**International Journal of Artificial Life Research**
Volume 7 • Issue 2 • July-December 2017

# Generation of UNL Attributes and Resolving Relations for Bangla EnConverter

Md. Nawab Yousuf Ali, Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh

Md. Sarwar Kamal, Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh

Md. Shamsujjoha, Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh

Mohammad Ameer Ali, Department of Computer Science and Engineering, Bangladesh University of Business and Technology, Dhaka, Bangladesh

Ghulam Farooque Ahmed, Computer Village Ltd., Dhaka, Bangladesh

## ABSTRACT

Conversion of Bangla language to another native language and another language to Bangla language using Universal Networking Language (UNL) is highly demanding due to rapidly increasing the usage of Internet-based applications. UNL has been used by various researchers as an inter-lingual approach for an Automated Machine Translation (AMT) scheme. This article presents a novel work on construction of EnConverter for Bangla language with a special focus on generation of UNL attributes and resolving relations of Bangla text. The architecture of Bangla EnConverter, algorithms for understanding the Bangla input sentence; resolution of UNL relations; and attributes for Bangla text/language are also explained in this article. This article highlights the analysis rules for EnConverter and indicates its usage in generation of UNL expressions. This article presents the results of implementation of Bangla EnConverter and compares these with the system available at Russian and English Language Server.

## KEYWORDS

Analysis Rule, Bangla Language Text, EnConverter, Machine Translation, Morphological Analysis, Universal Networking Language (UNL), Universal Words

DOI: 10.4018/IJALR.2017070102

## 1. INTRODUCTION

Internet has evolved from monolingual into a multilingual World Wide Web for more than 1000 languages shown in (Ali et al., 2012). The Internet usage by the vast number of users now varies from primarily academic purposes to widespread commercial, leisure, education, entertainment etc. purposes. In Internet, vast knowledge and information resources in different languages are scattered all over the world and remain mostly inaccessible due to non-machine representation and language barrier in information distribution as shown in (Ali et al., 2008). It has been showed that there is a great need to translate web pages and electronic mail messages into the native language for overcoming the language barrier (Kumar et al., 2011). Considering this issue, Universal Networking Language (UNL) has been developed by the United Nations University/Institute of Advanced Studies (UNU/IAS) to convert any language to other language presented in (Ali et al., 2013). The UNL system is introduced most of the languages like Arabic, Spanish, Japanese, Chinese, Russian and many more as shown in (Choudhury et al., 2005). A UNL based transfer system needs two components: one for conversion from source language to UNL expression and another for UNL expression to target language. The UNL system has 46 semantic relations and 87 attributes to express the semantic meaning of a sentence of a source language demonstrated by (Uchida et al., 1993). With this increasing pressure of providing information access without language and cultural barriers, Internet today has to deal with immense complexity of multilinguality.

Bangla is the 4$^{th}$ widely spoken language in the world with more than 250 million speakers, most of whom live in Bangladesh and the Indian state of West Bengal which is evidenced (Ali et al. 2008). Bangla text is converted to UNL expressions by the EnConverter of Bangla language. This UNL expression is converted to any other language that has its own DeConverter. This will indeed help to develop a multilingual machine translation system for Bangla language.

A statistical Bangla to English translation engine using only simple Bangla sentences that contain a subject, an object and a verb is developed by (Uddin et al., 2004). A low-cost English to Bangla (E2B)-ANUBAD translating, English text into Bangla text using both rule-based and transformation-based MT schemes along with three-level of parsing is shown in (Saha, 2005). MT Bangla dictionaries that address the organization, contents and details of the information have been developed by (Ali et al., 2002). The research has been done for morphological analysis of Bangla words for UNL by (Choudhury et al., 2005), parsing methodology for Bangla sentences by (Asaduzzaman et al., 2003), and dictionary development of Bangla words by (Ali et al., 2002; Islam, 2009). The suffix, prefix and inflexions of Bangla language are detailed in (Asaduzzaman et al., 2008; Khairunnahar, 2008).

A number of researches have been done in Bangla for UNL based machine translation scheme (Uddin et al., 2004; Khairunnahar, 2008). But so far no attempt has been made for the development of Bangla EnConverte and generation of UNL attributes and resolving the relation of Bangla EnConverter. This motivates us to develop a Bangla EnConverter for UNL expressions for resolving the above issues. A rigorous study on Bangla language grammar presented in (Uddin, 2004; Saha, 2005; Ali et al., 2002; Shahidullah, 2003; Shuniti, 1999; Remeswar 1996; Azad 1994) verb and roots (vowel ended and consonant ended) in (Ali et al., 2017) and morphological

analysis shown in (Shuniti, 1999; Rameswar, 1996; Azad, 1994) based on their semantic structures. The major components of this research touch upon: i) development the architecture of Bangla Enconverter ii) working procedure of Bangla Enconverter iii) algorithm for processing simple assertive Bangla sentence iv) algorithm for relation resolution and generation of attributes and more importantly v) comparative analysis of UNL expressions generated by Bangla Enconverter and UNL expressions generated by the Russian and English Language Server. The framework for designing the Bangla *EnConverter* is explained with special focus on generation of UNL attributes and relations from Bangla language text. The developed enconverter has been tested for various simple assertive Bangla sentences and the experimental results proves the excellent performance of it. These results have been compared with the results of the UNL expressions provided by the English-Russian Language Server. Both of the results are similar. This also proves the better performance of the developed EnConverter.

This paper is organized as follows: Section 2 explains the UNL systems and the design issues of the Bangla *EnConverter*. The architecture of Bangla *EnConverter* is shown in Section 3. The algorithm for processing the Bangla text is explained in Section 4. Section 5 includes the working procedures of Bangla *EnConverter*. The algorithm for resolution of relations and generation of attributes is focused in Section 6. Section 7 elaborates the working of the *Enconverter* with a Bangla input sentence. The experimentation analysis and results of the system is given in Section 8. Section 9 summarizes the paper with some concluding remarks and future directions.

## 2. UNIVERSAL NETWORKING LANGUAGE

The UNL project is concerned about developing an intermediary language system where any written text can be converted to many languages through UNL and simultaneously, all text written in different languages can be converted to that particular language as shown in (Ali et al., 2017; Uchida et al., 2002). The *EnConverter* and *DeConverter* are the main software in the UNL system. *EnConverter* converts any natural language texts into UNL expressions and *DeConverter* converts UNL expressions into target language texts. UNL language servers can be connected to the Internet to perform conversions between natural languages and UNL expressions certified by (Kumar et al., 2011). Each language server will thus contain an *EnConverter* and a *DeConverter* of the language as shown in Figure 1.

### 2.1. UNL Format for Representation of Information

Knowledge within a UNL document consists of Universal Words (UWs), UNL relations and UNL attributes. Word knowledge is expressed by Universal Words which are language independent. UWs constitute the UNL vocabulary and the syntactic and semantic units that are combined according to the UNL laws to form UNL expressions. They are tagged using restrictions describing the sense of the word in the current context. For example, *drink (icl>liquor)* denotes the noun sense of drink restricting the sense to a type of liquor. Here, *icl* stands for inclusion and form an *is-a* relation as in semantic nets (Uchida et al., 2005). These words are listed in the Universal Word Lexicon of UNL knowledge Base shown in (Uchida et al., 1999). Conceptual knowledge is captured by the relationship between UWs through a set of UNL relations. They

**Figure 1. Working procedure of Universal Networking Language (UNL)**

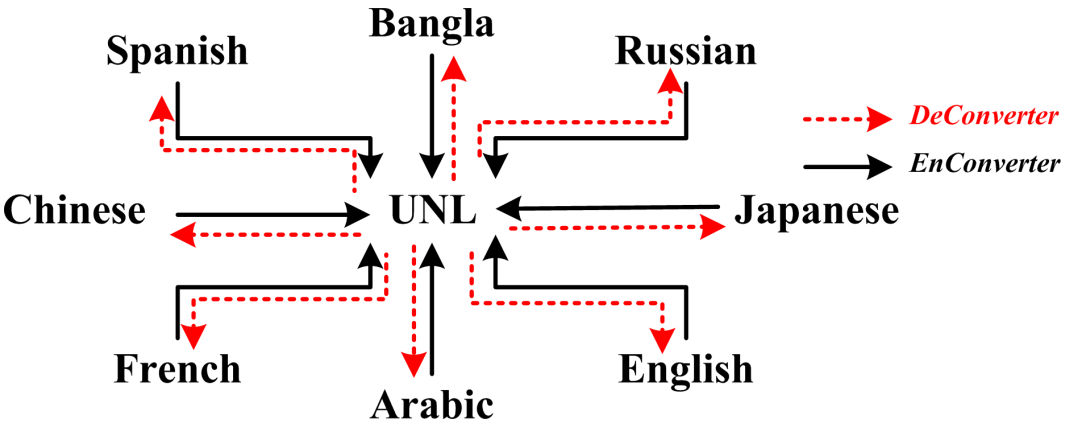
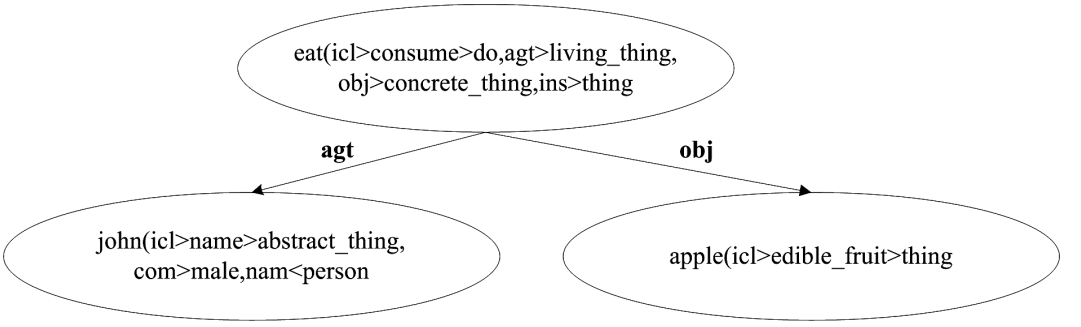



**Figure 2. Universal Networking Language (UNL) graph**




are the building blocks of UNL sentences. The relations between the words are drawn from a set of predefined relations as presented in (Uchida et al., 1993). The attribute labels are attached with universal words to provide additional information like number, tenses etc. For example, the sentence *John eats an apple* can be represented into UNL expression as follows:

```
{unl}
agt(eat(icl>consume>do,agt>living_thing,obj>concrete_
thing,ins>thing).@entry.@present,john(icl>name>abstract_
thing,com>male,nam<person))
obj(eat(icl>consume>do,agt>living_thing,obj>concrete_
thing,ins>thing).@entry.@present,apple(icl>edible_
fruit>thing).@indef)
{/unl}
```

### 2.2. Issues Related in Conversion of Bangla Texts to UNL

The important matters involved in conversion from Bangla to UNL are development of Bangla Word Dictionary based on the UNL format, creation of UNL attributes and

resolution of UNL relations. In the conversion process, we first search the lexicon/ morpheme to have a map between Bangla Words and UWs so that to have the root of word to establish the relevant relations and attributes. After finding the root words, we create the relevant attributes that will describe the subjective information of the texts. These attributes are to be generated based on the morphology of the Bangla language as per (Uchida et al., 1993). The binary relations between UWs are determined based on the different rules they play. A number of factors are to be considered in resolving a relation. The case maker of a language plays a vital role in the resolution of relations among UWs shown in (Dhanabalan et al., 2002).

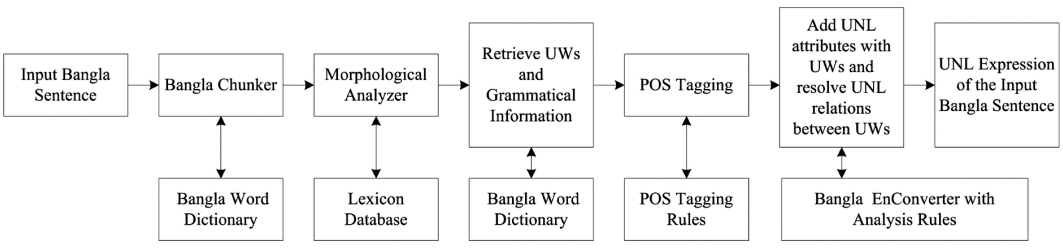
## 3. ARCHITECTURE OF BANGLA ENCONVERTER

The architecture of Bangla-UNL EnConverter is given in Figure 3. In this figure, the first phase of the module, the Bangla input sentence is divided into morphemes. The morphemes are automatically generated on the basis of Bangla Word Dictionary. Then they are processed in Bangla morphological analyzer to obtain the root and grammatical information of each morpheme in the form of tag set The UWs for each morpheme are retrieved from Bangla Word Dictionary. Some morphemes may return more than one tag set To resolve the ambiguity among these morphemes, Parts-of-Speech (POS) tagging is performed similar to (Uchida et al., 2002). The UNL relation resolution and attribute generation rule base is used to fire the appropriate rule according to the attributes of the word in the Bangla Word Dictionary identical to the work of (Kumar el. al., 2011). After applying the rules, UNL expressions is generated for the input Bangla sentence.

## 4. ALGORITHM FOR PROCESSING INPUT BANGLA SENTENCE

Algorithm for processing of input Bangla sentence involves the following steps: (a) Accept the input Bangla sentence and (ii) Dividing the input sentence into the morphemes with the help of Bangla Word Dictionary. The UWs from Bangla Word Dictionary that contain the target set of consecutive words from input sentence are considered as morphemes. An example of dividing the sentence into morphemes is described below: **Transliterated text:** Ami conference e ongshogrohon korar jonno Chin giechhilam. **Equivalent English text:** I went to China to attend a conference. This Bangla sentence will be divided into morphemes as:

**Figure 3. Architecture of Bangla-Universal Networking Language (UNL) EnConverter**



(Ami) (conference e) (ongshogrohon korar jonno) (Chin) (gie Chhilam)
(I) (a conference) (to attend) (China) (went to)

1. Process each morpheme obtained from step (ii) using Bangla Mophological Analyzer to get root and suffix for each morpheme. For example 'giechhilam' has the root word as 'gi' and suffix 'echhilam'.
2. Replace each of the morpheme in the input sentence with the root and suffix obtained from step. For example, now the final morphemes from the input sentence in step (ii) will be:

(Ami) (conference e) (ongshogrohon korar jonno) (Chin) (gi echhilam)
(I) (conference e) (to attend) (China) (went)

3. Construct the string from the modified morphemes The string for the given sentence will become,
4. Ami conference e ongshogrohon korar jonno Chin giechhilam.
5. Search the morphemes/case makers or verb ending suffix for the words that are available in Bangla Word Dictionary. For example, 'echhilam' is considered as morpheme as this has an entry in verb ending suffix.
6. Retrieve UWs and grammatical attributes corresponding to each morpheme from Bangla Word Dictionary.
7. Create a linked list for the input sentence. This linked list shall have nodes as morphemes of the input sentence with the attributes like Bangla word and UW.

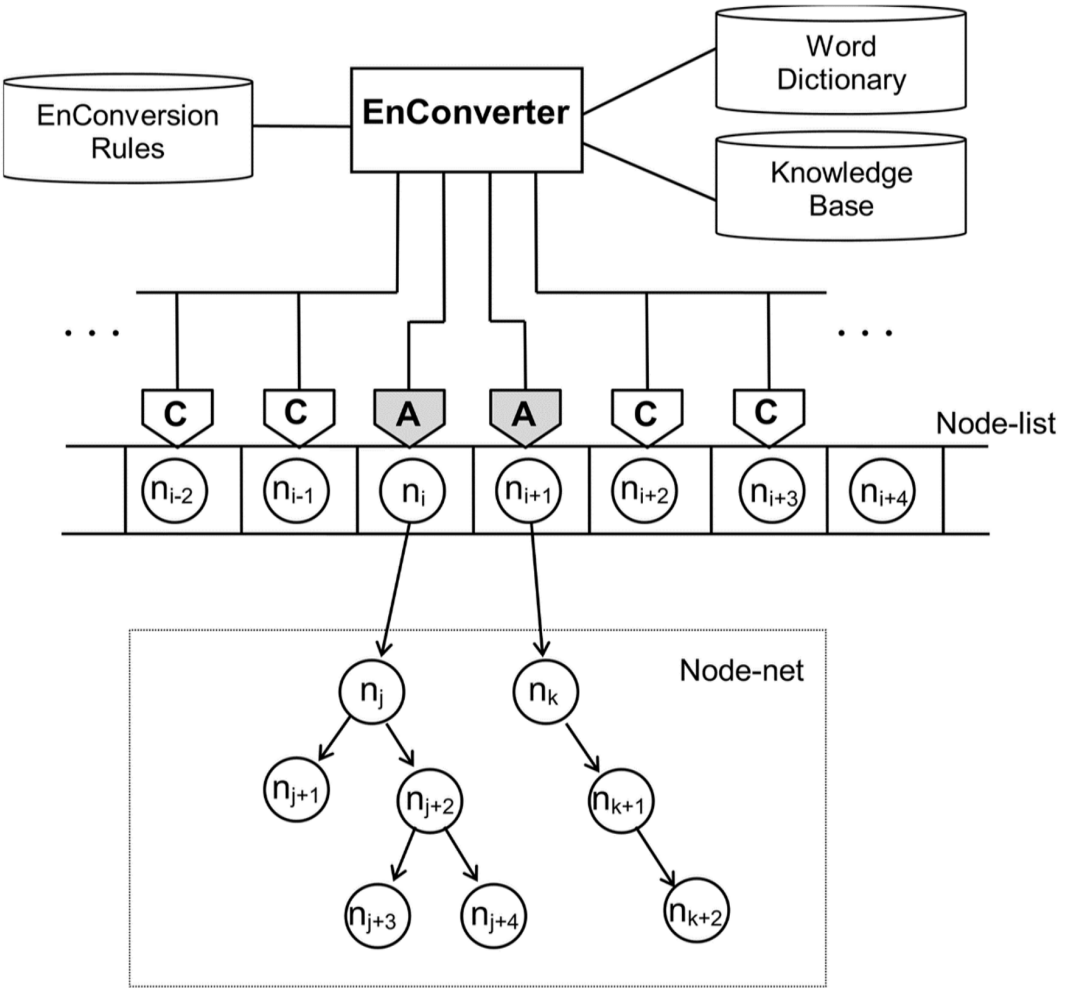
## 5. BANGLA ENCONVERTER

The Bangla EnConverter (EnCo) processes the given input sentence from left to right. When an input string is scanned, all matched morphemes with the same starting characters are retrieved from the dictionary and become the candidate morphemes according to the priority rule in order to build a syntactic tree and the semantic network for the sentence. It moves back and forth over the Node List, which contains words of the input sentence discussed in (Uchida et al., 1999; Dhanabalan et al., 2002; Uchida et al., 2002; Manoj et al., 2009).

In Figure 4, "A" indicates an *Analysis Window*, "C" indicates a *Conditional Window*, and *nn* indicates an *Analysis Node*. The *EnCo* uses the Conditional Windows for checking the neighboring nodes on both sides of the *Analysis Windows* in order to check whether the neighboring nodes satisfy the conditions to apply an *Analysis Rule*. The *Analysis Windows* are used to check two adjacent nodes in order to apply one of the Analysis Rules. If there is an applicable rule, *EnCo* adds lexical attribute to or deletes lexical attributes from these nodes, and creates a partial syntactic tree and UNL network according to the type of rule.

*EnConversion* rules are created on the basis of case makers and morphology of the Bangla language. These rules are designed according to the format provided by the UNL system of the UNDL Foundation (Ali et al., 2008). An *EnConversion* rule describes rule application conditions, a method to rewrite the attribute of node that satisfies the application condition, and construction methods of syntax tree. While applying rules, the *EnConverter* analyzes morphemes, syntax and semantics. Finally,

**Figure 4. Structure of Bangla EnConverter**




it generates a syntax tree and a semantic network. The description format of the analysis rules is as follows.

```
{<COND1>:<ACTION1>:<RELATION1>}{<COND2>: <ACTION2>:
<RELATION2>}
```

Where,

- <COND1> indicates Condition 1; it contains the lexicon attributes of the Left Analysis Window (LAW).
- <COND2> indicates Condition 2; it contains the lexicon attributes of the Right Analysis Window (RAW).
- <ACTION1> and <ACTION2> are used to indicate the actions performed if the corresponding conditions are true.
- <RELATION1> denotes the possible semantic relation of the RAW to LAW.
- <RELATION2> denotes the possible semantic relation of the LAW to RAW.

## 6. ALGORITHM FOR RELATION RESOLUTION AND GENERATION OF ATTRIBUTES

According to the algorithm explained in Section 4 and discussion from Section 5, a node-net is created from the input Bangla sentence. Then Bangla *EnConverter* system invokes the following algorithm for UNL relation resolution and generation of attributes.

1. Process each node of the node-list considering the first node as LAW and next node as RAW.
2. Search the required rule from *EnConversion Rules* file depending on the attributes of the left and right analysis windows.
3. Modify the node-list to resolve the UNL relations and generate UNL attributes according to the fired rule. If no rule is fired, go to step (v).
4. Consider the first node of modified node-list as LAW and next node as RAW. Go to step (ii) with new analysis windows. If the modified node-list contain only single node, consider that node as 'entry node' and stop further processing. It indicates that all the nodes are successfully processed by the system.
5. If no rule is fired in step (ii), shift the windows of *Enconverter* one step right and hence the right analysis node will become the left analysis window and next node will become the right analysis window.
6. Go to step (ii) with new analysis windows.

## 7. ENCONVERSION OF A BANGLA SENTENCE TO UNL EXPRESSION

The encoding process will be performed by the shift/reduce parsing (Earley, 1970). Both Bangla and Hindi (Indian Language) are derived from Sanskrit language. Hence, Bangla and Hindi languages have a lot of syntactic similarities. The conversion of UNL from Hindi language has been shown in (Bhattacharyya, 2001). Due to the syntactic similarities, this paper follows a similar approach as for Hindi language. In order to explain the encoding steps, we give an example of the analysis of a simple Bangla assertive sentence. Assertive simple sentences have only one main clause. We assume that analysis rules and the Bangla Word Dictionary for UNL are given to the analyzer system *EnCo*. Consider the following Bangla sentence.

Transliterated sentence: Dhakaye aaj khub gorom.

Equivalent English sentence: It is very hot in Dhaka today.

The input Bangla sentence is processed according to the algorithm discussed in Section 4. The chunks obtained from the input sentence are given below.

```
(Dhaka)   (ye)      (aaj)          (khub)     (groom).
```

There are five nodes in the input sentence are given below.

Node 1. Dhaka: Universal Word: "Dhaka(iof>place)", Dictionary attributes: (N, PLACE, CAP)

Node 2. ye: Universal Word: null, Dictionary attributes: (CASE)

Node 3. aaj:Universal Word: "today(icl>period, equ>nowadays)", Dictionary attributes: (N, TIM)

Node 4. khub: Universal Word: "very(icl>how, equ>extremely)", Dictionary attributes: (ADV)
Node 5. goroom: Universal Word: "hot(icl>state, ant>cold)", Dictionary attributes: (ADJ)

Here, N indicates noun, PLC indicates place, CAP for capital city, CASE is used for case maker in Bangla language, TIM represents time and ADJ represents adjective and NULL represents no dictionary attributes respectively. The analysis window is enclosed within "[" and "]".

```
[Dhaka]  [ye] [aaj] [khub] [goroom]
```

These nodes are processed by the algorithm given in Section 6. UNL relation resolution and generation of attributes are explained in the following steps.

Step 1: Initially, the system has following left and right analysis windows for the given input sentence.
Left node is: [Dhaka]: Universal Word: "Dhaka(iof>place)", Dictionary attributes: (N, PLACE, CAP)
Right node is: [ye]: Universal Word: null, Dictionary attributes: (CASE)

The analysis rule fired between left and right analysis window is given below.

```
+ {N, PLACE, CAP: null: null}{[ye]:@: null}
```

The above rule is preceded by "+" sign; this means that it is a left composition rule. It causes into concatenation of right node to the left node as a single composite node and the attributes of left node are inherited for further processing. If the operator "@" appears in the <ACTION> field of the rule for the left node, the attributes of the right node are also inherited. The linked list after application of above rule becomes,

```
 [dhakaye]       [aaj]       [khub]          [gorom]
```

Step 2: Now 'Dhakaye' becomes the left analysis window and 'aaj' becomes the right analysis window. In this step, the system right shifts twice to place adverb 'khub' on the LAW and predicate 'gorom' on the RAW, because there is no combination or modification rule between the nouns on the analysis windows.
Step 3: The system recognizes 'hot' as a predicate of the sentence. So, it generates *man* relation between the adverb 'khub' means *very* and the predicate 'goroom' means *hot*. The rule fired between right and left analysis windows is given below.

```
>{ADV:null:man}{ADJ:@:null}
```

The above rule is preceded by the ">" sign, indicates that it is a right modification rule. This rule is applicable, when left node modifies the right node. It deletes the left node from the node-list, while the right node remains in the node-list. It results the resolution of *man* relation between two nodes as shown below.

```
man(hot(icl>state, ant>cold), very(icl>how,
equ>extremely))
```

The linked list after the left node deletion becomes,

```
 [dhakaye]       [aaj]         [goroom]
```

**Table 1. Comparative analysis of UNL expressions generated by Bangla EnConverter Part I**

| TBS | EES | RR | RF | UEGBE | UEGRELS |
|---|---|---|---|---|---|
| Ami vat khai. | I eat rice. | obj, agt | +{ROOT,VEND,^ALT,^VERB:+VERB,ROOT:+@: }{VI,VEND:null:null:}<br>>{N::obj:}{VERB,#OBJ:null:null:}<br>>{HPRON,SUBJ::agt:}<br>{VERB,#AGT:null:null:} | agt(eat(icl>consume>do).@entry.@present,i(icl>person))<br>obj(eat(icl>consume>do,.@pentry.@present,rice(icl>grain>thing)) | {unl}agt(eat(icl>consume>do,<br>agt>living_thing,obj><br>concrete_thing,ins>thing).@entry.@present,i(icl>person))<br>obj(eat(icl>consume> do,agt>living_thing,obj><br>concrete_thing,ins>thing).@entry.@present,rice(icl>grain>thing))<br>{/unl} |
| Hasan bishobiddaloye jachhe. | Hasan is going to university. | agt, plt | +{ROOT,VEND,^ALT,^VERB:+VERB,-ROOT: +@:}{VI,VEND:null:null:}<br>+{N,CEND:+@, INF,+NP: null:}<br>{[ee]:null:null:}<br>>{NP::plt:}{VERB,#OBJ:null:null:}<br>>{N,SUBJ::agt:}{VERB,#AGT:null:null:} | plt(go(icl>move>do).@entry.@present.@progress,university(icl>body>thing))<br>agt(go(icl>move>do).@entry.@present.@progress,hasan)) | {unl}<br>agt(go(icl>move>do,plt>place,plf>place,agt>thing).@entry.@present.@progress,hasan))<br>plt(go(icl>move>do,plt>place,plf>place,agt>thing).@entry.@present.@progress,university(icl>body>thing))<br>{/unl} |
| John likhchhe. | John is writing. | agt | +{ROOT,CEND,^ALT,^VERB:+VERB,-ROOT: +@:}{VI,VEND:null:null:}<br>>{N,SUBJ::agt:}{VERB,#AGT:null:null:} | agt(write(icl>communicate>do).@entry.@present.@progress,john(icl>name>abstract_thing,nam<person)) | {unl}<br>agt(write(icl>communicate>do,agt>person,obj>information,cao>thing,ins>thing,rec>person).@entry.@present.@progress,john(icl>name>abstract_thing,com>male,nam<person))<br>{/unl} |

*In Table 1 to 3 we considered TBS: Transliterated Bangla Sentence, EES: Equivalent English Sentence, RR: Relations Resolved, RF: Rules Fired, UEGBE: UNL expressions Generated by the Bangla EnConverter, UEGRELS: UNL expressions Generated by the Russian and English Language Server.

**Table 2. Comparative analysis of UNL expressions generated by Bangla EnConverter Part II**

| TBS | EES | RR | RF | UEGBE | UEGRELS |
|---|---|---|---|---|---|
| Amra Japan giechhilam. | We went to Japan. | agt, plt | +{ROOT,VEND,^ALT,^VERB:+VERB,ROOT:+@: }{VI,VEND:null:null:}<br>>{N::plt:}{VERB,#OBJ:null:null:}<br>>{HPRON,SUBJ::agt:}<br>{VERB,#AGT:null:null:} | plt(go(icl>move>do.@entry.@past,japan(iof>archipelago>thing))<br>agt(go(icl>move>do).@entry.@past,we(icl>group).@pl) | {unl}<br>agt(go(icl>move>do,plt>place,plf>place,agt>thing).@entry.@past,we(icl>group).@pl)<br>plt(go(icl>move>do,plt>place,plf>place,agt>thing).@entry.@past,japan(iof>archipelago>thing))<br>{/unl} |
| Kamal kashemer sathe hatche. | Kamal walks with Kashem. | agt, ptn | +{ROOT,CEND,^ALT,^VERB:+VERB,-ROOT:+@:}{VI,CEND:null:null:}<br>+{N,CEND:+@,+NP:null:}{[er]:null:null:}<br>>{N::ptn:}{VERB,#PTN:null:null:}<br>>{N,SUBJ::agt:}{VERB,#AGT:null:null:} | ptn(walk(icl>do).@entry.@present,kashem))<br>agt(walk(icl>do).@entry.@present,kamal(icl>name>abstract_thing,nam<person)) | {unl}<br>agt(walk(icl>do,equ>stroll,agt>person).@entry.@present,kamal(icl>name>abstract_thing,com>male,nam<person))<br>ptn(walk(icl>do,equ>stroll,agt>person).@entry.@present,kashem))<br>{/unl} |
| Mery protidin cha pan kore. | Mery takes tea everyday. | agt, mod, obj | >{N,VEND:+NP,-N:obj:}{VERB,#OBJ:null:null:}<br>>{ADV, TIME:: mod:}{N,VEND: +NP,-N:null:}<br>>{N,VEND::agt:}{VERB,#AGT:null:null:} | obj(take(icl>consume>do).@entry.@present,tea(icl>beverage>thing))<br>mod(tea(icl>beverage>thing),everyday(icl>adj)<br>agt(take(icl>consume>do).@entry.@present,mery) | {unl}<br>agt(take(icl>consume>do,agt>living_thing,obj>concrete_thing).@entry.@present,mery)<br>mod(tea(icl>beverage>thing),everyday(icl>adj,equ>casual))<br>obj(take(icl>consume>do,agt>living_thing,obj>concrete_thing).@entry.@present,tea(icl>beverage>thing))<br>{/unl} |

**Table 3. Comparative analysis of UNL expressions generated by Bangla EnConverter Part III**

| TBS | EES | RR | RF | UEGBE | UEGRELS |
|---|---|---|---|---|---|
| Amra football khelbo. | We will play football. | agt, man, | +{ROOT,CEND,^ALT,^VERB:+VERB,-ROOT:+@:}{VI,CEND:null:null:}<br>>{N::obj:}{VERB,#OBJ:null:null:}<br>>{HPRON,SUBJ::agt:}<br>{VERB,#AGT:null:null:} | man(play(icl>compete>do).@entry.@future,football(icl>field_game>thing))<br>agt(play(icl>compete>do).@entry.@future,we(icl>group).@pl) | {unl}<br>agt(play(icl>compete>do,agt>thing,obj>uw,ptn>thing).@entry.@future,we(icl>group).@pl)<br>man(play(icl>compete>do,agt>thing,obj>uw,ptn>thing).@entry.@future,football(icl>field_game>thing))<br>{/unl} |

Here, right node remains in the right analysis window and noun 'aaj' (*today*) becomes on the left analysis window.

Step 4: Then the analyzer looks ahead further right beyond the noun phrase 'aaj' (*today*) to get the predicate 'goroom' (*hot*). A *tim* relation is created between 'aaj' (*today*), and 'goroom' (*hot*) and finally, 'aaj' (today) is deleted as discussed in previous rule.

The rule fired between right and left analysis windows is given below.

```
>{N,TIM::+@present:tim}{ADJ:@:null}
```

The resolution of *tim* relation between two nodes as shown below.

```
tim(hot(icl>state, ant>cold), today(icl>period,
equ>nowadays))
```

The linked list after the left node deletion becomes,

```
[dhakaye]    [gorom]
```

Step 5: By using *PLACE* attribute of the noun '*Dhaka'*, the system generates *plc* relation between the noun 'Dhaka' and the predicate 'goroom' and hence noun 'Dhaka is deleted.

The rule fired between right and left analysis windows is given below.

```
>{N,PLACE:null:plc}{ADJ:@:null}
```

The resolution of *plc* relation between two nodes as shown below.

```
plc(Dhaka(iof>place), hot(icl>state, ant>cold))
```

Finally, the modified linked list contains only one node as given below.

```
[goroom]
(gorom)
```

Being the last node, this is considered as root node and ".@entry" attribute is added to its UW as given below.

```
hot(icl>adj,ant>cold).@entry.@present.
```

Step 6: The UNL expression generated by Bangla EnConverter for input Bangla sentence is as follows.

Transliterated sentence: Dhakaye aaj khub gorom.
Equivalent English sentence: It is very hot in Dhaka today.
UNL expression generated by the system:

```
{unl}
man (hot (icl>state).@entry.@present, very (intensifier))
tim(hot(icl>state).@entry.@present, today(icl>period))
plc (hot (icl>state).@entry.@present, Dhaka (icl>place))
{/unl}
```

## 8. EXPERIMENTATION AND TESTING OF THE SYSTEM

We have tested our system on several Bangla sentences. It has been observed that the system successfully handles the resolution of UNL relation and generation of attributes for these sentences. The system has been tested with the help of English sentences available at Russian and English language server (UNL, 2017). We have manually translated the given English sentences at Russian and English Language server into equivalent Bangla sentences and then inputted those equivalent Bangla sentences to the designed Bangla-UNL EnConverter system. We have compared the UNL expressions generated by our system with the expression generated by Russian and English language server. This comparative analysis is given in Table 1 for seven sentences. We have tested the UNL expressions generated by our system with UNL parser and viewer available at Russian and English Language Server. All the seven sentences given in Tables 1 to 3 have successfully been parsed and viewed with the help of Russian English Language Server available at (UNL, 2017).

## 9. CONCLUSION AND FUTURE WORKS

In this paper, the Bangla *EnConverter* has been developed, implemented and tested. Bangla *EnConverter* uses the analysis rules for UNL relation and resolution and generation of attributes from the Bangla sentence. We have developed approximately five hundred analysis rules and three thousand Universal Words based on UNL format for the development of Bangla EnConverter. This *EnConverter* has been tested for its performance. The result of testing of our system and the Russian and English Language Server are in a very good agreement. Currently, the Bangla *EnConverter* is working well for conversion of any simple Bangla assertive sentence into UNL expressions. This will enhance the usage of Internet for Bangladeshi Community as they will be able to get need based information which are available in Internet in different languages. The conversion of complex, compound and interrogative sentences are yet to be completed. It is essential to check the effectiveness of the interlingua approach that has been used here, i.e., determining whether the meaning of the source language is being conveyed properly, so that a native speaker of the target language accepts it as natural.

## REFERENCES

Ali, M., & Ali, M. M. (2002). Development of Machine Translation Dictionaries for Bangla Language. In *Proceedings of Fifth International Conference on Computer and Information Technology*, Dhaka, Bangladesh.

Ali, M. N. Y., Das, J. K., Al-Mamun, S. M. A., & Choudhury, M. E. H. (2008). Specific Features of a Converter of Web Documents from Bengali to Universal Networking Language. In *Proceedings of 2008 International Conference on Computer, and Communication Engineering*, Dhaka, Bangladesh: IEEE. doi:10.1109/ICCCE.2008.4580700

Ali, M. N. Y., Das, J. K., Al-Mamun, S. M. A., & Nurannabi, A. M. (2008). Morphological Analyais of Bangla Words for Universal Networking Language. In *Proceedings of Third International Conference on Digital Information Management*. London, England: IEEE. doi:10.1109/ICDIM.2008.4746734

Ali, M. N. Y., Ripon, S. H., & Allayear, S. M. (2012). UNL Based Bangla Natural Text Conversion- Predicate Preserving Parser Approach. *International Journal of Computer Science Issues*, *9*(8(6)), 259–265.

Ali, M. N. Y., Shamsujjoha, M., Sorwar, G., & Ripon, S. H. (2017). Construction of Word Dictionary for Bangla Vowel Ended Roots and Its Verbal Inflexions in UNL Based Machine Translation. *Transactions on Machine Learning and Artificial Intelligence*, *5*(1), 1–25.

Ali, N. Y., & Ripon, S. H. (2013). UNL-Based Bangla Machine Translation Framework. In M. Karim, M. Kaykobad, & M. Murshed (Eds.), *Technical Challenges and Design Issues in Bangla Language Processing* (pp. 35–78). Hershey, PA: IGI Global. doi:10.4018/978-1-4666-3970-6.ch003

Asaduzzaman, M. (2008). Bangla Shdhito Prottoy: Punorbuchar. *The Dhaka University Journal of Linguistic*, *1*(1), 79–92.

Asaduzzaman, M. M., & Ali, M. M. (2003). A Knowledge Based Approach to Bangla-English Machine Translation for Simple Assertive Sentences. *International Journal of Translation*, *15*, 77–97.

Asaduzzaman, M. M., & Ali, M. M. (2003). Transfer Machine Translation - An Experience with Bangla English Machine Translation System. In *Proceedings of 2003 International Conference on Computer and Information Technology*. Dhaka, Bangladesh: IEEE.

Azad, H. (1994). Bakkotottoy.

Bhattacharyya, P. (2001). Multilingual Information Processing Through Universal Networking Language. In *Proceedings of the Indo UK Workshop on Language Engineering for South Asian Languages*.

Choudhury, M. E. H., Ali, M. N. Y., Sarkar, M. Z. H., & Razib, M. A. (2005). Bridging Bangla to Universal Networking Language- A Human Language Neutral Meta- Language. In *Proceedings of the Fifth International Conference on Computer and Information Technology*, Dhaka, Bangladesh.

Dhanabalan, K., & Saravanan, T. V. & Geetha (2002). Tamil to UNL EnConverter. In *Proceedings of of the International Conference on Universal Knowledge and Language*, Chennai, India.

Earley, J. (1970). An Efficient Context Free Parsing Algorithm. *Communications of the ACM*, *13*(2), 94–102. doi:10.1145/362007.362035

Islam, M. S. (2009). Research on Bangla Language Processing in Bangladesh: Progress and Challenges. In *Proceedings of Eight International Language and Development Conference*, Dhaka, Bangladesh.

Khairunnahar, K. (2008). Morphological Analysis of Bangla Prefix. *The Dhaka University Journal of Linguistic*, *1*(2), 157–168.

Kumar, P., & Sharma, R. K. (2011). Generation of UNL Attributes and Resolving Relations for Pubjabi EnConverter. *Malaysian Journal of Computer Science*, *24*(1), 34–46.

Manoj, J., & Damani, O. P. (2009). English to UNL (Interlingerua) Enconversion. In *Proc. of the Second Conference on Language and Technology*. Bombay, India.

Rameswar, D. S. (1996). *Shadharan Vasha Biggan and Bangla Vasha* (pp. 358–377). PA: Pustok Biponi Prokashoni.

Saha, G. K. (2005). The E2B Machine Translation: A New Approach to HLT. *Ubiquity Archive, Association of Computing Machineries, 6*(32).

Shahidullah, D. M. (2003). *Bangala Vyakaran*. PA: Maola Brothers Prokashoni.

Shniti, K. D. C. (1999). *Vasha-Prokash Bangla Vyakaran*. PA: Rupa and Company Prokashoni.

United Nations University. (1998). *The Universal Networking Language (UNL) Specification Version 3.0*.

Uchida, H., & Zhu, M. (1993). Interlingua for Multilingual Machine Translation. In Proceedings of Machine Translation Summit IV, Kobe, Japan.

Uchida, H., & Zhu, M. (2002). The Universal Networking Language beyond Machine Translation. In *Proceedings of the Third International Conference on Languages Resources and Evaluation*, Seoul, Korea.

Uchida, H., Zhu, M., & Senta, T. C. D. (1999). *A Gift for a Millennium*. Tokyo, Japan: The United Nations University.

Uchida H., Zhu M. & Senta T. C. D. (2006). Universal Networking Language. UNDL Foundation, Geneva, Switzerland.

Uddin, M. G., Ashraf, H., Kamal, A. H. M., & Ali, M. M. (2004). New Parameters for Bangla to English Statistical Machine Translation. In *Proceedings of the Third International Conference on Electrical & Computer Engineering*, Dhaka, Bangladesh.

UNL. (2017). Retrieved from http://www.unl.ru

*Md. Nawab Yousuf Ali was born in Rangpur, Bangladesh on April 05, 1966. He passed Secondary School Certificate from Rangpur High School, Rangpur, Rajshahi, Bangladesh in 1982 and Higher Secondary School Certificate from Carmichael College, Rangpur, Rajshahi, Bangladesh in 1984. He did his Master in Computer Science and Engineering from Lvov Polytechnic Institute, Lvov, Ukraine, USSR in1992 and obtained PhD in CSE from the Department of CSE, Jahangirnagar University, Dhaka, Bangladesh in 2012. He is currently serving as an Associate Professor in the Department of CSE, East West University, Dhaka, Bangladesh. His research interests include NLP, Universal Networking Language, Bangla text conversion to UNL. He has published one book, three book chapters, sixteen journals and 13 conference papers in international journals and conferences.*

*Md.Sarwar Kamal has completed his B.Sc in Computer Science and Engineering from University of Chittagong. He also completed M.Sc Engineering from Chittagong University of Engineering and Technology. Currently he is working as a Senior Lecturer at Computer Science and Engineering in East West University Bangladesh. His research interests include Bioinformatics and Big Data mining.*

*Md. Shamsujjoha is currently working as a Senior Lecturer at the Department of Computer Science and Engineering, East West University (EWU), Dhaka-1212, Bangladesh. In addition, he is also serving the EWU as an Assistant Proctor since Nov. 1, 2014. He received his B.Sc. and M.S degrees in Computer Science and Engineering from University of Dhaka, Dhaka-1000, Bangladesh. He has several research publications in the renowned international journal and conferences. His research interests include VLSI and Embedded System Design, Quantum/Reversible Logic Synthesis, Image Processing, Natural Language Processing, Low Power and Nano-Computing.*

*M. Ameer Ali has join as a Professor in the Department of Computer Science and Engineering, Bangladesh University of Business and Technology (BUBT) on May 02, 2017. He is also working as a Chairman of the same department. He completed his B. Sc. in Computer Science and Engineering in 2001 from Bangladesh University of Engineering and Technology (BUET), Bangladesh and PhD in IT in 2006 from Monash University, Australia. Before joining here, he was Visiting Research Fellow of Monash University, Australia for developing smart card for Australian Citizens and was working as an Associate Professor and Chairperson of the Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh. He has 42 published articles including a book, book chapters, international refereed journals, LNCS and international refereed conferences. His research interests are image processing, segmentation, fuzzy clustering, telemedicine, vendor selection, e-governance and corporate governance. He is also member of IEEE and IEB, Bangladesh and reviewer of several Journals and Conferences.*

*Mr. Ghulam Farooque Ahmed is the CEO of Computer Village Ltd. He is also the owner of Bangla Lekhani software. This software has been used in the several national newspapers of Bangladesh. He has more than 10 publications in the renowned journals and conferences. His research interest includes Bangla Fonts development, Natural Language Processing, Machine Intelligence Improvement etc.*