Morphological Rules of Bangla Repetitive Words for UNL Based Machine Translation

Md. Nawab Yousuf Ali[1(✉)], Golam Sorwar[2], Ashok Toru[1], Md. Anik Islam[1], and Md. Shamsujjoha[1]

[1] Department of Computer Science and Engineering, East West University, Dhaka 1212, Bangladesh
nawab@ewubd.edu, {ashoktoru,anik1400}@gmail.com, dishacse@yahoo.com
[2] School of Business and Tourism, Southern Cross University, Lismore, Australia
golam.sorwar@scu.edu.au

Abstract. This paper develops new morphological rules suitable for Bangla repetition words to be incorporated into an interlingua representation called Universal Networking Language (UNL). The proposed rules are to be used to combine verb roots and their inflexions to produce words which are then combined with other similar types of words to generate repetition words. This paper outlines the format of morphological rules for different types of repetition words that come from verb roots based on the framework of UNL provided by the UNL centre of the Universal Networking Digital Language (UNDL) Foundation.

Keywords: Morphological rules · Repetition words · UNL · Machine translation

1 Introduction

The UNL is an artificial language replicating the function of natural languages for human communications [1]. The motivation behind UNL is to develop an interlingua representation in which semantically equivalent sentences of all languages should have the same interlingua representation [2]. Rules play an important role in machine translation process. Bangla language processing research communities have been working on the development of morphological rules for morphological analyses of words, verbs and parsing methodology of Bangla sentences [3-7]. To the best of our knowledge, no attempts have been made to develop rules for Bangla repetition words in a concrete computational approach. To address the above limitation, this paper classifies different types of Bangla repetition words and developed formats of morphological rules for those words based on the structure of UNL. The core structure of UNL is based on the *Universal Words*, *Attributes* and *Relation* [3]. *Universal Words* (UW) are words that constitute the vocabulary of UNL. *Attributes* are for the purpose to describe the subjectivity information of sentences and also used to express range of

Y. Tan et al. (Eds.): ICSI-CCI 2015, Part III, LNCS 9142, pp. 401–408, 2015.
DOI: 10.1007/978-3-319-20469-7_43

concepts and logical expressions in strengthening the express ability of the UNL. The relation between UWs is binary with different labels according to the different roles [1].

2 Bangla Word Dictionary and Enconverter

A word dictionary is a collection of word entries. Each entry in a word dictionary is composed of three kinds of elements: the *Headword* (HW), the *Universal Word* (UW) and the *Grammatical Attributes* (GA) [1]. A headword is a notation of a word in a natural language composing the input sentence and to be used as a trigger for obtaining equivalent UWs from the word dictionary. An UW expresses the meaning of word and is to be used in creating UNL expression. Grammatical attributes are the information on how word behaves in a sentence and are to be used in making analysis rules for conversions. Each dictionary entry has the following format for any native language word [8-10].

Data format:
[HW]{ID}"UW"(Attributes)<FLG,FRE,PRI>

where, HW stands for head word, ID for identification of head word which is omitable, UW for universal word, FLG for language flag, FRE for frequency of head word, and PRI for priority of head word. The attributes part divided into three sub groups: *grammatical attributes*, *semantic attributes* and *UNL attributes*.

The EnConverter (EnCo) [1] converts a native language sentence/word into UNL expressions. It is a language independent parser provided by the UNL project, a multi-headed Turing Machine [11] providing synchronously a framework for morphological, syntactic and semantic analysis. The machine has two types of windows namely *Analysis Windows* (*AW*) and *Condition Windows* (*CW*).

The machine traverses input sentence back and forth, retrieves the relevant dictionary entry (UW) from the Word Dictionary (Lexicon), depending on the *attributes* of the nodes under the AWs and those under the surrounding CWs. It then generates the semantic relationships between the UWs and /or attaches speech act attributes to them. As a result, a set of UNL expressions are made equivalent of UNL graph. EnCo is driven by a set of analysis rules to analyzing a sentence using Word Dictionary and Knowledge Base. The enconversion rule has been described in [12].

Morphological analyses are performed by the left and right composition rules. The basic type to this group is "+". This type of rule is used primarily for creating a syntactic tree with two nodes on the Analysis Windows [12].

3 Format of Morphological Rules of Bangla Repetition Words

We have rigorously analysed the Bangla grammar [13-15] and found the following types of Bangla repetition words that come from verb roots. We have also outlined the format of morphological rules for constructing these types of repetition words.

3.1 Rules for the Repetition Words from Vowel and Consonant Ended Roots

The words that come from vowel and consonant ended roots used with present tense. Two types of morphological rules for creating these types of repetition words are:

3.1.1 Rules of Repetition Words from Consonant Ended Roots

These types of rules are usually used to make the repetition words of the present tense for second person singular number. For example, 'ধরধর' (pronounce as dhoro dhoro) means catch, etc. Example of dictionary entry of the root is given below.

[ধর]{}"catch(icl>hook>occur,obj>thing)"(ROOT,CEND,DITT,R011,R02,R051,R052, #OBJ)<B,0,0>

where, attributes ROOT denotes verb root, CEND for consonant ended root, semantic attribute, DITT for repetition words, R stands for rule, R01, R02, R05 are types of rules and R011 is the sub type of R01. We use attributes R011, R02, R051 and R052 with the above template because the repetition words under this template are fall into the above stated rule-groups. UNL attributes, #AGT indicates that an action is initiated by someone, #PLC denotes the place, where an action performs and #OBJ means which action is performed. FLG for language flag, here B for Bangla, FRE means how frequently a word used in a sentence and PRI denotes the priority of the word. <B,0,0> is appended with each of the avoid it for the next UWs in this paper.

- Rule for combining two roots for creating repetition words:

+{ROOT,CEND,DITT,R011:+cmp1::}{ROOT,DITT,CEND,R011:::}(BLK)

Description: The rule describes if one root is in LAW and another root is in RAW, two headwords of the left and right analyses windows are combined with each other to make a composite node. The new composite node is the desired repetition word that will be placed in the RAW.

3.1.2 Rules for Repetition Words from Vowel Ended Roots

These repetition words are usually used with present tense. For example, 'খাই খাই' (pronounce as khai khai) meaning eat. Example of dictionary entry is shown below.

[খা]{}"eat(icl>consume>do,agt>living_thing,obj>concrete_thing,ins>thing)"(ROOT, VEND,DITT,R012,#ART,#OBJ,#INS)

[ই]{}" "(INF,VINF,VEND,DITT,R012,#AGT,#PLF,#PLT,#OBJ,#INS,#EQU,#SRC)

- Rules for combining verb root and verbal inflexion:

+{ROOT,VEND,DITT,R012:+cmp::}{[ই]:::}(ROOT,DITT,VEND,R012:::)([ই])

- Rule for combining two verbs for making repetition words:

+{DITT,VEND,R012,cmp1:cmp2::}{DITT,VEND,R012,cmp1:::}(BLK)

3.2 Rules for the Repetition Words from Consonant Ended Roots with 'আ-কার' (akar) and 'ই-কার' (ikar)

In this case, 'আ-কার' (akar) is added with consonant ended root with the first word and 'ই-কার' (ikar) is added with the same root for the second word to create these kinds of repetition words. For example, 'বলাবলি' (pronounce as bola boli) means *tell*. An entry is as follows.

[বল]{}"tell(icl>say>do,cob>uw,agt>person,obj>uw,rec>person)"(ROOT,CEND, DITT, R011, R02, R051, R052, #COB,#ART,#OBJ,#REC)

[আ]{}" " (INF, VINF,CEND, DITT,R011,R02,R051,R052, #COB, #AGT,#REC, #OBJ,#AOJ)

[ই]{}" " (INF, VINF,CEND, DITT, R011,R02, R051,R052, #COB, #AGT,#REC, #OBJ,#AOJ)

- Rule for combining consonant ended root and verbal inflexion 'আ' to make verbal nouns:

  +{ROOT,DITT,R02,CEND:+cmp1::}{[আ:::]}(ROOT,DITT,R02,CEND:::)([ই])
- Rule for combining consonant ended root and verbal inflexion 'ই':
  +{DITT,R02,CEND,cmp1:cmp2::}{ROOT,CEND,DITT,R02:::}([ই])(BLK)
- Rule for combining two verbs to make repetition words:

  +{CEND,DITT,R02,cmp1,cmp2,ditt:-DITT,-CEND,-R02,-cmp1,-cmp2,+V::}
{CEND, DITT, R02, cmp1,cmp2,ditt:-DITT,-CEND,-R02,-cmp1,-cmp2, +V::}

3.3 Rule for Repetition Words from Verbal Noun

Two different types of words with comparatively similar meaning form this kind of repetition words. For examples, 'লেখাপড়া' (pronounce as lekha pora) means *reading and writing*. Dictionary entry is:

[লেখ]{}"write(icl>communicate>do,agt>person,obj>information,cao> thing, ins> thing,rec>person)"(ROOT,CEND,DITT, R02,R03,R051,R052 #ART, #OBJ, #CAO)

[পড়]{}"read(icl>see>do,agt>person,obj>information)"(ROOT,CEND,DITT,R02,

R03, R051, R052,#AGT, #OBJ)

[আ]{}""(INF,VINF,CEND,DITT,R02,R03,R051, #AGT,#OBJ,#CAO)

Morphological analyses of this kind of words are not possible for present enconverter. They can be combined with each other by applying semantic rules after semantic analyses.

3.4 Rule for Repetition Words with Sound উ(U), ই (I) and ও(O) in Initial Roots

This type of repetition words are formed with primary sounds উ (pronounce as U), ই (pronounce as (I) and ও (pronounce as O), where the last parts of the repetition words have no meanings. As a consequence, it is not possible to analyse these kinds of words on present enconverter. To solve the translation problems in the current work, we can create direct dictionary entries of these repetition words. Three types of such words are given below.

3.4.1 Repetition Words with Primary Sound উ (U)

If the primary sound of the root is উ (U) then we can construct these types of repetition words. Template of the words:

[HW] {} “UW” (ROOT, VEND, DITT, R041, #EQU, #OBJ)

A dictionary entry is as follows:

[চুপচাপ] {}"silent(icl>adj,equ>mum)”(DITT,CEND,R041, #EQU)

3.4.2 Repetition Words with Primary Sound ই(I)

If the primary sound of the root is ই (I) then we can construct following types of repetition words. Template of the words:

[HW]{}“UW”(ROOT,VEND,DITT,R011,R041 #EQU, #OBJ)

A dictionary entry is as follows:

[মিটমাট] {} "reconciliation (icl>cooperation>thing)” (DITT, CEND, R041, #OBJ)

3.4.3 Repetition Words with Primary Sound ও (O)

If the primary sound of the root is ও (O) then we can construct the following kinds of repetition words.

[HW]{}“UW”(ROOT,VEND,DITT,R043,#AGT,#OBJ,#PLC,#EQU,#AOJ, #OPL)

An entry is as follows:

[কোলাকুলি]{}"hug(icl>embrace>do,agt>person,obj>person,ins>thing)”(DITT, VEND, R043, #AGT, OBJ, #INS)

3.5 Repetition Words with Inflexion তে(te) and এ (e)

Inflexions তে (te) and এ (e) are added with both of the roots in creating such kind of repetition words. Two types of words are:

3.5.1 Repetition Words with Inflexion ত (te)

This kind of repetition words are formed by adding inflexion 'ত'(te) with both words.

Construction of repetition words:

repetition word= root + ত + *root* + ত

For example, হাসতেহাসতে (pronounce as haste haste) means *laugh* etc.

Template of the words:

[HW] { } "UW" (ROOT, VEND, DITT, R02, R051, #AGT, #OBJ,#INS,#COM)

Some dictionary entries are as follows:

[হাস] { } "laugh (icl>utterance>thing)"(ROOT, CEND, DITT, R02, R051, #OBJ)

[ত]{}""(INF,VINF,CEND,DITT,R02,R051,#COM, #AGT,#INS, #OBJ)

- Rule for combining root and verbal inflexion 'ত':
  +{ROOT,DITT,R051,CEND:+cmp1::}{[ত]:::}(ROOT,DITT,R051,CEND:::)([ত])
- Rule for combining two verbs to construct repetition word:

+{DITT,R051,CEND,cmp1:cmp2::}{DITT,CEND,R051,cmp1:::}(BLK)

Repetition Words with Inflexion এ (e)

Here, inflexion এ (e) is added with the first root and repeat it to form repetition words.

Construction of repetition words: *repetition word=root* + এ + *root* + এ

For example, নেচেনেচে (neche, neche) etc.

Some dictionary entries are as follows:

[নেচ]{}"dance(icl>move>do,com>grace,agt>person,obj>thing)"(ROOT,CEND, DITT, R052, #AGT, #OBJ)

[এ]{}""(INF,VINF,CEND,DITT,R011,R02,R051,R052,#AGT,#PLC,#OBJ, #COB)

- Rule for combining root and verbal inflexion 'এ' to create infinite verb:
  +{ROOT,DITT,R052,CEND:+cmp1::}{[এ]:::}(ROOT,DITT,R052,CEND:::)([এ])
- Rule for combining two infinite verb to create repetition word.
  +{DITT,R052,CEND,cmp1:cmp2::}{DITT,CEND,R052,cmp1:::}(BLK)

3.6 Repetition Words from Consonant Ended Roots with Inflexion এ (e) with the Second Word

This type of repetition words are formed by adding inflexion এ(e) with the second root while first same root remains unchanged .

Construction of the repetition words: *repetition word = root + root* + এ

For example, খিটখিটে (pronounce as khit khite) means *angry*, etc. An entry is:

[খিট]{}"annoy(icl>displease>do,agt>thing,obj>volitional_thin,met>thing)"(ROOT, CEND, DITT, R06,#AGT,#OBJ,#MET)

[এ] { }" " (INF, VINF ,CEND, DITT,R06,#AGT, #MET, #OBJ)

- Rule for combing second root and verbal inflexion 'এ' (e) to make a noun:
  +{DITT,CEND,R06,cmp1:cmp2::}{[এ]:::}

- Rule for combining first root and noun made by the previous rule to make repetition word:

+{ROOT,DITT,R06,CEND:+cmp1::}{N,DITT,R06,CEND:::}([এ])(BLK)

3.7 Repetition Words from Consonant Ended Roots with Inflexion আ (aa) with the First Word

This category of repetition words are formed by adding inflexion আ (aa) with the first root while second root remains unchanged.

Construction of the repetition words: *repetition word* = *root* + আ + *root*

For example, চলাচল (pronounce as chola chol) means *moving* etc. An entry is:

[চল]{}"move(icl>occur,equ>displace,plt>thing,plf>thing,obj>thing)"(ROOT,CEN D,DITT,R011,R051,R07,#EQU,#PLT #PLF,#OBJ)

[আ]{}""(INF,VINF,CEND,DITT,R011,R051,R07,#EQU,#PLT,#PLF,#OBJ)

- Rule for combining first root and verbal inflexion আ (aa) to make noun:

+{ROOT,DITT,R07,CEND:+cmp1::}{[আ:::]}(ROOT,DITT,R07,CEND:::)(BLK)

- Rule for combining noun formed by the previous rule and second root to make repetition word:

+{N,DITT,VEND,R07,cmp1:cmp2::}{ROOT,DITT,R07,CEND:::}(BLK)

3.8 Repetition Words from Consonant Ended Roots with Inflexion ঊ (U) U with Both Words

This type of repetition words are formed by adding inflexion ঊ (U) U with both first and second roots.

Construction of the repetition words: *repetition word* = *root* + ঊ +*root* + ঊ

For example, তুলতুল (pronounce as tulu tulu) means *softness* etc. Some entries are:

[তুল]{}"softness(icl>consistency>thing,ant>hardness)(ROOT, DITT, CEND,#ANT)

[ঊ]{}""(INF,VINF,CEND,DITT,R051,R052,R08,#ANT,EQU)

- Rule for combining first root and verbal inflexion ঊ (U) to make noun:

+{ROOT,DITT,R08,CEND:+cmp1::}{[ঊ]:::}(ROOT,DITT,R08,CEND:::)([ঊ])

Same rule will be applied to combine second root and verbal inflexion ঊ (U) to make noun.

- Rule for combining two nouns made by the previous rule to make repetition words.

+{N,DITT,R08,VEND,cmp1:cmp2::}{N,VEND,DITT,R08,cmp1:::} (BLK)

4 Conclusions

This paper has outlined the templates of morphological rules for different types of Bangla repetition words that are derived from verb roots. The developed templates are useful for morphological analyses of thousands of Bangla repetition words for Bangla-UNL language server. Theoretical analysis has proved that the proposed rules

perform well in morphological analyses of Bangla repetition words. The proposed rules can be equally applicable to repetitive words in other languages.

References

1. Uchida, H., Zhu, M., Senta T.D.: Universal Networking Language. UNDL Foundation, International environment house, 2nd edn. UNDL Foundation, Geneva(2006)
2. Bhatia, P., Sharma, R.K.: Role of punjabi morphology in designing punjabi-UNL enconverter. In: Int. Conference on Advances in Computing Communication and Control, pp. 562–566 (2009)
3. Asaduzzaman, M.M., Ali, M.M.: Morphological analysis of bangla words for automatic machine translation. In: Int. Conference on Computer, and Information Technology, Dhaka, pp. 271–276 (2003)
4. Ali, N.Y., Sarker M.Z., Farook, G.A., Das, J.K.: Rules for morphological analysis of bangla verbs for universal networking language. In: Int. Conference on Asian Language Processing (IALP), Harbin, China, pp. 31–34 (2010)
5. Ali, N.Y., Noor, S.A., Zakir, M.H., Das, J.K.: Development of analysis rules for bangla root and primary suffix for universal Networking language. In: Int. Conference on Asian Language Processing (IALP), Harbin, China, pp. 15–18 (2010)
6. Hoque, M.M., Ali, M.M.: A Parsing methodology for bangla natural language sentences. In: Int. Conference on Computer and Information Technology 2003, vol. I, Dhaka, Bangladesh, pp. 277–282 (2003)
7. Ali, N.Y., Das, J.K., Al-Mamun, S.M.A., Nurannabi, A.M.: Morphological analysis of bangla words for universal networking language. In: Third Int. Conference on Digital Information Management, London, England, pp. 532–537 (2008)
8. Ali, N.Y., Yousuf, K.B., Shaheed, F.I.: Development of Bangla Word Dictionary through Universal Networking Language Structure. Gobal Science and Technology Journal **1**(1), 53–62 (2013)
9. Ali, N.Y., Das, J.K., Al-Mamun, S.M.A., Choudhury, E.H.: Specific features of a converter of web documents from bengli to universal networking language. In: Proceedings of International Conference on Computer and Communication Engineering, Kuala Lumpur, Malaysia, pp.726–735 (2008)
10. Sarker, Z.H., Ali, N.Y., Das, J.K.: Dictionary Entries for Bangla Consonant Ended Roots in Universal Networking Language. International Journal of Computational Linguistics, 79-87 (2012)
11. Parikh, J., Khot, J., Dave, S., Bhattacharyya, P.: Predicate Preserving Parsing. Department of Computer Science and Engineering, Indian Institute of Technology, Bombay
12. Uchida, H., Zhu, M., Senta T.D.: EnConverter Specification, Version 3.3, UNL Center/UNDL Foundation, Tokyo Japan (2002)
13. Shohidullah, D.M.: Bangla Bayakaron. Ahmed Mahmudul Haque, Mowla Brothers Prokashoni, Dhaka, Bangladesh, (2003)
14. Shniti, K.C.: Vasha-Prokash Bangla Vyakaran. R and C. Prokashoni, Calcutta (1999)
15. Bangla, A.: Bengali-English Dictionary, Dhaka, Bangladesh (2004)