

Morphological Rule Formulation for Nouns in Assamese to Develop Morphological Tools

Palash Das, M.A.
Department of English and Foreign Languages
Tezpur University, Assam, India
dpalash1989@gmail.com
7638803977
Pin: 784028

Madhumita Barbora, Ph. D.
Department of English and Foreign Languages
Tezpur University, Assam, India
mmb@tezu.ernet.in
9435082112
Pin: 784028

Utpal Sharma, Ph. D.
Department of Computer Science and Engineering
Tezpur University, Assam, India
utpal@tezu.ernet.in
9435182047
Pin: 784028

1. Introduction

There is a great deal of debate regarding the level of efficacy between rule-based and data driven systems. In fact data-driven approach is preferred over the rule-based approach basically because rule-based approach takes more time and requires linguistically trained personnel. However, for languages which do not have large corpus of data, data-driven approach is very difficult to implement. This paucity of data is applicable in case of Assamese also. This can stand as a hurdle when one tries to develop a spell checker for Assamese. Rule-based approach is the only viable option to accomplish the task. Moreover, although some attempts have been made to do the same with the data-driven system, the rule-based one has not been tried in Assamese. In this paper attempt has been made to show how to formulate morphological rules for an inflectional language with the help of linguistic analysis. Another important reason to choose rule-based approach over a data-driven one is that a rule-based spellchecker has advantages such as its ability to handle the name-identity problem, i.e. it can absorb new words that are not included in the dictionary.

For Indian languages like Assamese, dictionaries covering all possible inflections, derivations and compounds obtainable from all root words do not exist. For example, for a noun in Assamese, numerous inflected forms may be possible. The morphological nature of the language makes a morphology-based approach more suitable.

1.1. Existing literature

There is not much literature available on morphological rule formation on Assamese as such. However, the following works lay the groundwork for morphological rule formation in Assamese.

Sharma et al. (2002) describes a method that is helpful for unsupervised learning of morphology for building lexicon. The work is important as the method described here is helpful to analyse words to identify the root form and the exact nature of derivation used in each case. Saharia et al. (2010) delineates implementation of a suffix based noun and verb tagging approach for Assamese. The work shows how to automatically assign lexical category to each lexical object occurring in a given text while performing parts of speech (PoS) tagging based on this approach. Saharia et al. (2013) describes a method which combines rule-based and HMM based algorithm that predicts single letter suffix for the purpose of stemming Assamese text.

2. The Mirror Principle

The Mirror Principle is immensely helpful in morphological rule formation. The Mirror principle is a part of a theory for the architecture of word grammar called the Distributed Morphology, propounded by Morris Halle (1993). The Mirror Principle refers to the particular approach to the architecture of language organ developed by Mark Baker (1985). This is one of the most important pillars of current linguistic theory which observes that syntactic and morphological orderings stand in a symmetrical relation. Baker (1985) further argues that the Mirror Principle is the result of the strict locality of Head Movement Constraint (HMC). As HMC posits that a head must move to the next head position in syntactic structure, similar constraints can also be observed in case of movement of morphemes in morphological structure. In a structure as in the movement of x to z, it can only take place in a roll-up fashion where y first attaches to x, yielding [x-y], which afterwards attached to z, yielding the morphological order x-y-z.

Mirror Principle (MP) will be helpful in determining constraints on morpheme order in Assamese words. Assamese words allow variable ordering, but within limits. The focus here is on examining the extent to which the Mirror Principle can be applied while determining the ordering of morphemes in Assamese words.

3. Analysis of relation between morphemes

Words are made of morphemes. Morphemes are smallest morphological unit that carry meaning. Morphemes are divided into two categories: free and bound morphemes.

The division is based on the ability of morphemes to express meanings independently. While the free morphemes can express meaning independently bound morphemes need help of the free morphemes to do that. However, the function of free and bound morpheme cannot be distinguished as easily as it appears. The interplay between free and bound morphemes becomes particularly important in case of languages like Assamese. Traditionally this interplay is treated as ‘word formation’ in linguistics which broadly includes inflection and derivation or inflectional and derivational morphology. However, Haspelmath & Sims (2010) use the term ‘morphological relation’ to refer to the same interplay between morphemes. Although there is no basic difference between the two classifications, the latter is more inclusive and orderly in the sense that it provides clear-cut scheme to link processes like compounding and reduplication with derivation. This term is adopted in the paper to refer to ‘word formation’ processes.

Inflectional morphology deals with word forms of a lexeme. The term ‘word paradigm’ is also used to denote the set of word forms related with a lexeme. For example DANCE is a lexeme and DANCES, DANCING, DANCED are its forms. Among them only DANCE is given individual entry in a dictionary.

A clear understanding of the difference between word-forms and lexemes, and between paradigms and word families is a must before analysing morphological relationship. This helps us to analyse the internal morphological structure of words in a language which is a prerequisite for developing a rule-based system. This paper will focus on the inflectional morphology of nominal category and the constituents of the Noun Phrase (NP).

3.0. Relation between morphemes in Assamese

In this section a brief discussion about the morpheme relations in case of inflection in nouns and adjectives in Assamese will be discussed.

Inflection:

Lexical words in Assamese like noun and adjective exhibit a wide range of inflectional complexity. This feature also makes morphology of Assamese worth analysis for a rule-based system.

3.1. Nominal inflection

Nominal in Assamese inflect for number, classifier, gender, and case markers. Role of these features in inflection in Assamese is discussed with some examples:

3.1.1. Number: Assamese has two grammatical numbers: singular and plural. The singular number is unmarked and plural number is marked by two plural morphemes *-bur* and *bilāk*. The difference between singular and plural number in Assamese is shown below in the Table 1.

Table 1: Singular and Plural number in Assamese

Root	Singular	Plural
<i>lorā</i> ‘boy’	<i>lorā</i> ‘boy’	<i>lorā-bur/bilāk</i> ‘boys’
<i>ām</i> ‘mango’	<i>ām</i> ‘mango’	<i>ām-bur/bilāk</i> ‘mangoes’
<i>kitāp</i> ‘book’	<i>kitāp</i> ‘book’	<i>kitāp-bur/bilāk</i> ‘books’
<i>mās</i> ‘fish’	<i>mās</i> ‘fish’	<i>mās-bur/bilāk</i> ‘fishes’

Personal pronouns in Assamese take the plural suffix *-luk* and *-h t* for second and third person which are distinct from the plural markers for the nominal category shown in Table 1. In Table 2 we have the personal pronouns inflected with the plural suffixes.

Table 2: Showing the subject pronouns taking plural suffix

Person	Subject Pronoun	
	Singular	Plural
1P	<i>mɔi</i> ‘I’	<i>āmi</i> ‘we’
2P casual	<i>tumi</i> ‘you’	<i>tumā-luk</i> you-plu ‘you all’
2P informal	<i>tɔi</i> ‘you’	<i>tɔ-hōt</i> you-plu ‘you-all’
2P formal	<i>āpuni</i> ‘you’	<i>āpunā-luk</i> you-plu ‘you all’
3P	<i>hi, tāi</i> ‘he, she’	<i>hi-hōt, tā-hāt</i> he-plu, she-plu ‘they’

In Table 2 we find first person singular subject pronoun *mi* does not take the plural suffixes *-bur/- bilāk*. Instead the first person plural subject pronoun *āmi* is an instance of suppletion. In case of the second person (casual) we find the subject pronoun being suffixed by *-luk* leading to a morpho-phonological change of the second syllable of *tumi* where /i/ → /ā/ to form *tumāluk*. Plural forms of the second person *tōi* ‘you’ (informal) is inflected by *-hōt*. This suffixation brings about a morpho-phonological change where /oi/ → /o/. The plural form of the second person (formal) pronoun is inflected by the *-luk* suffix and here too a morpho-phonological change takes place where /i/ → /ā/. Similarly, the third person subject pronoun is inflected by the *-hāt* suffix where *tāi* ‘she’ undergoes a vowel change /āi/ → /ā/.

3.1.2. Mass Noun: Mass nouns are uncountable nouns and are quantified by a word that signifies amount. Table 3 shows mass nouns in Assamese:

Table 3: Mass nouns in Assamese

Noun	Gloss
<i>kā^h</i>	Wood
<i>kəpā^h</i>	Cotton
<i>d^h ūwā</i>	Smoke
<i>pā^hni</i>	Water

In case of mass noun singular noun is preceded by the measured item like *e-gilās* (1a) *e-bōstā* (1b). Here, the cardinal *ek* ‘one’ takes the measure item ‘glass’ and ‘sack’ to form *e(k)-gilās* and *e(k)-bōstā* where the *k* sound is dropped from *ek*.

1. (a) *e-gilās pā^hni* (b) *e-bōstā sāul*
 one-glass water one-bag rice
 ‘A glass of water.’ ‘A sack of rice.’
2. (a) *pā^hni-bur pelā-i di - ā* (b) *gāk^hir-k^hini gəɾəm kəɾ - a*
 water-COL throw-NF give-2 milk- col hot do - 2
 ‘Throw away the water.’ ‘Boil the milk.’

In (2a) and (2b) we can see that plural morpheme *-bur* can be suffixed to uncountable nouns like *pā^hni* ‘water’. In (2b) the collective suffix *-k^hini* suffixed to *gāk^hir* ‘milk’. Both *pā^hni* and *gāk^hir* are mass nouns. *-bur* and *-bilāk* gives definite reading whereas *e-gilās* (1a), *e-bōstā* (1b) indicates a measured mass noun.

3.1.3. Classifier: Singular nouns in Assamese take classifiers *-tu*, *-zɔn*, *-zɔni*, *-dāl*, *-pāt* etc.

3. (a) *lorā-tu āzi āh-ib-ɔ* (b) *lorā-zɔn-e b^hāl p^hutbɔl k^hele*
 boy-CLF today come-FUT-3 boy-CL-ERG good football play-3
 ‘The boy will come today.’ ‘The boy plays football well.’
- (c) *xi kitāp-k^hɔn ni-l-e* (d) *kukur-tu-e duk^h pā-l-e*
 he book-CL take-PST-3 dog-CL-NOM pain get-PST-3
 ‘he took the book.’ ‘The dog is hurt.’
- (e) *sābi-pāt lu-ā*
 key-CLF take-2
 ‘Take the key.’

(3a-e) shows the use of classifiers with nouns. (3a and 3b) show that same noun *lorā* can take different classifiers like *-tu* and *-zɔn*. But *kitāp* (3c) and *sābi* (3e) can take only particular classifiers *-k^hɔn* and *-pāt* respectively. This shows that all classifiers cannot be suffixed with all nouns. Table 4 shows how nouns can be classified according to their ability to take different classifiers.

Table 4: Noun classification based on their classifier affixation ability

Noun roots	Singular classifier suffixes				
	<i>z n/z ni</i>	<i>tu</i>	<i>dāl</i>	<i>pāt</i>	<i>k^h n</i>
<i>lorā</i>	√	√			
<i>kukur</i>		√			
<i>sābi</i>				√	
<i>kitāp</i>					√

3.1.4. Collective morphemes: Nouns also take collective morphemes *-zāk*, *-xɔkɔl*, *mɔk^hā*, *pāl* etc.

4. (a) *bɔnd^hu-xɔkɔl āh-ib-ɔ* (b) *gɔru-zāk k^hed-u-ā*
 friend-COL come-FUT-3 cow-COL chase-CAUS-2
 ‘Friends will come.’ ‘Chase away the herd of cow.’
- (c) *lorā-hɔt-ɔk māt-ā*
 boy-COL-ACC call-2P
 ‘Call the group of boys.’

bɔnd^hu in (4a) *gɔru* in (4b) and *lorā* (4c) are nouns inflected by two different morphemes *-xɔkɔl*, *-zāk* and *-hɔt*.

3.1.5. Sequence of the plural markers, classifiers and collective markers

In the preceding sections we have observed the plural markers (Table 1), classifiers (3a-e) and the collective markers (4 a-c) suffix to the nominal head. These three bound morphemes can be substituted in the same environment. For example in (5a-c) below the bound morphemes occur in the same environment N-suffix; where the suffix could be a classifier, a plural marker or a collective marker showing that these suffixes can be substituted and this substitution is determined by the situation in which these markers are employed. In other words these markers are not compatible with each other as is seen in (5d-f).

5. a) *lorā-tu* (b) *lorā-bur* (c) *lorā-hōt*
 boy-CL boy-PL boy-COL
 ‘the boy.’ ‘boys’ ‘boys’
- (d) **lorā-tu-bur-hōt* (e) * *lorā-tu-hōt-bur* (f) **lorā-bur-tu-hōt*

(5a-c) shows the correct sequence of the classifiers, plural markers and collective markers that can be suffixed with a noun. From this observation we can formulate the following rule that the bound morphemes can occur optionally immediately after a Noun as shown in (6) below:

6. Noun - plural / classifier / collective

3.1.6. Gender: Gender is an inflectional category for a group of noun example of which is shown in (7a-d). In these examples we can see two other allomorphs of the feminine suffix – *ni* viz. - *ni* and –*uni*. Suffixation of these morphemes also results in the increase of syllables.

7. (a) *bɔruā* → *bɔruā-ni* (b) *nāti* → *nāti-ni*
 Mr. Barua Mrs. Barua grandson granddaughter
- (c) *sur* → *sur-uni* (d) *p^hukɔn* → *p^hukɔn-ɔni*
 thief female thief Mr. Phukon Mrs. Phukon

Another feminine suffix –*i* is suffixed to a classifier or an adjective but not to a noun. Adjectives suffixed with –*i* function as attributive adjectives of nouns. In (8a) we have instance of the masculine gender in *rupɔh* ‘handsome’ which does not take the –*i* whereas in (8b) the adjective *rupɔh* takes –*i* to form the feminine form *rupɔhi*. Similarly, in (8d) we see the instance of classifier –*ɔn* taking the feminine suffix –*i* to refer to a girl.

8. (a) *rup-ɔh dekā*
 beauty-ADJ.SFX young man
 ‘Handsome young man.’
- (b) *rup-ɔh-i zuwɔti*
 beauty-ADJ.SFX-F young lady
 ‘Beautiful young woman.’
- (c) *e-zɔn lorā*
 one-CLF boy
 ‘One boy.’
- (d) *e-zɔn-i suāli*
 one-CLF-F girl
 ‘One girl.’
- (e) *kukur-zɔn-i*
 dog-CL-F
 ‘The bitch.’
- (f) *suāli-zɔn-i leter.i*
 girl-CL-F dirty.F
 ‘The girl is dirty.’

3.1.7. Sequence of classifiers and collective m...

- Noun + feminine (7a)
- Noun + classifier + feminine (8f)
- Noun + adjective suffix + feminine (8b)
- Adjective +feminine (8f)

3.1.8. Case¹: Nouns in Assamese take six overt morphological case markers. They are: ergative, accusative, dative, genitive, instrumental, locative. The nominative case does not take an overt morphological case marker. Table 5 shows case markers in Assamese:

Table 5: Case markers in Assamese

Case	Case markers
Nominative	∅
Ergative	-e, -i
Accusative	-k/ɔk
Dative	-loi
Locative	-t/ɔt
Instrumental	-re, di
Genitive	-r/ɔr

¹ Case is a morphosyntactic phenomenon. Morphosyntax is the study of relation between morphology and syntax. Case is discussed here due to its relation with the form of a word. However, morphosyntax is not the focus of this paper. Palash Das, Madhumita Barbor, Ph.D. & Utpal Sharma, Ph.D. 349-364
 detailed discussion case marking is not included in this paper.

These cases are marked by adding a suffix after the noun. Case markers are the suffixes used to represent cases in Assamese. Case marking in Assamese nouns is shown with examples (9a-e):

9. (a) *gitā-i jāḍov-ɔk mā-t-il-e*
 gita-ERG jadov-ACC call-PST-3
 ‘Gita called Jadov.’
- (b) *xi pɔt^hār-ɔt k^hel-is-e*
 he.NOM field-LOC play-ASP-3
 ‘He is playing in the field.’
- (c) *mā-e mu-loi lāru pɔt^hiā-is-e*
 mother-ERG me-DAT sweet send-ASP-3
 ‘Mother has sent me sweet ball.’
- (d) *ām-tu kɔtāri-re kāt-ā*
 mango-CL knife-INS cut-2
 ‘Cut the mango with the knife.’
- (e) *rām-ɔr e-k^hɔn gāri as-e*
 ram-GEN one-CL car be-3
 ‘Ram has a car’

The suffixes *-i* and *-ɔk* (9a) are ergative and accusative case markers marking the subject and the direct object respectively. *-t* in (9b) is the locative case marker that inflects *pɔt^hār* ‘field’. The dative case marker *-loi* (9c) inflects the indirect object *mu* ‘me’.

A morpho-phonological change takes place when the diphthong /ɔi/ in the subject pronoun *mɔi* ‘I’ changes to /u/ in the object pronoun before taking the dative case marker *-loi* to become *mu-loi* ‘me’. The instrumental case marker *-re* (9d) inflects the noun *kɔtāri* ‘knife’. The genitive case marker *-ɔr* inflects the proper noun *rām* who is the possessor of the car.

3.1.9. Case inflection in personal pronoun

Table 6: Inflection of object pronoun

Person	Subject Pronoun	Object Pronoun	
		Singular	Plural
1P	<i>āmi</i>	<i>muk</i> me.ACC 'me'	<i>āmāk</i> us.ACC 'us'
2P casual	<i>tumā-luk-e</i>	<i>tumā-k</i> you-ACC 'you'	<i>tumā-luk-k</i> you-PLU-ACC
2P informal	<i>tɔ-hɔ̃t -e</i>	<i>tuk</i> you.ACC 'you'	<i>tɔ-hɔ̃t-ɔk</i> you-PLU-ACC 'you'
2P formal	<i>āpunā-luk-e</i>	<i>āpunā-k</i> you-ACC 'you'	<i>āpunā-luk-ɔk</i> you-PLU-ACC 'you'
3P	<i>xi- hɔ̃t -e, tā- hāt-e</i>	<i>tā.k, tāi-k</i> he.ACC, she-ACC 'him, her'	<i>hi- hɔ̃t -ɔk, tā- hāt-ɔk</i> 'them'

Table 6 shows that the first person object pronoun does not take any plural marker. The plural form of the subject pronoun *āmi* 'we' takes the accusative marker *-k* to form the object pronoun *āmāk* 'us'. The plural form of the second person (casual) takes the *-luk* suffix. The plural form of the second person (informal) takes the *-h t* suffix where the / u / → / ɔ / before taking *-hɔ̃t*. The plural form of the second person (formal) takes the *-luk* suffix. The third person object pronoun, singular is *tāk* 'him' which is a suppletion for *hi* 'he'. The plural object pronoun is formed when the plural suffix *-hɔ̃t* suffixes to *hi* 'he', Another third person object pronoun *tāi* takes the *-hāt* suffix and the morpho-phonological change / āi / → / ā / takes place before suffixing *-hāt* to form the third person plural object pronoun *tāhāt* 'they'.

In Table 7 we shall examine the case markers suffixing to a pronoun and its morphophonemic changes.

Table 7: Pronominal inflection by case markers

Case marker	Singular		Plural	
Accusative	<i>mɔi+k</i> = I-ACC	<i>muk</i> I.ACC 'me'	<i>āmi+k</i> we-ACC	<i>āmāk</i> we.ACC 'Us'
Dative	<i>mɔi+loi</i> = I-DAT	<i>mu - loi</i> 'to me'	<i>āmi+loi</i> we-DAT	<i>āmā-loi</i> 'to us'
Genitive	<i>mɔi+r</i> = we-GEN	<i>mur</i> 'my'	<i>āmi+r</i> = we-GEN	<i>āmār</i> 'our'

Table 7 shows that although both noun and pronoun are inflected by case markers there is a difference in terms of resulting words. Unlike a noun when a pronoun takes a case marker it undergoes a morpho-phonological change. Table 7 shows that when inflected by the accusative case marker *-k* 'mɔi' become 'muk' where /ɔ/ changes to /u/ and /i/ sound is replaced by the accusative case marker. Similar changes can be observed when suffixing the dative case marker *-loi* and the genitive case marker *-r*. Morpho-phonological changes can be observed in first person plural pronoun 'āmi' also. Pronouns do not take any prefix or derivative suffix in Assamese.

4.0. Noun phrase

Noun phrase (NP) is a phrase where the head is a noun which is modified by other elements. Quirk (1985) opines that the simplest noun phrase consists of an article and a head. The head may be modified by an adjective, participle, a noun, genitive, relative clauses, classifiers etc. The previous sections discussed how various grammatical categories inflect head of an NP and its modifier adjective. Following sections will show how these grammatical categories inflect non-lexical pronoun in an NP.

4.1. Inflection in non-lexical pronouns

4.1.1. Demonstrative pronoun: Assamese has three demonstrative pronouns: *ei*, *xei* and *xou*.

Table 8 shows the inflectional suffixes that a demonstrative pronoun can take.

Table 8: Inflection of Demonstrative pronouns

Demonstrative	Root	Classifier	Plural	Collective
Proximal	<i>ei</i>	<i>ei+k^hɔn/tu/dāl/pāt+he/u</i>	<i>ei+bur/bilāk+he/u</i>	<i>ei+zāk/t^hupā+he/u</i>
Medial	<i>xei</i>	<i>xei+k^hɔn/tu/dāl/pāt+he/u</i>	<i>xei+bur/bilāk+he/u</i>	<i>xei+zāk/t^hupā+he/u</i>
Distal	<i>xou</i>	<i>xou+k^hɔn/tu/dāl/pāt+he/u</i>	<i>xou-bur/bilāk+he/u</i>	<i>xou+zāk/t^hupā+he/u</i>

Table 8 shows that demonstrative pronouns in Assamese are inflected by classifier, plural marker and the collective marker. These suffixes are mutually exclusive in the same environment. The morphemes put within ‘/’ between two ‘+’ signs are optional. However, each of these suffixes can be accompanied by the emphatic markers *-he* and *-u* which is shown in Table 9.

4.1.2. Emphatic marker

Table 9: Inflection of demonstrative pronouns with emphatic marker

Demonstrative	Root	Classifier+emph	Plural+emph	Collective+emph
Proximal	<i>ei</i>	<i>ei+k^hɔn/tu/dāl/pāt+he/u</i>	<i>ei+bur/bilāk+he/u</i>	<i>ei-zāk/t^hupā+he/u</i>
Medial	<i>xei</i>	<i>xei- k^hɔn/tu/dāl/pāt+he/u</i>	<i>xei+bur/bilāk+he/u</i>	<i>xei+zāk/t^hupā+he/u</i>
Distal	<i>xou</i>	<i>xou- k^hɔn/tu/dāl/pāt+he/u</i>	<i>xou-bur/bilāk+he/u</i>	<i>xou+zāk/t^hupā+he/u</i>

Table 9 shows how classifier, plural marker and the collective markers are accompanied by the emphatic markers *-he* and *-u*.

4.1.3. Interrogative pronouns

Table 10: Inflection of interrogative pronouns

Pronouns	Classifier	Plural	Collective	Emph
<i>kun</i>	<i>kun+k^hɔn/tu/dāl/pāt</i>	<i>kun+bur/bilāk</i>	<i>kun+zāk/t^hupā</i>	-
<i>kei</i>	<i>kei+k^hɔn/ta/dāl/pāt</i>	-	<i>kei+zāk/t^hupā</i>	-

From Table 10 it can be seen that the pronoun *kun* ‘who’ is inflected by classifier, plural and collective morpheme and not by emphatic markers. The pronoun *kei* means ‘how many’ and this meaning is completely expressed only when *kei* is inflected by other morphemes. Table 6 shows that the pronoun *kei* is inflected by classifier and collective morphemes and not by plural and emphatic markers.

Table 11: Inflection of interrogative pronouns with case markers

Root	Nom	Acc	Dat	Loc	Gen
<i>k-</i>	<i>kun-e</i>	<i>k+-ak</i>	<i>k+-loi</i>	<i>k+-ɔt</i>	<i>k+-ar</i>

4.2. Affixation rule of noun roots in Assamese

In the above sections nominal inflectional categories in Assamese are discussed. Based on that the maximum morphemes an inflected noun root can take can be shown as below:

10. (a) without gender marker

R+CLF/PL/COL+CM+EMP → *lorā-tu-e-he* *k^hā-l-e*
 boy-CL-NOM-EMPH eat-PST-3P
 ‘Only, the boy ate.’

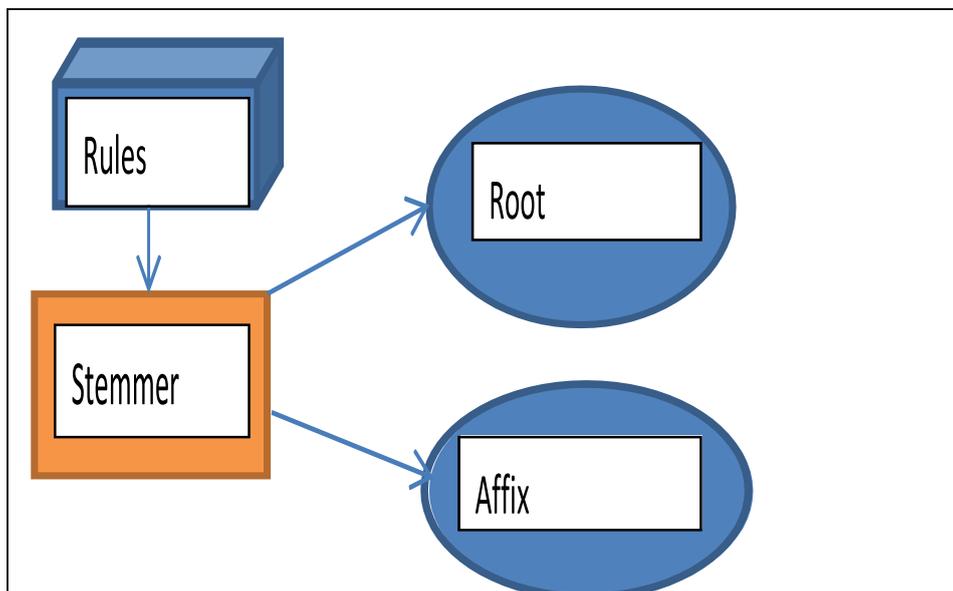
(b)with gender marker

R.F+CLF+F+CM+EMPH → *sur-uni-zɔn-i-e-he* *g^hɔri-tu ni-l-e*
 thief-F-CL-F-NOM-EMPH watch-CL take-PST-3P
 ‘The thief (female) is the one who stole the watch’.

Example (10b) shows an interesting aspect of double feminine gender marking in Assamese nouns. The example shows that apart from inflecting the root by one feminine marker *-uni* which is inherent in the noun another feminine marker *-i* is also used with the classifier to mark the feminine gender.

5. Application of this analysis

The analysis of morphology carried out in this paper is immensely helpful in finding out the correct morpheme combination patterns in Assamese. This analysis helps us to formulate rules regarding all possible and correct morpheme combination of nouns in Assamese. The rules can be used for various computational morphological operations like stemming and lemmatization which are prerequisites for applications like spell checking, information retrieval etc. In Fig. 1 the use of this method for stemming is explained:



Stemming is a process to reduce inflectional forms and sometimes derivationally related forms of a word to a common base form. As shown in Fig. 1, when an inflected word is put into the stemmer the stemmer will take help of the morphological rules to split the word into root and affixes. In an application like a rule-based spell checker this division of the word into root and affixes is very crucial. This helps the spell checker to check whether the affix combination pattern follows the valid affix combination pattern of the language. If it does then the given word is accepted as correct and otherwise the word is marked as incorrect.

6. Conclusion

Above mentioned methodology will be immensely helpful in rule formulation for inflectional languages. The rules can be used for various morphological applications like spell checker, morph analyser etc. Using methodologies like this for Natural Language Processing (NLP) purposes will help place Assamese on an equal footing with other data rich languages in an age of NLP boom.

List of abbreviation:

ADJ.SFX: adjectival suffix

ADV.SFX: adverbial suffix

ASP: aspect

CAUS: causative

CM: case marker

COL: collective morpheme

EMPH: emphatic marker

NMZ: nominalizer

PFX: prefix

SFX: suffix

VR: verb root

R: root

Reference

- Baker, Mark 1985. The Mirror Principle and Morphosyntactic Explanation. *Linguistic Inquiry* 16. 373-416.
- Bora, Saikia, Lilabati 2011. *Asamiya Bhasar Ruptattva* [Morphology of Assamese Language]. Banalata, Guwahati.
- Dabre, Raj; Amberkar Archana & Bhattacharyya, Pushpak 2012 *Morphological Analyzer for Affix Stacking Languages: A Case Study of Marathi*. Pp. 225-234. COLING. Proceedings, Mumbai.
- Haspelmath, Martin; & Sims D. Andrea 2010. *Understanding Morphology*. Hodder Education, London.
- Jurafsky, Daniel., & Martin H. James 2000. *Speech and Language Processing: An Introduction to Natural Language processing, Computational Linguistics, and Speech recognition*. Prentice Hall, USA.
- Kanuparthi, N; Inumella, A & Misra Sharma, D. 2012. Hindi Derivational Morphological Analyzer. Pp.10-16. *Twelfth Meeting of the Special Interest Group on Computational Morphology and Phonology (SIGMORPHON2012)*, Proceedings, Montré'al, Canada, June 7, 2012
- Saharia, N; Sharma Utpal, Kalita J. 2010. A Suffix-based Noun and Verb Classifier for an Inflectional Language. *International Conference on Asian Language Processing*. Harbin, China, December 28-30, pp: 19-22.
- Saharia, N; Konwar M. K, Sharma U, J. & K. Kalita 2013. An Improved Stemming Approach Using HMM for a Highly Inflectional Language. pp 164-173. International Conference on Intelligent Text Processing and Computational Linguistics. In: Gelbukh A. (eds) *Computational Linguistics and Intelligent Text Processing*. CICLing 2013. Lecture Notes in Computer Science, vol 7816. Springer, Berlin, Heidelberg
- Sharma U; Kalita, J, Das, Rajib 2002 Unsupervised Learning of Morphology for Building Lexicon for a Highly Inflectional Language. Pp:1-10. *6th workshop of the ACL Special Interest Group in Computational Phonology (SINGPHON)*

