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

[1] Md. Nawab Yousuf Ali, [2] Mohammad Arafat Ullah, [3] Md. Lizur Rahman, [4] Md. Golam Sarowar

[1, 2, 3, 4] Department of Computer Science and Engineering, East West University, Dhaka 1212, Bangladesh

[1] nawab@ewubd.edu, [2] arafat.ngc@gmail.com, [3] lizur.sky@gmail.com, [4] sojolewu6@gmail.com

**Abstract**

This paper displays the advancement of configuration for lexicon passages of thing, base action word root and their expressions to fuse them into an interlingua portrayal 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 centers to connect the holes among things and base action word establishes in the structure of UNL framework meaning to deliver a Bangla word lexicon for UNL. The paper analyzes the construction of compound verbs in a variety of approaches based on their structure and Bangla grammar. The paper additionally demonstrates the transformation methodology of a Bangla sentence with compound action word into UNL articulations utilizing the investigation rules which would then be able to be changed over into other local dialects utilizing the dialect explicit age rules.

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

1. **INTRODUCTION**

The Universal Networking Language (UNL) [1] is a machine-oriented language that permits computers to method info and information across the language barriers. It is a man-made language which in human communication reproduces the function of natural languages. The aim of UNL project is to create an intermediate language which can be converted into various languages where people from various nations can easily access information in their own language [2]. The aim of UNL is to develop a mediator language which can be easily converted into various languages. Similarly, sentences of various languages also can be converted into a common mediator language [3]. Learning and data in various dialects are dispersed everywhere throughout the world and stay difficult to reach to generally due to non-machine portrayal and dialect obstruction [4]. To manage the dialect hindrance, United Nations University/Institute of Advanced Studies (UNU/IAS) directed an audit of all globally accessible machine interpretation projects and began to devise an effective and useful system to build up a human dialect impartial meta-dialect for Internet. The consequence of the undertaking is Universal Networking Language (UNL) [1]. The point of this universally helpful activity is to kill the huge necessity of interpretation among dialects and diminish dialect to dialect interpretation to one time transformation to UNL. In UNL framework, a local dialect sentence is changed over into an UNL hypergraph by a system called "EnConverter" [5] following examination rules characterized in [6]. These hypergraphs are then converted into any local dialect, utilizing age rules characterized in [6], by another system called "DeConverter" [7]. The improvement of dialect explicit segments, for example, lexicon, investigation guidelines and age rules utilized by EnConverter and DeConverter, are the examination center over the world.

The general population in Bangladesh and two states (West Bengal and Tripura) in India use Bangla as their first dialect. Around one 6th populace of the world is talking in Bangla. Trade of data and sharing of information all around, it is fundamentally critical to devise change technique(s) for Bangla dialect writings into UNL and the other way around. Machine interpretation (MT) is a way to deal with deciphering writings starting with one normal dialect then onto the next naturally. Ali and Ali (2002) endeavored to create MT Bangla word references that address the association, substance and subtleties of the data [8]. Saha (2005) grew minimal effort English to Bangla (E2B)- ANUBAD making an interpretation of English content into Bangla content utilizing both guideline based and change based MT conspires alongside three dimensions of parsing [9]. Another endeavor by Uddin et al. (2004) was to build up a factual Bangla to English interpretation motor utilizing just basic Bangla sentences that contains a subject, an item and an action word [10].

As per the above, some research has been conducted in the development of automatic translation of Bangla noun based compound sentence into UNL documents and develops some automatic software for it. However, no research was devoted to define Bangla compound verbs which can convert Bangla language into UNL documents. This paper concentrates on the representation of compound verbs in the UNL Based machine translation system by developing dictionary entries. These action words can be utilized to change over characteristic Bangla sentences to UNL records and the other way around.

Whatever remains of the paper is sorted out as pursues. Section 2 portrays the starters of this work including structure of UNL and configuration of UNL-based Bangla word lexicon. Section 3 elaborates the structure of Bangla compound verbs and their classification. The proposed methodology which outlines the templates of compound verb roots and their verbal inflexions are represented in Section 4. Section 5 shows the result analysis and discussion of the developed system while some finishing up comments are displayed in Section 6.

1. **PRELIMINARIES**

Universal Networking Language (UNL) is a mechanism which can convert a language to other languages. Basically this conversion held via two steps: (i) conversion between source language to UNL, and (ii) conversion between UNL to target language. This section describes a detailed overview on UNL and UNL based Bangla word dictionary.

* 1. **Universal Networking Language (UNL)**

The UNL system expresses the meaning of native language from any unknown languages. Native language expressed as a hypergraph (group of nodes) with semantic relations. Basically Universal Words (UWs) or nodes are brought out from English and are detached based on knowledge based concept (KB) [1] by their positioning. The center structure of UNL depends on the Universal Words, Attributes and Relations [1]. All inclusive Words comprise the vocabulary of UNL. It is communicated as a hub in the semantic system of UNL articulation. *Attributes* are used in UNL to describe the subjectivity information of sentences as well as represent the logical expressions. Strings with maximum three characters are used to represent *Relations*. Relation between two Universal Words is binary with various labels based on their various rules [1-2].

To change over Bangla sentences into UNL expression, we use EnConverter (EnCo) [1], a widespread converter framework given by the UNL venture appeared in Figure 1. It is a dialect free parser; a multi-headed Turing Machine [11] gives synchronously a system to morphological, syntactic and semantic investigation. Regular dialect writings are broke down sentence by sentence utilizing an information rich dictionary and investigation rules. It filters an information string from left to right. At the point when an information string is filtered, every single coordinated morpheme with a similar string characters are recovered from the word reference and turn into the applicant morphemes as indicated by the need rule so as to fabricate a syntactic tree and the semantic system for the sentence. The left character string is checked from the earliest starting point following connected guideline. It moves forward and backward over the *Node List*, which contains expressions of the info sentence. In the figure, "A" shows an Analysis Windows (AW), "C" demonstrates a Condition Windows (CW) and " nn" demonstrates an Analysis Node. The machine navigates the info sentence forward and backward. Subsequently, it recovers the applicable word reference passage from the Word Dictionary (Lexicon) and relying upon the characteristics of the hubs under the AWs and those under the encompassing CWs. It at that point creates the semantic relations between the UWs and/or joins discourse act credits to them. Subsequently a lot of UNL articulations is made likeness UNL chart [12]. EnCo is driven by a lot of examination tenets to investigate a sentence utilizing Word Dictionary and Knowledge Base. The enconversion rules have been depicted in [5].



Figure 1: Structure of EnConverter

The machine crosses the information sentence forward and backward, recovers the significant word reference section (UW) from the Word Dictionary (Lexicon) contingent upon the characteristics of the hubs under the AWs and the encompassing of CWs. It at that point produces the semantic relations between the UWs and/or connects discourse act credits to them. Thus, a lot of UNL articulations is made what might be compared to UNL chart. EnCo is driven by set of investigation standards to examine a sentence utilizing Word Dictionary and Knowledge Base. The syntax of EnConversion rule is given below:

<TYPE>

[ "(" <PRE> ")" ["\*"] ]...

"{"|"""" [<COND1>] ":" [<ACTION1>] ":" [<RELATION1>] ":" [<ROLE1>] "}"|""""

[ "(" <MID> ")" ["\*"] ]...

"{"|"""" [<COND2>] ":" [<ACTION2>] ":" [<RELATION2>] ":" [<ROLE2>] "}"|""""

[ "(" <SUF> ")" ["\*"] ]...

"P(" <PRIORITY> ");"

Where, characters between quotes are predefined delimiters of the rule. Expressions in { } and “ ” are used to describe rules for the nodes designated by the Analysis windows.

At the point when the hub on the left Analysis Window fulfills the <COND1> characteristics, the hub on the right Analysis Window fulfills the <COND2> properties, and as applicable the hubs in the left, center and right sides of the Analysis Windows fulfill the conditions depicted in <PRE>, <MID> and <SUF> of the Condition Windows, the linguistic traits of the hubs in the Analysis Windows are revised by <ACTION1> and <ACTION2> individually. On the off chance that a connection is portrayed in either <RELATION1> or <RELATION2>, the paired connection between the two UWs of the hubs on the Analysis Windows are made utilizing the connection. The activities are completed on the Node-list contingent upon the kind of standard appeared in <TYPE> [1].

Morphological investigations are performed following the left and right arrangement rules. This kind of guideline is utilized essentially to make a syntactic tree with two hubs on the Analysis Windows. The semantic investigations are practiced after the left and right adjustment rules. They are utilized to make semantic relations between the words in a sentence [5].

* 1. **UNL Based Bangla Word Dictionary**

The mapping of Bangla words to UW and their corresponding lexical-semantic attributes are stored in word dictionary. Basically, the collection of entries in word dictionary is known as lexicon. Every entry is made out of three sorts of components: Universal Word (UW), Headword (HW) and Grammatical Attribute (GA). A HW is a documentation of a word in a local dialect making the information sentence. It is utilized as a trigger in acquiring equal UWs from a Word Dictionary during the time spent enconversion. A UW communicates the importance of a word is utilized in making UNL systems (i. e., UNL articulations) of yield. GAs are the data on how words carry on in a sentence and are utilized in enconversion rules. Every lexicon passage has the accompanying configuration partner with any local dialect word [1, 4].

Information Format: *[HW]{ID}"UW"(ATTRIBUTE1, ATTRIBUTE2,... )<FLG, FRE, PRI>*

Here,

HW: Head Word (Bangla Word)

ID: Head Word Identification (omissible)

UW: Universal Word

ATTRIBUTE: Head Word Attribute

FLG: Language Flag

FRE: Head Word Frequency

PRI: Head Word Priority

Some examples of Bangla Word dictionary 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 means noun, COMN for common noun, ANI for animal object, CEND means consonant ended word, VEND means vowel ended word, FLWR for flower, PLNT for plant, PRON alludes to Pronoun, 3SG for third person singular number, HPRON to Human Pronoun respectively.

1. **STRUCTURAL CLASSIFICATION OF BANGLA COMPOUND VERBS**

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.

Most of the UNL developers known English and therefore Universal Words (UWs) are mainly defined by English words [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 from 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. Table 3 shows some compound verbs that form UWs from idiometric meaning.

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 |

1. **METHODOLOGY**

To use compound verbs in translation scheme, we need to be represented them in the UNL system. A compound verb cannot 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.

* 1. **Compound Verb Representation with UW**

Noun part, base verb root part, and verbal inflexion part of dictionary entries forms compound verbs. We have developed templates for express those dictionary entries.

* The template for 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 *blank 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.

* 1. **Compound Verb Representation with Idiomatic Meaning**

Sometime noun, base verb, and inflexion of dictionary entry combines and form compound verb with idiomatic meaning. We have also developed templates to express the idiomatic meaning of compound verbs. Templates for dictionary entries of noun, base verb and inflexion for idiomatic meaning of compound verb are identical to those for generic compound verb shown in Section 4.1 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 idiomatic 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)

* 1. **Compound Verb Representation with Semantic Ambiguity**

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 entries 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.

* 1. **Morphological Rules for Construction of Compound Verb**

After analyzing the Bangla grammar [13 - 15] verbally, we have found various rules of compound verbs. We have marked out the structural format of these rules for the construction of compound verb. A brief discussion of these rules are given below-

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, attributes 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, attribute ROOT denotes base verb root, attribute VEND means vowel ended root, VEG2 denotes vowel ended group 2, #AGT for agent, #OBJ for object, V for verb means that verb root will combine with inflexion to construct verb, INF for inflexion and VI for verbal inflexion.

This rule is applied when left condition window (LCW) contains the attribute NCPND with noun, left analysis window (LAW) contains the attribute NCPND with root, and right analysis window (RAW) contains verbal inflexion. Following rule application, headwords of the LAW and RAW combined together in a composite node and form 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 Compound Verb:

+{N, ABS, NCPND:+@,-N,-ABS,-NCPND,+ncpnd, #OBJ,+VCPND::}{V, NCPND, [“ “]:::}

Where, attribute VCPND indicates compound verb. This rule is applicable when LAW contains a noun with attribute NCPND and RAW contains a verb with attribute NCPND. 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.

1. **RESULT ANALYSIS & DISCUSSION**

The conversion procedure will be performed by shift/reduce parsing [16]. To clarify the conversion steps, we present a case of a straightforward Bangla decisive sentence, which has just a single fundamental condition. We accept that investigation rules and the lexicon of Bangla to UNL are given to the analyzer framework 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.

In our previous work, we had developed an algorithom that can process any Bangla sentence which is used in this research [17]. The algorithom divides the input sentence into chunks based on the subject, noun, verb, root, and verbal inflexion of the sentence are given below.

(সে) (ঘুষ) (দে) (য়)।

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

We used an EnConverter[ 5] device for our test. The instrument takes a word reference document for the sentence shown in Table 6 and many rules for the examination appear as information in Table 7.

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

EnCo can enter a string or word list for a local dialect sentence. A list of morphemes or sentence expressions must be included with [<<] and [>>][ 1]. The sentence head (< <) in the LAW (Left Analysis Window), sentence writings or morphemes or words in the RAW (Right Analysis Window) and sentence tail (> >) in the RCW (Right Condition Window) appear in Figure 2 when the sentence is taken into EnCo.

|  |
| --- |
| …..  **A**  **C**  E  N  C  O  N  V  E  R  T TTTTT  E E  R  …..  **A**  <<  >> |
| Figure 2. Analysis Window State with Node List |

After inclusion of our given sentence the guidelines appeared Table 7 will be connected well ordered to finish the transformation procedures of the sentence to UNL articulations.

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 portrays when LAW contain the sentence head and RAW contain the subject ‘সে’, *she* (he) then after the rule application AWs will be moved to right. The EnCo will at that point recover the word ‘সে’ from the Word Dictionary document and the word, ‘সে’ stays in the LAW and then RAW will contain ‘ঘুষদেয়’, *ghush deye* (takes bribe). Rule 2 performs the deletion of the blank space between subject ‘সে’ and noun ‘ঘুষ’ (known as right node) when RAW contain the noun ‘ঘুষ’ (*ghush*) and RCW contain the verb ‘দেয়’ (*deye)*. Rule 3 describes the right shift operation of windows. Again Rule 4 performs the deletion of the space between the noun ‘ঘুষ’ (*ghush*) and the verb ‘দেয়’ (*deye)* so that the noun ‘ঘুষ’ (*ghush*) can be recovered from word lexicon and will be placed in LAW. After that the verb ‘দেয়’ (*deye)* can be divided into root ‘দে’ (*de*) which will be placed in RAW and verbal inflexion ‘য়’ (*ye*) which will be placed in RCW. Then, one step right shift is held to shift the windows of EnCo right by using right shift rule (Rule 5). Currently, LAW contains the root ‘দে’ (*de*) and RAW contains the verbal inflexion ‘য়’ (*ye*). 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 ‘ঘুষ’ and verb ‘দেয়’ (where LAW contains noun and RAW contains verb) to construct compound verb ‘ঘুষ দেয়’ which stores in the RAW.

Endless supply of morphological examination right change (Rule 8) is connected to perform semantic investigation between ‘সে’ (Se) and compound action word ‘ঘুষ দেয়’ by the agent connection, *agt*. Consequently, subject ‘সে’ is erased from the hub list, where compound action word ‘ঘুষ দেয়’ stays in the RAW, which is the primary predicate of the sentence. Standard 9 is then connected to move the windows to right and &@entry credit is added to the compound action word as compound action word ‘ঘুষ দেয়’ is the principle expression of the sentence.

At last, Rule 10 is connected to put the sentence tail (STAIL) on the LAW to finish the transformation procedure. After fruition the transformation procedure, the accompanying UNL expression shown in Table 8 will be created by the EnCo.

Table 8: UNL Expression of 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] |

Our method performs outwards in conversion of Bangla compound verb sentence into UNL. Many efficient translators also unable to detect the semantic ambiguity, but our system easily can detect the semantic ambiguity. Table 9 shows the efficiency of our proposed system by comparing our system with Google Translator [18].

Table 9: Result comparison with [18]

|  |  |  |  |
| --- | --- | --- | --- |
| Compound Verb | Universal Word (UW)/ English Word in [18] | Universal Word (UW)/ English Word in proposed system | Correct Universal Word (UW)/ English Word |
| সে ঘুষ খায় | He eats bribe | He takes bribe | He takes bribe |
| সে পানি খায় | He eats water | He drink water | He drink water |
| সে সিগারেট খায় | She ate cigarettes | She smoke cigarettes | She smoke cigarettes |
| আমি চা খাই | I eat tea | I take tea | I take tea |

1. **CONCLUSION & 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. At long last we have demonstrated the change techniques of a compound action word related sentence into the UNL expressions. A Bangla local dialect sentence with compound action word can be effectively changed over into UNL expression by investigation rules, which can later be meant some other dialects utilizing dialect explicit age rules. The proposed configuration can be similarly relevant to different dialects with compound action words. Our future research is to focus on the representation of infinite verb based compound verb in UNL based translation 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.
18. <https://translate.google.com.au/> (last accessed: 12 December, 2018)