Finally according to expectation of the object transitive verbs are classified into two categories- Mono-Transitive (I eat rice); Di-Transitive (I gave him a book) [11].

This knowledge has been represented using ontological method [12] with the help of knowledge database. At the time of developing an ontological tree corpora and verb requirements are studied. Next, the ontological tree has been developed to describe these entities and relations. Finally, accessing the ontological tree a knowledge base for Bangla nouns and pronouns

has been built up in the respect of most frequently using verbs in Bangla corpora.

7. Database of Verb Arguments Valence for Bangla

After defining ontological classes following databases have been developed.

7.1 Ontological Knowledge Database (OKD)

Ontological Knowledge Database (OKD) is a knowledge database with property information of each base word (right now Noun and Pronouns only) collected from corpora and Samsad Banan Abhidhan by Ashoke Mukhopadhyay [6]. The database has id, word, ontological information, number value of ontological information and updated date etc. in the following format:

Id	Word			Ontology
	Bangla	ITRANS	English	
1	গরু	garu	Cow	12
2	আমি	Ami	I	11
3	তুমি	tumi	You	11
4	সিংহ	si.nha	Lion	12

“Table-6 Ontological Knowledge Database (OKD)”

Where number sequence encodes the property set for the word. For example if the Number value of any word is 12 then, it will be *Animate* → *Animal*. If the number value of any word is 11 then *Animate* → *Human*.

7.2 Argument Valence Mapper Database (AVMD)

Argument Valence Mapper Database (AVMD) is subject and object mapper database. The database is in the following format:

Sr.	Verb Id	Subject	Object
1	VB1	11	31
2	VB1	12	O11
3	VB1	121	O111

“Table-7 Argument Valence Mapper Database (AVMD)”

All acceptable combinations of argument for the specific verb are given in the above table 7. Suppose VB1 is representing the Bangla verb “khA” (eat) and then the subject line represents all possible subjects for “khA” (eat) verb such as 11 for human, 12 for animal, 122 for animal → wild etc. Object line of Table-7 also represents the compatible objects for above mentioned subject such as if the subject occurs as 11 then object must be 31, 251, 253 etc. (follow the Figure 2 ontological tree) in the specific sentence according to “khA” (eat) verb.

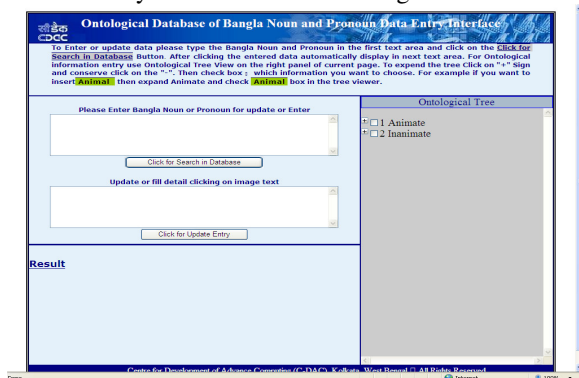
8. User Interface and Technical Details

9.1 User Interface

Web based interfaces have been developed for Data entry, analysis of verb expectation etc. The details are as follows-

8.1.1 Data Entry Interface

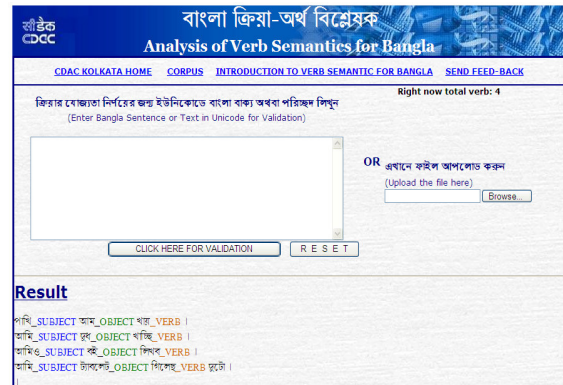
The Data entry user interface for data entry is web based. The interface allows the users to modify the existing entries and enter new words with ontological information in database with the help of ontological tree viewer, which is displayed on the right panel of the interface. In the interface there are two text areas and one tree viewer. The first text area for searching whether the entry exists or not, the second area is meant either for new entry that goes into the database or updation of an existing entry. The ontological tree is provided to alleviate the problem of data entry. Every nodes of the tree has been displayed with the check box () with the associated ontological value. After checking in the box of nodes system automatically returns the ontological value in second text area, so that the user don't have to remember every details of the ontological derivation. A sample of ontological class of the data entry interface is shown in Figure 3.



“Figure 3 Data Entry Interface”

8.1.2 Analysis of Verb Arguments Valence System Interface

This interface accepts Bangla Unicode sentence, text or corpora in text area. User can also upload a text file through file browsing. After clicking on button system returns analyzed data with acceptable subject and object tag. A sample of ontological class of the data entry interface is shown in Figure 4.



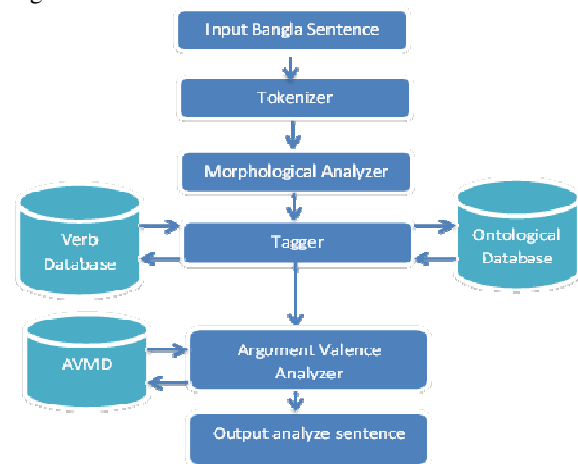
“Figure 4 Analysis of Verb Arguments Valence System Interface”

8.2 Technical Detail

Web based Data entry Interface has been developed in Python. My SQL database and text files are used for back end and Python Server Pages (PSP) are used for front end and Python for program. There are two options for input. The first one is for typing the data and search in the database. If entry found then user can update the information otherwise user can enter word with details. In case of second option user can upload file with all informations. After getting the necessary information system will automatically update the database. For Analysis of Verb Arguments Valence System (VAVS) has been developed in python, front end in PSP and back end is in My SQL Server.

9. Analysis of Verb Arguments Valence

The system can be understood by following Figure-



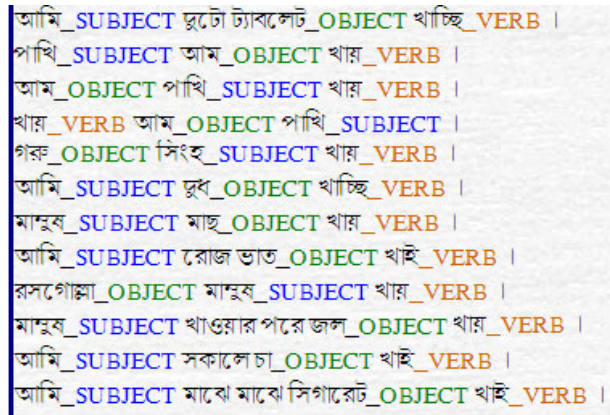
“Figure 5 Data Entry Interface”

Analysis of Verb Arguments Valence System (AVAVS) is under development stage. The analysis is

doing by the help of OKD and Argument Valence Mapper Database (AVMD). In the AVMD the specific verb has been defined with the help of valid argument-1 and argument-2 combinations. At first system gives the ontological information of each input words with the help of OKD, then check all ontological information through AVMD. If any combination of arguments found in AVMD then system returns as valid combination of argument otherwise it declares the combination of argument as invalid.

10. Result and Conclusion

Most frequently used 10 verbs (“khA” (eat), “bala” (speak), “likha” (write) etc.) among 10000 Bangla sentences have been studied and generated the ontological tree and rule. Then it tested over 2000 sentences. An instance of the result is shown in Figure 6 for “khA” (eat) verb.



“Figure 6 Result”

This is an effort to provide knowledge to Machine for verb argument valence. The current work is in developmental stage, so it does not cover all Bangla verbs. It works only in simple sentences. The whole problem cannot be solved by this approach only but it will be a very helpful model for language processing from the computational aspect. Besides it can be very useful system for knowledge representation to Machine for NLP, which is very challenging till date.

References:

[1] Shapiro LP, Zurif E, Grimshaw J. “Sentence processing and the mental representation of verbs”, *Cognition* 1987, 27:219-246.

[2] Ghosh Aniruddha, Das Amitava and Bandyopadhyay Sivaji, “Clause Identification and Classification in Bengali”, *Proceedings of the 1st Workshop on South and Southeast Asian Natural Language Processing (WSSANLP)*, The 23rd International Conference on Computational Linguistics (COLING), Beijing, August 2010, pages 17–25.

[3] Pabitra Sarkar, *Bhasha Jigyasa*, Vidya Sagar Pustak Mandir, Kolkata, 2003.

[4] Das D. and Choudhury M. . “A Valency Theoretic Approach to Shallow Parsing of Free Word Order Languages”, Presented at the Student Paper Contest at *ICON-KBCS 2004*. Hyderabad, India, Dec 2004.

[5] Majumdar Vijaya Chandra, *The history of the Bengali language*, Calcutta, Univ. of Calcutta, 1920.

[6] Mukhopadhyay Ashok, *Samsad Banan Abhidhan*, Sahitya Samsad, Kolkata, 1998.

[7] Jarmasz, M. and Szpakowicz. S. “Roget’s Thesaurus and Semantics Similarity”, *Proceedings of conference on Recent Advance in Natural Languages Processing (RANLP 2003)*. Borovets, Bulgaria, 2003, PP-212-219.

[8] <http://www.ilc.cnr.it/EAGLES96/rep2/node23.html>

[9] Christopher Welty and Barry Smith (eds.), *Formal Ontology in Information Systems*, ACM Press, New York, 2001, xvi + 348 pp

[10] D.Ch. Guha, *Navya-Nyaya System of Logic*, Second revised Edition, Motilal Banarsidass, Delhi, 1979, pp. 41-75.

[11] Bamandev Chakraborty, *Ucchotora Bangla Vyakarana*, Akhshay Malancha Prokashok O Pushtokbikreta, Kolkata, 2004.

[12] John F. Sowa, *Knowledge Representation: Logical, Philosophical and Computational Foundations*, Brooks Cole Publishing Co. Pacific Grove, CA, 1999.