

# Development Analysis Rules of Bangla for UNL Based Machine Translation

Khaled Bin Yousuf, Md. Fattah Ibne Shaheed, Md. Nawab Yousuf Ali

**Abstract**— *The aim of the UNL project is to allow people across nations to access information in the Internet in their own languages. A good number of researchers in computational linguistic all over the world have already started developing UNL system for their respective native language. But the UNL system is not developed for Bangla. Researchers have been working on this issue. But so far no immense attempt has been made for develop analysis rules of bangle for UNL based machine translation. this paper presents a work that develops some templates of analysis rules for converting Bangla sentence to UNL expressions which can later be converted into other languages using language specific generation rules.*

**Keywords**— *Universal Networking Language (UNL) Universal Words (UW) , Morphological Analysis , Semantic Analysis , Analysis Rules*

## I. Introduction

Universal Networking Language (UNL) is a project under the auspices of the United Nations University (UNU), Tokyo, Japan. The mission of the UNL project is to allow people across nations to access information in the Internet in their own languages [2].

English is the main language of the Internet. Understandably not all people know English. Teeming millions are deprived to access the information repositories directly in native language. On the other hand, vast information resources in different languages could not be shared. Knowledge and information are scattered all over the world and remain mostly inaccessible due to non-machine representation and language barrier [4]. Translation is the only means to disseminate information but only with much effort and involving direct and indirect cost.

Most translation works are done involving two languages. Usually translation involves native language and a foreign language like English or Bangla or Spanish. This is how the problem of inter language translation raised its new heights worldwide.

Among those who did their best to tackle this problem was the United Nations University/Institute of Advanced Studies (UNU/IAS) [11]. The institute conducted a review of all internationally available machine translation programs, and finally decided to start devising a better, more efficient and more workable technique via the Internet under the Universal Networking Language (UNL) project .The main aim of the UNL project is to overcome language barrier.

It is worth mentioning that UNU/IAS created the UNDL program which is now carried out by the non-profit UNDL Foundation. The Foundation was established at the beginning of the year 2001 with its headquarters in Geneva, Switzerland; as one of the UN organizations. It was officially registered three months later. By the way of illustration, the following Figure 1 shows the UNL system and how one language is converted to another:

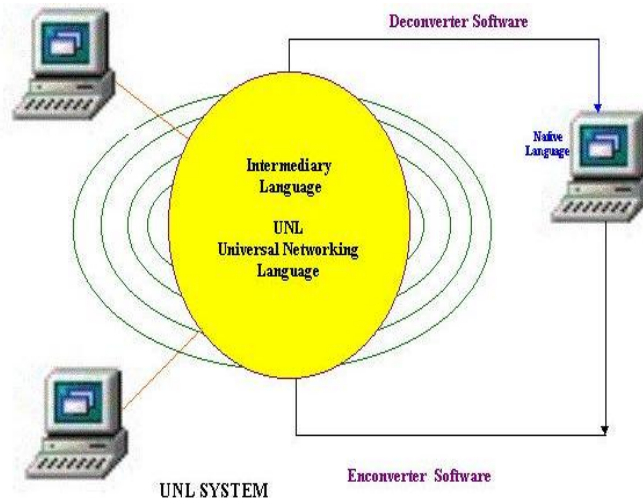


Figure 1. The UNL System

Bangla can also be the member of UNL society, because it is the 4<sup>th</sup> most widely spoken language with more than 250 million speakers, most of whom live in Bangladesh and the Indian state of West Bengal [12]. Therefore, it is essential to take into account a morphological and semantic analysis of Bangla Words for the UNL system to include Bangla as a member of UNL. It would allow a great number of people to access and share through the Internet a vast repository of knowledge.

**Khaled Bin Yousuf**  
East West University  
Dhaka, Bangladesh  
Khaled\_ewu@ovi.com

**Md. Fattah Ibne Saheed**  
East West University  
Dhaka, Bangladesh  
Tunan\_46@yahoo.com

**Dr.Md. Nawab Yousuf Ali**  
East West University  
Dhaka, Bangladesh  
nawab@ewubd.edu



formulate rules that model the knowledge of the speakers of the language. Morphological study comes here to help with rules for analyzing the structure and formation of the words [7].

Left (type is “+”) and right (type is “-“) composition rules are used to perform morphological analyses. By applying “+” type of rule the two headwords of the left and right nodes are combined into a composite node, the original left and right nodes are replaced with the composite node in the node-list and the sub-syntactic tree and attributes of the left node are inherited. If the operator “@” appears in the <ACTION> field of the rule for the left node, the attributes of the right node are also inherited [8]. “-“ type of rule works in reverse order of “+” type.

Both of the cases, original two nodes are deleted from the node-list and the insertion of new composite node into the node-list. The position of the new composite node is on the right analysis window. Templates of some proposed morphological rules are given below:

**Template 1:** Morphological rule for verb root and verbal inflexion:

--:C{<COND1>:<ACTION1>::}{<COND2>:<ACTION>::}  
 Where,  
 COND1::=ROOT, VEND/CEND  
 ACTION1::=@,-ROOT,-VEND/-CEND,+V  
 COND2::= VI, P1/P2/P3

Here, ROOT denotes verb root, CEND for consonant ended root, VEND for vowel ended root, V for verb, and P1, P2 and P3 for first, second and third persons respectively. Addition of V indicates that after morphological analysis a verb is formed combining verb root and verbal inflexion.

For example:

+:{ROOT,VEND/CEND:@::}{ VI ,P1/P2/P3::}

**Template 2:** Morphological rule for noun and preposition

+{ <COND1> :<ACTION2>::} { <COND2>::}  
 where,  
 COND1::=N, NPRO/ NCOM/ NMAT/ NCOL/ NABS,  
 VEND/CEND  
 ACTION1::=@, -N, -NPRO/ -NCOM/ -NMAT/ -NCOL/ -  
 NABS, -VEND/-CEND, +NP  
 COND2::= ABY, ANUS, #FRM

Where, N stands for noun, VEND for vowel ended noun, CEND for consonant ended noun, NP for noun phase, and NPRO, NCOM, NMAT, NCOL, NABS for proper noun, common noun, material noun, collective noun and abstract nouns respectively.

For Example:

+{ N,PRON :: @:} { ABY,ANUS, #FRM /#TO::}

**B. Semantic Rule:**

The purpose of semantic analysis is to determine the structure of the input sentence and how the relations between the words of the sentence are made.

Left (type is “<”) and right (type is “>”) modification rules are used to perform semantic analyses of a sentence. They create semantic tree and semantic relation for the two nodes on the analysis windows. After applying left modification rule, the right node becomes the modifier of the left node and it deletes the right node from the node-list, while the left node becomes the head of this partial syntactic-tree and remains in the node-list, whereas the right modification rule works in reverse order. In both cases, they create semantic relation according to the designation of the relation in the <RELATION> field with the node where the relation is described in the <RELATION> field as the to-node and the partner node as the from node of the semantic relation. They also add the semantic relation and output it in the result of UNL expressions when the enconversion is completed [2, 8].

If the operator “@” appears in the <ACTION>” field of the rule for the left/right node, the attributes of the right/left nodes are also inherited. Templates of some proposed semantic rules are given below.

**Template 1:** Semantic rule for pronoun and verb

>{<COND1>::<RELATION1>::}{<COND2>:<ACTION>::}  
 Where,  
 COND1::=N/PRON,SUBJ  
 RELATION1::=agt/cag/ptn/aoj/cao  
 COND2::= V  
 ACTION::=@, +&TENSE,+&@entry

Where, SUBJ denotes subject of the sentence, agt, cag, ptn, aoj, and cao denote the semantic relations between subject and verb of the sentence. Attribute TENSE presents different tenses and @entry is included with main predicate of the sentence.

For example,

{PRON, HPRON::agt:} {V::@present, @entry:}

**Template 2:** Semantic rules for noun phrase and verb.

>{<COND1>:<ACTION1>:<RELATION1>::}  
 {<COND2>:<ACTION2>::}

Where,  
 <COND1>:: = NP, #FRM  
 <ACTION1>:: = @  
 <RELATION1>:: = plf  
 <COND2>:: =V

Here, > denotes right modification rule, NP denotes Noun Phrase, #FRM denotes attribute of from and plf indicates relation plf (place from).

**Template 3:** Template of semantic rules for noun and verb.

>{<COND1>:<ACTION1>:<RELATION1>:}  
{<COND2>:::}

Where,

<COND1>:: = N, #PLC

<ACTION1>:: = @

<RELATION1>:: = to

<COND2>:: = V

### IV. Conversation of A Bangla Sentence By Analysis Rules

In this chapter, we convert the following sentence using the analysis rules developed in section III.

Sentence: “ আমি মুম্বাই হইতে ঢাকা যাচ্ছি। ”

Pronounce as “Aami Mumbai hoite Dhaka jachhi .”

Meaning: **I am going to Dhaka from Mumbai.**

Enconverter takes the following steps to convert the sentence into UNL expressions:

When we input the sentence into the enconverter, the sentence head (<<) is placed in the LAW, the sentence will be in the RAW and the sentence tail (>>) will be in the RCW shown in figure 3.

#### Initial State

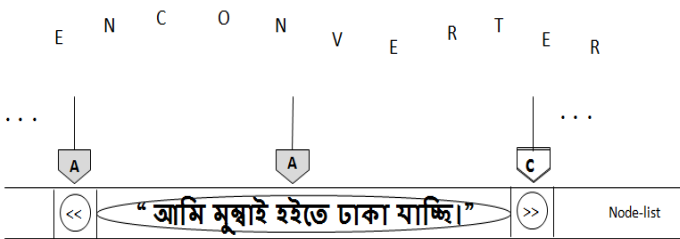


Figure 3. Initial State of the Enconverter

Then the input string is scanned from left to right and all matched morphemes with the same starting characters are retrieved from the word dictionary and become the candidate morphemes and are arranged according to the following a shown in figure 4.

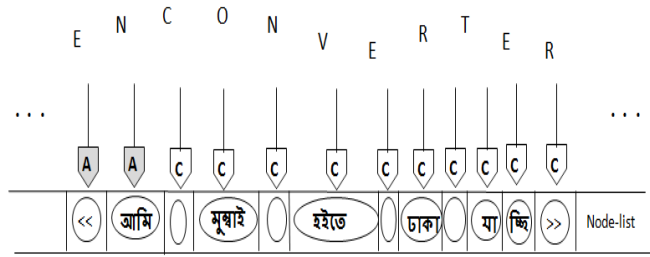


Figure 4. Distribution words and morphemes in the windows of enconverter.

**Rule 1:** R{SHEAD:::} {PRON,SUBJ:::}

**Rule 2:** R{PRON,SUBJ :::} {BLK:::}

**Rule 3:** R{BLK:::} {N, NPRO :::}

Now, the following right shift rules are applied to shift the windows of the enconverter to right to place the noun “মুম্বাই” in the LAW shown in figure 5.

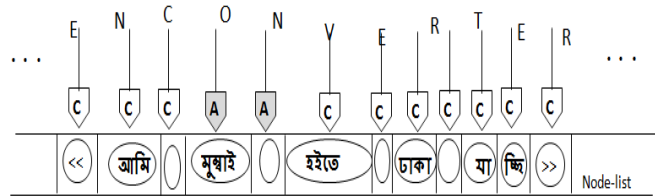


Figure 5. Condition of Enconverter after application rules 1, 2 and 3

**Rule 4:** DR {N, NPRO :::} {BLK:::}

In this situation, Rule 4, the right deletion rule (DR) will be applied to delete the space between “মুম্বাই” and “হইতে” shown in figure 6.

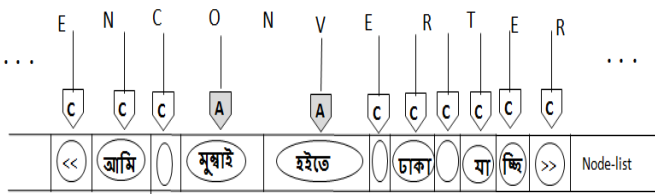


Figure 6. Enconverter after application of rule 4

**Rule 5:** +{ N,NPRO :: @:} {ABY,ANUS, #FRM:::}

Now, left composition rule that denotes “+” sign is applied to combine the two nodes “মুম্বাই” and “হইতে” in to a composite node “মুম্বাই হইতে” to complete the morphological analysis shown in figure 7.



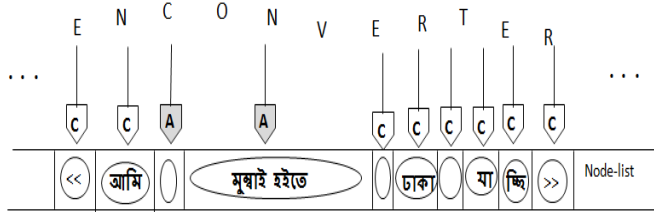


Figure 7. Enconverter after application of rule 5 and 6

**Rule 6:** R {BLK:::} {N,NPRO,ABY, ANUS, #FRM :-N, -NPRO, -ANUS, +NP :::}

Again , rules 7,8,9,10 are applied to shift the windows to right to place root is in LAW and “যা” to place root is in LAW and “ছি” is in RAW shown in figure 8 .

- Rule 7 :** R{NP:::} {BLK:::}
- Rule 8 :** R {BLK:::} {N, NPRO:::}
- Rule 9 :** R {N, NPRO:::} {BLK:::}
- Rule 10 :** R {BLK:::} {N, NPRO:::}

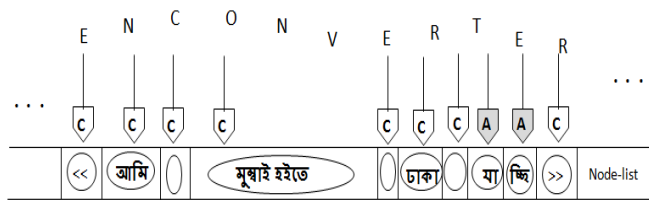


Figure 8. Enconverter after application of rule 7, 8, 9 and 10

Now morphological rule 11, is applied to combine the two head words “যা” and “ছি” to make “যাছি” to complete the morphological analysis shown in figure 9.

**Rule 11:** { +ROOT, VEND :::}{VI,1P:::}

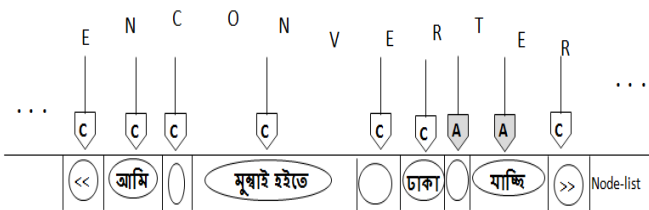


Figure 9. Enconverter after application of rule 11

Now , rule 12 (left shift rule) is applied to shift the windows one step left and modify the attributes of the newly formed verb “যাছি” .

**Rule 12:** L{BLK:::} {ROOT, VEND, VI,1P: -ROOT,-VEND,-VI,1P, +V, #TO:::}

**Rule 13:** DR{N:::}{BLK:::}

Then, in rule 13 the space between noun “ঢাকা” and verb “যাছি” will be deleted shown in figure 10.

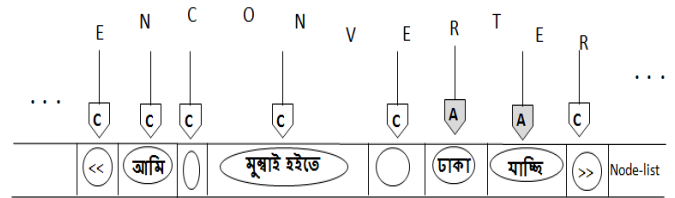


Figure 10. Enconverter after application of rule 12 and 13

At this moment, analysis rule 14 is applied to complete the semantic analysis between “ঢাকা” and “যাছি” using the relation ‘to’ and hence noun ঢাকা is deleted from verb “যাছি” will remain the node RAW shown in figure 11.

**Rule 14 :** >{N, NPRO::to:} {V, #TO:::}

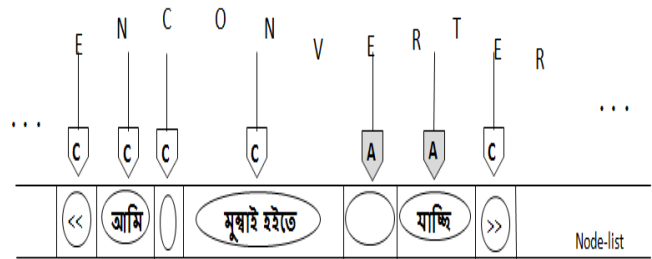


Figure 11. Enconverter after application of rule 14

Hence, rule 15 is applied to delete the space between noun phrase “মুন্স্বাই” and verb “যাছি” shown in figure 12.

**Rule 15 :** DL {BLK:::} {V:::}

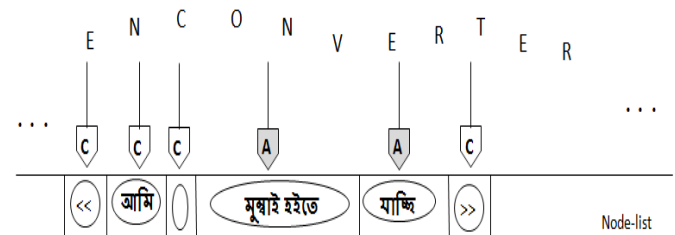


Figure 12. : Enconverter after application of rule 15

At present, Analysis rule 16 is applied to complete the semantic analysis between “মুন্স্বাই” and “যাছি” using relation ‘plt’ and noun phrase “মুন্স্বাই” is deleted shown in figure 13.

**Rule 16 :** >{NP, #FRM::plt:} {V:::}

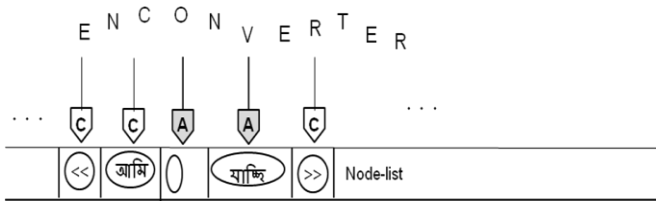


Figure 13. : Enconverter after application of rule 16

Again, rule 17 is applied to delete the node of LAW, so that pronoun “আমি” will be placed in the LAW.

**Rule 17 : DL {BLK:::} {V:::}**

Finally, rule 18 is applied to complete the semantic analysis between “আমি” and verb “যাচ্ছি” by relation “agt” and “আমি” will be deleted from the node-list and “যাচ্ছি” will remain in the RAW shown in figure 14.

**Rule 18: {PRON, HPRON:::agt:} {V:::@Progress, @entry:}**

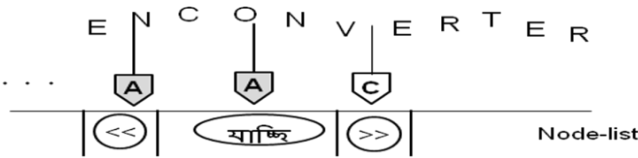


Figure 14. Enconverter after application of rule 17 and 18

When conversion is completed the following UNL expression is created shown in table 1:

TABLE I. UNL EXPRESSION

```
{unl}
agt(go(icl>move>do,plt>place,plf>place,agt>thing).@entry.@
present.@progress,i(icl>person))
to(go(icl>move>do,plt>place,plf>place,agt>thing).@entry.@p
resent.@progress,Mumbai)
plf(go(icl>move>do,plt>place,plf>place,agt>thing).@entry.@
present.@progress,Dhaka(iof>national_capital>thing))
{/unl}
```

**v. Conclusions and Future Work**

We have proposed some templates of analysis rules to convert Bangla sentences into UNL expressions. We have also shown the conversion procedures of a sentence into UNL expression. Theoretical experiment shows that the format of analysis rules work well to convert Bangla sentences into UNL expression.

Our future plan is to make more templates for other head words including their prefixes, suffixes and inflexions and templates for morphological and semantic rules based on the format provided by the UNL center of the UNDL foundation which will be helpful for the further research work on UNL based Bangla Machine Translation.

**References**

- [1] Earley, J., “An Efficient Context Free Parsing Algorithm”, Communications of the ACM, 13(2), 1970
- [2] Uchida H., Zhu M., “The Universal Networking Language (UNL) Specification Version 3.0”, Technical Report, United Nations University, Tokyo, 1998
- [3] Shibprashanna Lahiri, “Beboharik Vasha Bichitra”, Shahitya Prkashoni, Dhaka, 1999
- [4] Enconverter Specification Version 2.1, UNU Center, Tokyo 150-8304, Japan 2000
- [5] Bhattacharyya, “Multilingual Information Processing Using Universal Networking Language”, Indo UK Workshop on Language Engineering for South Asian Languages LESAL, Mumbai, India, 2001
- [6] M. M. Asaduzzaman, M.M. Ali, “Morphological Analysis of Bangla Words for Automatic Machine Translation”, International Conference on Computer, and Information Technology (ICCIIT), Dhaka,2003
- [7] Hiroshi Uchida, Meiyong Zhu, “The Universal Networking Language beyond Machine Translation”, UNDL Foundation, September 26, 2001.
- [8] Bangla Academy (2004), Bengali-English Dictionary, Dhaka
- [9] Dashgupta, Sajib, Khan, Mumit, Morphological Parsing of Bangla Words using PC KIMMO, International Conference on Computer, and Information Technology (ICCIIT), Dhaka, 2005
- [10] Nawab Yousuf Ali, Khaled Bin Yousuf and Fattah Ibne Shaheed, “Development of Bangla Word Dictionary through Universal Networking Language Structure”, Global Engineering, Science and Technology Conference (GIST), Dhaka, 2012
- [11] Md. Ershadul H. Choudhury, Md. Nawab Yousuf Ali, Mohammad Zakir Hussain Sarkar, “Framework for synthesis of Universal Networking Language”, East West Journal (Accepted)
- [12] <http://www.undl.org> (last access on 24-12-2012)
- [13] [http://en.wikipedia.org/wiki/Morphology\\_%28linguistic%29](http://en.wikipedia.org/wiki/Morphology_%28linguistic%29)(last access on 24-12-2012)
- [14] Sing, Prem Parkash, “Morphology and Punjabi Word Formation”, Patiala Madan Publications, 2002
- [15] [http://en.wikipedia.org/wiki/List\\_of\\_languages\\_by\\_number\\_of\\_native-speakers](http://en.wikipedia.org/wiki/List_of_languages_by_number_of_native-speakers) (last access on 24-12-2012)
- [16] [http://en.wikipedia.org/wiki/Semantic\\_analysis\\_\(linguistics\)](http://en.wikipedia.org/wiki/Semantic_analysis_(linguistics)) (last access on 24-12-2012)
- [17] T. Dhanabalan, K. Saravanan & T.V. Geetha, Tamil to UNL EnConverter, Proc. International Conference on Universal Knowledge and Language, Goa, India, 2002.

About Author (s):



**Khaled Bin Yousuf**, presently a B.Sc student of Dept. of computer science & engineering , East West University. His research paper “Development of Bangla Word Dictionary through Universal Networking Language Structure” awarded as best paper in Global Engineering, Science and Technology Conference which held 28 – 29 December 2012 at BIAM Foundation. Dhaka.



**Md. Fattah Ibne Shaheed**, presently a B.Sc student of Dept. of computer science & engineering, East West University.



**Dr. Md. Nawab Yousuf Ali**, presently act as chairperson and assistant professor of Dept. of computer science & engineering, East West University. At present he has three journal and fifteen conference papers published in journal and conference proceedings both home and abroad.

