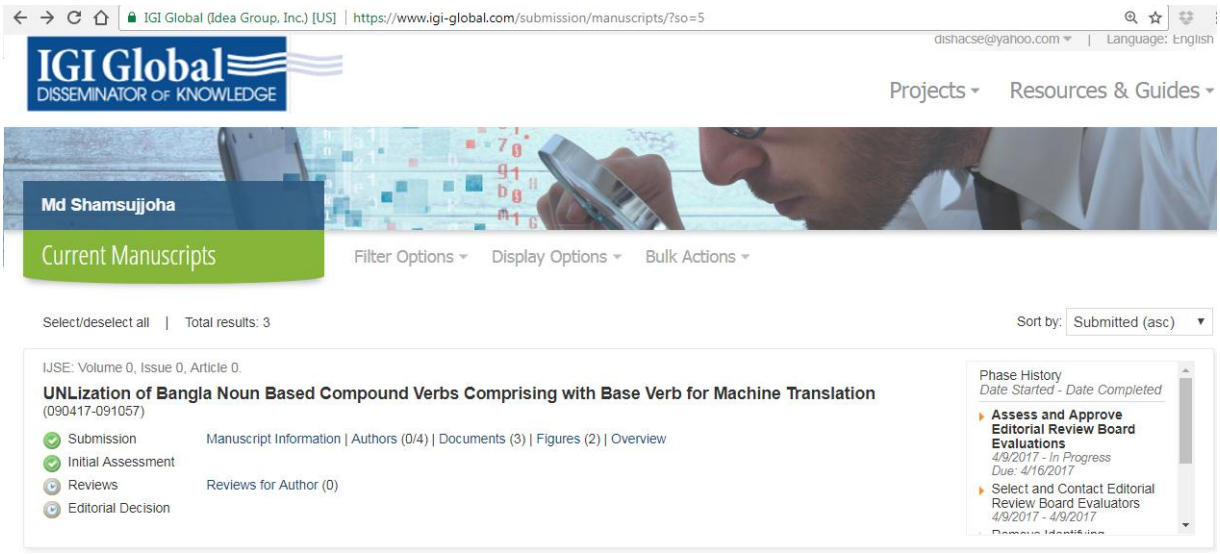
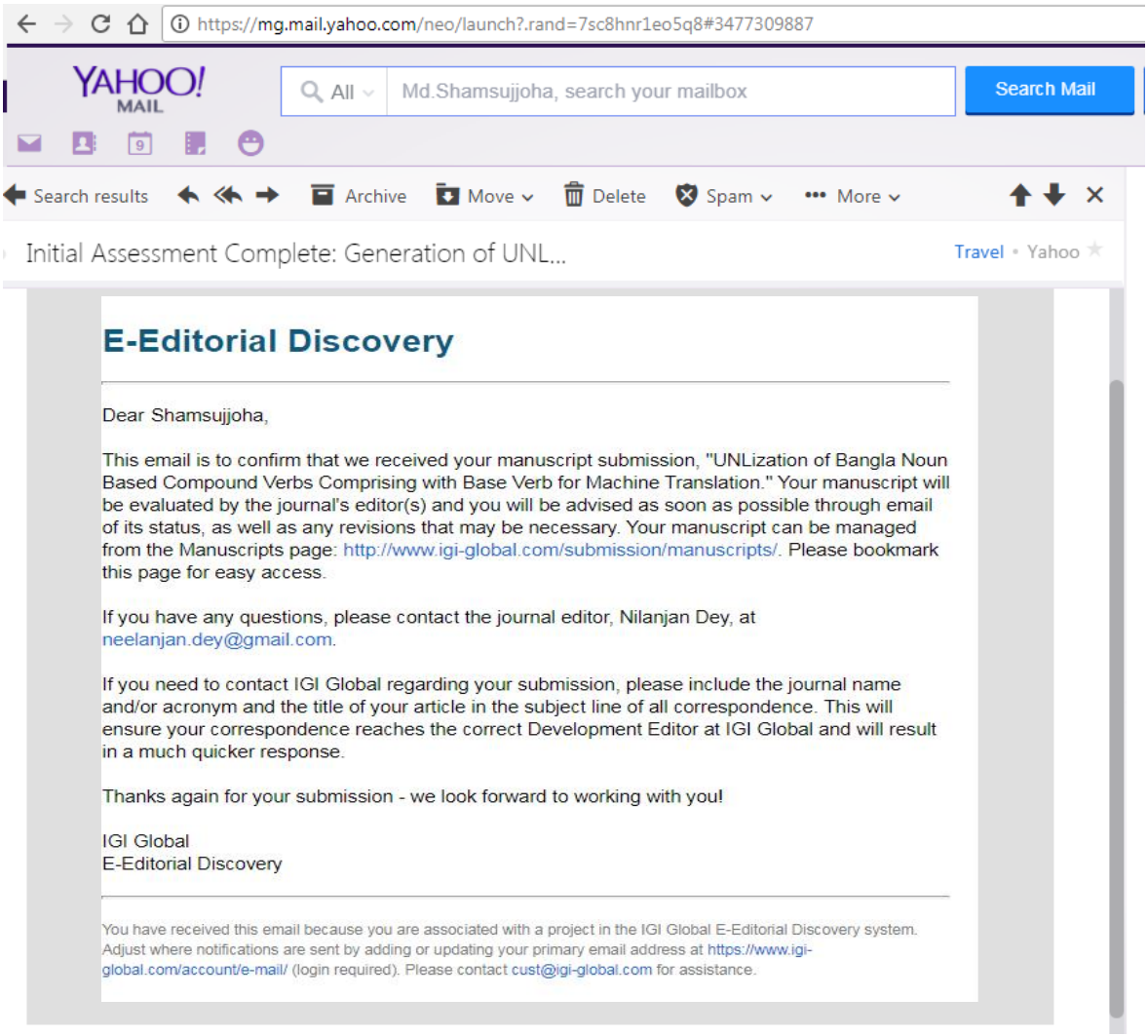

IGI Global (Idea Group, Inc.) [US] | https://www.igi-global.com/submission/manuscripts/?so=5
dishacse@yahoo.com | Language: English
IGI Global
DISSEMINATOR OF KNOWLEDGE
Projects
Resources & Guides
Md Shamsujjoha
Current Manuscripts
Filter Options
Display Options
Bulk Actions
Select/deselect all | Total results: 3
Sort by: Submitted (asc)
IJSE: Volume 0, Issue 0, Article 0.
UNLization of Bangla Noun Based Compound Verbs Comprising with Base Verb for Machine Translation
(090417-091057)
Submission
Initial Assessment
Reviews
Editorial Decision
Manuscript Information | Authors (0/4) | Documents (3) | Figures (2) | Overview
Reviews for Author (0)
Phase History
Date Started - Date Completed
Assess and Approve Editorial Review Board Evaluations
4/9/2017 - In Progress
Due: 4/16/2017
Select and Contact Editorial Review Board Evaluators
4/9/2017 - 4/9/2017


https://mg.mail.yahoo.com/neo/launch?.rand=7sc8hnr1eo5q8#3477309887
YAHOO! MAIL
All
Md.Shamsujjoha, search your mailbox
Search Mail
Search results
Archive
Move
Delete
Spam
More
Initial Assessment Complete: Generation of UNL...
Travel • Yahoo
E-Editorial Discovery
Dear Shamsujjoha,
This email is to confirm that we received your manuscript submission, "UNLization of Bangla Noun Based Compound Verbs Comprising with Base Verb for Machine Translation." Your manuscript will be evaluated by the journal's editor(s) and you will be advised as soon as possible through email of its status, as well as any revisions that may be necessary. Your manuscript can be managed from the Manuscripts page: http://www.igi-global.com/submission/manuscripts/. Please bookmark this page for easy access.
If you have any questions, please contact the journal editor, Nilanjan Dey, at neelanjan.dey@gmail.com.
If you need to contact IGI Global regarding your submission, please include the journal name and/or acronym and the title of your article in the subject line of all correspondence. This will ensure your correspondence reaches the correct Development Editor at IGI Global and will result in a much quicker response.
Thanks again for your submission - we look forward to working with you!
IGI Global
E-Editorial Discovery
You have received this email because you are associated with a project in the IGI Global E-Editorial Discovery system. Adjust where notifications are sent by adding or updating your primary email address at https://www.igi-global.com/account/e-mail/ (login required). Please contact cust@igi-global.com for assistance.

UNLization of Bangla Noun Based Compound Verbs Comprising with Base Verb for Machine Translation

Md. Nawab Yousuf Ali [1], Md. Shamsujjoha [2], Golam Sorwar [3], Md. Sarwar Kamal [4]

[1,2,4] Computer Science & Engineering Dept., East West University, Dhaka, Bangladesh

[3] School of Business & Tourism, Southern Cross University, Australia

[1] nawab@ewubd.edu, [2] msj@ewubd.edu, [3] golam.sorwar@scu.edu.au, [4] skamal@mail.ewubd.edu

Abstract

This paper presents the development of format for dictionary entries of noun, base verb root and their inflexions to incorporate them into an interlingua representation called Universal Networking Language (UNL) processors. Combination of base verb roots and verbal inflexions forms base verbs and hence nouns combine with base verbs produce compound verb. It focuses to bridge the gaps between nouns and base verb roots in the framework of UNL system aiming to produce a Bangla word dictionary for UNL. The paper analyzes the construction of compound verbs in a variety of approaches based on their structure and Bangla grammar. The paper also shows the conversion procedures of a Bangla sentence with compound verb into UNL expressions using the analysis rules which can then be converted into other native languages using the language specific generation rules.

Keywords: Bangla Compound Verb, Compound Verb roots, Verbal Inflexion, Bangla-UNL Dictionary, Universal Networking Language, Universal Words

1. INTRODUCTION

The Universal Networking Language (UNL) [1] is a computer language that enables computers to process information and knowledge across the language barriers. It is an artificial language that replicates the function of natural languages in human communication. The mission of the UNL project is to allow people across nations to access information in the Internet in their own languages [2]. The UNL project is concerned in developing an intermediary language system whereby any written text can be converted to different languages through UNL. Simultaneously, text written in different languages can be converted to that particular language [3]. Knowledge and information in different languages are scattered all over the world and remain inaccessible to mostly due to non-machine representation and language barrier [4]. To deal with the language barrier, United Nations University/Institute of Advanced Studies (UNU/IAS) conducted a review of all internationally available machine translation programs and started to devise an efficient and workable technique to develop a human language neutral meta-language for Internet. The result of the project is Universal Networking Language (UNL) [1]. The aim of this internationally cooperative initiative is to eliminate the massive requirement of translation among languages and reduce language to language translation to one time conversion to UNL. In UNL system, a native language sentence is converted into a UNL hypergraph by a tool called "Enconverter" [5] following analysis rules defined in [6]. These hypergraphs are then translated into any native language, using generation rules defined in [6], by another tool called "Deconverter" [7]. The development of language specific components, such as dictionary, analysis rules and generation rules used by Enconverter and Deconverter, are the research focus across the world.

The people in Bangladesh and two states (West Bengal and Tripura) in India use Bangla as their first language. About one sixth population of the world is speaking in Bangla. Exchange of information and sharing of knowledge globally, it is critically important to devise conversion technique(s) for Bangla language texts into UNL and vice versa. Machine translation (MT) is an approach to translating texts from one natural language to another automatically. Ali and Ali (2002) attempted to develop MT

Bangla dictionaries that address the organization, contents and details of the information [8]. Saha (2005) developed 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 [9]. Another attempt by Uddin et al. (2004) was to develop a statistical Bangla to English translation engine using only simple Bangla sentences that contains a subject, an object and a verb [10].

As per the above, some research has been conducted in the development of Bangla MT in the field of automatic translation, parsing and syntax analysis to develop software for translating English to Bangla (E2B) or vice-versa in Bangladesh. However, no research was devoted to define Bangla compound verbs for conversion of Bangla language sentences into UNL documents. This paper concentrates on the representation of compound verbs in the UNL Based machine translation system by developing dictionary entries. These verbs can be used to convert natural Bangla sentences to UNL documents and vice versa.

The rest of the paper is organized as follows. Section 2 describes the structure of UNL. Format of UNL-based Bangla word dictionary is presented in Section 3. Structure of Bangla compound verbs and their classification are elaborated in Section 4. Section 5 outlines the templates of compound verb roots and their verbal inflexions. Conversion of a Bangla sentence with compound verb into UNL expressions is shown in Section 6 while some concluding remarks are presented in Section 7.

2. UNIVERSAL NETWORKING LANGUAGE (UNL)

The meaning of a native language is expressed in UNL system as a hypergraph composed of nodes connected by semantic relations. Nodes or Universal Words (UWs) are extracted from English and disambiguated by their positioning in a knowledge base (KB) [1] of conceptual hierarchies. The core structure of UNL is based on the *Universal Words, Attributes* and *Relations* [1]. *Universal Words* constitute the vocabulary of UNL. It is expressed as a node in the semantic network of UNL expression. *Attributes* are used to describe the subjectivity information of sentences and to express logical expressions in strengthening the expressibility of the UNL. Relation label is represented as strings of three characters or less. The relation between UWs is binary with different labels according to the different roles they play [1-2].

To convert Bangla sentences into UNL expression, we use EnConverter (EnCo) [1], a universal converter system provided by the UNL project shown in Figure 1. It is a language independent parser; a multi-headed Turing Machine [11] provides synchronously a framework for morphological, syntactic and semantic analysis. Natural language texts are analyzed sentence by sentence using a knowledge rich lexicon and analysis rules. It scans an input string from left to right. When an input string is scanned, all matched morphemes with the same string 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. The left character string is scanned from the beginning following applied rule. It moves back and forth over the *Node List*, which contains words of the input sentence. In the figure, "A" indicates an *Analysis Windows (AW),* "C" indicates a *Condition Windows (CW)* and "n_n" indicates an *Analysis Node*. The machine traverses the input sentence back and forth. Hence, it retrieves the relevant dictionary entry from the Word Dictionary (Lexicon) and depending on the attributes of the nodes under the AWs and those under the surrounding CWs. It then generates the semantic relations between the UWs and /or attaches speech act attributes to them. As a result a set of UNL expressions is made equivalent of UNL graph [12]. EnCo is driven by a set of analysis rules to analyze a sentence using Word Dictionary and Knowledge Base. The enconversion rules have been described in [5].

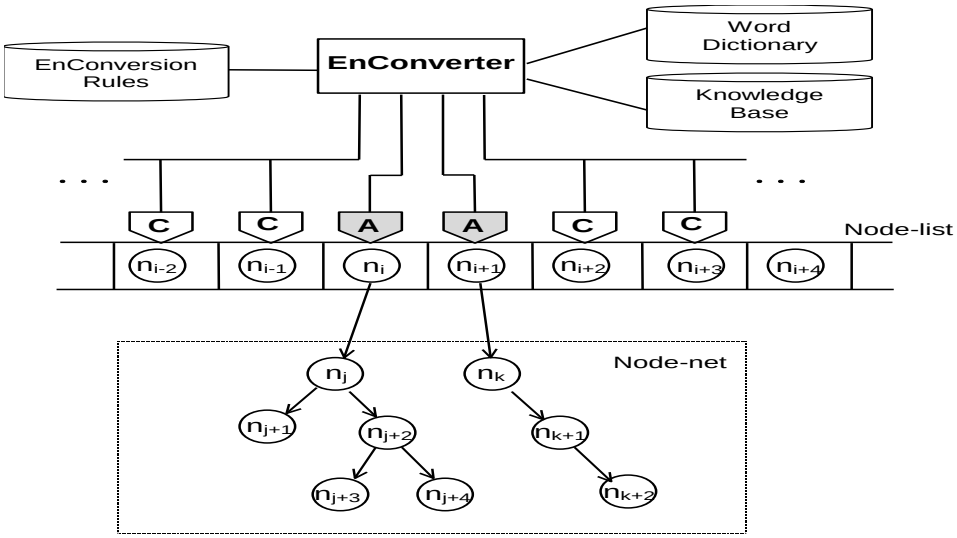


Figure 1: Structure of EnConverter("A" indicates an Analysis Window, "C" indicates a Condition Window, and "n_n" indicates an Analysis Node)

The machine traverses the input sentence back and forth, retrives the relevant dictionary entry (UW) from the Word Dictionary (Lexicon) depending on the attributes of the nodes under the AWs and the surrounding of CWs. It then generates the semantic relations between the UWs and /or attaches speech act attributes to them. As a result, a set of UNL expressions is made the equivalent of UNL graph. EnCo is driven by set of analysis rules to analyze a sentence using Word Dictionary and Knowledge Base. The enconversion rule has the following syntax:

<TYPE>
["(" <PRE> ")" ["*"]]...
"{"|""""" [<COND1>] ":" [<ACTION1>] ":" [<RELATION1>] ":" [<ROLE1>] "}"|"""""
["(" <MID> ")" ["*"]]...
"{"|""""" [<COND2>] ":" [<ACTION2>] ":" [<RELATION2>] ":" [<ROLE2>] "}"|"""""
["(" <SUF> ")" ["*"]]...
"P(" <PRIORITY> ");"

where, characters between quotes are predefiened delimiters of the rule. Expressions in { } and " " are used to describe rules for the nodes designated by the Analysis windows.

When the node on the left Analysis Window satisfies the <COND1> attributes, the node on the right Analysis Window satisfies the <COND2> attributes, and as relevant the nodes in the left, middle and right sides of the Analysis Windows fulfill the conditions described in <PRE>, <MID> and <SUF> of the Condition Windows, the grammatical attributes of the nodes in the Analysis Windows are rewritten according to <ACTION1> and <ACTION2> respectively. If a relation is described in either <RELATION1> or <RELATION2>, the binary relation between the two UWs of the nodes on the Analysis Windows are created using the relation. The operations are carried out on the Node-list depending on the type of rule shown in <TYPE> [1].

Morphological analyses are performed following the left and right composition rules. This type of rule is used primarily for creating a syntactic tree with two nodes on the Analysis Windows. The semantic analyses are accomplished following the left and right modification rules. They are used to create semantic relations between the words in a sentence [5].

3. UNL Based Bangla Word Dictionary

The word dictionary consists of mapping of Bangla words to Universal Words and lexical-semantic attributes describing the words. The lexicon is a collection of word dictionary entries. Each entry is composed of three kinds of elements: *Headword (HW), Universal Word (UW)* and *Grammatical Attribute (GA)***.** A HW is a notation of a word in a native language composing the input sentence. It is used as a trigger in obtaining equivalent UWs from a Word Dictionary in the process of enconversion. An UW expresses the meaning of a word is used in creating UNL networks (i. e., UNL expressions) of output. GAs are the information on how words behave in a sentence and are used in enconversion rules. Each dictionary entry has the following format associating with any native language word [1, 4].
Data Format: *[HW]{ID}"UW"(ATTRIBUTE1, ATTRIBUTE2,...)<FLG, FRE, PRI>*
Here,
HW: Head Word (Bangla Word)
ID: Identification of Head Word (omissible)
UW: Universal Word
ATTRIBUTE: Attribute of the HW
FLG: Language Flag
FRE: Frequency of Head Word
PRI: Priority of Head Word
Some examples of dictionary entries for Bangla words are presented below:
[কোকিল] {} "Cuckoo(icl>bird)" (N, COMN, CEND, BIRD, ANI, WILD)

[শিউলি] {}" Jasmine (icl>plant)" (N, COMN, VEND, FLWR, PLNT)

[সে] {} "he(icl>person)"(PRON, HPRON, SUB, MALE, ANI,3SG)

In the dictionary entries attributes N stands for noun, ANI for animate object, COMN for common noun, CEND for consonant ended word, VEND for vowel ended word, FLWR for flower, PLNT for plant, PRON refers to Pronoun, HPRON to Human Pronoun, 3SG for third person singular number respectively.

4. Structure of Bangla Compound Verbs and Their Structural Classification

A compound verb of Bangla is composed with a verb (base verb) and another word (semantic word). The semantic word may be either a noun or an infinite verb. The semantic word always remains unchanged. And hence we can structurally divide compound verbs into the following groups [13-15].

- Compound verbs formed by nouns
- Compound verbs formed by infinite verbs

Compound verbs formed by nouns: This type of compound verb is formed by a noun followed by a space and a verb as follows:

Compound verb = noun + " " + verb

For example: ঘুষ দেয় = ঘুষ + " " + দেয়

In this example, compound verb is ঘুষ দেয় (pronounce as ghush deye) meaning *to bribe*, where, ঘুষ (ghush) is noun and base verb is দেয় (deye). Again base verb is divided into base verb root and verbal inflexion as shown below.

Base verb = base verb root + verbal inflexion

In our example, 'দেয়' is base verb where 'দে' (de) is verb root and 'য়' (ye) is verbal inflexion. So, দেয় = দে + য়

Some compound verbs, noun parts, Base verbs, Base verb roots and inflexions are shown in Table 1.

Table 1. Some compound verbs, base verbs their roots and inflexions

Compound verb	Noun part	Base verb	Base verb root	Verbal inflexion
মাথা খাইতেছে	মাথা	খাইতেছে	খা	ইতেছে
ঘুষ দেয়	ঘুষ	দেয়	দে	য়
অনুসরণ করা	অনুসরণ	করা	কর্	আ
কসম খাওয়া	কসম	খাওয়া	খা	ওয়া

- Compound verbs formed by infinite verbs: This type of compound verb is formed by an infinite verb followed by a space and a verb as follows:

Compound verb = infinite verb + " " + verb

For example, ছুঁড়ে ফেলল = ছুঁড়ে + " " + ফেলল

Here, ছুঁড়ে ফেলল (pronounce as chhure fello) meaning *to throw* is a compound verb, where, ছুঁড়ে (pronounce as chhure) is infinite verb and ফেলল (pronounce as fello) is base verb. This paper focuses on only compound verbs fromed by noun.

Universal Words (UWs) are mainly defined by English words because English is known by all UNL developers and there are a lot of bilingual dictionaries between a local language and English [17]. A compound verb of Bangla mainly consists of two words. When a compound verb is translated into UW, the meaning of one or both words of the verb is changed into one English meaning. For example, সে ঘুষ দেয় pronounce as *Se ghush deye* meaning *He bribes*. In this sentence, ঘুষ দেয় is a compound verb. This compound verb is formed by two words: ঘুষ (ghush) and দেয় (deye). Here, the word 'ঘুষ' is noun and its English meaning is 'bribe' and the word ' দেয়' is base verb which means give. When these two words form the compound verb to construct a UW (English meaning), the meaning of the first word of the verb will be the meaning of the compound verb and the meaning of the second word will be ommited. However, the tense, person and subject-verb agreement of the sentence come form the second word of the verb. Some compound verbs that form UWs from one of the words of the compound verb shown in Table 2.

Table 2. Some roots of compound verbs, base verbs and their roots

Compound verb	Universal Word (UW)/English Word
ঘুষ দেওয়া (ghush de)	to bribe
অনুসরণ করা (onushoron kora)	to follow
কথা বলা (kotha bola)	to speak
কাজ করা (kaj kora)	to work

Sometimes UWs of Bangla compound verbs are formed without the usual meaning of the words of the compound verb. Meaning of these compound verbs generally come from idoms. For example, মাথা খাওয়া (pronounce as matha khawa) is a compound verb means *to pervert*, which has no relation with the meanings of the words of compound verb মাথা and খাওয়া. Another example is: গোল্লায় যাওয়া (pronounce as gollaye jawa) a compound verb meaning *to ruin* is not the meaning of the words of the

compound verb গোল্লায় and যাওয়া. For each combination of words of these types of compound verbs we define three dictionary entries: one for noun, second for base verbroot and third one for verbal inflexion. Second part of the verb will be created by combining base verb root and verbal inflexion. Noun will carry the ideometric meaning of the verb in the dictionary. Like as previous one second part of the verb represents tense, person and subject-verb agreement of the sentence. Some compound verbs that form UWs from idiometric meaning are shown in Table 3.

Table 3. Some roots of compound verbs, base verbs and their roots

Compound verb	Universal Word (UW)/English Word
মাথা খাওয়া (matha khawa)	to pervert
গোল্লায় যাওয়া (gollaye jawa)	to ruin
আছাড় খাওয়া (achhar khawa)	to tumble

5. Representation of Compound Verb in UNL Based Translation Scheme

To use compound verbs in translation scheme, they need to be represented them in the UNL sytem. A compound verb can not be represented directly as a dictionary entry in the UNL based Bangla word dictionary, rather dictionary entries are separately define for noun, base verb root and verbal inflexions.

5.1. Representation of compound verb froming UW from the noun part of the verb: we have developed templates for dictionary entries of noun, base verb and verbal inflexion combinely forming compound verb.

- The template of dictionary entry of noun part of the compound verb:

[HW]{}"UW(icl/iof…>concept1>concept2…,REL1>…,REL2>…,)"(N,NABS, VEND/CEND, BLK, NCPND, #REL1, #REL2, ..<FLG, FRE, PRI>
where, attributes N denotes noun, NABS for abstract noun, NCPND for compound verb which will be created by noun, VEND & CEND for vowel and consonant ended word, BLK for blank indicates that a blank space will be between noun and base verb for a compound verb.
Example of dictionary entry:
[ঘুষ] {}"bribe(icl>pay>do,agt>volitional_thing,obj>volitional_thing,gol>thing)" (N, NABS, NCPND, CEND, BLK, #OBJ)

- Template for dictionary entry of base verb root to construct compound verb:

[HW]{}"(ROOT, NABS, VEND/CEND, NCPND, ^VERB,BLK, #REL1, #REL2,..)"<FLG,FRE, PRI>
where, attributes ROOT denotes base verb root, ^VERB for not verb.
Example of dictionary entry: [দে] {} " " (ROOT, CEND,GEN,NCPND)

- Template for dictionary entry of inflexion to construct compound verb:

[HW]{}"(VI, VEND/CEND, NCPND, ^VERB, BLK, #REL1, #REL2,..)"<FLG,FRE, PRI>
Example of dictionary entry: [য়] {} " " (VI, VEND, NCPND)

Using the above three entries to create a compound verb, a morphological analysis needs to be done first between the base verb root 'দে' and inflexion 'য়' to form base verb 'দেয়'. The noun 'ঘুষ' and the base verb 'দেয়' is then combined to form compound verb 'ঘুষ দেয়' by another morphological analysis. Attribute NCPND in noun 'ঘুষ' indicates that this noun will create a compound verb and attribute BLK meaning *balnk space* denotes that a blank space will be created between noun 'ঘুষ' and verb 'দেয়' during construction of the compound verb. The same attribute NCPND will be used in other two entries to ensure that the entries with attribute NCPND will create compound verb. In this case of forming compound verb, the dictionary entries of the root and its alternative will have no UWs since root does not take part in forming the meaning of a compound verb.

5.2. Representation of compound verb forming idiomatic meaning: We have also developed templates for dictionary entries of noun, base verb and verbal inflexion that combinely form compound verb with idiometric meaning. Templetes for dictionary entries of noun, base verb and inflexion for ideometric meaning of compound verb are identical to those for generic compound verb shown in Section 5.1. except an attribute IDIOM and [[যা]] or [[খা]] will be in the entry of noun and only attribute IDIOM will be in other two entries as follows:

- Template for dictionary entry of noun part for the ideometric meaning of compound verb:

[HW]{}"UW(icl/iof…>concept1>concept2…,REL1>…,REL2>…,)"(N, NABS, VEND/CEND, NCPND,BLK#REL1, #REL2, ..<FLG, FRE, PRI>
Example of dictionary entry: খা
[গোল্লায়] {}"(ruin(icl>do,equ>destroy,agt>thing,obj>thing,met>thing)" (N, NABS, NCPND, VEND, BLK, IDIOM, [[যা]] #OBJ) <B, 1, 2>

- Template for dictionary entry of base verb root to construct the base verb:

[HW]{}"(ROOT,NABS,VEND/CEND,NCPND,^VERB,BLK,IDIOM,[[খা]], #REL1#REL2,..)"<FLG, FRE, PRI>
where, attributes ROOT denotes base verb root, ^VERB for not verb.
Example of dictionary entry: [যা]{}" " (ROOT, VEND, GEN, NCPND, IDIOM)

- Template for dictionary entry of inflexion to construct the base verb:

[HW]{}"(VI, VEND/CEND, NCPND, ^VERB, BLK, #REL1, #REL2,..)"<FLG,FRE, PRI>
Example of dictionary entry: [ইতেছে] {}" " (ROOT, CEND,GEN,NCPND,IDIOM, P1,PRG)

6. Analysis of Semantic Ambiguity in the Construction of Compound Verb

The verb root 'খা'(pronounce as kha) means *to eat*. To eat anything we use this verb root say 'ভাত খা' (pronounce as vat kah) meaning *to eat rice*, 'আম খা' (pronounce as aam kha) meaning *to eat mango*, 'সুপ খা' (pronounce as sup kha) meaning *to eat soup*. The dictionary entry of this verb root 'খা' is:

[খা]{}"eat(icl>consume>do,agt>living_thing,obj>concrete_thing,ins>thing)"(ROOT,VER,VEG1, #AGT, #OBJ,#INS) <B,0,0>

However, if we say 'পানি খা' (pronounce as pani kha), 'চা খা' (pronounce as cha kha), 'সিগারেট খা' (pronounce as cigarette kha) we cannot say *to eat water, to eat tea* and *to eat cigarate,* rather we say *to drink water, to take tea* and *to smoke* respectively.

Here, the meaning of the same root 'খা' (kha) in 'পানি খা' (pani kha) is *drink*, 'চা খা' (cha kah) is *take* and 'সিগারেট খা' (cigarette kha) is *smoke* respectively. To resolve these semantic ambiguities we define separate dictionary entries for root 'খা' for UWs drink, take and smoke in the word dictionary shown in Table 4.

Table 4. Dictionary entries of 'খা' for three different UWs

[খা]{}"drink(icl>consume>do,agt>living_thing,obj>concrete-thing)"(ROOT,VEND,#AGT,#OBJ, VEG1, LIQ)<B,1,2>
[খা]{} "take(icl>consume>do, agt>living_thing, obj>concrete_thing)"(ROOT,VEND, #AGT, #OBJ, LIQT) <B,1,1>
[খা]{} "smoke(icl>consume>do,agt>living_thing,obj>concrete_thing)"(ROOT, VEND, #AGT, #OBJ, TOBA) <B,1,2>

And the dictionary entries of 'পানি' (water), 'চা' (tea), and 'সিগারেট' (cigarette) are shown in Table 5.

Table 5. Dictionary entries of 'পানি', 'চা' and 'সিগারেট'

[পানি]{}"water(icl>matter,equ>h2o)"(N,NCOM,LIQ,VEND,#OBJ)< B,0,0>
[চা]{}"tea(icl>beverage>thing)"(N,NCOM,LIQT,VEND,#OBJ)< B,0,0>
[সিগারেট]{}"cigarette(icl>roll_of_tobacco>thing)"(N,NMAT,CEND,TOBA,#OBJ,#MOD)< B,0,0>

The dictionary entries of Table 5 can be combined sequentially with the entries of Table 4 to form 'পানি খা' (to drink water) 'চা খা' (to take tea) and 'সিগারেট খা'(to smoke cigarette) using morphological rules respectively. As they have separate dictionary enteries with UWs they are not considered as the compound verbs.

For example, in Table 4, Attributes LIQ indicates liquid water; NTCPND denotes not compound verb means when root 'খা' combines with 'পানি' on its left side they create 'পানি খা' (to drink water) which is not compound verb. Same procedure can be applicable to other two combinations.

7. Morphological Rules for Construction of Compound Verb

We have rigorously gone through the Bangla grammar [13-15] and found the following types of compound verbs. We have outlined the format of morphological rules for the construction of compound verb.

An abstract noun with 'NCPND' attribute and a verb with NCPND attribute make a compound verb which takes meaning from the noun. In this case, the dictionary entry of the root has no UW.

Rule 1: Morphological Rule for Compound Base Verb:
+ (N, ABS, NCPND) {ROOT, VEND, VEG2, NCPND, #AGT, #OBJ: +V: :}{ INF, VI :::}

where, atteibutes N denotes noun, ABS for abstract noun, NCPND for compound verb formed by noun-it indicates that this noun will construct a compound verb combining with base verb on the right side, ROOT for base verb root, VEND for vowel ended root, VEG2 means vowel ended group 2, #AGT for agent, #OBJ for object, V for verb mens that verb root will combine with inflexion to construct verb, INF for inflexion and VI for verbal inflexion. This rule is applied when noun with attribute NCPND is in the left condition window (LCW), the verb root with attribute NCPND is in the

left analysis window (LAW) and verbal inflexion is in the right analysis window (RAW). Following rule application, the two headwords of the left and right analysis windows are combined into a composit node forming a base verb placed in the RAW. Currently, the noun with attribute NCPND is placed in the LAW and the base verb with attribute NCPND is placed in the RAW.

Rule 2: Morphology Rule for Compuound Verb:
+{N, ABS, NCPND:+@,-N,-ABS,-NCPND,+ncpnd, #OBJ,+VCPND::}{V, NCPND, [“ “]:::}
Where, attribute VCPND indicates compound verb. This rule is applicable when a noun with attribute NCPND is in the LAW and and a verb with attribute NCPND is in the RAW. Following rule application noun in the LAW are combined with the verb in the RAW to construct compound verb with a space between them.

Rule 3: Attribute Changing
:{:::} {V, NCPND, ^ncpnd: -NCPND::}
The rule is applied to change the attribute of a word after completion of the morphological and semantic analysis.

8. Conversion of a Bangla Compound Verb Sentence into UNL Expression

The conversion process will be performed by shift/reduce parsing [16]. To explain the conversion steps, we present an example of a simple Bangla assertive sentence, which has only one main clause. We assume that analysis rules and the dictionary of Bangla to UNL are given to the analyser system *EnCo*.

[ঘুষ] {} “bribe(icl>payment>thing)” (N, ABS, NCPND, CEND, BLK, #OBJ)
[দে] {} “ ” (ROOT, CEND,GEN,NCPND)
[য়] {} “ ” (VI, VEND, NCPND)
Verb root ‘দে’ (de) has an alternative ‘দি‘ (di). The dictionary entry of ‘দি‘ is:
[দি]{} “ ” (ROOT, CEND, ALT1, NCPND)

Bangla sentence: সে ঘুষ দেয়।
Transliterated sentence: Se ghush deye.
Equivalent English sentence: He bribes.
The input Bangla sentence is processed according to the algorithm that we have developed in [17]. The chunks obtained from the input sentence are given below.

(সে) (ঘুষ) (দে) (য়)।

(Se) (ghush) (de) (ye)

We have used an EnConverter [5] tool for our experiment. The tool takes a dictionary file for the sentence shown in Table 6 and a set of analysis rules shown in Table 7 as its input.

Table 6. Dictionary entries of respective Bangla sentence

[সে]{} “he(icl>person)”(PRON, HPRON, P3, SG, SUBJ)
[ঘুষ] {} “bribe(icl>payment>thing)” (N, NABS, NCPND, CEND, BLK, #OBJ)
[দে] {} “ ” (ROOT, CEND,GEN,NCPND)
[য়] {} “ ” (VI, VEND, P3, PRS,IND, NCPND)

In Table 6, attributes PRON indicates pronoun, HPRON indicates human pronoun, P3 for third person, SG for singular, SUBJ for subject, N indicates noun, NABS for abstract noun, ROOT for verb root, VI for verbal inflexion, VEND for vowel ended root, PRS for present tense, IND for indefinite tense and NCPND indicates noun based compound verb respectively.

EnCo can input either a string or a list of words for a sentence of a native language. A list of morphemes or words of a sentence must be enclosed by [<<] and [>>] [1]. When the sentence is taken into EnCo, it places the sentence head (<<) in the LAW (Left Analysis Window), sentence texts or morphemes or words in the RAW (Right Analysis Window) and the sentence tail (>>) in the RCW (Right Condition Window) shown in Figure 2.

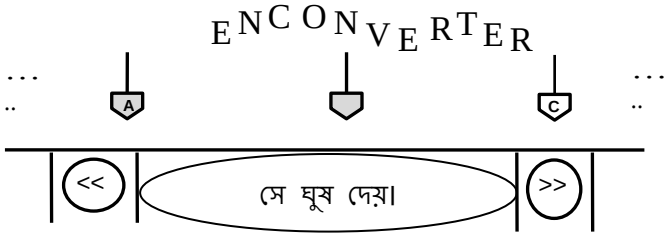


Figure 2. Initial state of the Analysis Windows and the node list

After insertion of our given sentence the rules shown in Table 7 will be applied step by step to complete the conversion processes of the sentence to UNL expressions.

Table 7. Analysis rules for converting the sentence into UNL expression

Rule	Description
Rule 1: R{SHEAD:::}{PRON,SUBJ:::}	Right Shift Rule
Rule 2: DR{SUBJ,^blk:blk::}{BLK:::}	Right Deletion Rule
Rule 3: R{PRON,SUBJ:::}{N:::}	Right Shift Rule
Rule 4: DR{N,^blk:blk::}{BLK:::}	Right Deletion Rule
Rule 5: R{N, NABS:::}{ROOT,^VERB:::}	Right Shift Rule
Rule 6: +{ROOT:::}{INF,VI,VEND,VEG1:::}	Left Composition Rule
Rule 7: +{N, NABS,NCPND:@::}{V,NCPND,P3:::}	Left Composition Rule
Rule 8: >{PRON, SUBJ::agt:}{VERB,#AGT:::}	Right Modification Rule
Rule 9: R{SHEAD:::}{VERB,^&@entry:+&@entry::}	Right Shift Rule
Rule 10: R{VERB:::}{STAIL:::}	Right Shift Rule

Rule 1 describes when sentence head is in the LAW and subject ‘সে’, *she* (he) is in the RAW then AWs will be shifted to right after rule application. The EnCo will then retrieve the word, ‘সে’ from the Word Dictionary file and the word, ‘সে’ remains in the LAW and ‘ঘুষদেয়’, *ghush deye* (takes bribe) will be in the RAW. Rule 2 is applied to delete the right node which is a blank space between subject ‘সে’ and noun ‘ঘুষ’, and only the noun ‘ঘুষ’ will be placed in the RAW, while the verb ‘দেয়’ (*deye)* will be placed in the RCW. Rule 3 is then applied to shift the windows to right, and the Rule 4 is

applied to delete the space between 'ঘুষ' (ghush) and 'দেয়' (deye) so that the word 'ঘুষ' (ghush) is retrieved from the Word Dictionary and remains in the LAW and the verb 'দেয়' (deye) is divided into root 'দে' (de) which remains in the RAW and verbal inflexion 'য়' (ye) remains in the RCW. Again a right shift rule (Rule 5) is applied to shift the windows of EnCo to one step right. Currently, root 'দে' (de) is in the LAW and inflexion 'য়' (ye) is in the RAW. A morphological analysis is held between the nodes on the left and right analysis windows to make the verb 'দেয়' (deye) after applying the rule 6 by retriving the base verb root 'দে' from the Word Dictionary, which remains in the RAW. A left composition rule (Rule 7) is now applied to perform morphological analysis between noun noun 'ঘুষ' (is in LAW) and verb 'দেয়' (is in RAW) to construct compound verb 'ঘুষ দেয়' which remains in the RAW.

Upon completion of morphological analysis right modification rule (Rule 8) is applied to perform semantic analysis between 'সে' (Se) and compound verb 'ঘুষ দেয়' by the agent relation, *agt*. Hence, subject 'সে' is deleted from the node-list, where compound verb 'ঘুষ দেয়' remains in the RAW, which is the main predicate of the sentence. Rule 9 is then applied to shift the windows to right and *&@entry* attribute is added to the compound verb as compound verb 'ঘুষ দেয়' is the main word of the sentence.

Finally, Rule 10 is applied to place the sentence tail (STAIL) on the LAW to complete the conversion process. After completion the conversion process, the following UNL expresson shown in Table 8 will be created by the EnCo.

Table 8. UNL Expression of the Converted Sentence

[S:00] {org:en} He bribes. {/org} {unl} agt(bribe(icl>pay>do,agt>volitional_thing,obj>volitional_thing,gol>thing).@entry.@present,he (icl>person)) {/unl} [/S]

9. Conclusions and Future Work

This paper has analyzed the Bangla compound verbs and grouped them into noun based and infinite verb based compound verbs. This paper has also developed the formats of noun, base verb roots and their inflexions required to represent noun based compound verbs in UNL structure. Finally we have shown the conversion procedures of a compound verb related sentence into the UNL expressions. A Bangla native language sentence with compound verb can be easily converted into UNL expression by analysis rules, which can later be translated to any other languages using language specific generation rules. The proposed format can be equally applicable to other languages with compound verbs. Our future research is to focus on the representation of infinite verb based compound verb in UNL based tranalation scheme.

REFERENCES

[1] H. Uchida, M. Zhu, and T. C. D. Senta, Universal Networking Language, *UNDL Foundation*, International environment house, 2005/6, Geneva, Switzerland.

[2] M. E. H. Choudhury, M. N.Y. Ali, M.Z.H. Sarkar, R. Ahsan, "Bridging Bangla to Universal Networking Language- a Human Language Neutral Meta- Language", International Conference on Computer and Information Technology (ICCIT), Dhaka, 2005, pp.104-109

[3] M.N.Y. Ali, J.K. Das, S.M. A. A. Mamun, A.M. Nurannabi, "Morphological Analysis of Bangla Words for Universal Networking Language", Third International Conference on Digital Information Management (ICDIM 2008), London, England. pp. 532-537

[4] M.N.Y. Ali, J.K. Das, S.M. A. 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.

[5] EnConverter Specification, Version 3.3, UNL Center/UNDL Foundation, Tokyo 150-8304, Japan 2002.

[6] UNDL Foundation: The Universal Networking Language (UNL) specifications version 3.2. (2003).

[7] DeConverter Specification, Version 2.7, UNL Center, UNDL Foundation, Tokyo 150-8304, Japan 2002.

[8] Ali, M., & Ali, M. M. (2002),"Development of machine translation dictionaries for Bangla Language", 7th International Conference on Computer and Information Technology (ICCIT), pp. 272-276.

[9] Saha, G. K. (2005). The E2B machine translation: A new approach to HLT. Ubiquity archive, Association of Computing Machineries (ACM), 6(32), New York.

[10] Uddin, M. G., Ashraf, H., Kamal. A. H. M, & Ali, M. M. (2004), "New Parameters for Bangla to English Statistical Machine Translation", Proceedings of 3rd International Conference on Electrical & Computer Engineering (ICECE 2004), pp. 545-548.

[11] J. Parikh, J. Khot, S. Dave & P. Bhattacharyya, "Predicate Preserving Parsing", Department of Computer Science and Engineering, Indian Institute of Technology, Bombay.

[12] Md. Nawab Yousuf Ali, Golam Sorwar, Ashok Toru, Md. Anik Islam, Md. Shamsujjoha, "Morphological Rules of Bangla Repetitive Words for UNL Based Machine Translation", 6th International Conference on Swarm Intelligence, ICSI, Beijing, China, June 25-28, 2015, pp-401-408.

[13] D. M. Shahidullah, "Bangala Vyakaran", Maola Brothers Prokashoni, Dhaka, August 2003, pp.110-130

[14] D. C. Shniti Kumar,"Vasha-Prokash Bangla Vyakaran", Rupa and Company Prokashoni, Calcutta, July 1999, pp.170-175.

[15] D. S. Rameswar, "Shadharan Vasha Biggan and Bangla Vasha", Pustok Biponi Prokashoni, November 1996, pp.358-377.

[16] M.E.H. Choudhury, M.N.Y. Ali, "Framework for synthesis of Universal Networking Language", *East West University Journal*, Vol. 1, No. 2, 2008, pp. 28-43

[17] N. Y. Ali, M. Ameer Ali, Abu Mohammad Nurannabi, J. K. Das, "Algorithm for Conversion of Bangla Sentence to Universal Networking Language", International Conference on Asian Language Processing (IALP), Harbin, China, 2010, pp.118-121