

# Finding Universal Word for Bangla Language: Attribute Analysis Approach

Md. Farid Hasan  
Bangladesh Livestock  
Research Institute, Savar,  
Dhaka, Bangladesh

Jugal Krishna Das  
Department of CSE,  
Jahangirnagar University,  
Dhaka, Bangladesh

## ABSTRACT

Universal natural language (UNL) deals with communication across nations of different languages. It covers many different related disciplines which are linguistic, epistemology etc. UNL helps to overcome the language barrier among people of different nations to solve problems of globalization and geopolitical interdependence. This paper examines bangla root words semantically to find out the universal words according to the usage of words into sentences in relation to the universal network language. The paper also shows different techniques to find the semantic attributes which are the indicators to make rule set. Also, it discusses semantic coverage issues for Bangla to explore the development of universal network language.

## Keywords

**Keywords:** bangla root words; universal words; universal network language (UNL).

## 1. INTRODUCTION

We know, Bangla is the 6th widely spoken language with more than 193 million speakers, most of them live in Bangladesh and the Indian state of West Bangle [1]. Therefore it is essential to take into account a conversion process of Bangla to UNL and vice versa, so that a vast people can be benefited to share information in the Internet. This paper deals with developing some rule set for bangla universal words by attribute analysis approach.

### 1.1 Semantic Analysis of Bangla words

If we want to find out the attributes for universal words, semantic analysis is found to be centered on analysis and generation of word forms. It deals with the implicit meaning of words and how words can be used [2]. It is applied to identify the actual meaning of the words [3, 4] identifying the usage of words from sentence to sentence. It is the field of linguistics that studies the meaning of words. Morphology of words help to find out the attribute needed for universal word list.

## 2. RESEARCH METHODOLOGY

In order to convert Bangla sentence to UNL expressions firstly, we have gone through Universal Networking Language (UNL) [5, 6] where we have learnt about UNL expression, Relations, Attributes, Universal Words, UNL Knowledge Base, Knowledge Representation in UNL, Logical Expression in UNL and UNL systems [7, 8]. All these are key factors for preparing Bangla word dictionary, enconversion and deconversion rules in

order to convert a natural language sentence (Bangla sentence) into UNL expressions. Secondly, we have rigorously gone through the Bangla grammar [9, 10], Verb and roots (Vowel ended and Consonant Ended), Morphological and Semantic Analysis [11], Primary suffixes, Cases and their inflexions, construction of Bangla sentence [12] based on semantic structure. Using above references we extort ideas about Bangla grammar for morphological and semantic analysis in order to prepare Bangla word dictionary (for root, root word suffix etc), morphological rules and enconversion rules in the format of UNL provided by the UNL center of the UNDL Foundation.

## 3. UNIVERSAL LANGUAGE RELATIONS AND SEMANTIC IDENTITIES

The words of Bangla language divided into root words and auxiliary words that creates a single meaningful word. The meaning of words varies depending upon the use of the word and their different placement in the sentence. For Example, “এ চলে বেশিদিন যাবে না”, which means “This rice is not sufficient.” Here the root word [যা] is acting as the word of “sufficient”. On the other hand if we say that “এ শার্টটি বেশিদিন যাবে না” which means “This shirt will not last long.” Here the meaning of root word [যা] is acting as the word of “last”. The parser must have to realize the application of the word in different situations accordingly. Otherwise meaning of the sentence after conversion will not be the accurate one.

One solution to this problem that can be applied to get the actual meaning that is; finding out the universal words as well as the attributes for different meanings and get the appropriate meaning. If we consider the sentence below:

- ট্রেন আজ যাবে না। means “The train will not depart today.” Here the root word [যা] used as the meaning of “depart”.

The universal word for the sentence

*obj(depart(icl>exit>occur, equ>go, obj>thing)).@entry.@not.@future, train(icl>public\_transport>thing)).@def)tim(depart(icl>exit>occur, equ>go, obj>thing)).@entry.@not.@future, today(icl>how, equ>nowadays))*

If we find out the relations for the sentence above we get:

**obj and tim** attributes

Other attributes that are needed to find out the specific word-application to get appropriate meaning here are:

**ROOT, INTR, VEND, CEND**

Here, INTR used for intransitive form of verb, VEND used for vowel ended morpheme and CEND for consonant ended morpheme. All these information will help you to get appropriate meaning of the sentence and for making rules for UNL. The approaches to find out the attributes are as follows:

**3.1.1 Attribute Analysis Approach**

The word [যা] which means “go” can be used in different ways in the Bangali sentence as in Table-1. Where single root word stands for various meaning for different usage.

**3.1.2 Finding the Universal Words for Bangla Words**

Finding universal words for bangla words is somewhat difficult case as this language has varying application in the sentence for the same words having different meanings.

So there should have some techniques to find out bangla universal words that will effectively make the conversion.

**Some rules are given below:**

সে স্কুলে যায়- means “He goes to school”. Here if we question the verb- যায় (যা) i.e; go in English- সে কোথায় যায়? means “Where does he go?” The answer is School. So, it makes the attribute ‘ple’ that shown in Table-2. Again, if we conduct the following techniques to find out the attributes, we will be able to get them more easily as to the Table-2. And, all these attribute have been checked with the Russian and Indian (Hindi) language servers that agrees with this principle. And this technique tremendously helps to make universal words for bangle language.

**3.1.3 UNL analysis for the verbs in Bangla**

If we go through the UNL roots of some bangla verbs that are widely used in the language processing for example, [যা] i.e. ‘go’; [কর] i.e; ‘do’; [খা] i.e. ‘eat’; we can find various complex relationships. And it is sometimes difficult to get the appropriate universal word that exactly maintain specific rule. For the variation of words used in sentences, their application and the way of expression defines the words meaning. In that case, conversion process becomes difficult with the meaning of sentences. But the rules of Table-2,

**Table 1. Different meaning of same root words**

| Root Word     | Meaning when used in sentence |
|---------------|-------------------------------|
| [যা]          | <i>leave/go</i>               |
|               | <i>depart</i>                 |
|               | <i>move</i>                   |
|               | <i>set out/start</i>          |
|               | <i>pass</i>                   |
|               | <i>proceed/advance</i>        |
|               | <i>terminated</i>             |
|               | <i>die</i>                    |
|               | <i>sufficient</i>             |
|               | <i>last</i>                   |
|               | <i>sell</i>                   |
|               | <i>be spent</i>               |
|               | <i>function/work</i>          |
|               | <i>remain</i>                 |
|               | <i>become</i>                 |
|               | <i>vanish</i>                 |
| <i>divide</i> |                               |
| <i>spread</i> |                               |

**Table 2. Attributes Identification Technique for Bangla Language**

| Word- (English meaning)                | Attribute |
|--|-----------|
| কে? - who                              | agt       |
| কি? - what                             | obj       |
| কেমন? - how                            | aoj       |
| কি দিয়ে? - by what                    | ins       |
| কেমন করে? - by how                     | man       |
| কিভাবে? - by how                       | man       |
| একটা হতে আরেকটা- from one to another   | fmt       |
| কি হলো? - result                       | gol       |
| কিভাবে শুরু? - how started             | frm       |
| কিভাবে শেষ- how ended                  | gol       |
| কোন? - which                           | mod       |
| কখন? - when                            | tim       |
| কখন থেকে? - from when                  | tmf       |
| কখন পর্যন্ত? - upto when               | tmt       |
| কোথায়? (অবস্থান)- where (location)    | plc       |
| কোথেকে? - from where                   | plf       |
| কোথায়? (গন্তব্য)- where (destination) | plt       |

If we consider following example-

নতুন বাড়িতে কবে যাচ্ছেন?

which means in English-

**When are you moving into the new home?**

We get the Attribute as below:

**tim, agt, plt**

and the Universal Word stands for:

```
{unl}
tim(move(icl>relocate>do,plt>thing,plf>thing,agt>volithional_
thing).@entry.@pl.@present.@progress.@interrogative,when
(icl>how,timagt(move(icl>relocate>do,plt>thing,plf>thing,agt>
volithional_thing).@entry.@pl.@present.@progress.@interroga
tive,you(icl>person).@pl)mod(home(icl>residence>thing,pos>
person).@def,new(icl>adj.ant>old))plt(move(icl>relocate>do,
plt>thing,plf>thing,agt>volithional_thing).@entry.@pl.@presen
t.@progress.@interrogative,home(icl>residence>thing,pos>
person).@def)
{/unl}
```

Here the root word is : [যা] and the meaning of this root word in this case is to

**leave**

Again, if we consider the following sentence:

এ গাছের শিকড় গভীরে যায় ।

which means in English: **Roots of this tree spread long.**

**The attributes we get are:**

**obj, mod, man**

Universal Word stands for:

```
{unl}
obj(spread(icl>occur,equ>distribute,obj>thing,scn>thing).@entr
y@past,root(icl>plant_organ>thing).@pl)mod(tree(icl>woody_
plant>thing),this(icl>determiner>adj,modbj(root(icl>plant_organ>thing).@pl,tree(icl>woody_plant>thing))
man(spread(icl>occur,equ>distribute,obj>thing,scn>thing).@ent
ry.@past,long(icl>how))
{/unl}
```

Where root word is : [যা] and the meaning of this root word in this case is

**spread**

So, it is clear that the same root word makes different meaning from sentence to sentence depending on the use of the word. And Attribute Analysis Approach makes it easy to find the attribute to make Universal Words. Same technique is true for the other words in making universal words.

## 4. CONCLUSION

This work presents an approach to find out bangla universal word and attributes. This is an effective approach to make the bangla UNL server for language generation. Rigorous thinking is needed for extensive UNL based bangla language structure implementation. We hope the present work on bangla UNL will provide necessary helps toward the difficulties.

## 5. REFERENCES

- [1] [http://en.wikipedia.org/wiki/List\\_of\\_languages\\_by\\_total\\_number\\_of\\_speakers](http://en.wikipedia.org/wiki/List_of_languages_by_total_number_of_speakers) Last Date of access August, 14, 2011
- [2] J. Parikh, J. Khot, S. Dave, P. Bhattacharyya, "Predicate Preserving Parsing", Department of Computer Science and Engineering, Indian Institute of Technology, Bombay.
- [3] M.N.Y. Ali, J.K. Das, S.M. Abdullah Al Mamun, M. E.H. Choudhury, "Specific Features of a Converter of Web Documents from Bengali to Universal Networking Language", International Conference on Computer and Communication Engineering 2008(ICCCE'08), Kuala Lumpur, Malaysia. pp. 726-731
- [4] M.N.Y. Ali, J.K. Das, S.M. Abdullah Al Mamun, A. M. Nurannabi, "Morphological Analysis of Bangla words for Universal Networking Language", International Conference on Digital Information Management, icdim, 2008, London, England, pp. 532-537
- [5] Lavoie B, Kittredge R, Korelsky T; Rambow O. (2000). "A Framework for MT and Multilingual NLG Systems Based on Uniform Lexico-Structural Processing". Proceedings of 6th Applied Natural Language Processing Conference. ACL, 2000 Seattle, USA.
- [6] Boitet C and Sérasset G. (1999). "UNL-French deconversion as transfer & generation from an interlingua with possible quality enhancement through offline human interaction". Machine Translation Summit 99. Singapore.
- [7] <http://www.undl.org> last accessed on January 30, 2011.
- [8] H. Uchida, M. Zhu, T.G. D. Senta "Universal Networking Language", 2005/6-UNDL Foundation, International Environment House.
- [9] D.M. Shahidullah, "Bangla Baykaron", Ahmed Mahmudul Haque of Mowla Brothers prokashani, Dhaka-2003.
- [10] D. C. Shuniti Kumar, "Bhasha-Prakash Bangala Vyakaran", Rupa and Company Prokashoni, Calcutta, July 1999, pp.170-175
- [11] D. S. Rameswar, "Shadharan Vasha Biggan and Bangla Vasha", Pustok Biponi Prokashoni, November 1996, pp.358-377
- [12] H. Azad, "Bakkotottoy", Second edition, 1994, Dhaka