

## Dictionary Entries for Bangla Consonant Ended Roots in Universal Networking Language

**Mohammad Zakir Hossain Sarker**

*Department of CSE  
Jahangirnagar University  
Savar, Dhaka, Bangladesh*

*zakir.publications@gmail.com*

**Md. Nawab Yousuf Ali**

*Department of CSE  
East West University  
Dhaka, Bangladesh*

*nawab@ewubd.edu*

**Jugal Krishna Das**

*Department of CSE  
Jahangirnagar University  
Savar, Dhaka, Bangladesh*

*drdas64@yahoo.com*

---

### Abstract

The Universal Networking Language (UNL) deals with the communication across nations of different languages and involves with many different related discipline such as linguistics, epistemology, computer science etc. It helps to overcome the language barrier among people of different nations to solve problems emerging from current globalization trends and geopolitical interdependence. We are working to include Bangla language in the UNL system so that Bangla language can be converted to UNL expressions. As a part of this process currently we are working on Bangla Consonant Ended Verb Roots and trying to develop lexical or dictionary entries for the Consonant Ended Verb Roots. In this paper, we have presented our work by describing Bangla verb, Verb root, Verbal Inflections and then finally showed the dictionary entries for the consonant ended roots.

**Keywords:** Universal Networking Language, Verb Root, Consonant Ended Verb Root, Vowel Ended Verb Root, Verbal Inflections, Dictionary Entry

---

### 1. INTRODUCTION

Universal Networking Language (UNL) is a declarative formal language specifically designed to represent semantic data extracted from natural language texts. It can be used as a pivot language in inter-lingual machine translation systems or as a knowledge representation language in information retrieval applications. Currently, the UNL includes 16 languages [1], which are the six official languages of the United Nations (Arabic, Chinese, English, French, Russian and Spanish), in addition to the ten other widely spoken languages (German, Hindi, Italian, Indonesian, Japanese, Latvian, Mongol, Portuguese, Swahili and Thai). In the last few years, machine translation techniques have been applied to web environments. The growing amount of available multilingual information on the Internet and the Internet users has led to a justifiable interest on this area. Hundreds of millions of people of almost all levels of education, attitudes and different jobs all over the world use the Internet for different purposes [2], where English is the main language of the Internet. But English is not understandable for most of the people. Interlingua translation programs are needed to develop. The main goal of the UNL system, which allows users to visualize websites in their native languages, is to provide a common representation for accessing Internet of multilingual websites by the majority of the people over the world. For this common representation, lexical (dictionary) knowledge is a critical issue in

natural language processing systems, where the development of large-scale lexica with specific formats capable of being used by distinguished applications, in particular to multilingual systems, has been given special focus. Our goal is to include Bangla in this system with less effort. To do so we have working on UNL, Bangla Grammar and then the step by step process for including Bangla in the UNL system. As a part of the process, in this paper, we have explained and described Bangla Verb, Verb Root especially Consonant Ended Verb Root, Verbal Inflections, etc. and dictionary entries of Consonant Ended Verb Roots and their Verbal Inflections.

The organization of this paper is as follow: In Section 2 we describe the Research Methodology, Section 3 has the detail about UNL, Section 4 describes Bangla grammar especially verb, verb roots, etc. in detail. In Section 5 and 6 we discuss about the dictionary entries which we have designed and developed to convert Bangla sentence. Finally, Section 7 draws conclusions with some remarks on future works.

## **2. LITERATURE REVIEW**

For converting Bangla sentence to UNL expressions firstly, we have gone through Universal Networking Language (UNL) [3, 4, 5, 6, 7, 8] where we have learnt about UNL expression, Relations, Attributes, Universal Words, UNL Knowledge Base, Knowledge Representation in UNL, Logical Expression in UNL, UNL systems and specifications of Enconverter. All these are key factors for preparing Bangla word dictionary, enconversion and deconversion rules in order to convert a natural language sentence (here Bangla sentence) into UNL expressions. Secondly, we have rigorously gone through the Bangla grammar [9, 10, 11], Verb and roots (Vowel ended and Consonant Ended) [9, 10, 12], Morphological Analysis [12, 13, 14], Primary suffixes [9, 10, 14, 15], construction of Bangla sentence [9] based on semantic structure. Using above references we extort ideas about Bangla grammar for morphological and semantic analysis in order to prepare Bangla word dictionary (for verb root, verbal inflections, etc)in the format of UNL provided by the UNL center of the UNDL Foundation.

## **3. UNIVERSAL NETWORKING LANGUAGE**

The UNL is an acronym for “Universal Networking Language”. It is a computer language that enables computers to process information and knowledge across the language barriers. It is an artificial language that replicates, in the cyber world, the functions of natural languages in human communication. As a result, it enables people to express all knowledge conveyed by natural languages. It also enables computer to intercommunicate, thus providing people with a linguistic infrastructure for distributing, receiving and understanding multilingual information. [4]

The UNL expresses information or knowledge in the form of semantic network with hyper-node. Different from natural languages, UNL expressions are unambiguous. In the UNL semantic network, nodes represent concepts, and arcs represent relations between concepts. Concepts can be annotated. Since the UNL is a language for computers, it has all the components of a natural language. It is composed of words expressing concepts called “Universal Words”, also referred to as UWs which are inter-linked with other UWs to form sentences. These links, known as “relations”, specify role of each word in a sentence. The subjective meaning intended by the speaker can be expressed through “attributes”.

The “Knowledge Base (UNLKB)” is provided to define semantics of UWs. The UNLKB defines every possible relation between concepts including hierarchical relations and inference mechanism based on inclusion relations between concepts. Thus, the UNLKB provides semantic background of the UNL to make sure the meaning of the UNL expressions is unambiguous.

In the last few years, Machine Translation (MT) techniques have been applied to web environments. The growing amount of available multilingual information on the Internet and the Internet users has led to a justifiable interest on this area. Hundreds of millions of people of almost all levels of education, attitudes and different jobs all over the world use the Internet for different purposes [3], where English is the main language of the Internet. But English is not understandable for most of the people. Interlingua translation programs are needed to develop.

The main goal of the UNL system, which allows users to visualize websites in their native languages, is to provide a common representation for accessing Internet of multilingual websites by the majority of the people over the world. A significant part of the development of any machine translation (MT) system is the creation of lexical resources that the system will use. Dictionaries are of critical importance in MT. They are the largest components of an MT system in terms of the amount of information they hold. Generation of natural language from a machine processable, precise knowledge representation has to grapple with the problem of redundancy and impreciseness inherent in any natural language. In the UNL System [4], natural language analysis has been carried out by the EnConverter (EnCo) [7] associated with Word Dictionary of a native language and language specific analysis rules that converts a native language text into UNL expression and DeConverter (DeCo) [8] also associated with Word Dictionary that converts UNL expression to a variety of native languages using language specific generation rules. Our goal is to include Bangla in this system with less effort and we have been working for the last 4 years to achieve our goal. As a part of this process we are presently working on the process to include Bangla Consonant Ended Verb Root in the dictionary so that any Bangla sentence having consonant ended verb roots can be converted to UNL expression. In the following sections we have described the process more elaborately.

#### **4. BANGLA VERB, VERB ROOT, VERB INFLECTIONS**

Bangla or Bengali (বাংলা, pronounce as bangla) is an eastern Indo-Aryan language. It is native to the region of eastern South Asia known as Bengal, which comprises present day Bangladesh, the Indian state of West Bengal and parts of the Indian states of Tripura and Asam. It is sixth widely spoken language in the world with more than 230 million speakers. Bengala is written in a script called the Bengali script. Like other Indian languages, the letters in the Bengali script are grouped together based on the way they are pronounced. The first 11 letters are all vowels followed by consonants and finally the semi vowels. The consonants are grouped based on how they are pronounced. First comes the velar consonants, then the palatal, the retroflex, the dental, and the labial consonants.

As in this paper we have concentrated on Consonant ended verb root, lets discuss on bangla verb and related terms before describing our work.

##### **2.1 Verb (ক্রিয়া)**

Bangla verbs are highly inflected and are regular with only few exceptions. They consist of a stem and an ending; they are traditionally listed in Bangla dictionaries in their "verbal noun" form, which is usually formed by adding –a(আ) to the stem, for instance, রাখা (rakha) = "to put or place". The stem can end in either a vowel or a consonant. Verbs are conjugated for tense and person by changing the endings, which are largely the same for all verbs. However, the stem vowel can often change as part of the phenomenon known as "vowel harmony", whereby one vowel can be influenced by other vowels in the word to sound more harmonious. An example would be the verb "to write", with stem lekh-: তোমরা লিখ (tomra likho) meaning "you (pl.) write" but আমরা লিখি (amra likhi) meaning "we write".

Bangla language has more than 30000 [9] verbs. Diversity of verb morphology in Bangla is very significant. For example, if we consider "লিখ" (likh means write) as a root word then after adding verbal inflexion "ইতেছি" (itechhi), we get a word "লিখিতেছি" (likhitechhi means am writing) which means a work is being doing in present (for first person). Similarly, after adding inflexion "ইতেছিলাম" (itechhilam) we get the word "লিখিতেছিলাম" (likhitechhilam means was writing) which means a work was being done in past. Here, one word represents present continuous tense of the root word "লিখ" (likh) and another represents past continuous tense. Therefore, by morphological analysis we get the grammatical attributes of the main word and other attributes. For this reason we have applied morphological analysis for different persons with different transformations to find out the actual meaning of the word. Morphological analysis of Bangla

verbs has been considered in different works [1, 2]. We show some data for root verbs যা (go) shown in Table 1.

Person/Tense	Verb as appears in a sentence	Inflectional Suffix
First		
Present	যাই (jai)	ই (i)
Present Continuous	যাচ্ছি (jachhi)	চ্ছি (chhi)
Present Perfect	গিয়েছি (giechhi)	এছি (chhi)
Past	জালাম (gelam)	লাম (lam)
Past Continuous	যাচ্ছিলাম (jachhilam)	চ্ছিলাম (chhilam)
Past Perfect	গিয়েছিলাম (giechhilam)	এছিলাম (echhilam)
Past Habitual	যেতাম (jetam)	এতাম (etam)
Future	যাবো (jabo)	বো (bo)
Second		
Present	যান (jan)	ন (n)
Present Continuous	যাচ্ছেন (jachhen)	চ্ছেন (chhen)
Present Perfect	গিয়েছেন (giechhen)	এছেন (echhen)
Past	জালেন (gelen)	এলেন (elen)
Past Continuous	যাচ্ছিলেন (jachhilen)	চ্ছিলেন (chhilen)
Past Perfect	গিয়েছিলেন (giechhilen)	এছিলেন (echhilen)
Past Habitual	যেতেন (jeten)	এতেন (eten)
Future	যাবেন (jaben)	বেন (ben)
Third		
Present	যায় (jae)	য় (e)
Present Continuous	যাচ্ছে (jachhe)	চ্ছে (chhe)
Present Perfect	গিয়েছে (giechhe)	এছে (echhe)
Past	জালো (gelo)	লো (lo)
Past Continuous	যাচ্ছিলো (jachhilo)	চ্ছিলো (chhilo)
Past Perfect	গিয়েছিলো (giechhilo)	এছিলো (echhilo)
Past Habitual	যেতো (jeto)	তো (to)
Future	যাবে (jabe)	বে (be)

**TABLE 1:** Morphology of root verb যা (go)

As verbs come from roots and verbal inflexions, for appropriate morphological analysis, we have divide Bangla verb roots into two categories.

- **Vowel Ended Roots:** In Bangla there are around 25 vowel ended roots e.g পা (pa), খা (kha), গা (ga), চা (cha), ছা (chha), নি (ni), দি (di), যা (ja), ছুঁ (chhu), থু (thu), শু (shu), ধু (ddhu),

ন (n), দু (dhu), নু (nu), রু (ru), হ (h), ধা (dha), না (na), বা (ba), ক (ko), ব (bo), র (ro) and স (sho).

- **Consonant Ended Roots:** There are around around 1500 consonant ended roots in Bangla Language. For examples, কর (kor), খেল (khel), গড় (gor), ঘষ (gosh), বখ (bokh), কহ (koh), গিল (gil), পিশ (pish), শিখ (shikh), লিখ (likh), ধর (dhor), উঠ (ut), বুজ (buj), ভুল (vul) etc. are consonant ended roots.

### 5. Our Proposed Template for Consonant Ended Verb Roots

For appropriate morphological analysis and designing template of verb roots, verb roots have been divided into two broad categories according to tenses and persons namely Vowel Ended Group (VEG) and Consonant Ended Group (CEG). Each of them again divided into sub-groups. Categorization of Consonant Ended Groups are shown in the following tables. (for first person only)

		Consonant Ended Roots
	Tenses	কম, কর, কশ, কষ, খেঁচ, খেপ, খেল, গড়, গল, ঘট, ঘষ, ঘঁষ, চট, চড়, চর, চল, চষ, চেত, ছড়, ছল, ছেঁক, ছেঁচ, জপ, জম, জর, জ্বল, ঝর, টক, টল, ঠক, ঠেল, ঠেস, ডল, ঢল, দম, দল, দেখ, ধর, ধস, পড়, পর, ফল, ফেল, বক, বখ, বন, বল, বস, বেচ, বেড়, বেল, ভজ, ভর, মজ, মল, মেল, রট, রস, রোপ, লড়, লেপ, সঁপ, সর, সেক, সেচ, হট, হর, হের, হেল, হেস, থস, ঘোষ
Present	Present Indef	ই
	Present Cont	ছি
	Present Perfect	এছি
	Imperative	*
Past	Past Indef	লাম
	Past Habitual	তাম
	Past Cont.	ছিলাম
	Past Perfect	এছিলাম
Future	Future Indef.	বো, ব
	Imperative	*
		Group CEG1

**TABLE 2:** Variations of Consonant Ended Roots and their Verbal Inflexions of CEG1 for First Person

		Consonant Ended Roots	
Tenses		কহ, দহ, বহ, রহ, সহ	নহ
Present	Present Indef	ই	ই
	Present Cont	ইতেছি	*
	Present Perfect	ইয়াছি	*
	Imperative	*	*
Past	Past Indefinite	ইলাম	*
	Past Habitual	ইতাম	*
	Past Continuous	ইতেছিলাম	*
	Past Perfect	ইয়াছিলাম	*
Future	Future Indefinite	ইব, ইবো	*
	Imperative	*	*
		<b>Group CEG2</b>	<b>Group CEG3</b>

**TABLE 3** Variations of Consonant Ended Roots and their Verbal Inflexions of CEG2 and CGE3 for First Person

		Consonant Ended Roots
	<b>Tenses</b>	আঁক, আঁচ, আঁট, আস, কাঁক, কাঁচ, কাঁড়, কাঁদ, কাঁপ, কাচ, কাট, কাড়, কাল্দ, কাশ, খাট, খাপ, গাঁজ, গাঁথ, গাড়াগাদ, গাব, গাল, ঘাট, ঘাম, চাঁচ, চাখ, চাট, চাপ, চাল, ছাঁক, ছাট, ছাঁদ, ছাড়, ছান, জাঁক, জাগ, জার, জাল, ঝাঁক, ঝাড়, ঝাঁপ, ঝাল, টান, ঠার, ঠাস, ডাক, ঢাক, ঢাল, তাত, তাপ, থাক, থাম, দাগ, দাব, ধাঁদ, ধাঁধ, ধার, নাচ, নাড়, নাম, পাক, পাড়, পাত, পার, ফাঁদ, ফাঁপ, ফাঁস, ফাট, ফাড়, বাঁক, বাঁচ, বাঁট, বাঁধ, বাছ, বাজ, বাট, বাড়, বাস, ভাঁজ, ভাগ, ভাঙ, ভাপ, ভাজ, ভান, ভাব, ভালবাস, ভাস, মাখ, মাগ, মাপ, মাজ, মাড়, মাত, মান, মাপ, মার, যাচ, রাঁধ, রাখ, রাগ, লাগ, লাদ, শান, শাস, সাজ, সাঁট, সাধ, সান, সার, হাঁক, হাঁচ, হাজ, হাট, হান, হার, হাস
Present	Pre Ind	ই
	Present Cont	ছি
	Prese Perf	আঁক>এঁক এছি
	Imperative	*
Past	Past Indef	লাম
	Past Habit	তাম
	Past Conti	ছিলাম
	Past Perfect	আঁক>এঁক এছিলাম
Future	Future Indef	ব
	Imperative	*
		<b>Group CEG4</b>

**TABLE 4:** Variations of Consonant Ended Roots and their Verbal Inflexions of CEG4 for First Person

Consonant Ended Roots			
	<b>Tenses</b>	কিন্,খিঁচ্,গিল্,ঘিন্ন্,চিন্,চিন্ন্,ছিঁড়্,জিত্,জিন্,টিক্,টিপ্,নিব্, পিজ্,পিট্,পিশ্,ফিন্ন্,বিধ্,ভিজ্,ভিড়্,মিট্,মিল্,মিশ্,লিখ্,শিখ্,সিচ্	<b>আছ্</b>
Present	Pres. Indef.	ই	ই
	Pres. Cont	ছি	*
	Pres. Perf.	এছি	*
	Imperative	*	*
Past	Past Indef	লাম	*
	Past Habit.	তাম	*
	Past Cont.	ছিলাম	*
	Past Perf.	এছিলাম	*
Future	Fut Indef	ব	*
	Imperative	*	*
<b>Group CEG5</b>			<b>Group CEG6</b>

**Table 5:** Variations of Consonant Ended Roots and their Verbal Inflexions of CEG5 and CEG6 for First Person

Consonant Ended Roots				
	<b>Tenses</b>	উঠ্,উড়্,উব্,কুঁদ্,কুদ্,কুট্,কুর্,খুঁজ্,খুট্,খুঁড়্,খুদ্,খুল্,গুঁজ্,গুগ্,গুল্,ঘুট্,ঘুচ্,ঘুর্,চুক্, চুন্,চুশ্,ছুঁড়্,ছুট্,ছুড়্,ছুল্,জুট্,জুড়্,ঝুল্,টুক্,টুট্,ঠুক্,ঠুস্,ডুব্,চুক্,টুঁড়্,তুল্,তুল্, তুশ্,দুল্,দুশ্,ধুঁক্,ধুল্,পুঁত্,পুঁত্,পুছ্,পুড়্,পুর্,পুশ্,ফুঁক্,ফুঁড়্,ফুট্,ফুল্,বুঁজ্, বুজ্,বুঝ্,বুল্,বুল্,ভুগ্,ভুল্,মুছ্,মুড়্,মুত্,মুদ্,রুখ্,রুধ্,লুট্,লুফ্,শুঁক্,শুধ্,শুন্,শুশ্		
Present	Present Indef.	ই		
	Present Cont	ছি		
	Present Perfect	এছি		
	Imperative	*		
Past	Past Indef	লাম		
	Past Habitual	তাম		
	Past Cont.	ছিলাম		
	Past Perfect	এছিলাম		
Future	Future Indef.	ব		
	Imperative	*		
<b>Group CEG7</b>				

**Table 6:** Variations of Consonant Ended Roots and their Verbal Inflexions of CEG7 for First Person

We have also developed such template for second and third person also.

## 6. OUR PROPOSED TEMPLATE FOR DICTIONARY ENTRIES

The template that has been developed in this paper for Bangla verb roots is depicted bellow:

[HW] {} “UW(icl/iof...>concept1>concept2..., REL1>...,REL2>...,” (ROOT, VEND/CEND [,ALT/ ALT1/ALT2..] VEGn/CEGn, #REL1, #REL2, ... <FLG, FRE, PRI>

where,

- HW← Head Word (Bangla Word; in this case it is Bangla root);
- UW← Universal Word (English word from knowledge base);
- icl/iof/... means inclusion/instance of ...to represent the concept of universal word
- REL1/REL2., indicates the related relations regarding the corresponding word.
- ROOT ← It is an attribute for Bangla roots. This attribute is immutable for all Bangla roots.
- CEND and VEND are the attributes for consonant ended and vowel ended roots respectively as every root is ended either with consonant or vowel;
- VEGn ← means attribute for the group number of vowel ended roots
- CEGn ← means attribute for the group number of consonant ended roots
- ALT, ALT1, ALT2 etc. are the attributes for the first, second and third alternatives of the vowel or consonant ended roots respectively. If the root is default, then no alternative is used.
- #REF1, #REF2 etc. are the possible corresponding relations regarding the root word.

Here, attributes say, ROOT, CEND/VEND are fixed for all Bangla roots whereas ALT or ALT1 or ALT2 etc. does not necessary for all roots, they are used only for alternative roots.

In the following examples we are constructing the dictionary entries for some sample Bangla roots using our designed template:

[কর,]{ } “do(icl>do, agt>thing, obj>process)” (ROOT, CEND, CEG1, #OBJ, #AGT,#PLC)<B,0,0>

[আঁক্]{ } “draw(icl>get>do,equ>reap,src>thing,agt>thing,obj>thing)”(ROOT,CEND, CEG4,#AGT,#OBJ,#PLC) <B,0,0>

[কিন্]{ } “buy(icl>get>do,cob>thing,src>thing,agt>person,obj>thing)”(ROOT, CEND,CEG5,#AGT,#OBJ,#PLC)<B,0,0>

where, ROOT denotes Bangla root, CEND for consonant ended root, CEG for consonant ended group. #AGT, #OBJ and #PLC indicate that concerned headwords can be used to make agent (agt), object (obj) and place (plc) relation respectively. Similarly, other entries have been developed according to the format discussed above.

## 7. CONCLUSION AND FUTURE WORK

We have shown dictionary entries for Bangla Consonant Ended Roots that are useful for conversion of Bangla sentences to UNL expression. So with this we have developed dictionary entries for both i.e. vowel ended and consonant ended verb roots. Now in future we'll work on developing rules for these and then we'll use the en- converter to see the result.

## 8. REFERENCES

1. Md. Ershadul H. Choudhury, Md. Nawab Yousuf Ali, Mohammad Zakir Hossain Sarker, Ahsan Razib. "Bridging Bangla to Universal Networking Language- A Human Language Neutral Meta-Language", In proceedings International Conference on Computer, and Information Technology (ICIT), Dhaka, 2005 pp. 104-109
2. Md. Nawab Yousuf Ali, Mohammad Zakir Hossain Sarker , Ghulam Farooque Ahmed , Jugal Krishna Das."Conversion of Bangla Sentence into Universal Networking Language Expression", International Journal of Computer Science Issues, Vol. 8, Issue 2, March 2011, pp. 64-73
3. H. Uchida, M Zhu. T.G. Della Senta. Universal Networking Language, 2005/6-UNDL Foundation, International Environment House.
4. H. Uchida, M. Zhu. The Universal Networking Language (UNL) Specification Version 3.0 Edition 3 , Technical Report, UNU, Tokyo:, 2005/6-UNDL Foundation, International Environment House, 2004
5. Enconverter Specification Version 3.3, UNU Centre, Tokyo 150-8304, Japan 2002
6. <http://www.undl.org> last accessed on 29 July 2012
7. EnConverter Specification, Version 3.3, UNL Center/UNDL Foundation, Tokyo 150-8304, Japan 2002
8. DeConverter Specification, Version 2.7, UNL Center, UNDL Foundation, Tokyo 150-8304, Japan 2002
9. D.M. Shahidullah. Bangla Baykaron, Dhaka: Ahmed Mahmudul Haque of Mowla Brothers prokashani, 2003
10. D. C. Shuniti Kumar. Bhasha-Prakash Bangala Vyakaran, Calcutta : Rupa and Company Prokashoni, July 1999, pp.170-175
11. Humayun Azad. Bakkotottoyo - Second edition, Dhaka: Bangla Academy Publishers, 1994
12. D. S. Rameswar. Shadharan Vasha Biggan and Bangla Vasha, Pustok Biponi Prokashoni, November 1996, pp.358-377
13. M.N.Y. Ali, J.K. Das, S. M. Abdullah 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 (ICCC'E'08), Kuala Lumpur, Malaysia. pp. 726-731
14. M.N.Y. Ali, J.K. Das, S.M. Abdullah Al Mamun, A. M. Nurannabi. "Morphological Analysis of Bangla words for Universal Networking Language", International Conference on Digital Information Management, icdim, 2008, London, England, pp. 532-537
15. M.N.Y.Ali, A. M. Nurannabi, G. F. Ahmed, J.K. Das. "Conversion of Bangla Sentence for Universal Networking Language", International Conference on Computer and Information Technology (ICIT), Dhaka, 2010 pp. 108-113