

A Brief Analysis of Conversion Process of Bengali Case Structure (Dirukto Shobdo) for Universal Networking Language

Md. Afzalur Rahaman
Lecturer
Department of CSE
Hamdard University Bangladesh

Mahmudul Hasan
Assistant Professor
Department of CSE
Cumilla University

ABSTRACT

The internet now provides us with incredible tools for communicating and accessing information on a variety of topics from anywhere in the world. Despite the abundance of information, language has been considered a great challenge since many countries develop their online information with their native languages. Universal Networking Language (UNL) enables a computer to process and share information across the language barrier. It replicates the functionalities of natural languages processes and allows users to get perceptible information.

This paper describes the conversion process of Bengali sentences into UNL expressions as well as grammatical rules, morphemes, UW, relational model, hyper-graph of attributes, language dictionary, case structures, and expressions of “Dirukto Shobdo” with the aim to develop an infrastructure of Bengali UNL.

General Terms

Language processing, Machine translation

Keywords

Bangla Case Structure, Bangla Word Dictionary, Dirukto Shobdo, Hyper Graph, UNL

1. INTRODUCTION

It is necessary to develop an environment conducive to education and information sharing in order to prepare a country for globalization. International organizations such as UNESCO and the EU are working hard to improve communication between nations where language is a major issue. Even though only one-fourth of the world's population speaks English as their first language, the majority of scientific research, education, and information materials are currently written in English. Instantly, English is recognized as the primary communication language, and individuals devote a significant amount of effort and money to become fluent in it. Furthermore, due to a shortage of communication resources, many trainees may not receive sufficient training, environment, or study materials. Many nations have lost their traditional culture, their own language, and their identity in the pursuit of English proficiency; after all, language is the foundation of culture and social identity. As a result, non-English speakers still face a communication barrier while undertaking business, entertainment, scientific, or government-related topics.

Various countries are currently employing translators to address the communication challenge, despite the fact that their capabilities are limited in comparison to the demand. In this case, a computer translation system has made significant

progress. However, after translating, the user has to work to edit the output document. In addition, it requires language knowledge to edit the translation of the document in the other language. The computer must distinguish between lexical and syntactic ambiguity in order to analyze a sentence in its semantic representation and then derive the correct semantic representation. UNL can overcome these challenges dramatically.

The next section is composed of the major contributions of Bengali UNL researcher. Then UNL architecture and its components like expression, Relation, Attribute, Universal Word, Knowledge Base, and Knowledge Representation are presented followed by Bengali grammars, sentence conversion procedure, a case structure (Dirukto Shobdo), and database for UNL dictionary and lastly challenges and future scopes of Bengali language for UNL are presented.

2. LITERATURE REVIEW

Despite the fact that more than 250 million people speak Bengali, the language has yet to be included in the UNL. This project is still in the works, and the following are some of the contributions:

N. Y. Ali et al. proposed a system to automatically generate semantic expressions from text documents [7]. The objective of the study is to establish semantic relations between the syntactic units of a sentence. In this system, EnConverter provides a framework for morphological, syntactic, and semantic analysis. The researchers also explained the morphological rules of Bengali words for UNL. These rules usually expose modifications in parts of speech in terms of tense, person, and subject of a sentence. Morphology is the field of linguistics that studies the structure of words. It focuses on patterns of word formation across languages and attempts to formulate rules that model the knowledge of the speakers of those languages. In this work, researchers broadly analyzed the morphological analysis of compound words.

M. F. Mridha et al. analyzed the structure of dictionary entries of Bengali morphemes for UNL [3]. It deals with the internal structure of words and how words can be formed. In this paper, separate word dictionary entries for all prefixes and words are developed to combine meaningful words applying rules. The morphological analysis describes that every word is derived from a root word. Because different morphemes are added as suffixes to a root word, it can take on numerous forms. So, the meaning of a word varies for its different transformations.

K. Dey and P. Bhattacharyya developed the Bengali case structure (karok) for UNL [4]. The “karoks” are broadly classified into 6 types, each having a finer categorization into sub-types. In this paper, they have broadly discussed different

categories of karok and classified the UNL relationship for each of them.

M. F. Mridha et al. [27] proposed EnConversion analysis rules for UNL relation by resolution and generation of attributes derived from the Bengali ambiguous sentences. At this juncture, they developed around five hundred EnConversion analytical rules. This EnConverter has been tested and examined for its performance in a public domain hosted by the Russian language server. In the processing, the suggested system used an analysis window and a condition window. The system consists of the tasks of processing of Bangla sentence by a Bangla parser, creation of a linked list of nodes on the basis of an output of parser, extraction of UWs, and generation of UNL expression for the input sentence.

P. Das and A. Das [6] developed a Bengali Noun morphological analyzer. This development gives a very detailed linguistics description of the nominal suffixes that get attached to the regular nominal roots. This tool also identifies the grammatical categories of the nominal tokens accurately. In Bengali, the linguistic analysis includes a list of noun suffixes as well as the grammatical properties of these suffixes.

3. UNL: SYSTEM ARCHITECTURE

UNL is an artificial language that allows information to be processed across linguistic barriers and was created to convey natural language phrases for machine translation purposes. Such information is expressed in an unambiguous way through a semantic network with hyper-nodes, where nodes (that represent concepts) and arcs (that represent relations between concepts) compose the network. UNL has a central repository for all languages, as well as individual repositories for each language server. These are utilized by the UNL's parser module, which consists of the EnConversion and DeConversion processes, which are now referred to as UNLization and NLization, respectively. UNLization translates any natural language to UNL, whereas NLization uses a set of grammatical rules to convert UNL to the appropriate natural language. Nodes, also known as Universal Terms (UWs), are words borrowed from the English language that are distinguished by their placement in a knowledge base of conceptual hierarchies. To offer more information, function words such as determiners and auxiliaries are expressed as attributes to UWs or nodes. The core architecture of UNL is shown in Figure 1.

3.1 Universal Word

Universal words are the basic units of the UNL vocabulary; those are considered as words in any normal language and represented by the node of the universal expression. UW is considered as the lexical item of UNL. The headword in the UW can be English expression, word, phrase, compound word, or a sentence. A constraint list restricts the interpretation of a UW. Constraint list is followed by a list of attributes, which provides information on how the concept is used in a sentence. To reduce the inherent ambiguity of natural language vocabulary, UNL employs English words that have been changed by a set of semantic limitations. In this regard, the language gets an expressive richness from the natural languages but without their ambiguity. For example, the English word "construction" meaning "the action of constructing" and the "final product". Thus, the word "construction" will be paired with two different universal words.

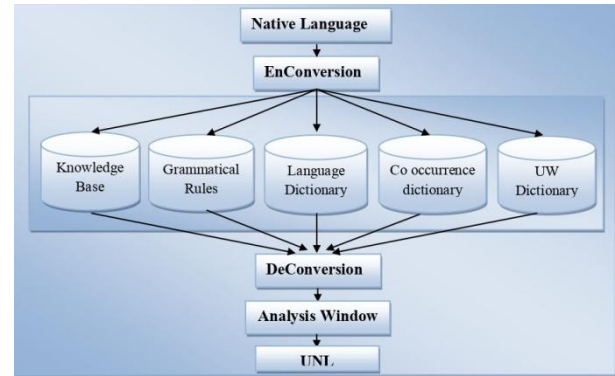


Figure 1: Basic architecture of UNL system

3.2 Attribute Level

The attributes represent the grammatical properties of words. Attributes of UWs are used to describe subjectivity of sentences and the speaker's point of view. This covers things like speech, acts, propositional attitudes, truth values, and so on. Conceptual relations and UWs are used to describe objectivity of sentences. For example, the corresponding UW of play is "play(icl>do)". If the word "play" is in the past form in the sentence an attribute @past is tagged with "play(icl>do)". If it is the main word in the sentence then @entry is tagged such as "play (icl>do),@entry,@past". Following attributes are extensively used in development of UNL for Bengali language.

Table 1: Commonly used attribute list of Bengali UNL

Attribute List	Meaning
BIV/INF	Bivokti (inflexion).
I TH	Inflexion for i th number (i=0 to 7)
VI	Verbal inflexion
PRON	Pronoun
HPRON	Pronoun of human category
HUMN	Noun of human category
SUBJ	Subjective pronoun
#TIM	Time
FUT	Future tense
#PLC	Place
VEND	Vowel ended
CEND	Constant ended
ABY/ANUS	Preposition
NPRO	Proper noun
NULL	No universal word
NMOR	Noun morphology
VMOR	Verbal morphology
ADJMOR	Adjective morphology
ANI	Noun of animal category
3PG	3 rd person general
PRI	Present indefinite
PRC	Present continuous
PRP	Present perfect
PERSON	Form of verb changes according to tense
SUF	Suffix
PRE	Prefix

3.3 Relational Label

There are 46 types of relation in UNL. Relationships between UWs are binary, with different labels for the various roles they play. Strings of three characters or less are used to represent relation labels. When deciding on a relational inventory, many criteria are taken into account. The syntax of UNL relation is defined as follows:

```
agt["<Compound UW-ID>"] "(" {<UW1>|"":<Compound UW-ID>} ", " {<UW2>|"":<Compound UW-ID>} ")"
```

For example, *agt* (*agent*) defines a thing which initiates an action: *agt* (*do, thing*).

Examples and readings:

agt(break(icl>do), John(icl>person)) John breaks.

agt(translate(icl>do), computer(icl>machine)) computer translates .

3.4 UNL Expressions

The UNL represents knowledge in the form of a semantic network, which consists of a collection of binary relationships. UNL expression is identified with the following tags:

{unl} beginning of UNL expression

{/unl} end of UNL expression

There are two kinds of UNL expression, one is table form and the other is list form. Although the table form of a UNL expression is more understandable than the list form, the list form is more compact. There are three sorts of information in UNL expression: binary relations, UWs, and encoded binary relations.

3.5 UNL Document Format

The structure of a UNL document is expressed using the following tags:

[D] beginning of document

[/D] end of document

[P] beginning of paragraph

[/P] end of paragraph

[S] beginning of sentence

[/S] end of sentence

Finally, applying the UNL relations, attributes and document template a sample of UNL expression for the sentence "Monkey eats banana" is shown as follows:

[S]

{unl}

[W]

eat(icl>do).@present.@entry:00

monkey(icl>animal).@generic:01

banana(icl>food).@generic:02

[/W]

[R]

00agt01

00obj02

[/R]

[/S]

3.6 Knowledge Base

The UNL Knowledge Base maintains a database of potential binary relationships between UWs. The knowledge base is a collection of entries in the knowledge base. The syntax of knowledge base entries is mentioned as follows:

```
<Knowledge Base entry> ::= <Binary relations> "=" <degree of certainty>
```

```
<Binary Relation> ::= <Relation Label> "(" <UW1> "," <UW2> ")"
```

```
<degree of certainty> ::= "0" | "1" | ... | "255"
```

The relationship between two UWs is false when the degree of certainty is "0." When the degree of confidence exceeds "1," the relationship between two UWs is true, and the higher the number, the more credible the relationship.

3.7 DeCoverter

A deconverter generates natural language from UNL. The basic architecture of the deconverter is widely shared throughout the world, in order to treat all languages with the same quality. DeCo is a deconversion software developed by UNU/IAS that includes a deconverter, a word dictionary, a co-occurrence dictionary, and conversion rules for a language. This DeCo is a language-independent program that can be used with any language. The fact that the deconverter can accurately convey UNL information is quite important.

3.8 Language Rule

A word dictionary provides information for computers to understand and express information in natural language. There are two kinds of conversion rules. One is an enconversion rule and another is a deconversion rule. The enconversion rule is used to analyze sentences, and the de-conversion rule is to generate sentences. The language rule is expressed by the following tags:

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

Here, characters inside double quotes are the predefined delimiters of the rule. Under the left processing window there is a node for satisfying <COND1> and under the right processing window a node for satisfying <COND2> attributes. The <PRE>, <MID>, and <SUF> nodes with the order of left, middle and right sides of processing windows respectively used to fulfill other conditions. The operations are done on the node-list is depend on the <TYPE> of the rule. The value of <PRIORITY>, which ranges from 0-255, describes the rules interpretation order. A higher priority is indicated by a larger number. For several matching rules, the highest priority matching rule is chosen. Depending on the sentence circumstance, sequences of such rules are activated.

3.9 UNL Dictionary

UNL dictionary is a collection of the word entries. Each entry of the word dictionary is composed of three elements: Headword (HW), Universal Word (UW) and Grammatical Attributes (GA). HW is made up of the input sentences and serves as a trigger for EnConversion to retrieve corresponding UWs from the word dictionary. An UW expresses the meaning of a word and is used to build UNL networks. GA is used in enconversion rules to describe how a word acts in a sentence.

3.10 UNL Hyper-Graph

To mark relations appearing inside this node, hyper-nodes are denoted by a hyper-node ID. We can represent any sentence's logical relationship using a hyper graph. For example, "I saw you yesterday in Dhaka" is represented by the UNL graph as shown in Figure 2.

agt:01(see(icl>do).@past, i(icl>person).@def.@entry)

obj:01(see(icl>do).@past, you(icl>person))

tim(yesterday(icl>time).@past.@entry, :01)

plc(dhaka(icl>plc).@place)

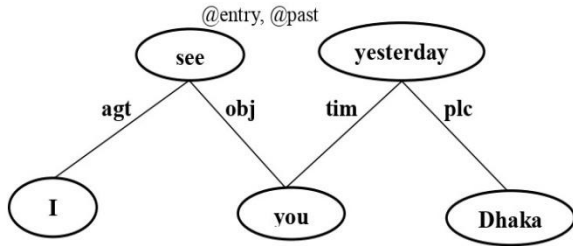


Figure 2: Sentence representation into UNL Hyper-graph

4. UNL FOR BENGALI CASE STRUCTURE: DIRUKTO SHOBDO

Bengali isn't a new language; it was formed from chorjapod roughly a thousand three hundred years ago, and Bangladesh has been ruled by various rulers from Europe, England, and other parts of the world since its inception. Similarly, people of all religions, cultures, and races coexisted intimately with the inhabitants of this subcontinent. As a result, Bengali is more diverse and structured than other languages. This language has a large number of grammatical rules that are beyond the scope of a single research paper.

4.1 Bengali Grammar

Bangla is a morphologically rich semantic language in which the majority of words are generated from roots. To complete the morphological analysis between noun and case inflexion, a morphological rule is applied. By using the 'plc' connection, semantic rule is used to perform semantic analysis between noun and verb. At sentence conversion into UNL expression, INF denotes attribute for inflexion, KROK for case inflexion, N for noun, NP for noun phase, and temporary attributes ('inf' and 'krok') are used to prevent recursive operations. Inflections and derivations are generated by changing vowels and insertion of consonants. In terms of numbers, gender, definitiveness, case, and person, Bengali sentences are distinguished by great agreement between their parts, verb and noun, noun and objective. The following are the linguistic characteristics of roots that have been used in the dictionary: SORANT, BANJANT, and CASE MARKER are three different types of case markers.

In Bengali language prefix words used to express various meanings of the same words. Morphological analysis describes that every word is derived from a root word. A root word may have different transformations. This is due to the addition of distinct morphemes as suffixes. The noun category is used to demonstrate morphology, which is described as follows:

4.1.1 Noun Morphology

From the parts of speech, noun morphology can be accusative, nominative and genitive. Gender and number are also important to identify proper categories of nouns. Noun morphology can be formed as follows:

- [] {} "" (3P, SUF, N) <B,0,0>
- [] {} "" (3P, SUF, N, HUMN, SG) <B,0,0>
- [] {} "" (3P, PL, SUF, N, HUMN) <B,0,0>
- [] {} "" (N, SG, SUF, 3P) <B,0,0>
- [] {} "" (N, PL, SUF, 3P) <B,0,0>
- [] {} "" (N, SG, SUF, 3P) <B,0,0>

4.1.2 Verb Morphology

If (go) is considered as a root, it can be represented in the dictionary as follows:

- [] {} "go (icl>move>do)" (V, @present) <B,0,0>

The verb, adjective, and pronoun morphology have the same dictionary structure.

4.1.3 Bengali Root

ROOT means an attribute for BANJONANT/SORANT. Every root is ended with vowel or consonant. For Bengali roots, URoot is optional, and only a few Roots are featured in URoots. KPG1 and KPG2 are attributes for the names of the kritprottoy groups 1 and 2, respectively. The following is the syntax for kritprottoy:

[HW] {} "UW" (ROOT, BANJONANT/SORANT, URoot, KPG1, KPG2...) <FLG, FRE, PRI>

4.1.4 Kria bivokti

To make kria bivokti, a list of alphabets is added as a suffix to the Bangla root. The PERSON and TENSE attributes in the attribute list are changed according to the Bengali case person and tense, respectively. The following is the syntax of kria bivokti:

[HW] {} "" (BIV, V, PERSON, TENSE ...) <FLG, FRE, PRI>

Bengali kria Bivokti(case and inflexion):

[HW] {} "" (BIV, V, PERSON, TENSE ...) <FLG, FRE, PRI>

In Bengali sentence the structure of kria bivokti for dictionary is given as follows: ঐ (3PG, PRI), ইতেছে (3PG, PRC), ইয়াছে (3PG, PRP), উক (3PG, IMPS), এন (3PR, 2PR, PRI), ইতেছেন (3PR, 2PR, PRC), ইয়াছেন (3PR, 2PR, PRP), উন (3PR, 2PR, IMPS), অ (2PG, PRI), ইতেছ (2PG, PRC), ইয়াছ (2PG, PRP), ও (2PG, IMPS).

4.1.5 Krit Prottoy

Person and inflexion are added to the root verb in Bengali to form verbal pars of speech. The addition of an additional voice to the root verb during sentence generation is known as krit prottoy. For example chal (noun) + onto (krit prottoy) = cholonto. Around 80 rules are used to generate krit prottoy. The UNL dictionary of krit prottoy is given below:

[HW] {} "UW" (KPROT, BENJONANT/SORANT, NOUN/ADJECTIVE/PROJOJOK KRIA, Gname.....) <FLG, FRE, PRI>

The following is a list of regularly used krit prottoy: আ(BANJONANT, NOUN), ওয়া(SORANT, NOUN), ই(BANJONANT, NOUN), আও(BANJONANT, NOUN), আনো(BANJONANT, NOUN), অন্ট(BANJONANT, ADJECTIVE), অন(BANJONANT, NOUN), ভি(BANJONANT, NOUN, ADJECTIVE), ইয়ে(BANJONANT, ADJECTIVE), ও(BANJONANT, NOUN, ADJECTIVE), উয়া(BANJONANT, NOUN, ADJECTIVE)

BENJONANT/SORANT is an attribute in this case, and KritProttoy is added to it. To make Projojok Kria or any combination of them, NOUN/ADJECTIVE qualities are applied to root Kritprottoy.

4.2 Dirukto Shobdo (দ্বিগুণ শব্দ)

In Bengali, "dirukto shobdo" means "two repetitions of a word." We frequently use certain words in conversation to express the meaning of a specific thing. The meaning of those words is changed if we repeat the same term (successive word) or repeat with the negative connotation of the first word. These repeated words are extensively used to represent the depth, grandiose, incompleteness, uncertainty, strength, and phrase word. For example, the sentence "আমার জ্বর জ্বর লাগেছে" means "I'm feeling feverish today". In English, "জ্বর" means "fever" and when we use the word "জ্বর" in Bengali for two times it means someone (subject) feels like fever. A word dictionary is required to construct UNL platform for Bengali language since the translation of repeated words differs. In

Bengali there are several types of dirukto sobdo. The categories of dirukto shobdo, relationships between entries, agents, times, the UNL lexicon, and expressions of certain Bengali sentences with dirukto shobdo are all discussed in the following subsections.

4.2.1 Shobder Dirukti (শব্দের দ্বিগুণিত)

Shobder dirukti (repetition of word) and its UNL dictionary structure is composed as followings:

- In this case, the primary word (main verb) is repeated without any change. Simply “ভাল” means “good”. To strengthen the goodness of a particular thing in Bengali we use “ভাল ভাল” which means “very good or qualified”. Similarly, to express many amounts with big size we use “বড়বড়”.
[ভালভাল] {} “good(iof>quality)” (N) <B,0,0>
[বড়বড়] {} “big(aoj>thing)” (ADJ) <B,0,0>
- A similar word is added with the primary word (main verb), which has no specific meaning in English but is a fairly prevalent Bengali practice. For example, in English “লালন পালন” means foster, where “লালন” and “পালন” have the same meaning but none of them used individually.
[লালন পালন] {} “foster(icl>service)” (V) <B,0,0>
- Sometimes posterior word (verb) is slightly changed to form a sentence; which has no distinct meaning. For example, in English, “বকা-ঝকা” means “chide”, where “ঝকা” (verb) has no individual meaning.
[বকা-ঝকা] {} “chide(icl>do)” (V) <B,0,0>
- Sometimes negative form of primary verb is added to express bidirectional outcomes of anything. For example, Bengali word “ধনী-গরীব”, “লেনদেন”, “ভাল মন্দ” means “rich or poor”, “transaction”, and “good or bad” respectively.
[লেনদেন] {} “transaction (icl>do)” (V) <B,0,0>

4.2.2 Poder Dirukti (পদের দ্বিগুণিত)

Poder dirukti (repetition of parts of speech) can be formed in the following way:

- Applying the same inflection on both words
- Slightly changing the vocal but keeping the inflection and word unchanged.

In Bengali language poder dirukti can be applied on the following scopes:

Noun (বিশেষ্য) পদের দ্বিগুণিত: In Bengali we use some words as noun; those are changed into adjective when applied. For example,

- “রাশি রাশি ধান” means “huge amount of rice”, where simply “রাশি” is a noun.
[রাশিরাশি] {} “many(iof>amount)” (ADJ) <B,0,0>
- “জ্বর জ্বর” means “feel like fever”, where simply “জ্বর” is a noun.
[জ্বর জ্বর] {} “feverish(icl>state)” (N) <B,0,0>
- “দিন দিন” means the change or repetition of an event after a certain interval/sequence of days. For example “সে দিন দিন ভাল হচ্ছে” means “He is being good day by day”. Again, “তিনি মাঝে মাঝে আসেন”
[মাঝেমাঝে] {} “sometimes(icl>time)” (ADJ) <B,0,0>
- “ধীরে ধীরে যাও” means “move slowly”. In this case noun is transforming into verbal adjective.

- “সকাল সকাল” means “early morning”. To start something at very first time or early morning we say, “সকাল সকাল শুরু করবে”, which means “start at early morning”.
[সকাল সকাল] {} “early morning (icl> time)” <B,0,0>

UNL Representation:

Bengali Sentence: আমি জ্বর জ্বর বোধ করছি

Bengali Phonetic: ami jor jor bodh korchi

UNL dictionary

[আমি] {} “I” (icl>person) (1SG, HPRON, PRON, SUB)
<B,0,0>;

[জ্বরজ্বর] {} “feverish(icl>state)” (ADJ)

[বোধ] {} “feel(icl>thing)”

UNL expression

[s]

{unl}

[w]

agt(I(icl>person)).@entry:00

con(feverish(icl>thing)).@entry:01

aoj(feel(icl>thing)).@entry:02

[/w]

[/S]

[R]

00 agt 02

01 met 02

[/R]

{unl}

Adjective (বিশেষণ) পদের দ্বিগুণিত: Repetition of a word can also be used to create an adjective as shown below:

- “ভাল ভাল আম” means “qualified mangoes”
- “গরম গরম জিরাপি” means “hot sweets”
[গরমগরম] {} “hot(icl>state)” (ADJ) <B,0,0>

Pronoun (সর্বনাম) পদের দ্বিগুণিত: Dirukto sobdo can also form pronoun as follows:

- “কেউ কেউ বলে” means “someone says”
[কেউকেউ] {} “someone(icl>person)” (PRON)
<B,0,0>
- “কে কে যাবে” means “who will go”

Verb (ক্রিয়া) পদের দ্বিগুণিত: The repetition of verb can also form an adjective as well as verbal adjective as follows:

- “তার নেই নেই ভাব গেল না” means “he is never satisfied”. Here repetition of verb forming an adjective. “রোগীর যায় যায় অবস্থা” means “The patient is in serious condition”. Here “যায়যায়” means “serious condition” and again forming an adjective.
[যায়যায়] {} “serious (icl>situation)” (ADJ) <B,0,0>
- “ঘুমিয়ে ঘুমিয়ে তুমি শুনলে কিভাবে?” means “How did you hear in sleeping?” The repetition of verb forming verbal adjective.

UNL Representation:

Bengali Sentence: ঘুমিয়ে ঘুমিয়ে কিভাবে শুনলে

Bengali Phonetic: ghumiye ghumiye kivabe sunle

UNL dictionary

[ঘুমিয়ে ঘুমিয়ে] {} “sleeping(icl>state)”

[কিভাবে] {} “how(ict>)”

[শুনল] {} “hear (icl>do)”

[৭] “” (7TH)

UNL expression

[s]

{unl}

[w]
Con(asleep(icl>state)).@entry:00
Cob(hear(icl>occur)).@entry:01
[/w]
[S]
[R]
00 agt 02
01 met 02
[R]
{/unl}

Interjection (অব্যয়) পদের দ্বিগুণিত: The repetition of word can form an interjection like,

- “হায়হায় লোকটি মারা গেল” means “Alas! The man has died”. Repetition of interjection representing the depth of emotion.
- “সে বার বার ডাকছিল” means “He was calling again and again”. Repetition of interjection representing the depth of seriousness.
- “মিট মিট তারা জ্বলছে” means “the stars are twinkling”. Repetition of interjection representing the depth of brightness.
[মিট মিট] {} “twinkle(icl>light)” (ADJ) <B,0,0>
- “সকাল থেকে গুড়ি গুড়ি বৃষ্টি হচ্ছে” means “It has been drizzling since morning”. Repetition of interjection representing the depth of amount.

UNL Representation:

Bengali Sentence: সকাল থেকে গুড়ি গুড়ি বৃষ্টি হচ্ছে

Bengali Phonetic: sokal theke guri guri bristi hocce

UNL dictionary

[সকাল] {} “morning(icl>time)”

[থেকে] {} “since” (5TH)

[গুড়ি গুড়ি] {} “drizzling(icl>quantity)”

[বৃষ্টি] {} “rain(icl>thing)”

[হচ্ছে] {} “has been” (PREP)

UNL expression

[s]

{unl}

[w]

tim(morning(icl>time)).@entry:00

qua(drizzling(icl>thing)).@entry:01

agt(rain(icl>thing)).@entry:02

[/w]

[S]

[R]

00 agt 02

02 met 01

[R]

{/unl}

4.2.3 Dhonattok Dirukti (ধনাত্মক দ্বিগুণিত)

In Bengali, we make up terms to express the meaning of certain sounds when the genuine sounds aren't the same as the phonetic equivalents. We called them phonetic repeated word, which can be formed as follows:

- Without changing the repeated word (phonetic), sounds of human/animal/others is formed. For example, “হি হি” is the sound formed by human that means “sound of laugh” in Bengali. Similarly, “ঘেউঘেউ” means the “barking of dog” and “ঝমঝম” means the “sound of rainfall”.
[ঝমঝম] {} “(icl>sound_of_rainfall)” <B,0,0>
- This form is developed by adding “আ” (vocal of a) with the repeated word. For example, “গগাগগ” and

“পটাপট” means quicker on having food and work respectively.

- This form is developed with adding “ই” (vocal of e) with the repeated word. For example, “ধরাধরি”, “ঝনঝনি” means “holding with attacking mind” and “attractive sound” respectively.
- Dual diversity of phonetic can also form dirukto sobdo. For example, we say “কিচির মিচির”, “টাপুরটাপুর”, “হাপুসহপুস” in Bengali to express the sound of bird, rain and devour respectively.
- Adding “আনি” (ani) with the suffix (প্রত্যয়) phonetic repeated word is formed. For example “my heart was melted on her suffering” means “তার ছটফটানিতে আমরা মনটা খারাপ লেগেছিল”
[ছটফটানি] {} “suffering (icl>state)” (ADJ) <B,0,0>

4.2.4 Repeated word in idioms

Repetition of word can also form idiom as follows:

- To say extra care we use “চোখে চোখে”. For example, “ছেলেটিকে সবসময় চোখে চোখে রাখবে” means “always keep an eye on the boy”.
- To express very much devil we use “হাড়ে হাড়ে”. For example, “লোকটি হাড়ে হাড়ে শয়তান” means “The man is very much devil”.
- To express the topmost quality we use “বাছা বাছা”. For example “বাছা বাছা ফুল এনেছ” means “you have taken the topmost flowers”.
- To express pureness, we also use repeated word as idioms. For example “চকচক করিলেই সোনা হয় না” means “All glitters are not gold”.

5. CONCLUSIONS

Developing a UNL platform for Bengali is a massive and difficult project. The notion of UNL, its functioning principle, sentence construction method, and Bengali grammars for UNL system have all been examined in this study. The various forms, meanings, and rules of a case structure (Dirukto sobdo), as well as the conversion procedure to Universal Networking Language expression, are thoroughly examined. Some Bengali sentences are used to demonstrate how conversion rules are applied to transform any Bengali Dirukta Shobdo statement into a UNL expression. The work's successful implementation will make it easier for us to communicate with others who speak other language over the internet.

6. REFERENCES

- [1] H. Uchida, M. Zhu, "The Universal Networking Language (UNL) Specification Version 3.0", Technical Report, United Nations University, Tokyo, 1998.
- [2] M. N. Y. ALI, S. M. A. A. Mamun, J. K. Das, A. M. Nurannabi, “Morphological Analysis of Bangla Words For Universal Networking Language”, International Conference on Digital Information Management(ICDIM), London, IEEE, 2008.
- [3] M. F. Mridha, K. M. Nur, M. Banik, M. N. Huda, “Structure of Dictionary Entries of Bangla Morphemes for Universal Networking Language (UNL)”, International Journal of Computer Information Systems and Industrial Management Applications, vol. 3, pp. 746-754, 2011.

- [4] K. Dey and P. Bhattacharyya, "Universal Networking Language based analysis and generation of Bengali case structure constructs", *Universal Network Language: Advances in Theory and Applications Research on Computing Science*, pp. 215-229, 2005.
- [5] P. Kumar, , K. Goel, "Universal networking language: A framework for emerging NLP applications", *India International Conference on Information Processing*, 2016. P. Das and A. Das, "Bengali Noun Morphological Analyzer", in *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2013.
- [6] M. N. Y. Ali, S. H. Ripon and S. M. Allayear, "UNL Based Bangla Natural Text Conversion - Predicate Preserving Parser Approach", *International Journal of Computer Science Issues*, 2012.
- [7] M. N. Y. Ali, S. A. Noor, M. Z. Hossain, J. K. Das, "Development of Analysis Rules for Bangla Root and Primary Suffix for Universal Network Language", *International Conference on Asian Language Processing*, IEEE, 2010.
- [8] M. M. Asaduzzaman, M. M. Ali, "Morphological Analysis of Bangla Words for Automatic Machine Translation", *International Conference on Computer, and Information Technology (ICCIT)*, Dhaka, 2003, pp. 271-276
- [9] M. N. Y. Ali, M. Z. H. Sarker, G. F. Ahmed, G. F. Ahmed, "Conversion of Bangla Sentence into Universal Networking Language Expression" , *International Journal of Computer Science Issues*, vol. 8, 2011.
- [10] M. N. Y. Ali, M. A. Ali, A. M. Nurannabi, J. K. Das, "Algorithm for Conversion of Bangla Sentence to Universal Networking Language", *International Conference on Asian Language Processing*, 2010.
- [11] M. Z. H. Sarker, M. N. Y. Ali, J. K. Das, "Outlining Bangla Word Dictionary for Universal Networking Language", *IOSR Journal of Computer Engineering*, vol. 4, pp 14-19, 2012
- [12] M. N. Y. Ali, M. Z. H. Sarker, J. K. Das, "Analysis and Generation of Bengali Case Structure Constructs for Universal Networking Language" *International Journal of Computer Applications*, vol. 18, 2011.
- [13] M. Ershadul, H. Choudhury, M. N. Y. Ali, M. Z. H. Sarker, A. Razib, "Bridging Bangla to Universal Networking Language- A Human Language Neutral Meta-Language", *International Conference on Computer, and Information Technology*, Dhaka, 2005.
- [14] M.N.Y. Ali, J.K. Das, S.M. A.A.Mamun, M. E. H. Choudhury, "Specific Features of a Converter of Web Documents from Bengali to Universal Networking Language", *International Conference on Computer and Commmication Engineering*, pp. 726-731, 2008.
- [15] J. Parikh, J. Khot, S. Dave, P. Bhattacharyya, "Predicate Preserving Parsing", *Department of Computer Science and Engineering*, Indian Institute of Technology, Bombay.
- [16] J. Earley, "An Efficient Context Free Parsing Algorithm", *Communications of the ACM*, 13(2), 1970.
- [17] P. Bhattacharyya, "Multilingual Information Processing Using Universal Networking Language", *Indo UK Workshop on Language Engineering for South Asian Languages LESAL*, Mumbai, India, 2001
- [18] H. Uchida, M. Zhu, "The Universal Networking Language beyond Machine Translation", *UNDL Foundation*, September 26, 2001.
- [19] H. Uchida, M. Zhu, "The Universal Networking Language beyond Machine Translation", *UNDL Foundation*, September 26, 2001.
- [20] M. M. Asaduzzaman, M.M. Ali, "Morphological Analysis of Bangla Words for Automatic Machine Translation", *International Conference on Computer, and Information Technology*, 2003.
- [21] M.E.H. Choudhury, M.N.Y. Ali, "Framework for synthesis of Universal Networking Language", *East West Journal*, Vol 2, 2008.
- [22] N. Adly and S. Alansary, "Evaluation of Arabic machine translation system based on the Universal Networking Language", *Natural Language Processing and Information Systems Lecture Notes in Computer Science 5723*, pp. 243-257, 2009.
- [23] V. Agarwal and P. Kumar, "UNL-ization of Numbers and Ordinals in Punjabi with IAN", *International Journal on Natural Language Computing (IJNLC)*, vol. 2, no. 3, June 2013.
- [24] Choudhury, M.E.H., et al. Bridging Bangla to Universal Networking Language- A Human Language Neutral Meta-Language. in *International Conference on Computer, and Information Technology (ICCIT)*. 2005. Dhaka, Bangladesh.
- [25] M.Z. Hossain, S. A. Noor, M. F. Mridha, "Some Proposed Standard Models for Bangla Dictionary Entries of Bangla Morphemes for Universal Networking Language", *International Journal of Computer Applications (0975-8887) Volume 12- No. 6*. December 2010
- [26] M. F. Mridha, M. N. Huda, C. M. Rahman, J. K. Das, "Development of Morphological Rules for Bangla Root and Verbal Suffix for Universal Networking Language". *6th International Conference on Electrical and Computer Engineering, ICECE 2010*, Dhaka, Bangladesh, 18-20 December 2010, pp.570-573.
- [27] M. S. Rahman, S. R. Poddar, M. F. Mridha, M. N. Huda, "Open Morphological Machine Translation: Bangla to English", *NWESP'2010*, November, India, page, 460-465, ISBN: 978-1-4244-7817-0.
- [28] Md. Zakir Hossain, Shahid Al Noor, Muhammad Firoz Mridha, "Some Proposed Standard Models for Bangla Dictionary Entries of Bangla Morphemes for Universal Networking Language", *International Journal of Computer Applications (IJCA) (0975 –8887), Volume 12– No.6*, December 2010.
- [29] Muhammad Firoz Mridha, Md. Zakir Hossain, Shahid Al Noor, "Development of Morphological Rules for Bangla Words for Universal Networking Language" *IJCSNS International Journal of Computer Science and Network Security*, VOL.10 No.10, October 2010.