The Korea Academic Society of Digital Business Administration (KASDBA)

(사)한국디지털경영학회

ICIDB-2018
Conference Program

International Conference on Information and Communication Technology and Digital Convergence Business
(ICIDB-2018)
Hanoi University of Science and Technology
Hanoi, Vietnam, January 19-21, 2018

Presentation Session 2-3, ICT Convergence

Session Chair: Dr. Nguyen Dinh Thuan
Place: Room No. 303
Date: Friday, January 19th, 2018
Time: 15:15 – 17:15

Cinematic Storytelling Game "Noah Ark": Implementation and Design (Woo-Hyun Park, Yun-Gyung Cheong)
SitBuCE: A Small Data-Driven Generalization Strategy for Technology Maturity Status Mapping (Adigun M.O, Kabanda S.K, Thinyane M)
An Efficient Conversion Approach of the Bangla Infinite Verb Sentence into UNL (Md. Nawab Yousuf Ali, Golam Sorwar, Md. Shamsujjoha)
The Structural Influences of Suicide-related Keywords Search on Suicide Rate in Korea (Hye-Ryeon Young, Ha Na Kang, Ghee-Young Noh, Hyun-Seok Hwang)
A Novel Similarity Measure for Trace Clustering Based on Normalized Google Distance (Hong-Nhung BUI, Quang-Thuy HA, Tri-Thanh NGUYEN)
Factors Affecting Social-Presence and Self-Presence in Virtual Reality Media (Sang-Min Park, Min-Ji Choe, Jeong-Min Park, Ghee-Young Noh)

AN EFFICIENT CONVERSION APPROACH OF THE BANGLA INFINITE VERB SENTENCE INTO UNL

Md. Nawab Yousuf Ali, Golam Sorwar[‡], Md. Shamsujjoha[‡]

[*‡]Department of Computer Science and Engineering
East West University, Dhaka-1212, Bangladesh
Email: [‡]nawab@ewubd.edu, [*]dishacse@yahoo.com

[‡]School of Business & Tourism
Southern Cross University, Australia
Email: [‡]golam.sorwar@scu.edu.au

Abstract

This paper presents conversion procedure of Bangla infinite verb sentences to assimilate them into an interlingua representation called Universal Networking Language (UNL). It focuses the analysis of infinite verbs and develops the morphological rules to resolve morphological analysis between the infinite and finite verbs. This paper also develops semantic rules to perform semantic analysis between the words in a sentence for the EnConverter to convert infinite verb sentence into UNL Expression. Finally, we have shown the conversion procedures of a Bangla infinite verb sentence into UNL.

Keywords :

UNL, NLP, Infinite Verb, EnConverter, Universal Words.

JP Journal of Heat and Mass Transfer

http://www.pphmj.com
http://dx.doi.org/10.17654/
Volume …, Number …, 2018, Pages … **ISSN: 0973-5763**

AN EFFICIENT CONVERSION APPROACH OF THE BANGLA INFINITE VERB SENTENCE INTO UNL

Md. Nawab Yousuf Ali[1,*], Golam Sorwar[2] and Md. Shamsujjoha[2]

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

[2]School of Business and Tourism
Southern Cross University
Australia

Abstract

This paper presents conversion procedure of Bangla infinite verb sentences to assimilate them into an interlingua representation called Universal Networking Language (UNL). It focuses the analysis of infinite verbs and develops the morphological rules to resolve morphological analysis between the infinite and finite verbs. This paper also develops semantic rules to perform semantic analysis between the words in a sentence for the EnConverter to convert infinite verb sentence into UNL Expression. Finally, we have shown the conversion procedures of a Bangla infinite verb sentence into UNL.

I. Introduction

UNL is developed by the United Nations University/Institute of

Received: January 16, 2018; Accepted: January 31, 2018
Keywords and phrases: UNL, NLP, infinite verb, EnConverter, universal words.
A preliminary version of this paper was presented at the ICIDB-2018, Hanoi, Vietnam.
*Corresponding author
Communicated by Gyanendra Prasad Joshi

Advanced Studies (UNU/IAS) to convert a native language text to UNL expressions, and hence this UNL expression can be converted to any other native languages [1]. A UNL based conversion system uses two components: one is EnConverter [1-2] that converts source language to UNL Expression, and the other one is DeConverter [1] for conversion from UNL Expression to target language. The UNL system has 46 semantic relations and 87 attributes to express the semantic meaning of a sentence in a source language [1]. Bangla text is converted to UNL expressions by the EnConverter of Bangla language. This UNL expression is converted to any other language that has its own DeConverter. This will indeed help to develop a multilingual machine translation system for Bangla language. This paper is organized as follows: The formation of dictionary entries of Bangla Word is presented Section 2. Analysis of infinite verbs based on UNL format along with the development of analysis rules is explained in Section 3. Section 4 shows the experimental analysis of a Bangla infinite verb sentence into UNL expression by applying analysis rules. Section 5 summarizes the paper with some concluding remarks and future directions.

II. UNL Based Bangla Lexicon

The lexicon is a collection of word dictionary entries. Each entry is composed of three kinds of elements: Headword (HW), Universal Word (UW) and Grammatical Attribute (GA). Each dictionary entry has the following format associating with any native language word [1-3].

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

Here, HW stands for Head Word (Bangla Word), ID for Identification of Head Word (omissible), UW for Universal Word, ATTRIBUTE for attribute of the HW, FLG for Language Flag, FRE for Frequency of Head Word, PRI for Priority of Head Word respectively. We use transliterated Bangla text in this paper and use *tt* for transliterated text. For example:

[kha]{}"eat(icl>consume>do,agt>living_thing,obj>concrete_thing,ins>thing "(ROOT, VEND,VEG1)

[kolom] {} "pen(icl>writing_implement>thing)" (N, NCOM, CEND, OBJ),

where, attributes ROOT stands for verb root, VEND for vowel ended root, VEG for vowel ended group 1, N for noun, NCOM for common noun, CEND for consonant ended word, OBJ for object respectively.

III. Analysis of Infinite Verbs and the Development of Analysis Rules for UNL Conversion

Unlike finite verb an infinite verb (tt: *oshomapika kria*) does not reflect the person and tense and can't use independently from other verbs to complete a sentence. Infinite verbs can be formed by adding three different verbal inflexions (VI): tt: *te*, tt: *e*, tt: *le* or with the verb roots (vowel or consonant ended roots). For example: tt: *kha* meaning eat is a vowel ended root which is added with the inflexions tt: *e*, tt: *le* or tt: *te* to form three different infinite verbs as follows [4].

- Inflexion Infinite Verb (tt: *tumorthok kria*)
- Perfect Conjunctive Infinite Verb (tt: *puraghotito songjojok kria*)
- Conditional Conjunctive Infinite Verbs (tt: *shortogapok shonjojok kria*)

This paper focuses only on the inflexion infinite verbs that can be formed with inflexion tt: *te*. Some specific roots are primarily used to construct finite verbs in the sentences where infinite verbs are formed with inflexion tt: *te*. The roots are: (a) tt: *di* meaning to give, (b) tt: *ja* meaning to go, (c) tt: *cha* meaning to want, (d) tt: *ni* meaning to take, (e) tt: *ho* meaning auxiliary verb (am/is/are/has/have/was/were, (f) tt: *par* meaning can, (g) tt: *jan* meaning to know and (h) tt: *dhor* meaning to catch. Some of these roots explain the views of the speaker only. No Universal words (UWs) are used for these roots in UNL expressions rather UNL attribute that describes the speakers view is added in the attribute list of UW of the root of infinite verb. Any infinite verbs with inflexion tt: *te* that are used with finite verbs made by the above roots indicate the 'to permit for action' of the infinite verbs. If we consider a sentence, say tt: *Se take khete dilo* meaning He gave him to eat. In

the sentence, tt: *khete* is the infinite verb where, tt: *khe* (to eat) is the deform root of root tt: *kha* & tt: *te* is the verbal inflexion and tt: *dilo* is the finite verb where, tt: *di* (to give) is the root and tt: *lo* is the verbal inflexion. Here, tt: *khete dilo* meaning 'gave to eat' denotes 'gave permission for eating'. To implement in UNL, the UWs of these verbs are directly added into the condition field of RAW (Right Analysis Windows) of EnConverter [1] and corresponding UNL attributes are to be added in action field of RAW at the UW of the root of infinite verb. We have thoroughly analyzed all the roots and constructed the dictionary entries, contents of the condition fields, UNL attributes of the actions fields, semantic rules to resolve the analysis for the roots, and UNL relations in the UNL expressions shown in the Table 1 and Table 2. In this process, we use a tool presented in [6]. Table 1 shows UNL based.

Table 1. Dictionary entries of the roots of finite verb

Roots of finite verbs	Dictionary entries of the roots
tt: di	**[**di**]{}"**give(icl>do,equ>handover,agt>living_thing,obj>concrete_thing)"(V, VER)
tt: ja	[ge]{}"go(icl>move>do,plt>place,plf>place,agt>thing)"(V, VER,VEG3, ALT2)
tt: cha	**[**cha**]{}"**want(icl>desire>be,obj>uw,aoj>volitional_thing,pur>thing))"(V, VER)
tt: *ni*	**[**ni**]{}"**(take(icl>perceive>be,obj>thing,aoj>person,man>uw"(V, VER,VEG2)
tt: ho	**[**ho**]{}"**be(icl>be)"(V, VER,VEG8)
tt: par	**[**par**]{}"**can**(**icl>do,agt>thing,obj>process)"(VERB,CER, CEG4)
tt: jan	**[**jan**]{}"**know(icl>be,equ>cognise,obj>uw,aoj>volitional_thing)"(V, CER, CEG4)
tt: *dhor*	**[**dhor**]{}"**catch(icl>hook>occur,obj>thing)"(V, CER, CEG1)

Bangla Dictionary entries of the roots of finite verbs. Table 2 shows the contents of the condition fields, contents of UNL attributes in the action field, the UNL relations between the words in the sentence, and rules for dropping finite verbs in the UNL expressions.

Table 2. Contents of the condition fields of the rules for the roots

RFV	Contents of the condition fields	UNL attributes action field of the rule	UNL Relations between the words in the sentence	Rules for dropping finite verbs
tt: *di*	{condition, ..., [[give]]: : :}	{ : +&@grant::}	pur(give, UW of the root of infinite verb)	+{VERB,INF,INFT,blk:+@,+&@grant,-INF,-INFT,+inf,+inft::} {VERB,[[give]]:::}
tt: ja	{condition, ...,[[go]]: : :}	{ : +&@attempt}	pur(go, UW of the root of infinite verb)	+{VERB,INF,INFT,ALT1,blk:+@,+&@attempt,@past,-INF,-INFT,+inf,+inft::}{VERB,[[go]]:::}
tt: cha	{condition, ...,[[want]]: : :}	{ : +&@want::}	obj(want, UW of the root of infinite verb)	+{VERB,INF,INFT,ALT1,blk:+@,+&@want,,-INF,INFT,+inf,+inft::} {VERB,[[want]]:::}
tt: ni	{condition, ...,[[take]]: : :}	{ : +&@wish::}	pur (take, UW of the root of infinite verb)	+{VERB,INF,INFT,ALT1,blk:+@, +&@wish,-INF,-INFT,+inf,+inft::} {VERB,[[take]]:::}
tt: ho	{condition, ...,[[has]]: : :}	{ :+&@obligation-strong::}	agt (take, UW of the subject)	+{VERB,INF,INFT,blk:+@,+&@obligation-strong,-INF,INFT,+inf,+inft::} {VERB,[[]]:::}
tt: par	{condition, ...,[[can]]: : :}	{ : +&@ability::}	agt (take, UW of the subject)	+{VERB,INF,INFT,blk:+@,+&@ability,INF,INFT,+inf,+inft::}{VERB,[[can]]:::}
tt: jan	{condition, ...,[[know]] : : :}	{ : +&@topic::}	cao (know, UW of the root of infinite verb)	+{VERB,INF,INFT,blk:+@,+&@topic,@present,-INF,INFT,+inf, +inft::}{VERB,[[know]]:::}
tt: dhor	{condition, ..., [[catch]]: : :}	{ : +&@generic::}	pur (catch, UW of the root of infinite verb) obj (catch, UW of the subject)	+{VERB,INF,INFT,blk:+@,+&@generic,@present,-INF,-INFT, +inf, +inft::}{VERB,[[catch]]:::}

IV. Conversion of an Infinite Verb Sentence into UNL Expression

The shift/reduce parsing [7] method will be applied to perform the conversion procedure. We take a simple Bangla infinite verb sentence for conversion into UNL expression. Bangla Transliterated sentence: *She jete pare*. Equivalent English sentence: He can go.

The input sentence is processed according to the algorithm that we developed in [8]. We have used an EnConverter (EnCo) [6] tool for our experiment. The tool takes a dictionary file for the input sentence and a set of analysis rules as its input. When the sentence is taken into EnCo, it places the sentence head (<<) in the LAW, sentence texts in the RAW (Right Analysis

Window) and the sentence tail (>>) in the RCW (Right Condition Window) shown in Figure 1.

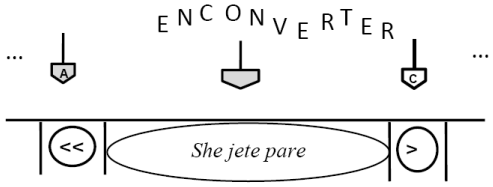

Figure 1. Initial state analysis windows and the node list.

It then automatically divides the sentence and distributes the chunks into the following five nodes. Here, nodes obtained from the input sentence are

tt: (Se) (je) (te) (par) (e).

HWs, UWs, and attributes of the nodes are given in Table 3. In Table 3, attributes PRON indicates pronoun, HPRON for human pronoun, 3P for third person, SG for singular number, SUBJ for subject, ROOT for verb root, ALT1 for alternative or deform root 1, VI for verbal inflexion, INFT for inflexion tt: *te*, CEND for consonant ended root respectively. The analysis window is enclosed within "[" and "]". Therefore, the nodes for the sentence would represent as:

tt: [Se] [je] [te] [par] [e].

These nodes are processed by the algorithm. The resolution of UNL relation and attributes generation are discussed in the steps below.

Table 3. Contents of the condition fields of the rules for the roots.

Head Word	Universal Word	Dictionary Attribute
tt: Se	"he(icl>person)"	(PRON, HPRON, 3P,SG,SUBJ)
tt: je	"go(icl>move>do, plt>place,plf>place, agt>thing)"	(ROOT, VEND, VEG3, ALT1)
tt: te	null	(VI, INFT)
tt: par	"can(icl>do, agt>thing,obj>process)"	(ROOT, CEND, CEG4)
tt: e	null	(VI, CEND)

Step 1. Initially, the EnCo has following left and right analysis windows for the input sentence.

Left node: Head word: tt: *Se*, Universal word: "he(icl>person)", Dictionary Attribute: (PRON, HPRON, 3P,SG,SUBJ).

Right node: Head word: tt: *je*, Universal Word: "go(icl>move>do, plt>place,plf>place, agt>thing)" Dictionary attribute: (ROOT, VEND, VEG3, ALT1).

The following analysis rule is now applied to shift the windows of the EnCo one step right.

R: {PRON,HPRON, SUBJ: null: null: null}{ROOT,VEND: @: null: null}.

Step 2. After rule application node 2 becomes on the LAW and node 3 becomes on the RAW. The analysis rule fired between left and right analysis windows is as follows: + {ROOT, VEND :::}{[te]:@:null}.

The above rule is proceeded by "+" sign; indicates that it is a left composition rule. It performs morphological analysis by combining the left and right node into a composite node to construct infinite verb tt: *jete* and the attributes of the left node are inherited for further processing. Two new attributes: V indicates verb and INFV for infinite verb are added with the UNL expression with this node. The operator @ is added in the <ACTION> field of the rule for the left node to inherit the attribute of the right node. The linked list after application of above rule becomes, tt: [Se] [jete] [par] [e]. Here, subject tt: *Se* becomes on the LAW and infinite verb tt: *jete* is on the RAW.

Step 3. In this step, the analyzer right shifts twice like as step 1 to place verb root tt: *par* on the LAW and inflexion tt: *e* on the RAW. The analysis rule fired between left and right analysis rule is + {*ROOT*, *CEND*:::}{[*e*]:@:*null*}. This rule concatenates the nodes of the LAW and RAW into a composite node to form the finite verb tt: *pare*. The attributes V is added with the node of the verb in the UNL expression.

Step 4. Now, LAW contains infinite verb tt: *jete* (go) and RAW contains finite verb tt: *pare* (can). The rule fired between the left and right analysis windows is, +{*V*,*INFV*,*blk*:+@,+@*ability*,-*INF*,-*INFT*,:*inf*,+*inftl*:*null*}

{*V*,[[*can*]]:*null*:*null*:*null*}*P*10. It deletes the right node from the node-list, while the left-node remains in the node-list. It causes the addition of UNL attribute &@ability in the condition field of the left node and the universal word of right node is added to the UNL expression of the node-list.

Step 5. Then the analyzer recognizes subject tt: *Se* on the LAW and main predicate of the sentence tt: *jete* on the RAW. It now generates *agt* relation between subject tt: *Se* (he) predicate tt: *jete* (go). The rule fired between the analysis windows is, >{*SUBJ*,*PRON*,*HPRON*:*null*:*null*:*null*} {*V*,*INFT*:&@ *ability*,[[*can*]]:*null*:*null*}. This rule is applicable when left node modifies the right node and hence the left-node tt: *Se* is deleted from the node-list while the right-node remains in the node-list. Finally, the modified linked list contains only one node is tt: *jete* (go). Being the last node, this is considered as root node and UNL attribute '@entry' is added to the UW. The UNL expression generated by Bangla EnConverter for the input sentence is as follows.

{*unl*}

agt(*go*(*icl*>*move*>*do*,*plt*>*place*,*plf*>*place*,*agt*>*thing*).@*entry*.@*present*.@ *ability*,*he*(*icl*>*person*))

{/*unl*}

V. Conclusion and Future Works

This paper has analyzed the formation of Bangla infinite verbs based on UNL structure. We have developed analysis rules for specific roots of finite verbs associating with infinite verbs in sentences to perform morphological analysis between the infinite and finite verbs. We have also shown the conversion of a Bangla infinite verb sentence into UNL expression. Our experimental results show that Bangla native language sentences with infinite verbs can now be easily converted into UNL expressions by analysis rules. The proposed format can be equally applicable to other languages. Our future research is to analyze infinite verbs with other inflexions in all tenses and persons.

References

[1] M. N. Y. Ali, M. S. Kamal, M. Shamsujjoha, M. A. Ali and G. F. Ahmed, Generation of UNL attributes and resolving relations for Bangla EnConverter, Int. J. Artificial Life Research 7 (2017), 21-37.

[2] M. N. Y. Ali, M. Shamsujjoha, G. Sorwar and S. H. Ripon, Construction of word dictionary for Bangla vowel ended roots and its verbal inflexions in UNL based machine translation scheme, Transaction of Machine Learning and Artificial Intelligence 5 (2017), 1-25.

[3] M. N. Y. Ali, G. Sorwar and M. Shamsujjoha, Formation of word dictionary of Bangla vowel ended roots for first person for universal networking language, Int'l Conf. on Information and Knowledge Engineering and World Congress in Computer Science, Computer Engineering and Applied Computing (2015), 201-206.

[4] D. S. Rameswar, Shadharan Vasha Biggan and Bangla Vasha, Pustok Biponi Prokashoni, Dhaka, 1996.

[5] D. M. Shahidullah and Bangala Vyakaran, Maola Brothers, Dhaka, 2003.

[6] http://www.undl.org/, Last access: 14-11-2017.

[7] J. Earley, An efficient context free parsing algorithm, Communications of the ACM 13(2) (1970).

[8] M. N. Y. Ali, M. A. Nurannabi, M. A. Ali and J. K. Das, Algorithm for conversion of Bangla sentence to universal networking language, Int. Con. Asian Language Processing (2010), 118-121.