

Bengali POS Guide Line
Taruviithi Project
Jadavpur University
11.05.2015

1. Noun (N)

The category noun is a top level category with two subcategories namely common and proper nouns. A third category considered which are mainly temporal, spatial and situation noun with location.

1.1 Common noun (NN)

A noun denotes generic expression is a common noun. Thus, boi , nadI, ghaTanA, mATh etc. are common noun. All common nouns would be assigned a tag N_NN.

1.2 Proper noun (NNP)

Like other Indian languages Bengali proper names also do not provide any cue as it is noted in English. Besides the personal proper names have different types, consisted of more than one constituent elements with white space, and the abbreviated stops following English like C. R. Das which should be put together as these don't have separate value.

However, it was argued that since it is easy for the annotators to identify proper names in the context, it might be annotated in the corpora.

Each constituent elements of a proper name is considered as an isolated NNP as in the structure like

শঙ্খ ঘোষ sankha ghoS

The first, middle and surname etc. will receive the separate NNP [N_NNP] tagging. Even in the combination of

কলকাতা পুলিশ kolkAtA puliS

The first word will get the কলকাতা kolkAtA [N_NNP] পুলিশ puliS [N_NN] .

The Abbreviated or acronyms form like সিআইডি CID, the words for WHO, HIV, SBI etc. will receive the same status.

Like this sTeT byAnk ab inDiya needs different strategies in the POS level as the first two lexical items can form a unit, on the other hand the *ab* has no special functional role in the formation of *ab inDiyaa* from Bengali grammatical notion. This phrase will be annotated like

sTeT byAnk ab InDia

[[sTeT [N_NN] byAnk [N_NN]]NP [ab [RD_UNK] inDiya [N_NNP]]NP.

This string has singular functional role forming a single chunk as tagged above.

1.3 N loc – particular nouns of locations – space and time (NST)

The third subtype under the category noun is Nloc. This category has been included to register the distinctive nature of some of the locational nouns which also function as part of a complex postpositions.

The category Nloc has been included to cover an important phenomenon of Indian languages. Certain expressions such as '*upare*' (above/up), '*nIce*' (below) '*Age*' (front), *pAs'e* etc. are content words denoting time and space. These expressions, however, are used in various ways. For example,

These words often occur as temporal or spatial arguments of a verb in a given sentence taking the appropriate *vibhakti* (case marker):

মাঠের	পাশেই	নদী ।	[পাশেই N_NST]
mATher	pAs'ei	nadI	
field-of	beside	river the	

The river is beside the field.

Apart from the functioning like an argument of a verb, these elements also modify another noun taking possessive marker.

আগের	লেখাটি	পড়ে ফেল ।	[আগের N_NST]
earlier-of	write up the	read complete.	
Complete reading the write up of earlier.			

or	baDo	bhaai	<u>oparer</u>	aMs'e	thAke [oparer N_NST]
his	elder	brother	upstairs-of	portion -in	live-pres
His elder brother lives in the upper portion of the house.					

Apart from such use of a nominal expression, it functions as postposition after a genitive noun And takes the –e / zero.

<i>Tebiler</i>	<u>Upara / upare</u>	<i>bai</i>	<i>rAkhA Ache.</i>	<i>Upar/upare [NST]</i>
Table-of	above	book	kept is.	
The book is kept on the table.				

In Some cases it occurs always with –e referring a type of location.

<i>tomAr</i>	<i>ciThir</i>	<i>madhye</i>	<i>ki</i>	<i>lekha</i>	<i>Ache?</i>
your	letter of	inside	what	written is	

What is written inside of your letter?

'Upara' / 'upare', 'madhye' etc are used as postpositions with possessive nouns in the above structure respectively which can be translated into English prepositions on, inside etc.

2. Pronoun (PR)

The need for inclusion of a separate category, pronouns was also discussed which is a variable and functionally a noun only. The tag for pronouns will be helpful for anaphora resolution. It should be a separate category and to be classified under the category Noun. It was decided to keep it at the top level as it is not a subtype of noun but is rather a variable and it can hold a value which need not necessarily be a noun.

In Bengali Pronoun has seven categories like –

Pronoun [PR] : This is a general category which is different from all other subcategories like *sekhAne* 'there' *takhan* 'then' *seTi* 'that one' etc.

Personal pronouns [PR_PRP]: All forms which replace the person like *Ami* I, *tumi* you, *se* he/she etc, Reflexive [PR_PRF] are under this group. The forms expressed the self like e.g., myself, herself, himself etc in Bengali *nije* self, *svayam* 'self' *ApnAke* 'ownself' etc, are highly closed class lexical elements in Bengali like other languages.

Relative [PR_PRL]: These forms function as the relative elements in the sentence structure to link another part. Eg. *ye* 'who' *yArA* 'who', *yakhan* 'when' *yemanTi* 'as' etc.

Reciprocal [PR_PRC] : *paraspar* 'each-other'.

Interrogative or Wh-words [PR_PRQ] : Eg. *ke* 'who' *kakhan* 'when' *kI* 'what' *kothAy* 'where' *kAke* 'whom' etc are used for asking to enquire the arguments of a verb.

Bengali has Indefinite pronominal forms which are marked as [PR-PRI]. *keu* 'anybody' *kothaao* 'anywhere' *kono* 'anything' etc are available in language.

3. Demonstratives (DM)

The category of Demonstratives is another major top level category.

Deictic [DM_DMD] is relating to or denoting a word or expression whose meaning is dependent on the context in which it is used. As for example *ei* 'this', *oi* 'that', *sei* 'that' etc . When a deictic is used before a noun as a specifier and should be present before that. It should not have any antecedent and has function only to demonstrate a head noun.

Eg.

<i>oi</i>	<i>boiTi</i>	<i>daao.</i>
that	book	give

He/she saw the cinema escaping from the school.

Finite helping verb in compound Verb [V_VAUX]: In compound verbal structure the helping verb receives the TAM marker and treated is finite one like yAbe in the sentence below, the main verb which is structurally non-finite should be marked with [V_VM] will be marked as yAbe in
se bidesh cale [V_VM] yAbe [V_VAUX].
he foreign-to will go.
He will go to foreign.

In the following structure Se skul theke paaliye [V_VM] giye sinemaa dekheche. *giye* will be marked as [V_VAUX] . In the discussion it is mentioned that in the chunk level the non-finite function will be mentioned in both the cases.

In simple verb one verb root is involved and functions as the main verb as in (i) above to complete the sentence.

Compound : This type of verbs have main and one or more than one helping verb like in

Majhi naukATa bhAsiye niye yAcche.
The boatman boat-the is sailing away by floating.
The boatman is sailing away the boat

Here *niye* and *yAcche* are two helping verbs which expresse the aspectual feature of the main verb.

The main verb has sub types of Finite, Non-finite, Infinitives and Gerund.

4.2 Verb Non-finite: (V_VM_VNF)

In Bengali enormous of Non-finite forms are available which are non-finite in character. These are perfective in nature.

- i) Se khele (V_VM_VNF) bARi gela.
He/she played home went.
He went home having played.
- ii) Se balle (V_VM_VNF) aami yaabo.
He speaks-if I go-shall
If he said I shall go.

4.3 Verb Infinitive: (V_VM_VINF)

IN this structure the verbal form is purposive but treated as infinitive which is in tradition a type of non-finite.

Se mAch Ante (V_VM_VINF) bAjAre gela.
 He/she fish bring-to market went
 He went to the market to bring the fish.

4.4 Verbal Noun (VN)

The category Verbal noun generated could be interpreted in several ways, including an alternative term for 'gerunds'. The major points of discussion for this category were

- (i) How it was different from the category gerund. Therefore, the discussion on the category of verbal noun naturally lead to a discussion on 'gerund'?
- (ii) Why should the category gerund be not classified under Nouns rather than under Verbs where it falls right now?
- (iii) Whether the information cannot be obtained from the morph analysis – if yes why should it be included as a separate category at the POS level

The answer that emerged from the discussions for (i) above is that the category of verbal noun is to be viewed differently from gerunds on the following basis .

Both a gerund and a verbal noun are nouns derived from a verb. So, to that extent they are similar. However, gerunds retain their verbal properties and can still carry their arguments with proper grammatical case. For example, *lekhA* (writing), *paRA* (reading) *s'onA* (hearing), *khAoyA* (eating) etc or *lekhAno* (make to write), *paRAno* (make to read) in Bengali are gerunds. As a gerund it functions as a noun and can also occur as an argument of a verb taking appropriate case markers. These are marked (V_VM_VNG) during the tagging.

- a) *Ramer paRA / paRAno* (V_VM_VNG) *bhAlo lAge*
 Ram-gen reading/teaching like + pres
 Trans-1 : Ram likes reading / teaching. (1)
 Trans-2: (sub) like/s Ram's reading/ teaching. (2)
 (in the translation *paRA* – is object and in Bengali has a '0' marker which is available here)

- b) *sakAle dauRno* (V_VM_VNG) *s'arIrer pakSe bhAlo*.
 morning-in running health for good be+ pres
 Running in the morning is good for health.

(*dauRAno* – subj, therefore in nominative case)

- c) *Ram lekhAy* (V_VM_VNG) *Ananda pAy*
 Ram-nom writing-in / from happiness get / enjoy-pres
 Ram enjoys happiness in writings or receives happiness from writing
 (*lekhAy* – ViSay location or source of happiness (?), in Bengali which receives –ya suffix)

The above examples show that *lekha*, in Bengali derived from verbs behaves as noun. However, these nominalized verbs can also take their own arguments syntactically, i.e., they take arguments and these arguments have appropriate syntactic case marking. For example,

- a) *RAmke mArA* (V_VM_VNG) *khub anyAy.*
 [[*RAmke*]_{obj} *mArA*]_{sub} (*hay*) *khub anyAy.*
 Ram-obj beating-sub highly injustice.

Beating Ram is highly injustice.

(*khAoyAno rAmke* has obj , obj in Bengali normally has a 0 or –ke case marker)

- b) *rAm kAgaj paRAr* (V_VM_VNG) *janya klAbe Ase.*
 Ram newspaper reading for club-to comes.
 Ram comes to the Club for reading newspaper.
 (*paRA* receives the genitive and form a post positional phrase by taking *janya* as postposition like other noun)
- c) *ei s'rIre bhAt khAoyA* (V_VM_VNG) *darker.*
 This health-in rice-obj eating (is) essential.
 Eating rice is essential in this condition of Health.

In Bengali there is another form of nominal structure in Bengali using –an with the root which is little different from other like *calan*, *balan* etc in some cases

- Or *calan* (V_VM_VNG) *bhAlo nay.*
 His behaviour good not
 His behaviour is not good.

4.5 Auxiliary verb (VAUX)

A sub-type of Auxiliary verbs with the tag VAUX has been included to annotate auxiliary verbs. The auxiliary verbs also have sub types of finite, non-finite, infinitive and gerund.

- i) Aux in finite verb. In Bengali the compound verb is productive where the Auxiliary element receives the TAM. As a result frequent use of such structure available here and there.
Tumi bides' cale yAbe [V_VAUX].
 You foreign-to go will.
 You will go to foreign country.
- ii) The auxiliary verb of the passive should be included as [V_VAUX]

Aj baiTa paRA hala.

Today book-the read was.
Today the book was read.

iii) Aux in non-finite verb

a) Skul theke pAliye giye [V_VAUX] sinemA dekheche.
Skul from escaping gone cinema seen-has.
Escaping from the school he/she has seen the cinema.

b) kathATi bale pheltei [V_VAUX] tini rege Agun.
Matter-the spoken-thrown as soon as he angry-became fire-like.
As soon as the matter have been spoken he became angry like fire.

c) ebhAbe bale phelATA [V_VAUX] ucit hayni.
in this speaking is not justified.
The speaking in this way is not justified.

Like this in all other context when the second element compound verb either in finite or non finite the [V_VAUX] is marked in POS level.

5. Adjective (JJ)

Adjectives denoting some properties of the nouns fall under this category.

i) Ami tomAke bhAlo[JJ] ghaRi upahAr debo.
I you-obj nice wrist-watch present give-future
I shall present a nice wrist-watch to you.

ii) eTA ekTA yAcchetAi [JJ] byApAr.
This (be+present) one degraded matter
This is a degraded matter.

Another type of adjective is noted

i) sei man-marA [JJ] cheleTA eseche.
that disheartened boy-specifer has come
That disheartened boy has come.

6. Adverb (RB)

The term adverb in this scheme refers only to manner adverbs.

Two types of Tags are listed

i) RB ii) RB_RBQ

i) s'is'uTi haThAt [RB] ke~de uThla.
 Baby-the suddenly crying out
 The baby crying out suddenly.

ii) Ei jhaRer rAte kIbhAbe [RBQ] yAbe?
 In this stormy night how go-will
 How will you go in this stormy night?

Adverbs of time and space are not included here as in Bengali structure these are mostly noun in function having unique role in the sentence structure. Due to this these are included in another group [NST], discussed above.

7. Postposition (PSP)

Postposition, with a tag PSP, has been included for annotation which has a syntactic relation as pan Indian Feature.

tAr samparke [PSP] kichu balis nA
 him about something say no.
 Don't say anything about him.

cheler janya [PSP] skule gela.
 Son for school-to went.
 He went to the school for his son.

gAchTi theke [PSP] phulgulo tule An.
 Tree-the from flower-plu having plucked bring.
 Bring the flowers having plucked from the tree.

rAjnIti niye [PSP] kathA hacche.
 Politics taking discussion is going on.
 Discussion is going on selecting politics.

Oke Aj cAmac diye [PSP] khAiyeche.
 Him today spoon-the with fed-has.
 (She) has fed him today with the spoon.

8. Conjuncts (CC)

Conjuncts in Bengali can be distributed in three groups like other languages like co-ordinatives, subordinative and the quotatives which are marked CC_CCD , CC_CCS and [convention of Tag is not yet decided after Manipal discussion in March 2015]

i) a) rAm ebang [CC_CCD] s'yAm tArpar berubo.
 Ram and Shyam after-that go out - future.
 Ram and Shyam will go out after that.

b) Tumi Asbe Ar [CC_CCD] Ami tArpar berubo.
 You come-will and I after-that go out - future.
 You will come and I shall go out after that.

c) Tumi Asbe , [CC_CCD] se yAbe. [the punctuation (,) has played a role of conjunct]
 You come-future, he go-future.
 You will come and she/he will go.

ii) a) Se balla ye [CC_CCS] Aj keu Asbe nA.
 He told that today anybody come-future no
 He told that nobody will come today.

b) Se balla, [CC_CCS] Aj keu Asbe nA. [the punctuation (,) has played a role of conjunct]
 He told, today anybody come-future no
 He told that nobody will come today.

c) yadi [CC_CCS] se Ase tAhale [CC_CCS] Ami yAbo.
 If he comes then I go-future
 If he comes, I shall go.

9. Particles (RP)

The 'Particles' is a cover term even in Bengali which was traditionally known as the part of indeclinable and may function classifier [RP_CL], intensifier [RP_INTF], interjections [RP_INJ], negation [RP_NEG], expression of interrogation, vocative particles etc as well as syntactic particles.

[RP_RPD] Default

Se to [RP_RPD] bollo.

He also tell-past.

He also told.

[RP_CL]

Tin jan [RP_CL] lok esechila rAtre.
 Tree-classifier people came at night.
 Three people came at night.

[RP_INJ]
 Ki [RP_INJ] sundar phul!
 How beautiful flower!

[RP_INTF]
 Khub [RP_INTF] sundar to phulTi!
 Very beautiful really flower-the
 Very beautiful really the flower!

10. Quantifiers (QT)

In general the quantifiers are with modifying character. QT_QTC has been used for the lexical item which is cardinal numbers and the QTO is for the Ordinals. The QTF is used for other cases with similar function.

i) QT_QTC

a) AmAr pA~c [QT_QTC] TAkA hAriye geche.
 My five rupees lost has gone
 My five rupees has lost.

If the QT_QTF is not followed by a noun, it should be tagged as N_NN.

b) boimelA theke duTo [N_NN] kinechi.
 Book-fair from two purchase-perfect
 I have purchased two from the book fair.

but

c) boi melA theke duTo [QT_QTF] boi [N_NN] kinechi.
 Book-fair from two books purchase-perfect.
 I have purchased two books from the book-fair.

ii)

a) pratham [QT_QTO] cheleTi bharti hayeche.
 First boy-the admit-perfect
 The first boy has admitted.

Similarly the with without a noun followed by it will be treated as noun or N_NN.

b) prathamTi [N_NN] esechi.
 First-the come-has.
 The first (one) has come.

iii) Others receives QT_QTF

- a) ebAr bhUmikampe bahu [QT_QTF] mAnuS mArA gechen.
This time earthquake-in many people die-perfect
In this earthquake many people have died.

Even the noun is followed by it the same Tag will be used.

- b) ebAr bhUmikampe bahu [QT_QTF] mArA geche.
This time earthquake-in many die-perfect
In this earthquake many have died.

11 Residuals (RD)

Finally several types of categories are available there which don't come under this category. Like other language Bengali has also category of Residuals which are different form said categories.

These are

Foreign word (RDF)

Symbol (SYM)

Punctuation (PUNC)

Echo-words (ECH)

Unknown (UNK)

Examples will be sent shortly.

This document is prepared by

Mahidas Bhattacharya
Director
School of Languages and Linguistics
Jadavpur University, Kolkata - 32
On 11.05.2015

Sankar Ram Barman
M Phil. Student. JU

Sanchari Boral
M Phil. Student.