Syntactic Part of Speech Tagging Guidelines for Bangla Text

Altaf Mahmud*

Center for Research on Bangla Language Processing (CRBLP), BRAC University, Dhaka, Bangladesh

Mumit Khan†

Center for Research on Bangla Language Processing (CRBLP), BRAC University, Dhaka, Bangladesh

*altaf.mahmud@gmail.com
†mumit@bracu.ac.bd

Abstract

Recently, several techniques have been tested to automatically assign part-of-speeches to Bangla texts using different tag sets. But there is always a need for a standard tagset for Bangla that has been formally published for syntactical bracketing, along with a details POS tagging guideline for the annotators which shows how a word should be tagged in a particular context. This document presents a guideline for annotating Bangla text by part-of-speech to assist the syntactical bracketing task. This tagset consists of total 55 tags tried to precisely distribute all of the required syntactic categories and encode necessary syntactic information to facilitate advanced linguistic analysis of a morphologically rich and flexible word ordered language. After trained a simple Brill tagger on a manually tagged corpus consists of around 25,000 words, overall accuracy has been achieved 70.6% which is comparable to minimal standard set by different experimental results using any simple supervised learning method on Bangla text.

Contents

iv

1.   Introduction

While Bangla is still limited by lack of large annotated corpus, some recent works on experimenting stochastic models [Dandapat et al. 2004; Dandapat et al. 2007; Ekbal et al. 2008] have achieved higher accuracy in automatic POS tagging. It has been shown that the accuracy of the POS tagger can be significantly improved by integrating morphological analyzer, prefix/suffix information, name entity recognizer etc. These advanced experiments placed part-of-speech tagging as an established technology for Bangla language. Different tagsets have been considered to carry out these experimental models. Since part-of-speech tagging can form a basic step towards building a syntactically annotated corpus, it would be practical for an annotator to have a details POS tagging guideline for disambiguation process of a word in a particular context, where a word will be tagged according to its syntactic distribution. The work reported in this document intends at presenting a POS tagging guideline for annotating Bangla text to form a syntactical Treebank. The tagset presented here has been considered as a *Flat Tagset* that just list down the categories applicable only for a particular language. It has been designed mostly based on the SPSAL tagset [SPSAL 2007] which has been proposed as a common framework for all Indian languages. Analyses for disambiguating the confusing examples have primarily been influenced by Penn tagset for English [Santorini 1990]. The corpus used for this project consists of texts taken from a daily Bangla newspaper 'Prothom Alo' (web: http://www.prothom-alo.com). Most of the examples and linguistic issues addressed in this document come from the scope of this corpus. So far, more than 30,000 words have been manually tagged, and among them around 25,000 words have been used as training set, and rest 5,000 are considered as test set. For testing, a simple supervised learning method using Brill tagger previously experimented on Bangla and other two Indian languages [Hasan et al. 2007] has been trained, and it demonstrated an accuracy of 70.6%. This precision level is closely comparable to other precision levels of experimenting any baseline model using a simple supervised learning method on in between 20K-25K amounts of Bangla tagged words [Hasan et al. 2007; Dandapat et al. 2007].

Due to its rich morphological nature, Bangla is a language with a high inflectional system. Inflections include postpositions, number, gender and case markers on nouns, and inflections on verbs include person, tense, aspect, honorific, non-honorific, pejorative, finiteness and non-finiteness. Since syntactical bracketing is a task of shallow processing and size of the tagset is one of the important factor, only postpositions, accusative and possessive case markers on nouns have been incorporated in this tagset. To reflect only these characteristics of morphology, a separate category 'Suffixes' has been included to denote the inflections. When a noun or a pronoun is inflected by a suffix, the base form and inflections are separated by a plus sign (+). Verbs are categorized according to their form such as finite, non-finite   etc.   The   entire   tagset   is listed in section 2. Although, syntactic distribution has been considered as main criteria while designing this tagset, there is also consciousness about encoding precise syntactic information. For example, the word 'কেন' (why) has been tagged as separate category Question Adverb (QRB), rather than being included into Question Word (QW) as SPSAL tagset [SPSAL 2007]. The sub-categorizations has been done because when 'কেন' (why) relates a relative phrase and simply tagged as QW, it can't be shown that that relative phrase is a 'Relative Adverbial Phrase' (QADVP). Thus, loss of useful syntactic information in case of syntactically formalizing a relatively free word ordered language.

The tagset presented here along with the POS tagging guidelines have been revised and corrected several times based on user feedback during the tagging process. Discussions for all of the confusing examples and resolution for disambiguation of some specific words and collocations have been

prepared throughout this process of hand-tagging. The confusing examples that not yet been found for those part-of-speeches, have been marked as **NA** (Not Analyzed). Thus, this document intends to present first version of the POS tagset with guidelines.[1]

1.1   Syntactic Function vs. Lexical Category

While tagging, pure lexical category of a word has been preferred to be taken into consideration so far [Bharati et al. 2006; SPSAL 2007], because it ensures the consistency in tagging and reduces the confusion involved in manual tagging. It is also helpful for a machine to establish a word-tag relation which leads to efficient machine learning. But in this document we prefer to give much importance to the syntactic function of the word rather than pure lexical category, although it is vulnerable for decreasing the performance and accuracy of automatic POS tagging process. Because categorizing a word according to its syntactic distribution is the most reliable way leads to further linguistic analysis. For example, dependency trees the nodes of which are labeled by the word forms of the sentence (i.e. each word form constitutes a node of the tree) together with tags representing the syntactic relations between the governing and the dependent node (such as subject, object, adjunct, adverbial etc.). Dependency structures, which act like a 'bridge' between the gap of syntactic and semantic analysis and have become popular for syntactically representing flexible word order languages, are the results of syntactic tagging. That is, once a significant amount of data has been categorized according to their syntactic distribution, it would be a valuable resource of a linguistic representation for syntactic/semantic modeling of the language.

1.2   Organization of the document

This document is organized as follows. Including introduction in section 1, section 2 enlists the summary of the entire tagset. In section 3, each tag is discussed and analyzed in more detail. Confusing examples for each tag with others and guidelines about how to distinguish them have also been discussed. The entire section 3 has been designed following the template similar to Part of speech Template of PAN Localization project (http://panl10n.net/wiki/PartOfSpeechTemplate). Every example in that section has been given as following format:

> *Example sentence with tags*
> *Word by word English translation*
> *"Sentence level English translation"*

Section 4 contains list of specific problematic examples with guidelines on how to handle such cases and section 5 has the conclusion.

---

[1] Thanks to Rabia and Fahmina for useful suggestions and comments documented in their thesis report, while they were using this tagset for their undergraduate thesis.

2. Summary of Tagset

Category	Tag Description	Tag Label	Examples
Noun	Common	NN	মানুষ (man), পানি (water)
	Compound Common Noun	NNC	ছেলে/NNC মেয়ে/NN (boys and girls), স্বরাষ্ট্র/NNC মন্ত্রনালয়/NN (home ministry)
	Proper	NNP	মতিউর (Motiur), ঢাকা (Dhaka)
	Compound Proper Noun	NNPC	কাজী/NNPC নজরুল/NNPC ইসলাম/NNP (Kazi Nazrul Islam)
	Verb Root	NNV	গোসল/NNV করা (taking bath), পান/NNV করা (drink)
	Temporal	NNT	গতকাল (yesterday), আজ (today), এখন (now)
	Question Temporal	QNT	কখন (when), যখন (relative 'when')
	Locative	NNL	উপর (up), নিচ (down), আগে (front)
	Question Locative	QNL	কোথায় (where), যেখানে (relative 'where')
Pronoun	Personal Pronoun	PRP	আমি (I), আমরা (We), তুমি (You), সে (s/he), আপনি (honorific 'you')
	Question Pronoun	QPR	কে (who), কারা (plural 'who'), যে (relative 'who')
Adjective	Simple	JJ	সুন্দর (beautiful), লাল (red), শ্রেষ্ঠ (best), শ্রেষ্ঠতম (the

			best)
	Verb Root	JJV	লাল/JJV হওয়া/VBM (to redden), দূর্বল/JJV হওয়া/VBM (to weaken)
	Question Adjective	QJJ	কেমন (how), যেমন (relative 'how')
Vocatives	Vocatives	VOC	ওরে (vocative)
Verb	Main Finite Verb	VB	করি (first person 'do'), করব (first person 'do' in future tense), কর (second person or imperative 'do'), করে (third person 'do')
	Nonfinite Nominal	VBM	করা (doing), করানো (causative 'doing'), পড়া (reading)
	Nonfinite Perfective	VBT	করে (after doing), গিয়ে (after going)
	Nonfinite	VBF	করতে (to do)
	Subjunctive	VBC	করলে (if done)
	Auxiliary	VBX	করে ফেললাম/VBX (have done), হেসে উঠলো/VBX (burst into laughter)
	Finite Existential	VBE	হয় (be), হবে (will be)
	Nonfinite Existential	VBEF	হতে (to be)
Adverb	Adverb	RB	দ্রুত (fast), হয়তো (may be), না (not), খুব (very)
	Question Adverb	QRB	কেন (why), কিভাবে (how), যেভাবে (relative 'how')
Conjunction	Co-ordinating	CC	এবং (and), অথবা (or), নতুবা (nor)
	Compound Co-	CCC	না/CCC হয়/CC

	ordinating		(neither)
	Suspicion	CN	যদি (if), পাছে (if)
	Eternal Joining	CET	যেমন/CET ... তেমন/CET (like … like), যখন/CET ... তখন/CET (when … then)
	Subordinating	CS	যে (complementizer 'that'), এইজন্য (for this)
	Compound Subordinating	CSC	তাই/CSC বলে/CS (that's why), এই/CSC কারণে/CS (for this reason)
Postposition	Postposition	ON	দ্বারা (by), থেকে (from), জন্য (for), চাইতে (than)
Interjection	Interjection	UH	ওহ্! (oh!), হায়! (alas!)
Particle	Particle	RP	না (negative particle), তো (particle)
	Question Particle	QRP	কি (question particle)
Demonstratives	Common	DM	এই (this), ওসব (those), কোন (any), যেকোন (relative 'any')
	Singular	DMS	এটি (this), ওটি (that)
	Question Demonstratives	QDM	কোনটা (which one), যেটা (relative 'which')
Quantifiers	Quantifier	QF	সকল (all), আরও (more), কম (less), কিছু (some), অনেক (many)
	Quantifier Numbers	QFNUM	১ (1), ২ (2), এক (one), তিন (three), একটি (definite

			‘one’)
	Question Quantifiers	QQF	কত (how much), যত (relative 'how much')
Foreign Word	Foreign Word	FW	যেকোন বিদেশী শব্দ (any foreign word)
Symbol	Symbol	SYM	বৈজ্ঞানিক বা অংকশাস্ত্রীয় যেকোন চিহ্ন (any scientific or mathematical sign), অন্যান্য (others)
List Item Marker	List Item Marker	LS	A, b, (a), 1, 2.3.1, ক, ৩.১৩
Suffixes	Postpositional	SFON	এ, য়, তে
	Accusative	SFAC	কে, রে, এরে, দিগকে, দিগেরে
	Possessive	SF$	এর, দের
Punctuation Marks	Sentence Final Punctuation	.	| (full stop), ?, !
	Comma	,	,
	Colon, Semi-colon	:	:, ;
	Dash, Double-Dash	-	-, --
	Left Parenthesis	(	({ [
	Right Parenthesis	)	) }]
	Opening Left Quote	LQ	', "
	Closing Right Quote	RQ	', "

3. List of Part of Speech with corresponding Tag, description and confusing examples

Language: Bangla

ID No: 1
Part of Speech: Common Noun
Tag: NN
Category: Noun
Example: এই/DM **পানিতে**/**NN+SFON** অনেক/QF **জীবাণু**/**NN** থাকে/VB

        this water-at many germs stay-3P
        "There are many germs in this water"

        গরু/**NN** **ঘাস**/**NN** খায়/VB
        cow grass eat-3P
        "Cow eats grass"

Description and Analysis:
A common noun, that can be preceded by the demonstratives and that represents one or all of the members of a class. Other nouns such as Material Noun, Collective Noun, Abstract Nouns etc. are also subsumed under this category.

Possible confusing tags:
    NN or NNP—See **NNP or NN**
    NN or JJ—See **JJ or NN**
    NN or NNT—See **NNT or NN**
    NN or NNV—See **NNV or NN**
    NN or DMS—See **DMS or NN**
    NN or QF—See **QF or NN**
    NN or QFNUM—See **QFNUM or NN**

Child Tags:

Sub-tag	Example	Description
–	–	–

ID No: 2
Part of Speech: Compound Common Noun
Tag: NNC
Category: Noun
Example: ছেলে/**NNC** মেয়ে/**NN** সবাই/PRP খেলতে/VBF গিয়েছে/VB

        boys girls all to-play gone-3P
        "All boys and girls have gone to play"

        বাংলাদেশের/NNP+SF$ প্রত্যেকটি/DM জেলায়/NN+SFON রয়েছেন/VB একজন/QFNUM
        জেলা/**NNC** প্রশাসক/NN

        Bangladesh's each district-at stay-3P one-SUFF district commissioner
        "There is one district commissioner at each of the district of Bangladesh"

প্রস্তাবটি/NN **স্বরাষ্ট্র/NNC** মন্ত্রনালয়ে/NN+SFON পেশ/NNV করা/VBM হয়েছে/VBE
proposal-the home ministry-at submit doing been-3P
"The proposal has been submitted to home ministry"

Description and Analysis:
This part has been taken from SPSAL tagset [SPSAL 2007].

In this tagset, the tag NNC is used for compound nouns. This tag has been introduced in order to identify unhyphenated compound words as one unit. All words except the last one, of compound words will be marked as NNC. Thus any NNC will always be followed by another NNC or an NN. This category helps identify these words as one unit although they are not conjoined by hyphen.

Possible confusing tags:
    NNC or NNPC—See **NNPC or NNC**
    NNC or JJ—See **JJ or NNC**

Child Tags:

Sub-tag	Example	Description
-	-	-

ID No: 3
Part of Speech: Proper Noun
Tag: NNP
Category: Noun
Example: মতিউর/**NNP** একজন/QFNUM যোদ্ধা/NN
    Motiur one-SUFF warrior
    "Motiur is a warrior"

ডিসেম্বর/**NNP** আমাদের/PRP+SF$ বিজয়ের/NN+SF$ মাস/NN
    December our victory's month
    "December is our month of victory"

Description and Analysis:
A noun belonging to the class of words used as names for unique individuals, events or places. Also called *proper name*.

Proper nouns are not usually referred by demonstratives.

Possible confusing tags:
    NNP or NN
    Compounds containing proper nouns as their second constituent, such as 'দক্ষিণ-এশিয়া' (south-Asia) should be tagged as proper nouns (NNP).

    When either quotation marks or italic type can emphasize that an instance of a word refers to the word itself rather than its associated concept, the word should be tagged as proper noun (NNP).
    Ex: '/' কেদারা/NNP '/' একটি পুরনো বাংলা শব্দ যার ইংরেজী হচ্ছে 'চেয়ার/NNP'
    ' kedara ' one-the old Bangla word which's English being 'chair'
    "'Kedara' is an old Bangla word which is called 'chair' in English"
    But,
    আমার কামরায় দু'টো কেদারা/NN আছে

my room-at two chair are

"There are two chairs in my room"

NNP or NNT—See **NNT or NNP**
NNP or NNL—See **NNL or NNP**
NNP or JJ—See **JJ or NNP**
NNP or NNP+SFON—See **NNP+SFON or NNP**

Child Tags:

Sub-tag	Example	Description
-	-	-

ID No: 4
Part of Speech: Compound Proper Noun
Tag: NNPC
Category: Noun
Example: কাজী/**NNPC** নজরুল/**NNPC** ইসলাম/NNP

Kazi Nazrul Islam
"Kazi Nazrul Islam"

গণপ্রজাতন্ত্রী/**NNPC** বাংলাদেশ/**NNPC** সরকার/NNP
People's-republic Bangladesh government
"Government of People's Republic of Bangladesh"

Description and Analysis:
This part has been taken from SPSAL tagset [SPSAL 2007].

This tag is also an addition. All words in a compound proper noun will be marked as NNPC excluding the last one. Just as the NNC tag this tag too helps identify a compound proper noun as one unit and not confuse it with a list of proper nouns.

Possible confusing tags:

NNPC or NNC
The name of a degree or a title that appears at the starting or ending of a name of a person should be tagged as compound common noun (NNC). All such titles will always be followed or preceded by a Proper Noun. In order to indicate that these are parts of proper nouns but are nonetheless nouns themselves, they will be tagged as NNC [Bharati et al. 2006].

Ex: এডভোকেট/NNC আব্দুর/NNPC রহমান/NNP

advocate Abdur Rahman
"Advocate Abdur Rahman"

ডঃ/NNC রাবেয়া/NNPC খাতুন/NNPC এমবিবিএস/NN
Dr. Rabeya Khatun M.B.B.S.
"Dr. Rabeya Khatun M.B.B.S."

জেনারেল/NNC আতাউল/NNPC গণি/NNPC ওসমানী/NNP
General Ataul Gani Osmani
"General Ataul Gani Osmani"

But sometimes the name of a degree or a title attached at the beginning/end of a name to become an unavoidable part of the whole name. In that case it can be tagged as proper noun

9

(NNP).

Ex: আব্দুর/NNPC রহমান/NNPC এডভোকেট/NNP

   Abdur Rahman Advocate

   "Abdur Rahman Advocate"

শের-এ-বাংলা/NNPC এ/NNPC কে/NNPC ফজলুল/NNPC হক/NNP

Sher-e-bangla A K Fazlul Haque

"Sher-e-bangla A K Fazlul Haque"

In the example above, 'শের-এ-বাংলা' is a title, but it became the unavoidable part of the rest of the name, thus tagged as Compound Proper Noun (NNPC).

Child Tags:

Sub-tag	Example	Description
-	-	-

ID No: 5
Part Of Speech: Nominal Verb Root
Tag: NNV
Category: Noun
Example: আমি/PRP **গোসল/NNV** করেছি/VB

   I bath done-1P

   "I have taken a bath"

সে/PRP এখন/NNT চা/NN **পান/NNV** করছে/VB

he now tea drink doing-3P

"He is now taking tea"

Description and Analysis:

This part of the following paragraph has been taken from SPSAL tagset [SPSAL 2007].

This tag has been introduced to account for the concept of kriyamuls of Bangla language. Kriyamuls are verbs formed by combining a noun or an adjective with a helping verb. The kriyamuls formed by joining a noun will be NNV. As in the above example 'গোসল' (bath) is a noun which is joined to the verb 'করেছি' (done) to express the sense of the verb 'to bathe'. So here 'গোসল' (bath) is marked as NNV and the main verb is marked as VB.

Possible confusing tags:

   NNV or NN

A 'kriyamul' (verb root) is generally followed by a helping verb and in that case it will be tagged as verb root (NNV) as well. But when it occurs as a separate word or takes a postpositional or a possessive marker, it acts as a common noun (NN).

Ex: আমরা সবাই গোসল/NNV করছি/VB

   we all bath doing-1P

   "We all are taking bath"

   But,

   গোসলের/NN+SF$ পানি/NN নাই এখানে

   bath's water not here

"No water available here for taking bath"

সে গিয়েছে গোসলে/NN+SFON

he went-3P bath-at

"He went to take bath"

Moreover, when it also in possessive relation (possessee of possessor), or preceded by a demonstrative, it tends to be tagged as common noun (NN), not nominal verb root (NNV).

Ex: আমার/PRP+SF$ গোসল/NN হয়নি এখনো

my bath been-not-3P yet

"I don't have taken bath yet"

Child Tags:

Sub-tag	Example	Description
-	-	-

ID No: 6
Part Of Speech: Temporal Noun
Tag: NNT
Category: Noun
Example: সে/PRP **আজ/NNT** আসবে/VB

he today will-come-3P

"He will come today"

শিবলু/PRP **আগামীকাল/NNT** তার/PRP+SF$ বাসায়/NN+SFON ফিরে/VBT যাবে/VB

Shiblu tomorrow his home-at after-returning will-go-3P

"Shiblu will go back to his home tomorrow"

Description and Analysis:

Nouns that are belonging to a class of words and denote a time span of events occured. It has been distinguished from noun due to a different syntactic distribution i.e. it is not necessary to be inflected by postpositional markers when contains the time of event been occurred. At the example given above, if the word 'আজ' (today) is inflected by a postpositional marker 'কে', both the syntactic and semantic distribution of the sentence won't be altered.

Ex: সে আজকে/NNT আসবে

He today will-come-3P

"He will come today"

If we want to tag it only by its syntactic distribution then they can be adverb: সে আজ/RB আসবে

And they can also behave like nouns when followed by postposition: সে আজ/NN হতে/ON আসবে না

he today from will-come-3P not

"He will not come from today"

Above examples show that these are most likely to be nouns, but they are not pure nouns. So they have been categorized as 'Temporal Noun' (NNT).

Possible confusing tags:

NNT or NN

Sometimes temporal nouns are appeared as simple nouns; don't denote a time span of events. In that case it should be tagged as common noun (NN).

Ex: সে আজ/NNC কাল/NN করে দিন গুনছে

he today tomorrow after-doing day counting-3P

11

"He is counting days by everyday"

NNT or NNP

When the name of a month or a day itself denotes a time or duration of an event, it will be tagged as temporal noun (NNT), not as proper noun (NNP).

Ex: গত/JJ ১৪ই/JJ এপ্রিল/NNT উৎযাপিত হলো বাংলা নববর্ষ

    last 14th April held was Bangla new-year

    "Bangla New Year was held at last 14th April"

সে গত/JJ সোমবার/NNT এসেছিল

he last Monday came-3P

"He came last Monday"

আগামীকাল/NNT রোববার/NNT আমার ছুটি আছে

tomorrow Sunday my holiday being-3P

"I have a holiday at Sunday tomorrow"

 But,

আজ শনিবার/NNP

today Saturday

"Today is Saturday"

বছরের প্রথম মাস জানুয়ারী/NNP

year's first month January

"First month of the year is January"

However, it can also be tagged as proper nouns (NNP) when followed by postpositions (ON).

Ex: গত সোমবার/NNP থেকে/ON বৃষ্টি হচ্ছে

   last Monday from rain being-3P

   "It has been raining since Monday last"

NNT or NNL—See **NNL or NNT**

 NNT or RB—See **RB or NNT**

Child Tags:

Sub-tag	Example	Description
-	-	-

ID No: 7
Part Of Speech: Question Temporal Noun
Tag: QNT
Category: Noun
Example: সে/PRP **কখন/QNT** আসবে/VB ?/.

    he when will-come-3P ?

    "When will he come?"

আমরা/PRP অপেক্ষা/NNV করছি/VB **কখন/QNT** সে/PRP ব্যাপারটা/NN ব্যখ্যা/NNV করে/VBT বলবে/VB

we wait doing-1P when he matter-the explain after-doing will-tell-3P

"We are waiting when he will explain the matter"

আমাকে/PRP+SFAC বলো/VB **যখন/QNT** তিনি/PRP আসবেন/VB
me tell when he will-come-3P
"Tell me when he will come"

Description and Analysis:
Question temporal nouns (QNT) are the question forms of simple temporal nouns (NNT) which are used to form interrogative sentences or to join relative clauses. Thus, temporal nouns used for interrogatives and relative categories have been subsumed under a single category.
Possible confusing tags: NA
Child Tags:

Sub-tag	Example	Description
QNT-Q	কখন	Temporal noun that can be used for both interrogative sentences and relative clauses
QNT-R	যখন	Temporal noun used for only relative clauses

ID No: 8
Part Of Speech: Locative Noun
Tag: NNL
Category: Noun
Example: তুমি/PRP **উপরে/NNL** যাও/VB
        you up-at go-2P
        "You go upside"

এই/VOC ছেলে/NN ,/, **এদিকে/NNL** এসো/VB
hey boy, here-at come-2P
"Hey boy, come here"

Description and Analysis:
Following concept and analysis has been taken from SPSAL tagset [SPSAL 2007].

This is another entirely new tag as 'Locative Noun', introduced to cover an important phenomenon of Bangla language. Words like 'আগে' (front), 'উপর' (up), 'নিচ' (down) are used in various ways in Bangla.
1. They act as postposition.
    Ex. আমি বইটির/NN+SF$ **উপর/ON** কলম রেখেছি
       I book's on pen put-1P
       "I have put the pen on the book"
Here 'উপর' is a post position which is the direct equivalent of the English preposition 'on'.
2. They also act as adverbs.
    Ex. তুমি **উপরে/RB** যাও
       you up-at go-2P
       "You go up"
Here 'উপর' is an adverbial of place.
3. These words also take post positions themselves and so in some sense behave like nouns.

Ex. সে **উপর**/**NN** হতে এসেছে

    he up from come-3P

    "He has come from upside"

4. As pointed out in 3. above, these words take postpositions and act as arguments of the verb in the sentence. And they also take possessive suffixes to join with another noun. So in that sense, they also behave like nouns.

Ex. **উপরের**/**NN+SF$** দিক/NN

    up's portion

    "upper portion"

To tag such words one option is to tag them according to the category to which they belong in the given sentence. For example in 1. above, the word is occurring as a postposition so can be marked as a postposition. In example 2. above, it is an adverb so can be marked as an adverb and so on.

But we feel that these words are more like nouns as is evident from 3. and 4. above, and also if we consider for examples, 'আগে', 'উপর', etc. as places which are in front, up, etc then we can tag them as nouns.

But these are not pure nouns as well. They are nouns which indicate a location. These also function as adverbs or postpositions in a context. So a new tag NNL is introduced for such words. But, sometimes locative nouns can also be used as postpositions. Please see the discussion of confusing example **'ON or NNL'** for further details.

In addition to analysis above, we also include that locative nouns denote locations or places literally or figuratively. In both cases they are tagged as NNL.

Ex: উপরে/NNL অনেক বাতাস (literally denotes a place)

    upside-at a-lot-of air

    "There is lots of air at upside"

জীবনে অনেক উপরে/NNL উঠবে তুমি (figuratively denotes a place to represent 'prosper')

life-at very upside-at will-reach-2P you

"You will greatly prosper in life"

Possible confusing tags:

    NNL or NNT

    In case of some words like 'আগে', 'পরে' can be used as both temporal noun (NNT) and locative noun (NNL). Nevertheless, in case of distinction between confusing tagset, determining the part of speech of a word prominently depends on the entire context. So one should consider the entire sentence containing the word that someone is unsure of, not just the word in isolation. At the example below, a vertical slash '|' as Penn [Santorini 1990] is used to denote two confusing tags for the word 'আগে'.

    Ex: সে/PRP আগে/**NNL | NNT** বসেছে/VB

        he front-at/first-at sat-3P

        "He sat at front / He sat first"

        But,

    সে/PRP আগে/**NNT** দেখেছে/VB

    he first-at saw-3P

    "He saw first"

NNL or NNP

When a proper noun that is name of a place denotes a location for an event described by an intransitive verb, it will be tagged as locative noun (NNL), not as NNP.

Ex: নাফিদ ঢাকা/NNL যাবে

    Nafid Dhaka will-go-3P

    "Nafid will go to Dhaka"

In this example, 'যাবে' (will go) is an intransitive verb and 'ঢাকা' (Dhaka) is tagged as NNL.

NNL or ON—See **ON or NNL**

Child Tags:

Sub-tag	Example	Description
-	-	-

ID No: 9
Part Of Speech: Question Locative Noun
Tag: QNL
Category: Noun
Example: আমরা/PRP **কোথায়/QNL** যাচ্ছি/VB ?/.

    we where going-1P ?

    "Where are we going?"

দীপা/NNP শুনেছে/VB **কোথায়/QNL** ঘটেছিল/VB ঘটনাটি/NN

Deepa heard-3P where occurred occurrence-the

"Deepa have heard about where the occurrence occurred"

আপনাকে/PRP+SFAC আমি/PRP দেখাব/VB **যেখানে/QNL** তিনি/PRP এসেছিলেন/VB

you-to I will-show-1P where he came-3P

"I will show you where he came"

Description and Analysis:

When locative nouns are used to form interrogative sentences or being used to join relative clauses will be tagged as question locative nouns (QNL). Thus, interrogatives and relative categories have been subsumed under a single tag.

Possible confusing tags: NA
Child Tags:

Sub-tag	Example	Description
QNL-Q	কোথায়	Locative noun used for both interrogative sentences and relative clauses
QNL-R	যেখানে	Locative noun used for only relative clauses

ID No: 10
Part Of Speech: Personal Pronoun
Tag: PRP
Category: Pronoun
Example: তিনি/**PRP** আমাকে/**PRP+SFAC** ব্যাকরণ/JJ বইটি/NN দিয়েছিলেন/VB

he me grammar book-the gave-3P

"He gave me the grammar book"

সে/PRP নিজে/PRP এসে/VBT কাজটি/NN করে/VBT গিয়েছে/VB

he himself after-coming work-the after-doing gone-3P

"He has come and done the work himself"

Description and Analysis:

One of the basic parts of speech, a pronoun takes the place of a noun, often serving as a subject or as an object in a sentence. Penn tagset for English makes a distinction between personal pronouns and possessive pronouns. But here for Bangla, all pronouns are marked as PRP and possessive pronouns are marked with a suffix (SF$) information.

Reflexive and Inclusive pronouns are also subsumed under this category. At the second sentence of the above example, the word, 'নিজে' is a reflexive pronoun and has been tagged as 'PRP'.

Possible confusing tags:

      PRP or QF—See **QF or PRP**

Child Tags:

The following sub-tags will not be used for tagging, but it is important to mention the subject-verb agreement in case of pronouns. In Bangla, subject-verb agreement occurs only for person variations. These sub-categorizations are recoverable from lexical analysis. So far, this category subsumes all of the following sub-categories shown in the table:

Sub-tag	Example	Description
PRP-P1	আমি, আমরা	1st person
PRP-P2	তুমি, তোমরা	2nd person
PRP-P3	সে, তারা	3rd person
PRP-J	তুই, তোরা	2nd person pejorative form
PRP-H	আপনি, তিনি, তাঁরা	Honorific form
PRP-X	নিজ, নিজে	Reflexive
PRP-IC	সব, উভয়	Inclusive

ID No: 11

Part Of Speech: Question Personal Pronoun

Tag: QPR

Category: Pronoun

Example: তোমাকে/PRP+SFAC বইটি/NN **কে/QPR** দিয়েছে/VB ?/.

      you book-the who gave-3P ?

      "Who gave you the book?"

      আমি/PRP দেখেছি/VB **কে/QPR** এসেছিল/VB গতকাল/NNT

      I saw who came-3P yesterday

16

"I saw who came yesterday"

সেই/DM লোকটা/NN **যে/QPR** আমাকে/PRP+SFAC চিনে/VB
that man-the who me know-3P
"That man who knows me"

Description and Analysis:
Question pronouns (or interrogative pronouns) are used as question words in an interrogative sentence, and can also act as relative pronouns to join a relative clause. Therefore, question and relative pronouns are subsumed under a single tag QPR. This category is similar to Wh-Pronoun (WP) in penn tagset for English.

Possible confusing tags: NA

Child Tags:

Sub-tag	Example	Description
QPR-Q	কে	Pronoun that is used for both interrogative sentences and relative clauses
QPR-R	যে	Pronoun that is used for only relative clauses

ID No: 12
Part Of Speech: Simple Adjective
Tag: JJ
Category: Adjective
Example: এটা/DMS একটা/QFNUM **ভাল/JJ** বই/NN
　　　　　this one-the good book
　　　　　"This one is a good book"

আমাকে/PRP+SFAC **সুন্দর/JJ লাল/JJ** জামাটি/NN দাও/VB
me beautiful red dress-the give-2P
"Give me the beautiful red dress"

Description and Analysis:
Adjectives are words that describe or modify a noun or a pronoun, or another adjective in the sentence. Unlike adverbs, which often seem capable of popping up almost anywhere in a sentence, adjectives nearly always appear immediately before the noun or the noun phrases that they modify. In a noun phrase, sometimes adjectives can be used without the presence of head noun.

In addition to regular adjectives, hyphenated compounds that are used as modifiers are tagged as adjectives. Ordinal numbers are also tagged as adjectives.

Possible confusing tags:

　　JJ or NN
　　Following analysis in intended texts have been taken from the concept of Penn [Santorini 1990] and has been adapted for Bangla.

　　Modifiers with determinative inflection (–টি, –টা) that are gradable (i.e. they can be modified by a degree adverb or form a comparative or superlative) should be tagged as adjectives (JJ), not as nouns (NN).

Ex: আমি ভালটা/JJ চাই
   I good-the want-1P
   "I want the good one"

cf. আমি সবচেয়ে/JJ ভালটা/JJ চাই   আমি উত্তমটি/JJ চাই
   I most good-the want-1P   I best-the want-1P
   "I want the best one"   "I want the best one"

Color words should be tagged as nouns (NN) when they are used as names since they have the distribution of nouns— i.e. they can be modified by adjectives and they have an overt plural.
Ex: ওই সুন্দর একটা লাল/NN
   that nice one-the red
   "That's one a nice red"

লালগুলো/NN খুব বেশি মানাবেনা এই নীলের/NN+SF$ সাথে
   reds very much won't-match-3P this blue's with
   "These reds won't match that much with these blues"

Also note the following contrast:
Ex: এই গাছগুলো গাড়ো/JJ সবুজ/JJ
   these trees deep green
   "These trees are deep green"
   But,
   এই গাছগুলো এক/QF গাড়ো/JJ সবুজ/NN
   these trees a deep green
   "These trees are a deep green"
Although adjectives are allowed to be used pronominally (without the head noun), but when they are preceded by demonstratives or quantifiers they will be considered as nouns. In the second example above, the word 'সবুজ' (green) is tagged as noun, because only nouns can be quantified by quantifiers, not adjectives.

Generic adjectives should be tagged as adjectives (JJ) and not as common nouns (NN), even when they trigger subject-verb agreement, if they can be intensified by another adjective or an adverb.
Ex: [খুব] ধনীরা/JJ এই দেশে খুব কম আয়কর দেয়
   [very] riches this country-at very less tax give-3P
   "The [very] riches in this country pay too few taxes"

JJ or NNC
Following linguistic analysis for annotating compound nouns has been acclimatized from the guidline to identify compound nouns in English [Séaghdha 2008]. It has been assumed that compounds are copulative or semantically right headed. Some rules are described here with those words given as examples which can also be found to be used somewhat frequently in Bangla, and cause confusions in case of identifying whether they are compounds or not. If collections of nouns are composed of two nouns such as N1 and N2, they will be annotated as compound if:
X is N1 and X is N2
Ex: পশম/NNC কাপড়/NN vs. পশমী/JJ কাপড়/NN

wool cloth vs. woolen cloth

মহিলা চালক,   পরিশোধন প্রক্রিয়া
woman driver   purification process

This rule does not admit sequences such as: ডেপুটি চেয়ারম্যান (deputy chairman), সহকারী সম্পাদক (assistant editor), where it is not correct to state that *an [N1 N2] is an N1* (an assistant editor is not an assistant). Such sequences are not to be considered compounds, and their modifiers are to be considered as adjectives (JJ).

N1/N2 owns N2/N1 or has exclusive rights or the exclusive ability to access or to use N2/N1 or has a one-to-one possessive association with N2/N1.
Ex: সামরিক ঘাঁটি,   সেনাবাহিনী সদরদপ্তর,   কাষ্টমার একাউন্ট
    military base   army headquarter   customer account

N1/N2 has N2/N1 as a part or constituent.
Ex: গাড়ী দরজা,   ইঞ্জিন নৌকা,   মনুষ্য রক্ত
    car door   engine boat   human blood
This is the case even when the substance has been extracted: নারিকেল তেল
                                                         coconut oil

N1/N2 is a group/society/set/collection of entities N2/N1.
Ex: ডাকটিকেট সংগ্রহ,   পদাতিক সৈন্য
    stamp collection   infantry soldier

N1/N2 is an object spatially located in or near N2/N1.
Ex: স্কুল শিক্ষক,   শীর্ষ খেলোয়াড়,   পার্শ্ব পথ
    school teacher   top player   side street

N1/N2 is an event or activity spatially located in N2/N1.
Ex: ডাইনিং রুম,   ফুটবল মাঠ,   ক্রিকেট স্টেডিয়াম
    dinning room   football field   cricket stadium

N1/N2 is a collection of items whose descriptive, significative or propositional content relates to N2/N1 or an event that describes or conveys information about N2/N1.
Put simply, when relation between N1 and N2 is *ABOUT*, annotate them as compound.
Ex: চলচ্চিত্র প্রদর্শনী   cf. চলচ্চিত্র বিষয়ক প্রদর্শনী
    film exhibition   film about exhibition
                      "exhibition about films"

বিজ্ঞান শিক্ষা,   বোমা আতঙ্ক,   নিরাপত্তা কর্মী
science education   bomb fear   security personnel
Compounds with conjoined modifiers such as: ফল ও শবজি বিক্রেতা
                                        fruit and vegetable seller
can be treated as valid compounds so long as the conjunction is elliptical (*fruit and vegetable seller* has the same meaning *fruit seller and vegetable seller*) and will be tagged as ফল/NNC ও/CC শবজি/NNC বিক্রেতা/NN.

Hyphenated modifiers, on the other hand, should always be tagged as adjectives (JJ). Thus, we have different part-of-speech assignments in examples like the following—depending on the orthographic convention used. This part of current analysis has again adapted from Penn [Santorini 1990].

Ex: ইঙ্গ-মার্কিন/JJ বাহিনী;   ইঙ্গ/NNC মার্কিন/NNC বাহিনী/NN
    engo-markin force   engo markin force
    "engo-markin force"   "engo markin force"

JJ or NNP

Following analysis in intended texts have been adapted for Bangla from the concept of Penn [Santorini 1990].

Words that refer to languages or nations, like 'ইংরেজ' (English), 'জাপানী' (japanese), can be either adjectives (JJ) or proper nouns (NNP). In prenominal position, such words are almost always adjectives (JJ). Do not be led to tag such words as proper nouns just because they occur in idiomatic collocations.

Ex: ইংরেজী/JJ কবিতা লিখেছিলেন মাইকেল মধুসূদন
    english poem wrote-3P Michael Madhusudan
    "Michael Madhusudan wrote English poem"
    But,
    সে খুব ভাল ইংরেজী/NNP বলতে পারে
    he very good English to-tell can-3P
    "He is very fluent in English"

Hyphenated compound proper nouns acting as modifiers, such as 'ঢাকা-চট্টগ্রাম মহাসড়ক' (Dhaka-Chittagong highway), should be tagged as proper nouns (NNP) rather than as adjectives (JJ).

JJ or JJV—See **JJV or JJ**
JJ or QJJ—See **QJJ or JJ**
JJ or RB—See **RB or JJ**
JJ or VBM—See **VBM or JJ**
JJ or QFNUM—See **QFNUM or JJ**
JJ or NN+SF$**—See **NN+SF$ or JJ

Child Tags:

Plain adjectives can also be sub-categorized by degree i.e. comparative, superlative. Sentence containing comparative adjectives can be syntactically formed by postpositions i.e. থেকে, অপেক্ষা (than), to state the comparison. For example: ছাত্র হিসেবে নাফিদ অভীক থেকে/ON শ্রেষ্ঠতর/JJ
                        student in-account-of Nafid Oveek than better
                        "As a student, Nafid is better than Oveek"

Sub-tag	Example	Description
JJ	শ্রেষ্ঠ	Bare Form
JJ-R	শ্রেষ্ঠতর	Comparative Form
JJ-S	শ্রেষ্ঠতম	Superlative Form

ID No: 13
Part Of Speech: Adjectival Verb Root
Tag: JJV
Category: Adjective
Example: উত্তপ্ত/JJ করাতে/VBF লোহাটি/NN **লাল/JJV** হয়ে/VBT যাচ্ছে/VBX

        heated because-of-doing iron-the red after-being going-3P

        "The iron is being redden because of been heated"

        ঘূর্ণিঝড়টি/NN আস্তে/RB আস্তে/RB **দুর্বল/JJV** হয়ে/VBT পড়ছে/VBX

        cyclone-the slowly slowly weak after-being falling-3P

        "The cyclone is being weakened slowly"

Description and Analysis:
This part has been taken from SPSAL tagset [SPSAL 2007].

As discussed at verb root in noun category earlier, the kriyamuls are being combined with a helping or light verb to express the sense of a complete event. The kriyamuls formed by joining an adjective will be tagged as JJV. As in the above first example, 'লাল' (red) is an adjective which is joined to the light verb 'হয়ে' (been) to express the sense of the verb 'to redden'. So here 'লাল' (red) is marked as JJV and the main verb is marked as VBT.

Possible confusing tags:

    JJV or JJ

    If an adjective co-occur with a verb to express a complete process, it will be tagged as verb root (JJV) as well. But in a pragmatically flexible word order language like Bangla, they can be used as predicate adjectives, rather than precede the light verb to become a verb root. In these cases they are tagged as simple adjectives (JJ).

Ex: উত্তপ্ত করাতে লোহাটি লাল/JJV হলো/VBE

    heated because-of-doing iron-the red been-3P

    "The iron has been redden because of been heated"

    গত ১১ই জুন এই সংবাদটিই প্রকাশিত/JJV হয়েছিল/VBE

    last 11ᵗʰ June this news-the-definitely published was

    "Definitely this news was published at last 11ᵗʰ June"

    But,

    উত্তপ্ত করাতে লোহাটি হলো/VBE লাল/JJ

    heated because-of-doing iron-the been-3P read

    "The iron has been redden because of been heated"

    এই সংবাদটিই হয়েছিল/VBE প্রকাশিত/JJ গত ১১ই জুন

    this news-the-definitely was-3P published last 11ᵗʰ June

    "Definitely this news was published at last 11ᵗʰ June"

Child Tags:

Sub-tag	Example	Description
-	-	-

ID No: 14
Part Of Speech: Question Adjective
Tag: QJJ
Category: Adjective
Example: কেমন/**QJJ** ছেলে/NN সে/PRP ?/.

        what-type boy he ?
        "What type of boy is he?"

নাফিদ/NNP আমাদের/PRP+SFAC বলেছে/VB **কেমন/QJJ** ছিল/VB লোকটি/NN দেখতে/VBF

Nafid us told-3P how was man-the to-see
"Nafid have told us how the man was look like"

আজ/NNT আবহাওয়া/NN অতো/RB সুন্দর/JJ নয়/VB **যেমন/QJJ** আমি/PRP ভেবেছিলাম/VB

today weather that-much beautiful not as I thought-3P
"Today's weather is not as that much beautiful as I thought"

Description and Analysis:
Question adjectives modify nouns and form interrogative sentences or relative clauses. Relative clauses are dependent clauses and must be joined to an independent clause. Unlike simple adjectives that usually precede nouns, relative adjectives follow nouns that they modify to attach a relative clause to main clause. Both question and relative adjectives have been subsumed under this category.

Possible confusing tags:

    QJJ or JJ
    Question adjectives can also be used as simple adjectives. In these cases, they are not used to form interrogative or relative clauses.

    Ex: লোকটা কেমন/JJ যেন

      man-the what-type like
      "The man doesn't appear to be all right"

    তার কথাবার্তা কেমন/JJ কেমন/JJ মনে হচ্ছে

      his conversation what-type what-type mind-at being
      "His conversation seems to be doubtful"

    QJJ or QRB—See **QRB or QJJ**

Child Tags:

Sub-tag	Example	Description
QJJ-Q	কেমন	Adjective that can be used for both question types and relative clauses
QJJ-R	যেমন	Adjective that only used for relative clauses

ID No: 15
Part Of Speech: Vocatives
Tag: VOC
Category: Vocatives

Example: ওহে/**VOC** করিম/NNP !/. কোথায়/QNL যাচ্ছ/VB তুমি/PRP ?/.

      hey Karim ! where going-2P you ?

      "Hey Karim! Where are you going?"

      ওরে/**VOC** ,/, তোরা/PRP শুনে/VBT যা/VBX

      hey , you after-listening go-2P

      "Hey all of you, come and listen"

Description and Analysis:

Vocatives are words in Bangla used to address someone, or get attention from somebody. Vocatives are generally put just before the target noun or pronoun.

Possible confusing tags:

      VOC or DM—See **DM or VOC**

Child Tags:

Sub-tag	Example	Description
-	-	-

ID No: 16

Part Of Speech: Finite Main Verb

Tag: VB

Category: Verb

Example: আমি/PRP খেলতে/VBF **যাই/VB**

      I to-play go-1P

      "I go to play"

      নাফিদ/NNP এখন/NNT **ঘুমাচ্ছে/VB**

      Nafid now sleeping-3P

      "Nafid is sleeping now"

      সুমী/NNP রহমানকে/NNP+SFAC বইটি/NN **দিয়েছিল/VB**

      Sumee Rahman-to book-the gave3P

      "Sumee gave the book to Rahman"

Description and Analysis:

The part of speech that describes an action or occurrence or indicates a state of being, and usually the main element of a predicate-argument structure is called verb. Any finite verb that constitutes the main verb of a clause has been categorized as finite main verb (VB), and subsumes all the sub-categorizations based on relevant tense, aspect and mood (TAM) features. Bangla doesn't have auxiliary or helping verbs (although in a verb pair of compound verb analysis, the second verb is called 'auxiliary' or 'vector' verb, but that is not like 'helping' verb in English or other languages), so any change in tense information is only reflected by different form of main verbs. Thus in Bangla, tense information doesn't change the syntactical structure of a sentence. However, although Bangla has subject-verb agreement only for person variation, this information for morphological form of the verbs according to person, has not been included in the tagset. That is recoverable only by further lexical or morphological analysis. That's why all finite main verbs have been subsumed under a single tag (VB).

Possible confusing tags:

      VB or VBC—See **VBC or VB**

      VB or VBX—See **VBX or VB**

Child Tags:

In case of agreement, since Bangla has subject-verb agreement only for person information, finite main verb can be sub-catagorized by following five categories:

Sub-tag	Example	Description
VB-P1	করি, করছি, করলাম	1st Person
VB-P2	কর, করছ, করলে	2nd Person
VB-P3	করে, করছে, করল	3rd Person
VB-H	করেন, করছেন, করলেন	Honorific
VB-J	কর, করছিস, করলি	Pejorative

ID No: 17
Part Of Speech: Non-finite Nominal Verb
Tag: VBM
Category: Verb
Example: সমস্যাটার/NN+SF$ সমাধান/NN **করা/VBM** হচ্ছে/VBE

      problem-the-GEN solution doing being
      "Solution of this problem is being done"

      তোমার/PRP+SF$ বাসায়/NN+SFON **যাওয়ার/VBM+SF$** দরকার/NN নাই/VB
      your home-at going-GEN need don't
      "You don't need to go home"

      এটা/DMS ওকে/PRP+SFAC **দেখানো/VBM** হয়েছে/VBE
      this him-to shown been
      "This has been shown to him"

Description and Analysis:
This category is introduced to cover an important phenomenon of Bangla language. These words like 'করা' (doing), 'দেখা' (seeing), 'দেখানো' (show) act like a verb when they take a nominal complement such as 'সমাধান' (solution) at the first example. On the other hand, they also act as nouns when inflected by postpositional or possessive markers. But this is apparently conceivable that these words are actually verb, because they constitute an event in a sentence. However, these verbs also syntactically act as nouns and they themselves can be complements of other finite verbs. Thus, a new tagset non-finite nominal verb (VBM) has been devised for such kind of verbs.

Possible confusing tags:
    VBM or JJ
    In many cases nominal verbs precede nouns or pronouns to act as modifiers, they should be tagged as adjectives (JJ) in these cases.
    Ex: করা/JJ কাজ/NN, দেখা/JJ লোক/NN, হাঁটা/JJ পথ/NN
      done work, seen man, walking path
      "completed work", "man who has been seen before", "walking distance"

Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 18
Part Of Speech: Non-finite Perfective Verb
Tag: VBT
Category: Verb
Example: সে/PRP নাস্তা/NN **করে/VBT** বিদ্যালয়ে/NN+SFON গিয়েছে/VB

        he breakfast after-doing school-at went-3P
        "He went to school after taking breakfast"

       এটা/DMS **পড়ে/VBT** শেষ/NNV কর/VB
       this after-reading finish do-2P
       "Finish it by reading"

       আপনি/PRP অফিস/NN **করে/VBT চলে/VBT** যেতে/VBF পারেন/VB
       you office after-doing after-going to-go can-2P
       "You can go after doing your office"
Description and Analysis:
Perfective verbs entail another following clause that will be accomplished after the verb itself containing an action whose goal already been achieved. It is most likely to be 'past participle' verb in English in case of event accomplishment. A perfective verb is followed by a finite clause or by a non-finite verb or by another perfective verb.
Possible confusing tags:
    VBT or RB—See **RB or VBT**
Child Tags:

Sub-tag	Example	Description
-	-	-

ID No: 19
Part Of Speech: Non-finite Verb
Tag: VBF
Category: Verb
Example: আমাকে/PRP+SFAC কিছু/NN **বলতে/VBF** হবে/VBE

        me something to-tell will-be
        "I have to tell something"

       সে/PRP **নাচতে/VBF নাচতে/VBF** যাচ্ছে/VB
       he dancing dancing going-3P
       "He is going while dancing"
Description and Analysis:
A simple non-finite verb can introduce and followed by both finite and non-finite clauses. But a finite main verb must end a clause to denote a complete event which was left unfinished by the non-finite verb. Non-finite verbs are sometimes repetitive to denote a continuous event, such as at second example above, and both of them have been tagged as VBF as well.
Possible confusing tags: NA

Child Tags:

Sub-tag	Example	Description
-	-	-

ID No: 20
Part Of Speech: Subjunctive Verb
Tag: VBC
Category: Verb
Example: তুমি/PRP খেলতে/VBF **গেলে/VBC** তারপর/CS আমি/PRP আসব/VB

        you to-play if-go then I will-come-1P
        "If you go to play then I will come"

        আপনি/PRP এটা/DMS **বললে/VBC** সে/PRP বুঝতে/VBF পারবে/VB
        you this if-tell he to-understand will-can-3P
        "If you tell him this, he will understand"

Description and Analysis:
Subjunctive mood of verbs generally subordinate a clause which has the content that being doubted, supposed or feared true. Here, this syntactic categorization has been done only when the verbs are inflected with a suffix to take the subjunctive form in Bangla (such as বল+লে → বললে), because subjunctive mood of verbs can also be constructed by a suspicion conjunction such as 'যদি' (if). The second example given above can also be written as: আপনি যদি/CN এটা বলেন/VB সে বুঝতে পারবে

        you if this tell he to-understand can-3P
        "If you tell him this, he will understand"

Possible confusing tags:
    VBC or VB
    Lexical ambiguities may arise because subjunctive form of the verb is orthographically identical to '2nd person present perfect' form of the verb.
    Ex: তুমি আমাকে কথাটা বললে/VBC আমি যাব

        you me word-the if-tell I will-go-1P
        "If you tell me the word I will go"

    তুমি আমাকে কথাটা বললে/VBC তারপর/CS আমি যাব
    you me word-the if-tell then I will-go-1P
    "If you tell me the word then I will go"

    But,
    তুমি আমাকে কথাটা বললে/VB এবং/CC আমি যাব
    you me word-the told and I will-go-1P
    "You have told me the word and I will go"

Child Tags:

Sub-tag	Example	Description
-	–	-

26

ID No: 21
Part Of Speech: Auxiliary Verb
Tag: VBX
Category: Verb
Example: আমি/PRP ঘরটি/NN প্রায়/RB গুছিয়ে/VBT **এনেছি/VBX**

    I house-the almost after-arranging brought-1P
    "I have almost arranged the house"

    নাফিদ/NNP সবার/PRP+SF$ মাঝে/ON কথাটা/NN বলে/VBT **বসলো/VBX**
    Nafid all-GEN among word-the after-telling sit-3P
    "Nafid went and said it in the middle of the crowd"

    বুড়ো/JJ লোকটা/NN মরতে/VBF **বসেছে/VBX**
    old man-the to-die sit-3P
    "The old man is about to die"

Description and Analysis:

Following description has been taken from a Ph.D. dissertation on Bangla Compound Verb [Paul 2004].

In Bangla, a Compound Verb sequence is a two-member construction in which the first member chooses between the usual conjunctive participle form –এ and the (rare, for this construction) infinitive form –তে while the second member bears the inflection in whole context. This second verb has been categorized here as auxiliary or vector verb.

In a compound verb sequence both verbs express a completion or continuity of a single event rather than two separate sequential events.

Possible confusing tags:

  VBX or VB

Following analysis has been developed from the Ph.D. dissertation on Bangla Compound Verb [Paul 2004].

Compound verbs are mono-clausal units. The participant-verbs in a compound verb sequence constitute one functional-semantic unit.

Ex: সুমী খাবারটা খেয়ে/VBT নিলো/VBX
  Sumee meal-the after-eating taken-3P
  "Sumee took her meal"

Above example shows that, it does not express two separate events, 'eating' and 'taking'. Instead the verbal sequence indicates that Sumee took her meal; the activity is completed. The event represented by compound verb cannot be substituted by two conjoined independent events:

Ex: সুমী খাবারটা খেলো এবং নিলো
  Sumee meal-the ate and took-3P
  "Sumee ate the meal and took it"

But in the following example, it expresses two separate sequential events, 'eating' and then 'getting up'.

Ex: সুমী খাবারটা খেয়ে/VBT উঠলো/VB

Sumee meal-the after-eating got-up-3P

"Sumee has got up after eating the meal"

The verb 'উঠলো' (got up) doesn't lose its primary lexical meaning and has been tagged as a finite verb (VB). Both of the verbs can be substituted by two conjoined independent events:

Ex: সুমী খাবারটা খেলো/VB এবং উঠলো/VB

Sumee meal-the ate and got-up-3P

"Sumee ate the meal and got up"

And both of the events represented by the two verbs can be modified by separate independent adverbial phrases:

Ex: সুমী খাবারটা ভালকরে/RB খেয়ে/VBT তাড়াতাড়ি/RB উঠলো/VB

Sumee meal-the well after-eating quickly got-up-3P

"Sumee has got up quickly after eating the meal well"

The first adverb 'ভালকরে' (well) taking the scope over 'খেয়ে' (after-eating), and 'তাড়াতাড়ি' (quickly) over the second verb 'উঠলো' (got up). That's why, it is a not a compound verb phrase.

But,

Ex: মেয়েটা হঠাৎ হেসে/VBT উঠলো/VBX

girl-the suddenly after-laughing got-up-3P

"The girl has suddenly burst into laugher"

কাজ করে/VBT চলো/VBX , ফলের আশা করো না

work after-doing go-3P , result's expectation do-3P not

"Keep on working, don't expect the result"

Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 22
Part Of Speech: Existential Verb
Tag: VBE
Category: Verb
Example: তোমাকে/PRP+SFAC কিছু/NN বলতে/VBF **হবে/VBE**

you-to something to-say will-be-3P

"You have to say something"

নাফিদ/NNP **হচ্ছে/VBE** খুব/RB ভাল/JJ একটা/QFNUM ছেলে/NN

Nafid being-3P very good one-the boy

"Nafid is a very good boy"

Description and Analysis:

Existential or 'Copula' are verbs that link the subject of a sentence with the predicate of the sentence. Although it might not express itself an action or condition, it serves to equate or associates the subject with the predicate.

Unlike any other finite verbs, existential verbs allow its corresponding subjects to be inflected with accusative/dative or genitive case marker in order to involve a duty or an obligation into the event.

For example: তোমাকে কিছু বলতে হবে/VBE

        you-to something to-say will-be-3P

        "You have to say something"

Here, the syntactic nominal subject of the existential verb 'হবে' (will be) is 'তোমাকে' (to you), although it has been inflected by an accusative case marker 'কে'. If there is any other verb rather than the existential verb, the word 'তোমাকে' (to you) will be a nominal object of the verb 'বলতে' (to tell) instead of a subject.

For example: তোমাকে কিছু বলতে যাবে/VB

        you-to something to-tell will-go-3P

        "(Somebody) will go to tell you something"

These two examples show that, they have different translations because of different meanings. And their corresponding syntactic constructions will be different as well, and it would also help machine translation if existential verbs are tagged with different POS category.

But note one thing that, only the finite existential verbs (হয়, হবে, হচ্ছে etc.) and non-finite existential verbs (হতে) will be tagged as VBE and VBEF respectively. Other non-finite forms of existential verbs should be tagged according to their forms (VBT, VBC, VBM etc.).

Possible confusing tags: NA

Child Tags:

Sub-tag	Example	Description
-	-	-

ID No: 23
Part Of Speech: Existential Nonfinite Verb
Tag: VBEF
Category: Verb
Example: তোমাকে/PRP+SFAC অনেক/QF কিছু/NN বলতে/VBF **হতে/VBEF** পারে/VB

        you-to many something to-say to-be can-3P

        "You may have to say many things"

        আমাকে/PRP+SFAC অনেক/JJ ধনী/JJ **হতে/VBEF** হবে/VBE

        me much rich to-be will-be-3P

        "I have to be very rich"

Description and Analysis:

NonFinite Existential verb is the non-finite form of existential verb or 'Copula'. Like existential, non-finite existential verbs also take its syntactic nominal subject being inflected with accusative/dative or genitive case marker to involve a duty or an obligation into the event.

Considering the first example above, if there is any other verb rather than the non-finite existential verb, the word 'তোমাকে' (to you) will be a nominal object of the verb 'বলতে' (to tell) instead of a subject.

For example: তোমাকে কিছু বলতে যেতে/VBF পারে

        you-to something to-tell will-go can-3P

        "(Somebody) may go to tell you something"

As mentioned for existential, except the finite existential verbs (হয়, হবে, হচ্ছে etc.) and non-finite existential verb (হতে) will be tagged as VBE and VBEF respectively, other non-finite forms should be tagged according to their form (VBT, VBC, VBM etc.).

Possible confusing tags: NA

Child Tags:

Sub-tag	Example	Description
-	-	-

ID No: 24
Part Of Speech: Adverb
Tag: RB
Category: Adverb
Example: ঘোড়াটি/NN **বেশ/RB দ্রুত/RB** দৌড়ায়/VB

        horse-the very fast run-3P
        "The horse runs very fast"

    সে/PRP পড়াশোনা/NN করে/VB **না/RB**
    he study do-3P not
    "He doesn't do study"

    শিশুটি/NN **আস্তে/RB আস্তে/RB** হাঁটতে/VBF শিখছে/VB
    child-the slowly slowly to-walk learning-3P
    "The child is learning slowly how to walk"

Description and Analysis:

An adverb can modify a verb, a clause, a phrase or another adverb or adjective by expressing time, place, manner, degree, cause etc. Bangla adverbs often end with suffix 'ভাবে'. Words that intensify adverbs or adjectives (like 'খুব', 'বেশ') are also categorized as adverbs. In the collocation of repetitive adverbs both of them will be tagged as RB as shown in the third example above.

Unlike adjectives, adverbs can popup at any place in a sentence or a clause. But intensifier adverbs always co-occur with other adverbs or adjectives that they intensify.

Possible confusing tags:

    RB or NNT

Repetitive temporal nouns, such as 'মাঝে মাঝে' (sometimes) will be tagged as temporal nouns (NNT), don't be led to tag these as adverbs.

Ex: আমরা মাঝে/NNT মাঝে/NNT এ ব্যাপারটা নিয়ে ভাবছি

        we sometimes sometimes this issue-the about thinking-1P
        "Sometimes we are thinking about this issue"

But sometimes temporal nouns are used as adverbs to mean 'frequently', 'very often' etc. In that case it will be tagged as adverb (RB).

Ex: আজকাল/NNT দেশব্যাপী যখন-তখন/RB হরতাল ডাকা হচ্ছে

        now-days throughout-the-country very-often strike calling being-3P
        "Now days, strike is being called very often throughout the country"

    RB or JJ

This contrast in intended lines has been developed from the concept of Penn [Santorini 1990]

and has been adapted for Bangla.

While most adverbs formed from adjectives end in -ভাবে (ly), not all do. The crucial criterion is whether a word modifies a noun, in which case it is an adjective (JJ), or a non-noun, in which case it is an adverb (RB).
Ex: দ্রুত/JJ বিকাশ/NN
  rapid growth
  "rapid growth"

  দ্রুত/RB বেড়ে/VBT ওঠা/VBM বৃক্ষ
  rapid grown rise tree
  "rapid grown up tree"

Also note the following contrast:
Ex: সুন্দর(ভাবে)/RB গেয়েছে মেয়েটি
  beautiful(ly) sung-3P girl-the
  "The girl sung beautifully"
  But,
  সুন্দর/JJ গান/NN গেয়েছে মেয়েটি
  beautiful song sung-3P girl-the
  "The girl sung a beautiful song"

Take care not to tag predicate adjectives as adverbs. Thus, in 'জীবন সহজ/JJ কর'
                       life simple do-2P
                       "make life simple"
Here 'সহজ' (simple) is an adjective; cf. the unacceptability of 'জীবন সহজে/RB গড়'
                       life simply build-2P
                       "build life simply"

RB or VBT
In a given context, perfective form of a verb acts as adverbial participle [Biswas et al. 2008] by modifying the following verb, where both of them indicate two parallel events. In such case it is an adverb (RB), or else if there are two sequential events it will be a perfective verb (VBT) as well. A usual test case is, replace the perfective verb by repetitive non-finite verbs (VBF) and see if the sentence infers the similar sense of the event (two parallel processes) as previous, then tag it as an adverb (RB). For example:
তিনি দৌড়ে/RB এলেন vs. তিনি দৌড়াতে/VBF দৌড়াতে/VBF এলেন
He/she after-running come-3P vs. He/she to-run to-run come-3P
"He/she has come by running" vs. "He/she has come running"

সে হেসে/RB দরজা খুলল vs. সে হাসতে/VBF হাসতে/VBF দরজা খুলল
He/she after-smiling door opened-3P vs. He/she to-smile to-smile door opened-3P
"He/she has opened the door with a smile" vs. "He/she has opened the door smiling"

But in the following examples the verb pair denotes two sequential events—
তিনি কথাটা বলে/VBT এলেন vs. তিনি কথাটা বলতে/VBF বলতে/VBF এলেন
He/she word-the after-telling come-3P vs. He/she word-the to-tell to-tell come-3P

"He/she has come after telling the word" vs. "He/she has come telling the word"

সে লোকটিকে দেখে/VBT দরজা খুলল vs. সে লোকটিকে দেখতে/VBF দেখতে/VBF দরজা খুলল

He/she man-the after-seeing door opened-3P vs. He/she man-the to-see to-see door opened-3P

"He/she has opened the door after seeing the man" vs. "He/she has opened the door seeing the man"

RB or CC—See **CC or RB**
RB or RP—See **RP or RB**
RB or QF—See **QF or RB**

Child Tags:
Bangla does have two prominent sub-categorizations for adverb as other language: plain and negative adverbs. Another type of sub-categorization includes the intensifiers. But this tagset subsumes all of these categories into adverb (RB).

Sub-tag	Example	Description
RB	আস্তে, দ্রুত	Plain
RB-N	না, নাই	Negative
RB-INTF	খুব, বেশ	Intensifiers

ID No: 25
Part Of Speech: Question Adverb
Tag: QRB
Category: Adverb
Example: তুমি/PRP আজ/NNT স্কুলে/NN+SFON যাওনি/VB **কেন/QRB** ?/.

        you today school-at went-not-2P why ?
        "Why didn't you go to school today?"

      নাফিদ/NNP **কিভাবে/QRB** অংকটি/NN করেছে/VB ?/.
      Nafid how math-the did-3P ?
      "How did Nafid solve the math?"

      কাজটি/NN হয়নি/VBE আমি/PRP **যেভাবে/QRB** বলেছিলাম/VB
      work-the been-not-3P I how told-1P
      "The work has not been done as I told"

Description and Analysis:
Question adverbs modify verbs to form an interrogative sentence or to form a relative clause that attached to the main clause. Question adverbs and relative adverbs are subsumed into a tag QRB. Relative adverbs join only relative clauses and question adverbs can be used as both question words and relative adverbs.

Question adverbs cannot be intensified, such as 'খুব/RB কিভাবে/QRB' (very how) is wrong.
Possible confusing tags:

QRB or QJJ

When distinguishing between these two types, it has to be observed that whether that word modifies a noun or a verb. If that modifies a noun (can also appears pronominally) then it will be tagged as question adjective (QJJ), otherwise it will be tagged as question adverb (QRB).

Ex: কেমন/QJJ মানুষ সে ?

what-type man he ?

"What type of man is he?"

আমি জানি আমি কেমন/QJJ চাই (pronominal use of question adjective)

I know I what-type want-1P

"I know what type I want"

But,

কেমন/QRB আছ তুমি ?

how are you ?

"How are you?"

Child Tags:

Sub-tag	Example	Description
QRB-Q	কেন, কিভাবে	Adverbs that are used for both interrogative sentences and relative clauses
QRB-R	যেভাবে	Adverbs that is only used for relative clauses

ID No: 26
Part Of Speech: Coordinating Conjunction
Tag: CC
Category: Conjunction
Example: নাফিদ/NNP বই/NN পড়ছে/VB **এবং/CC** খেলা/NN দেখছে/VB

Nafid book reading-3P and game watching-3P

"Nafid is reading book and watching game"

তুমি/PRP অবশ্যই/RB যাবে/VB **নতুবা/CC** তাকে/PRP+SFAC আসতে/VBF হবে/VBE

you must will-go-3P otherwise him to-come will-be

"You must go otherwise he has to come"

Description and Analysis:

An uninflected word used to connect words, phrases, clauses, or sentences. Conjunction which joins two equal level phrases or clauses is called coordinating conjunction. Coordinating conjunctions subsumes all other types of conjunction such as adversative, disjunctive, exclusion, conclusive etc. as Penn Tagset for English.

Possible confusing tags:

CC or RB

The words 'আর', 'আবার' usually used as conjunctions (CC) when join two clauses, but in some cases they modify verb to express continuity or repetitiveness of an action. In this case it is tagged as adverb (RB). It is often difficult to decide. Again it is worth mentioning that in case

of distinguishing of part of speech; consider the whole sentence that contains it, not the specific words or collocation. Since, context is a vital issue for deciding relative part-of-speeches.

Ex: ছেলেটি পড়াশোনা করেছে আর/CC ছবিও এঁকেছে

      boy-the study done-3P and picture-too drawn-3P

      "The boy has finished his study and also drawn picture"

    তুমি অংক করেছ আবার/CC কবিতাটিও পড়েছ?

    you math done-2P again poem-the-too read-2P?

    "You have practiced math and again read the poem?"

    But,

    আমি আর/RB আসবনা   cf. *আমি আর/CC আসবনা

    I more come-not-1P

    "I won't come any more"

    ধন্যবাদ তোমাকে, আবার/RB এস

    thank you-to, again come-2P

    "Thank you, come again"

    সে বাসায় গিয়েছে, আবার/RB আসবে

    he home-at gone-3P, again will-come-3P

    "He has gone home, will come again"

    CC or RP—See **RP or CC**

Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 27
Part Of Speech: Compound Coordinating Conjunction
Tag: CCC
Category: Conjunction
Example: নাফিদ/NNP বই/NN পড়ছে/VB **না/CCC** হয়/CC খেলা/NN দেখছে/VB

    Nafid book reading-3P not be game watching-3P

    "Nafid is either reading book or watching game"

    আমি/PRP তাকে/PRP+SFAC কিছু/NN বলবনা/VB **যদি/CCC** না/CC সে/PRP কথাটা/NN বলে/VB

    I him anything will-tell-not-1P if not he word-the tell-3P

    "I won't tell him anything unless he tell the word"

Description and Analysis:

This is another new tagset introduced for coordinating conjunction (CC) with same categorization as compound common noun (NNC) and compound proper noun (NNPC) as discussed for nouns. Some words in Bangla compounded to act as a single unit of coordinating conjunction (CC). All words except the last one will be marked as CCC. Thus any CCC will be followed by another CCC or a CC. This strategy will help to recognize those words as a single unit of conjunction.

Possible confusing tags: NA

Child Tags:

Sub-tag	Example	Description
-	-	-

ID No: 28
Part Of Speech: Suspicion Conjunction
Tag: CN
Category: Conjunction
Example: আমি/PRP **যদি/CN** যাই/VB সে/PRP হয়তো/RB আসবে/VB

        I if go-1P then he may-be will-come-3P
        "If I go then may be he will come"

    তাকে/PRP+SFAC তুমি/PRP কিছু/NN বলোনা/VB, **পাছে/CN** সে/PRP
তোমাকে/PRP+SFAC ভুল/NNV বুঝে/VB

    him you something don't-tell-2P, if he you-to wrong understand-3P
    "Don't tell him anything, he may misunderstand you"

Description and Analysis:
Suspicion conjunctions [Saha et al. 2004] are conjunctions which are used to express a suspected action that may happen or not by joining another dependent clause to the main clause.

Since suspicion conjunctions join subordinating clauses, thus act as subordinating conjunctions. They also act as eternal joining conjunction in a pair along with another coordinating or subordinating conjunction.

Possible confusing tags:
      CN or CET—See **CET or CN**

Child Tags:

Sub-tag	Example	Description
-	-	-

ID No: 29
Part Of Speech: Eternal Joined Conjunction
Tag: CET
Category: Conjunction
Example: যখন/CET ডাক্তার/NN আসল/VB **তখন/CET** রোগী/NN মারা/NNV গেল/VB

     when doctor has-come-3P at-that-time patient dying has-gone-3P
     "The patient has died when doctor has come"

    যেই/CET সে/PRP বাসা/NN থেকে/ON বের/NNV হলো/VBE **অমনি/CET** বৃষ্টি/NN
নামলো/VB

    when he home from come-out been-3P as-soon-as rain downed-3P
    "As soon as he has come out of home, rain started"

Description and Analysis:
When two clauses are joined by double coordinating conjunctions, are called eternal joined conjunctions [Saha et al. 2004]. Actually, when two conjunctions are paired to form eternal joined conjunctions, both of them are considered as 'coordinating conjunctions'.

Possible confusing tags:

CET or CN

Suspicion conjunctions (CN) when paired with another subordinating conjunction to form eternal joining conjunctions will be tagged as CET, not CN.

Ex: আমি যদি/CET বলি তাহলে/CET সে আসবে

  I if say-1P then he will-come-3P

  "If I say then he will come"

পাছে/CET কেউ শুনে ফেলে এইজন্য/CET সে আস্তে কথা বলছে

if someone after-hearing complete-3P for-this-reason he quietly word talking-3P

"He is talking quietly if someone could hear"

CET or DM—See **DM or CET**

Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 30
Part Of Speech: Subordinating Conjunction
Tag: CS
Category: Conjunction
Example: রহমান/NNP আমাকে/PRP+SFAC জিজ্ঞেস/NNV করল/VB **যে/CS** আমি/PRP চলে/VBT যাচ্ছি/VB কিনা/RP

  Rahman me ask did-3P that I after-leaving going-1P whether

  "Rahman asked me whether I am leaving or not"

নাফিদ/NNP পরীক্ষায়/NN+SFON আশানুরূপ/JJ ফলাফল/NN করতে/VBF পারেনি/VB **কেননা/CS** সে/PRP ভালভাবে/RB পড়াশোনা/NN করেনি/VB

  Nafid exam-at expected result to-do couldn't-3P because he well study didn't-3P

  "Nafid couldn't do expected result at the exam because he didn't study well"

আজ/NNT বৃষ্টি/NNV হতে/VBEF পারে/VB **কারণ/CS** আকাশ/NN মেঘাচ্ছন্ন/JJ

  today rain to-be can-3P because sky cloudy

  "It may rain today because sky is cloudy"

Description and Analysis:

A subordinating conjunction usually introduces a subordinate clause. A subordinate clause depends on the rest of the sentence for its meaning. It does not express a complete thought, so it does not stand alone. It must always be attached to a main clause that completes the meaning.

Possible confusing tags: NA

Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 31
Part Of Speech: Compound Subordinating Conjunction
Tag: CSC
Category: Conjunction

Example: আমি/PRP তাকে/PRP+SFAC কিছু/NN বলিনি/VB **এই**/**CSC** কারণে/CS সে/PRP আসেনি/VB

      I him something told-not-1P this reason-at he came-not-3P

      "I didn't tell him anything that's why he didn't come"

      আজ/NNT বৃষ্টি/NNV হচ্ছে/VB, **তাই**/**CSC** বলে/CS স্কুলে/NN+SFON যাবনা/VB

      today rain falling-3P, that's why school-at will-go-not-1P

      "It is raining today, that's why I won't go to school"

Description and Analysis:

Same as compound coordinating conjunction, subordinating conjunctions are also formed by more than one word such as 'এই কারণে' (for this reason), 'তাই বলে' (that's why). So, in that case all words except the last one will be marked as CSC. Thus any CSC will be followed by another CSC or a CS. This strategy will help to recognize those words as a single unit of subordinating conjunction.

Possible confusing tags: NA

Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 32
Part Of Speech: Postposition
Tag: ON
Category: Postposition
Example: আমি/PRP বাবার/NN+SF$ **সাথে**/**ON** ঢাকা/NNT গিয়েছিলাম/VB

      I father's with Dhaka went-1P

      "I went to Dhaka with father"

      তোমার/PRP+SF$ **দ্বারা**/**ON** এ/DM কাজ/NN হবে/VBE না/RB

      your by this work will-be-3P not

      "This work cannot be done by you"

      নাফিদ/NNP **থেকে**/**ON** অভীক/NNP লম্বা/JJ

      Nafid than Aveek tall

      "Aveek is taller than Nafid"

Description and Analysis:

A word that indicates the relationship between a noun or a pronoun, or a syntactic construction to another element in a sentence is called postposition. Unlike preposition in English that precedes a clause, in case of Bangla, postpositions usually follow nouns, pronouns or clauses. That's why, its being named as 'Postposition' (or 'Adposition').

Possible confusing tags:

    ON or NNT

    The words 'আগে' (before), 'পরে' (after) will be always tagged as temporal noun (NNT) when denotes time. Don't be led to tag them as postpositions (ON), because they can also be followed by another postposition.

    Ex: এক বছর/NN আগে/NNT   cf. এক বছর/NN আগে/NNT থেকে/ON

      one year before   one year before from

      "One year before"   "From one year before"

আমি তোমার/PRP+SF$ পর/NNT এসেছিলাম

    I your after came-1P

    "I came after you"

ON or NNL

Locative nouns not always used as nouns only, they can also be used as postpositions (ON).

Ex: সে আমার উপর/ON রেগে আছে   cf. সে আমার সাথে/ON রাগ করেছে

    he my on after-being-angry is-3P   he my with angry done-3P

    "He is angry at me"   "He has become angry with me"

আমি তার কাছে/ON চিঠি লিখেছি   cf. আমি তার জন্য/ON চিঠি লিখেছি

    I his to letter written-1P   I his for letter written-1P

    "I have written a letter to him"   "I have written a letter for him"

এই পোশাকটির মুল্য ১৫০০ টাকার নিচে/ON নয়

    this dress-the-GEN price 1500 taka's less not

    "The price of this dress is not less than 1500 taka"

And they can also be tagged as postposition even if followed by another postposition—

Ex: তোমার উপর/ON থেকে/ON আস্থা উঠে গেছে আমার

    your upside from reliance after-going-off gone-3P my

    "I don't have any more reliance on you"

Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 33

Part Of Speech: Interjection

Tag: UH

Category: Interjection

Example: বাহ্/UH !/. কী/RB সুন্দর/JJ দৃশ্য/NN !/.

    wow! what beautiful scenery!

    "Wow! What a beautiful scenery!"

হায়/UH !/. তুমি/PRP একি/QDM করলে/VB ?/.

    alas! you what done-2P ?

    "Alas! What you have done?"

Description and Analysis:

Interjections are words used to express strong feeling or sudden emotion. They are included in a sentence usually at the start to express a sentiment or a strong emotion such as surprise, disgust, joy, excitement or enthusiasm.

They generally come at the starting of a sentence followed by an exclamation point, or by a comma if the feeling's not as strong.

Possible confusing tags: NA

Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 34
Part Of Speech: Particles
Tag: RP
Category: Particles
Example: ছেলেটি/NN **তো/RP** এখনও/NNT এলোনা/VB
    boy-the PARTICLE yet come-not-3P
    "The boy has not come yet"

   আমি/PRP অতোটা/DMS জানিনা/VB যতোটা/QDM **নাকি/RP** আপনি/PRP জানেন/VB
   I that-much know-not-1P as-much PARTICLE you know-2P
   "I don't know as much as you know"

   তুমি/PRP **না/RP** খুব/RB দুষ্টু/NN
   you PARTICLE very naughty
   "You are very naughty"

Description and Analysis:
An uninflected item that has grammatical function but does not clearly belong to one of the major parts of speech is called particle. It is mostly used for words that help to encode grammatical categories (such as negation, mood). Negative particles are also merged into this category.
Possible confusing tags:
  RP or RB
Negative word 'না' (not) found to be used both as an adverb for negation and as a particle for reaffirmation. A word is a particle if it bears a stress at any clausal position. On other hand, a word is an adverb if it modifies a verb and contributes to the sentence's final semantic interpretation.
Ex: তুমি না/RP আমাকে কথাটা বলবে?
  you PARTICLE me word-the tell-2P
  "Aren't you supposed to tell me that?"
  But,
তুমি কখনো আমাকে কথাটা বল না/RB
  you ever me word-the tell-2P not
  "You never tell me that"

Sometimes 'না' appear to express a request rather than modifies a verb, although it appears at the same position as adverb. In that case it is a particle.
Ex: তুমি আমাকে বল না/RP
  you me tell-2P PARTICLE
  "You tell me please"
  RP or CC
Particles can also be used as coordinating conjunctions (CC) when join another clause to indicate the effect of current event in extend.
Ex: এটা হয়নি তো/CC কি হয়েছে?
  it been-not so what happened-3P ?

"It has not been done, so what happened?"

But be aware not to tag it as a coordinating conjunction (CC) when a clause is already attached by a subordinating (CS) or suspicion conjunction (CN). In these cases it would be a particle (RP).

Ex: এটা যদি/CN না হয় তো/RP কি হবে?

    it if not be PARTICLE what will-happen-3P?

    "If it is not done then what will happen?"

Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 35
Part Of Speech: Question Particle
Tag: QRP
Category: Particles
Example: তুমি/PRP উপন্যাসটি/NN পড়েছ/VB **কি/QRP** ?/.

    you novel-the read-2P QUESTION-PARTICLE ?

    "Have you read the novel?"

    আপনি/PRP **কি/QRP** আমাকে/PRP+SFAC চিনেন/VB ?/.

    you QUESTION-PARTICLE me know-2P ?

    "Do you know me?"

Description and Analysis:
Question particles are used to form the sentences that are used for basic yes/no question types. This type has been included mostly in grammatical aspect, rather than for pure syntactic distribution. The reason is to distinguish the syntactic constructions of interrogative sentences (basic yes/no question) from assertive sentences.

Possible confusing tags:
    QRP or QDM—See **QDM or QRP**

Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 36
Part Of Speech: Common Demonstratives
Tag: DM
Category: Demonstratives
Example: এই/**DM** বইগুলো/NN আমার/PRP+SF$

    these books my

    "These books are mine"

    কোন/**DM** বইতে/NN+SFON লেখাটা/NN পেলাম/VB না/RB

    any book-at writing-the found-1P not

    "I didn't find the writing at any of the books"

Description and Analysis:

A demonstrative used to demonstrate the identity of the thing to restrict their meaning by limiting the reference of the following noun. This tagset categorizes all types of demonstratives (proximal and distal) that usually precede nouns as 'common demonstratives' or 'simple demonstratives', because common demonstratives refer the following noun that can be either singular or plural and can be inflected by any suffixes in contrast to singular demonstratives (DMS).

Possible confusing tags:

DM or VOC

Demonstrative words are sometimes used as vocatives to call or address someone. So in that case tag it as VOC, not DM.

Ex: এই/VOC নাফিদ, স্কুলে যাও   cf. *এই/DM নাফিদ, স্কুলে যাও

     hey Nafid , school-at go-2P

     "Hey Nafid, go to school"

DM or CET

The concept to describe the following contrasts has been taken from Penn [Santorioni 1990] and has been adapted for Bangla.

When there are first members of the double conjunctions "যেকোন...অথবা", 'যেকোন' (either) and 'অথবা' (or) are tagged as eternal joining conjunctions (CET), not as demonstratives (DM).

Ex: যেকোন/DM বালক গাইতে পারত

     either boy to-sing could-3P

     "Either boy could sing"

     But,

     যেকোন/CET একজন বালক অথবা/CET একজন বালিকা গাইতে পারতো

     either one-SUFF boy or one-SUFF girl to-sing could-3P

     "Either a boy or a girl could sing"

Be aware that 'যেকোন' or 'অথবা', 'নতুবা' can sometimes function as demonstratives (DM) even in the presence of 'অথবা' or 'নতুবা'.

Ex: যেকোন/DM বালক অথবা/CC বালিকা গাইতে পারতো

     either boy or girl to-sing could-3P

     "Either boy or girl could sing"

DM or QFNUM—See QFNUM or DM

Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 37
Part Of Speech: Singular Demonstrative
Tag: DMS
Category: Demonstratives
Example: এটি/DMS বল/NN

     this ball

     "This is ball"

     ওটি/DMS এখানে/NNL পাবে/VB না/RB

     that here-at will-find-2P not

"You won't find it here"

কে/QPR **ওটা/DMS** ?/.
who that?
"Who is that?"

Description and Analysis:
Singular demonstratives demonstrate nouns or nominal phrases, in which the head noun is singular in number and cannot be inflected by any suffix. For example, in a nominal phrase: কলমের কালি (ink of pen), the word কালি (ink) is the head noun and কলমের (pen's) is a possessive modifier. If the head noun is inflected, it can have different syntactic distribution.

Ex: এটি/DMS কলম/NN
    this pen
    "This is pen"

এটি/DMS কলমের কালি
this pen's ink
"This is ink of pen"

But,
এটি/DMS কলমে/NN+SFON
this pen-at
"This is at pen"

এটি/DMS কলমের/NN+SF$ কালিতে/NN+SFON
this pen's ink-at
"This is at ink of pen"

Singular demonstratives can syntactically pop up without the presence of head noun and can also appear after the noun that it demonstrates, such as কলম এটি
                                                                pen this
                                                                "This is pen"

Possible confusing tags:
    DMS or NN
    When demonstratives are used pronominally, i.e. without a head noun, they should still be tagged as demonstratives (DMS)—not as nouns (NN).
    Ex: আমি এটা/DMS ধরে রাখতে পারছিনা
        I this held to-keep can't-1P
        "I can't hold this up"

Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 38
Part Of Speech: Question Demonstratives
Tag: QDM
Category: Demonstratives
Example: সুমী/NNP **কোন/QDM** পোশাকটি/NN কিনতে/VBF চায়/VB ?/.

Sumee which dress-the to-buy want-3P ?
"Which dress Sumee wants to buy?"

আমি/PRP জানি/VB **কি/QDM** চাও/VB তুমি/PRP
I know what want-2P you
"I know what you want"

আমাকে/PRP+SFAC কথাটা/NN বল/VB **যেটা/QDM** তুমি/PRP শুনেছ/VB
me word-the tell-2P what you heard-2P
"Tell me that that you have heard"

Description and Analysis:
Demonstratives that are used to form interrogative sentences and to join relative clauses are called question demonstratives. Both relative and question demonstratives are subsumed under a single category (QDM).

Possible confusing tags:

QDM or QRP

When the word 'কি' is used to construct yes-no questions it is Question Particle (QRP), on the other hand when it demonstrates a nominal object in an interrogative sentence it is a Question Demonstrative (QDM). Since they are orthographically similar and can also appear in similar position in a sentence, one has to consider the entire sentence to disambiguate between these two tags.

When 'কি' is used as question particle—

Ex: সে কি/QRP দেখেছিলো?
   he QUESTION-PARTICLE saw-3P ?
   "Did he see?"

When 'কি' is used as question determiner—

Ex: সে কি/QDM দেখেছিলো?
   he what saw-3P ?
   "What did he see?"

Child Tags:

Sub-tag	Example	Description
QDM-Q	কি, কোন, কোনটা	Demonstratives that used for both interrogative sentences and relative clauses
QDM-R	যে, যেটা	Demonstratives that used for only relative clauses

ID No: 39
Part Of Speech: Quantifiers
Tag: QF
Category: Quantifiers
Example: নাফিদ/NNP **কিছু/QF** পাখি/NN দেখেছে/VB
        Nafid some birds saw-3P
        "Nafid saw some birds"

আমি/PRP কিন্তু/RP তাদেরকে/PRP+SFAC **সব/QF** কথা/NN বলিনি/VB

    I but them all words told-not-1P

    "I didn't tell them everything"

Description and Analysis:

Quantifiers are type of demonstratives but they can also denote the quantity of nouns. All words that quantify the nominal objects have been categorized as quantifiers (QF). In a noun phrase, quantifiers may occur without head noun.

Possible confusing tags:

    QF or NN

    Although quantifiers can be used without the head noun, but sometimes their pronominal use should be considered as noun (NN).

    Ex: আমি কিছু/NN বলতে চাই

        I something to-say want-1P

        "I want to say something"

    তুমি কি কিছু/NN দেখেছ?

    you QUESTION-PARTICLE something saw-2P?

    "Did you see anything?"

    But they will be tagged as quantifiers (QF) when quantify nominal object that is already mentioned in a given context.

    ছবিগুলো রাখা আছে ওখানে; কিছু/QF দেখতে পার

    pictures keeping is-3P there; some to-see can-2P

    "The pictures have been kept there; you can see some of those"

    QF or PRP

Inclusive pronouns: সবাই (all), সকলে (all), উভয় (both) should be tagged as PRP, not as quantifier (QF).

Ex: চল সবাই/PRP

    go-2P all

    "Let's go all"

    তোমরা সব/PRP এদিকে এসো

    you all here-at come-2P

    "All of you come here"

    QF or JJ

Quantifiers can also be quantified by another quantifier; they shouldn't be tagged as adjectives.

Ex: তুমিতো আরো/QF অনেক/QF পাচ্ছ

    you more many getting-2P

    "You are also getting many more"

    অল্প/QF কিছু/QF কাগজ/NN নিয়ে আস   cf. *অল্প/JJ কিছু/QF কাগজ/NN নিয়ে আস

    little few papers after-bringing come-2P

    "Bring a few papers"

    QF or RB

The words, such as 'অনেক' (many), 'বহু' (many), 'প্রচুর' (lots) are quantifiers (QF) when

they are used pronominally and being the complement of transitive verb types.

Ex: আমি ওকে অনেক/QF বলেছি cf. আমি ওকে অনেক/QF কথা/NN বলেছি
 I him many told-1P   I him many word told-1P
 "I told him a lot"   "I told him a lot of things"

 আমি প্রচুর/QF খেয়েছি   cf. আমি প্রচুর/QF ভাত/NN খেয়েছি
 I lot ate-1P   I lot rice ate-1P
 "I ate a lot"   "I ate lots of rice"

But these words also used as adverbs (RB) by modifying verbs to denote the repetitiveness of an action from the context's perspective. So one should consider the whole context instead of the specific words or collocations.

Ex: আমি ওকে অনেক/RB বলেছি cf. আমি ওকে অনেকবার/RB বলেছি
 I him many told-1P   I him many-times told-1P
 "I told him a lot of times"   "I told him a lot of times"

They are also adverbs when modify intransitive verb types:

Ex: আমি অনেকটা/RB ঘুমিয়েছি/VB
 I a-lot slept-1P
 "I have slept a lot"

 সে এতটা/RB রেগে/VBT গেছে বুঝতে পারিনি
 he so-the after-getting-angry gone-3P to-understand could-not-1P
 "I didn't understand he had become so angry"

Exception is, the cognate objects of intransitive verbs can also be quantified by quantifiers (QF or QFNUM).

Ex: আমি অনেক/QF ঘুম/NN ঘুমিয়েছি/VB
 I a-lot sleep slept-1P
 "I have slept a lot of sleep"

 সে একটা/QFNUM নাচ/NN নাচল/VB বটে!
 he one-the dance danced-3P no-doubt!
 "He has no doubt danced a great dance!"

Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 40
Part Of Speech: Quantifier Numbers
Tag: QFNUM
Category: Numbers
Example: এক/QFNUM লোক/NN এসেছিলো/VB
 one person came-3P
 "A man came"

এই/DM বইটির/NN+SF$ মূল্য/NN ২০০/QFNUM টাকা/NN

this book-the-GEN price 200 taka

"Price of this book is 200 taka"

Description and Analysis:

This category is a sub-categorization of quantifiers (QF) that represent numerical numbers to quantify nominal objects. Number-number combinations are also tagged as quantifier numbers if they don't have same distribution as adjectives.

Possible confusing tags:

   QFNUM or NN

In case where in some constructions Quantifiers Numbers are used as noun, they will be tagged as noun (NN).

Ex: প্রতি পাঁচজন/NN থেকে একজন/NN আসবে

   each five-SUFF from one-SUFF will-come-3P

   "One person will come from each of the five persons"

When there is an expression for counting event where quantifiers are appeared pronominally, they will be tagged as quantifiers as well, not nouns. The nominal phrase that consists of quantifiers is followed by a postposition (ON).

Ex: একজন/QFNUM একজন/QFNUM করে/ON

   one-SUFF one-SUFF by

   "One person by one person"

   দুজন/QFNUM দুজন/QFNUM দ্বারা/ON

   two-SUFF two-SUFF by

   "By each of two persons"

N.B. here the word 'করে' is tagged as postposition (ON). Please see discussion for the word 'করে' under **specific words and collocation section**.

If preceded by quantifying adjectives such as 'মোট' (total) 'প্রায়' (approximately) or enforced to be a quantifier when in the scope of some words like 'সংখ্যা' (number), then it is a numeral quantifier (QFNUM).

Ex: মোট/JJ পাঁচজন/QFNUM

   total five-SUFF

   "Total five persons"

   প্রায়/JJ পাঁচজন/QFNUM

   approximately five-SUFF

   "Approximately five persons"

   শিক্ষার্থীর সংখ্যা ছিল দশজন/QFNUM

   apprentice's number was ten-SUFF

   "Number of apprentices was ten"

   QFNUM or JJ

Number-number combinations should be tagged as adjectives (JJ) if they have the same distribution as adjectives.

Ex: সেই ৯-১১/JJ সন্ত্রাসী হামলা   cf. সেই ভয়াবহ/JJ সন্ত্রাসী হামলা

that 9-11 terrorist attack   that terrified terrorist attack

"That 9-11 terrorist attack"   "That terrifying terrorist attack"

But if number-number combinations are used to numerically quantify nominal objects, they will be tagged as quantifier number (QFNUM) as well. In this case, they are generally followed by suffixes or definite markers to indicate plurality of objects, such as suffix 'জন' (to indicate plurality of humans) or definite markers 'টি', 'টা', 'খানা', 'খানি' (to indicate plurality of materials).

Ex: ১০-১২/QFNUM জন/JJ লোক
      10-12 SUFF man
      "10 to 12 numbers of peoples"

      ৫-৬টি/QFNUM বই
      5-6-the book
      "5 to 6 numbers of books"

N.B. Note that the suffix 'জন' has been tagged as an adjective (JJ). Please see the discussion for the word 'জন' in **specific words and collocations** section for further details.

In SPSAL tagset [SPSAL 2007] all ordinal numbers are also considered as quantifier numbers, but here the ordinal numbers will be tagged as adjective (JJ).

Ex: গত ৫/JJ জুলাই/NNT ছিল আমার ভাইয়ের জন্মদিন
      last 5 July was my brother's birthday
      "Last 5ᵗʰ July was my brother's birthday"

QFNUM or DM

If quantifier numbers are used as counterparts of another demonstratives such as 'অন্য' (another), 'আরেক' (another one) etc. they will be tagged as demonstratives (DM).

Ex: এ হচ্ছে এক/DM দিক, অন্য/DM দিক হচ্ছে এই দিকে
      this being-3P one side, another side being-3P this side-at
      "This is one side, another side is here"

      এক/DM ভাই ছোট, আরেক/DM ভাই বড়
      one brother younger , another brother elder
      "One brother is younger, another brother is elder"

Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 41
Part Of Speech: Question Quantifier
Tag: QQF
Category: Quantifier
Example: এই/DM জামাটির/NN+SF$ মুল্য/NN **কত/QQF** ?/.
      this dress-the-GEN price how-much ?
      "What is the price of this dress?"

      ইকবাল/NNP সাহেব/NN দোকানদারকে/NN+SFAC বলেছেন/VB **কতটুকু/QQF** চাল/NN

তিনি/PRP কিনতে/VBF চান/VB

Iqbal Mr. to-shopkeeper told-3P how-much rice he to-buy want-3P

"Mr. Iqbal told shopkeeper how much rice he wants to buy"

আমি/PRP শুনেছি/VB তুমি/PRP তাদেরকে/PRP+SFAC **যতটুকু/QQF** বলেছ/VB

I heard-1P you them how-much told-2P

"I heard how much you have told them"

Description and Analysis:

Question quantifiers are also a type of quantifiers which are used to form interrogative sentences or being used to join relative clauses. Thus, question words and relative categories for quantifiers have been subsumed under a single tag (QQF).

Possible confusing tags: NA

Child Tags:

Sub-tag	Example	Description
QQF-Q	কত, কতটুকু	Quantifiers that used for both interrogative sentences and relative clauses
QQF-R	যত, যতটুকু	Quantifiers that only used for relative clauses

ID No: 42

Part Of Speech: Foreign Word

Tag: FW

Category: Foreign Word

Example: '/' **Bengali/FW** '/' আসলে/RB একটা/QFNUM বিদেশী/JJ শব্দ/NN

' Bengali ' actually one-the foreign word

"Bengali is actually a foreign word"

গতকাল/NNT উইকিপিডিয়ার/NNP+SF$ প্রথম/JJ পৃষ্ঠার/NN+SF$ শিরোনামে/NN+SFON লিখা/VBM ছিল/VB **International/FW Mother/FW Language/FW Day/FW**

yesterday Wikipedia's first page's headline-at writing was-3P International Mother Language Day

"Yesterday the writing of the headline at the first page of Wikipedia was International Mother Language Day"

Description and Analysis:

The words taken from another language used in Bangla and not being used by native speaker are tagged as Foreign Word (FW). But due to long historical consequences, Bangla has absorbed countless words from other languages which have been fully integrated into the core lexicon. These words are no more considered as foreign words, for example 'টেবিল' (table), 'চেয়ার' (chair) from English and 'আদালত' (court), 'সুফি' (a person of Muslim mystics) from Persian. One should use his/her judgment to identify what is a foreign word.

Possible confusing tags: NA

Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 43
Part Of Speech: Symbol
Tag: SYM
Category: Symbol
Example: x/**SYM** ∞/**SYM** yz/**SYM** ,/, যেখানে/QNL **y/SYM** এবং/CC **z/SYM** উভয়ই/PRP সচল/JJ

    x ∞ yz, where y and z both variable
    "x ∞ yz, where both y and z are variables"

    θ =90°/SYM হলে/VBC **sin θ =/SYM** কত/QQF হবে/VBE ?/.
    θ =90° if-been sin θ = how-much will-be-3P ?
    "If θ =90° then sin θ = what?"

    আমার/PRP+SF$ খরচের/NN+SF$ পরিমাণ/NN প্রায়/JJ ১৭,০০০/QFNUM **৳/SYM**

    my expense's amount around 17,000 ৳

    "The amount of my expense is around 17,000 ৳"

Description and Analysis:
Following description has been adapted from Penn tagset guidelines [Santorini 1990].

This tag should be used for mathematical, scientific and technical symbols or expressions that aren't words of Bangla. Sign of currencies like ৳ (Taka), $ (Dollar) are also tagged as symbol (SYM).

It should not be used for any and all technical expressions. For instance, the names of chemicals, units of measurements (including abbreviations thereof) and the like should be tagged as nouns.
Possible confusing tags: NA
Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 44
Part Of Speech: List Item Marker
Tag: LS
Category: List Item Marker
Example: এডিসনের/NNP+SF$ মতে/NN+SFON সাফল্য/NN হলো/VBE –/-
    ১/**LS** ./. ১/QFNUM ভাগ/NN উদ্দীপনা/NN ,/, এবং/CC
    ২/**LS** ./. ৯৯/QFNUM ভাগ/NN পরিশ্রম/NN
    Edison's opinion-by success is –
      1. 1 quantity inspiration , and
      2. 99 quantity perspiration
    "According to Edison, success is –
      1. 1 percent inspiration, and
      2. 99 percent perspiration"

প্রধানত/RB দুটি/QFNUM অঞ্চলে/NN+SFON বাংলা/NNP ভাষা/NN প্রচলিত/JJ –/-

    ক/LS ./. বাংলাদেশ/NNP

    খ/LS ./. পশ্চিমবঙ্গ/NNP ,/, ভারত/NNP

mainly two region-at Bangla language usual –

    a . Bangladesh

    b . West-Bengal, India

"Bangla language is usually spoken mainly at two regions –

    a. Bangladesh

    b. West-Bengal, India"

Description and Analysis:

Following description has been adapted from Penn tagset guidelines [Santorini 1990].

This category includes letters and numerals when they are used to identify items in a list.

Possible confusing tags: NA

Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 45
Part Of Speech: Postpositional Suffixes
Tag: SFON
Category: Suffixes
Example: শিশুটি/NN **বিছানায়/NN+SFON** ঘুমাচ্ছে/VB

    child-the bed-at sleeping-3P

  "The child is sleeping at bed"

এতে/DM+SFON কি/QRP পুরোটা/NN চলবে/VB ?/.

this-by QUESTION-PARTICLE whole-the continue-will-3P ?

"Will it be enough for the whole thing?"

Description and Analysis:

Description of the case system is prominent factor to obtain syntactical analysis of any language that is rich in morphology and case system. That is, detail formalism of the case system entails formalism of the syntax [Azad 1994]. In case of Bangla which is rich in morphology and case system, a noun or pronoun is inflected with morphology/case suffix. Since these suffixes inflect a nominal object and convert it to have different features and syntactic category, these can be assigned a single POS category which is 'Suffixes'. The motivation behind this: if different tag is introduced for each inflected nominal object, the size of the tag set will be large. Rather, it will be concise decision to have another POS category for suffixes and put a '+' sign when a noun or pronoun is inflected. This will make a compact tagset and information can be easily retrievable from the assigned tag.

Postpositional suffixes attached with nouns or pronouns to form phrases that are categorically and structurally equivalent to postpositional phrases. These postpositional phrases have same syntactic distribution as those which are constructed with a normal postposition placed after noun.

Possible confusing tags:

    NNP+SFON or NNP

    In Bangla, sometimes the subject is being inflected with a postpositional marker (usually common noun and proper noun) to amplify that subject as an 'agent' with 'instrumental' case

form, while it's syntactic distribution remains as a nominal subject of the sentence. So only in this case, we can tag that nominal subject simply as proper noun (NNP), not NNP+SFON.

Ex: নাফিদে/NNP বল খেলছে   cf. *নাফিদে/NNP+SFON বল খেলছে

  Nafid-AGENT ball playing-3P
  "Nafid is playing ball"

Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 46
Part Of Speech: Accusative Suffixes
Tag: SFAC
Category: Suffixes
Example: আমি/PRP **তোমাকে/PRP+SFAC** ব্যাগটি/NN দিয়েছি/VB

  I you bag-the gave-1P
  "I gave you the bag"

  ভিক্ষুকদের/NN+SFAC কিছু/QF টাকা/NN দাও/VB

  beggars-to some money give-2P
  "Give some money to beggars"

  আজ/NNT **আমাকে/PRP+SFAC** অনেক/QF কাজ/NN করতে/VBF হবে/VBE

  today I-to a-lot-of work to-do will-be-3P
  "Today I have to do a lot of works"

Description and Analysis:
Accusative inflectors are added as suffix to denote a noun's or pronoun's case to be either accusative or dative. The suffix or case marker only gives the lexical information about a noun's inflection; it doesn't define the actual case distribution of an inflected noun. As shown in the third example above [also see the discussion for existential verbs (VBE and VBEF)], 'আমাকে' is a pronoun inflected with accusative case marker but it remains as a subject because the main finite verb of the sentence is an existential verb (VBE) 'হবে'.

The inflector 'দের' denotes both a nominal object's case to be accusative (or dative) and the number to be plural. Since the tagset described here doesn't include any information about plural inflections on nouns, this suffix has been included only into accusative inflector category.

Possible confusing tags:
  SFAC or SF$**-- See **SF$ or SFAC

Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 47
Part Of Speech: Possessive Suffixes
Tag: SF$
Category: Suffixes

Example: **করিমের/NNP+SF$** কলমটি/NN আমি/PRP নিয়েছি/VB

       Karim's pen-the I taken-1P

       "I have taken Karim's pen"

অনুগ্রহ/NN করে/VBT **আপনার/PRP+SF$** নামটি/NN আমাকে/PRP+SFAC বলুন/VB

please after-doing your name-the me tell-2P

       "Please tell me your name"

Description and Analysis:

According to linguistics, possessive suffix is a suffix attached to a noun to indicate its possessor, much in the manner of possessive adjectives. Since Bangla is referred into Indo-Aryan language family, it also supports possessive suffix. Recent linguistic analysis also categorizes possessive nouns as possessive determiners.

Possible confusing tags:

    NN+SF$ or JJ

In many cases, nouns are inflected with possessive suffix (SF$) to act as modifiers and describe the type or properties of the target noun. Since they are used as modifiers, they will be tagged as adjectives (JJ), not as nouns with possessive inflection (NN+SF$), because they don't possess, they rather modify.

Ex: কাঠের/JJ চেয়ার,   মোমের/JJ পুতুল,   মাটির/JJ ঘর,   প্রাণের/JJ বন্ধু

    wood's chair,   wax's doll,   mud's house,   soul's mate

      "wooden chair"   "doll of wax"   "mud house"   "soul mate"

    SF$ or SFAC

There is a case of lexical confusions when the case suffix 'দের' is being used as short form of accusative case marker 'দেরকে'. So the inflector 'দের' can be an accusative/dative marker and also a possessive marker. The only way to disambiguate this case is to analyze the meaning of the whole sentence or the context.

Ex: ভিক্ষুকদের/NN+SFAC টাকাগুলো দিয়ে দাও

    beggars-to moneys after-giving give-2P

    "Give the moneys to beggars"

    But,

    আমাদের/PRP+SF$ টাকাগুলো দিয়ে দাও

    our moneys after-giving give-2P

    "Give us our moneys"

Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 48

Part Of Speech: Sentence Final Punctuation Marks

Tag: .

Category: Punctuation Marks

Example: আমি/PRP যাচ্ছি/VB │/.

       I leaving-1P .

       "I am leaving."

ইস্ !/. যদি/CN যেতে/VBF পারতাম/VB !/.

oops ! if to-go could-1P !
"Oops! I wish I could go."

তুমি/PRP কি/QDM চাও/VB **?/.**
you what want-2P ?
"What do you want?"

Description and Analysis:
Sentence final punctuation marks are the punctuation commonly placed at the end of several different types of sentences in Bangla and many other languages.

Sentence final punctuation marks are consist of . (full stop), ! (exclamation sign) and ? (question mark) which are placed at the end of a line or a text, to denote the end of a sentence or an expression.
Possible confusing tags: NA
Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 49
Part Of Speech: , (Comma)
Tag: ,
Category: Punctuation Marks
Example: আপনি/PRP দাঁড়ান/VB **,/,** আমি/PRP আসছি/VB
আপনি wait-2P , I coming-1P

you wait-2P , I coming-1P
"Wait, I am coming"

আমি/PRP **,/,** তুমি/PRP **,/,** সে/PRP আর/CC নাফিদ/NNP শিক্ষকের/NN+SF$ কাছে/ON পড়তে/VBF যাব/VB
I , you, he and Nafid teacher's with study-to will-go-1P
"You, he, Nafid and I will go to teacher to study"

Description and Analysis:
Comma is a mid-separator and is being used principally for separating objects and clauses in many contexts.
Possible confusing tags: NA
Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 50
Part Of Speech: ; (Semicolon), : (Colon)
Tag: :
Category: Punctuation Marks
Example: সুমী/NNP করেছে/VB ৭০/QFNUM পয়েন্ট/NN **;/:** রেহানা/NNP **,/,** ৬০/QFNUM **;/:** এবং/CC দীপা/NNP **,** ৪০/QFNUM ।/.
Sumee done-3P 70 point; Rehana, 60; and Deepa, 40 .
"Sumee scored 70 points; Rehana, 60; and Deepa, 40."

বাংলাদেশের/NNP+SF$ স্বাধীনতা/NN ৹ঃ/: এক/QFNUM রক্তক্ষয়ী/JJ সংগ্রামের/NN+SF$ ফসল/NN

Bangladesh's independence : one gory struggle's outcome
"Independence of Bangladesh: An outcome of a gory struggle"

Description and Analysis:
Colons and semicolons are kinds of punctuation marks and they have distinct set of contributions of their own within a context. But here, they have been subsumed to have same syntactic category. The reason behind this, they both actually tend to bind two sentences than they would be if separated by a full stop.

Possible confusing tags: NA

Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 51
Part Of Speech: - (Dash), -- (Double Dash)
Tag: -
Category: Punctuation Marks
Example: পৃষ্ঠা/NN দ্রষ্টব্য/JJ :/: ১১/QFNUM -/- ২৮/QFNUM

    page reference : 11 - 28
    "Page reference: 11-28"

দ্বিতীয়/NNPC বিশ্বযুদ্ধ/NNP (/(১৯৩৯/QFNUM -/- ১৯৪৫/ QFNUM)/)

2nd World-War (1939 – 1945)
"2nd World-War (1939–1945)"

Description and Analysis:
Dashes (-) or Double Dashes (--) are widely used in different languages to illustrate a relationship or connection between two things, or denote a range of values. Both of the categories are subsumed using a single tag (-).

When Dashes are used to define only ranges, they are explicitly tagged. In any other cases, they act as a connector such as compound adjectives. For example: পিতা-পুত্র/JJ সম্পর্ক/NN

              Father-son relationship
              "Father-son relationship"

But Double Dashes (--) are most likely to be tagged explicitly, because they are used to join sentences, indicates a parenthetical thought or some similar interpolation.

Possible confusing tags: NA

Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 52
Part Of Speech: ((Left Parenthesis), { (Left Curly Brace), [(Left Square Bracket)
Tag: (
Category: Punctuation Marks

Example: তৎকালীন/JJ পূর্ব/NNPC পাকিস্তান/NNP (/(বর্তমানে/NNT বাংলাদেশ/NNP)/) সরকার/NN

   former East Pakistan (present-at Bangladesh) Government
   "Former East Pakistan (Bangladesh at present) Government"

   সেই/DM ব্রিটিশ/JJ আমল/NN (/(১৭৫৭/ QFNUM -/- ১৯৪৭/ QFNUM)/)
   that British period (1757 – 1947)
   "That British period (1757 – 1947)"

Description and Analysis:
Brackets are punctuation marks used in pairs to set apart on interject text within other text. Left brackets are the left or the starting parts of that pair.

In this tagset, all kinds of left brackets, such as ((left parentheses), { (left curly brace), [(left square bracket) are being tagged using left parenthesis [( ].

Possible confusing tags: NA
Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 53
Part Of Speech:) (Right Parenthesis), } (Right Curly Brace),] (Right Square Bracket)
Tag:)
Category: Punctuation Marks
Example: আলোর/NN+SF$ গতি/NN প্রতি/DM সেকেন্ডে/NN+SFON ২৯,৯৭,৯২,৪৫৮/QFNUM মিটার/NN (/(ঘন্টায়/NN+SFON ১০৭,৯২,৫২,৮৪৯/QFNUM কিলোমিটার/NN)/)

   light's speed every second-at 299,792,458 meter (hour-at 1,079,252,849 kilometer)
   "The speed of light is 299,792,458 meters per second (1,079,252,849 kilometers per hour)"

   ভাষাবিজ্ঞান/NNP (/(Linguistics/FW in/FW English/FW)/) একটি/QFNUM কঠিন/JJ ক্ষেত্র/NN

   Linguistics (Linguistics in English) one-the difficult field
   "Linguistics is a difficult field"

Description and Analysis:
Right brackets are the right or the ending parts of the bracket pair.

In this tagset all kind of right brackets, such as) (right parentheses), } (right curly brace),] (right square bracket) are being tagged using right parenthesis [ )].

Possible confusing tags: NA
Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 54
Part Of Speech: ', " (Opening Left Quotes)
Tag: LQ

Category: Punctuation Marks

Example: **'/LQ** চেয়ার/NNP '/RQ শব্দটা/NN এসেছে/VB ইংরেজী/NNP ভাষা/NN থেকে/ON

        'chair' word-the has-come-3P English language from

        "The word 'chair' has come from English language"

        সে/PRP বলল/VB ,/, **"/LQ** আজ/NNT বৃষ্টি/NNV হতে/VBEF পারে/VB "/RQ

        he said-3P , " today raining be-to can-3P "

        "He said, "It may rain today""

Description and Analysis:

Quotation marks or inverted commas (occasionally speech marks) are punctuation marks used in pairs to set off speech, a quotation, a phrase or a word. The pair consists of an opening quotation mark and a closing quotation mark, which may or may not be the same character.

Both single and double opening quotation marks are tagged as LQ.

Possible confusing tags: NA

Child Tags:

Sub-tag	Example	Description
-	–	-

ID No: 55

Part Of Speech: ', " (Closing Right Quotes)

Tag: RQ

Category: Punctuation Marks

Example: '/LQ ২/NNP **'/RQ** সংখ্যাটা/NN কি/QRP মৌলিক/JJ ?/.

        '2' number-the what prime ?

        "Is the number '2' prime?"

        "/LQ যেমন/CET কর্ম/NN ,/, তেমন/CET ফল/NN **"/RQ** একটা/QFNUM বাগধারা/NN

        " like work , like result " one-the phrase

        ""As you sow, sow you reap" is a phrase"

Description and Analysis:

Closing quotes are the right or ending quotes of the quotation pair. Both single and double closing quotation marks will be tagged as RQ.

Possible confusing tags: NA

Child Tags:

Sub-tag	Example	Description
-	–	-

4. Specific words and collocations

অনেকে : It is tagged as common noun (NN), since it has same distribution as nouns.

    Ex: অনেকে/NN যাচ্ছে

        many-people going-3P

        "Many people are going"

But it is also used as an inclusive pronoun and tagged as personal pronoun (PRP).

    Ex: তাদের/PRP+SF$ অনেকে/PRP যাচ্ছে

        their many-people going-3P

        "Many of them are going"

        তারা/PRP অনেকে/PRP যাচ্ছে

        they many-people going-3P

        "Many of them are going"

কখন : It is a simple temporal noun (NNT) when denotes an indefinite time span—

    Ex: সেতো কখন/NNT চলে গেছে

        he-PARTICLE long-before after-leaving gone-3P

        "He left long before"

        লোকটি সেই কখন/NNT থেকে বসে আছে

        man-the that long-before from after-sitting is-3P

        "The man has been sitting since long before"

But it is tagged as question temporal noun (QNT) as well if the sentence is interrogative—

    Ex: আজ কখন/QNT থেকে বৃষ্টি শুরু হলো?

        today when from raining start been-3P?

        "Since when did the rain start today?"

করে : This word exhibits various morpho-syntactic functions. This is found to be the most frequent word in Bangla corpus and hence the most ambiguous lexical item which requires special treatment in any NLP application. A corpus based theoretical analysis of the word 'করে' [Biswas et al. 2008] shows 11 morpho-syntactic functions and recommends total 5 types of Part-Of-Speech categories for tagging which are—Noun with locative case marker, Finite verb, Non-finite verb, Postposition and Particle. The category 'Particle' has been used when the preceding word is always being an adjective, adverb or quantifier. But in our analysis we categorize this also as Postposition. To facilitate the understanding of readers following are the 11 morpho-syntactic categories enlisted with examples from that theoretical analysis [Biswas et al. 2008]. Each time, the word 'করে' has been tagged according to corresponding POS category.

Type 1 (Noun)

Ex: আয়-করে/NN+SFON ১০% বৃদ্ধি হয়েছে

    income-tax-at 10% increase been-3P

    "Income-tax has been increased by 10%"

Type 2 (Finite Verb)

Ex: রাম রোজ একই ভুল করে/VB

    Ram everyday same mistake do-3P

    "Ram does the same mistake everyday"

Type 3 (Perfective Participle)

Ex: রাম হোমওয়ার্ক করে/VBT স্কুলে গেলো

    Ram homework after-doing school-at gone-3P

    "After doing the homework, Ram went to the school"

Type 4 (Adverbial Participle)

Ex: চুপ করে/VBT শোনো

    quiet after-doing listen-2P

    "Listen quietly"

Here 'চুপ করে' is a complex predicate analysis that means "to keep quiet". The non-finite form of this complex predicate act as an adverbial participle, meaning 'quietly', rather than a perfective particple.

Type 5 (Action Instrumentalized)

Ex: সে বল প্রয়োগ করে/ON দরজা খুলল

    he force application after-doing door opened-3P

    "He opened the door by applying force"

The translation of this sentence to many languages will surface these types of sentences as prepositional phrases. Thus it would help during machine translation if it is tagged as postposition (ON). But it is also possible to treat this type as an adverbial participle, in which case it should be annotated as a non-finite perfective verb (VBT). It could be possible that this scheme will lead to a more accurate POS tagger.

Ex: সে বল প্রয়োগ/NNV করে/VBT দরজা খুলল

    he force application after-doing door opened-3P

    "He opened the door after applying force"

Type 6 (Adverbializer)

Ex: মাটির সঙ্গে ভালো/JJ করে/ON মিশিয়ে দিন

    soil's with good after-doing after-mixing give-3P

    "Mix well with soil"

It is tagged as Postposition, although in the analysis [Biswas et al. 2008], it has been recommended to tag it as a Particle, because the phrase 'ভালো করে' is an adverbial phrase and 'করে' act as a particle. But we think that, as mentioned earlier, adjectives and quantifiers are able to be used pronominally (without head noun) in a Noun Phrase and can be followed by postpositions. For example: লোকটি গরীব/JJ থেকে/ON ধনী/JJ হয়েছে

                          man-the poor from rich been-3P

                      "The man has become rich from being poor"

The word 'গরীব' has been tagged as an adjective not as a noun while followed by a Postposition 'থেকে' (from). The reason is, both of the words 'গরীব' and 'ধনী' can be intensified such as 'খুব গরীব' (very poor) and 'খুব ধনী' (very rich). Similarly, 'ভালো' is tagged as an adjective followed by a postposition 'করে' in the phrase 'ভালো/JJ করে/ON'

which can be interpreted as 'in a good manner'. That is, postpositional phrase functioning as adverbs. Another thing is, in this context 'করে' can be replaced by 'ভাবে' as in 'ভালো/JJ ভাবে/NN+SFON' where 'ভাবে' is an inflected from of 'ভাব' (manner) [ভাবে → ভাব+এ]. The phrase 'ভালো ভাবে' is also a postpositional phrase i.e. inflected with postpositional suffix (SFON) and has the same interpretation: 'in a good manner'.

But when 'ভাবে', 'করে' are written together with adjectives as a suffix such as 'ভালকরে', 'ভালভাবে' can simply be tagged as adverbs (RB).

Type 7 (Postposition)
Ex: থালাতে/NN+SFON করে/ON খাবারটা নিয়ে এসো
　　plate-at after-doing food-the after-bringing come-2P
　　"Bring the food in a plate"

　　গাড়ী করে/ON বাড়ী গেলাম
　　car after-doing home went-1P
　　"I went home by a car"

Type 8 (Hedged Adverbs)
Ex: অল্প/QF করে/ON ভাত দিন
　　little after-doing rice give-2P
　　"Give rice a little"
As discussed in **Type 5 (Adverbializer)**, like adjectives, quantifiers are also used pronominally and can be followed by postpositions. Here the word 'অল্প' is a quantifier and 'অল্প করে' is a postpositional phrase functioning as adverbs.

Type 9 (Hedged Adjectives)
Ex: লম্বা/JJ করে/ON ছেলেটা
　　tall after-doing boy-the
　　"The tallish boy"
Same as **Type 5 (Adverbializer)** adjectives are here followed by postposition. The word 'লম্বা' is an adjective and 'লম্বা করে' is a postpositional phrase functioning as adjectives that modify the noun 'ছেলেটা'.

Type 10 (Distributive or Iterative Event)
Ex: প্রত্যেক ক্ষেত্রে দুটি/QFNUM করে/ON বিকল্প তৈরী হয়েছে
　　every case-at two after-doing option create been-3P
　　"There are two options in each of the cases"

　　একজন/QFNUM একজন/QFNUM করে/ON দেখে আসব
　　one-SUFF one-SUFF after-doing after-seeing will-come-1P
　　"(We) will go and see one by one"

Type 11 (Frozen Expressions)
Ex: অনেক/QF করে/ON থাকতে বলেছেন

many after-doing to-stay said-3P

"He requested several times to stay back"

কত/QF করে/ON থাকতে বললেন

how-much after-doing to-stay said-3P

"Requested so many times to stay back"

কারো : It is tagged as a possessive inflected personal pronoun (PRP+SF$).

Ex: রহমান ছাড়া এ বিষয়ে আর কারো/PRP+SF$ ভাল ধারণা নেই

Rahman without this matter-at else anyone's good concept not-exist

"No one else has a good concept about this matter except Rahman"

কেউ : It is tagged as personal pronoun (PRP) when used as an indefinite pronoun.

Ex: কেউ/PRP আসেনি

anybody came-not-3P

"Nobody came"

But, it will be tagged as question pronoun when used in interrogative sentences.

Ex: কেউ/QPR এসেছিলো?

anybody came-3P?

"Did anybody come?"

কি : It is a question demonstrative (QDM) or question particle (QRP), but also acts as coordinating conjunction to join two verb phrases.

Ex: আমি এটা পাবো/VB কি/CC না/RB জানিনা

I this will-get whether not know-not-1P

"I don't know whether I will get it or not"

কিন্তু : This word belongs to coordinating conjunction (CC) but it is also widely used as a particle (RP).

Ex: আমি চলে যাচ্ছি কিন্তু/RP

I after-leaving going-1P PARTICLE

"I am leaving then"

ছেলেটা কিন্তু/RP কিছুই শুনতে পাচ্ছেনা

boy-the PARTICLE nothing to-hear can't-3P

"The boy can't hear anything at all"

কী : It is generally a question demonstrative (QDM), but sometimes its being used as an question adjective (QJJ) to describe or modify a situation.

Ex: কী/QJJ পরিস্থিতিতে এটা হলো ?

what-type situation-at this happened-3P ?

"At what type of situation did this happen?"

And sometimes, it can also be an adverb (RB) when intensifies a modifier, generally in a sentence that expresses stronger emotions.

Ex: কী/RB সুন্দর/JJ দৃশ্য!

what beautiful scenery!

"What a beautiful scenery!"

জন : This suffix is used after a quantifier number (QFNUM) to denote the plurality of human.

Ex: সাতজন/QFNUM মানুষ

seven-SUFF man
"Seven men"

But sometimes it's being used after the quantifier number (QFNUM) as a separated word. In this case tag it as an adjective (JJ), because in such cases it itself co-occurs after the quantifier to modify nouns that are human beings.

Ex: সাত/QFNUM জন/JJ মানুষ
    seven SUFFIX man
    "Seven men"

    ১০/QFNUM জন/JJ বালক
    10 SUFFIX boy
    "10 boys"

And it will be tagged as noun (NN) if used pronominally

Ex: আহত হয়েছে ৭ জন/NN
    wounded been-3P 7 people
    "7 peoples have been wounded"

তখন : This is generally a temporal noun (NNT) when simply denotes the time of an event.

Ex: তখন/NNT ঢাকার নাম ছিল জাহাঙ্গীরনগর
    at-that-time Dhaka's name was Jahangirnagar
    "At that time the name of Dhaka was Jahangirnagar"

But when attaches a subordinating clause, it is tagged as subordinating conjunction (CS).

Ex: সে আসলে তখন/CS ব্যবস্থা করা যাবে
    he if-come then arrangement doing wil-go-3P
    "If he comes, then arrangement could be make"

    দাওয়াত পেলাম, তখন/CS গেলাম
    invitation got, then went-1P
    "I have gone when got the invitation"

দরকার : See 'প্রয়োজন'

দেখাদেখি : Generally a common noun, denotes a process or event. But sometimes it also popup as postposition (ON), because in that case it has same distribution as other postpositions.

Ex: ওরা দেখাদেখি/NN করে লিখছে
    they copying done writing-3P
    "They are writing by copying each other's script"
    But,

তোমার দেখাদেখি/ON সেও গিয়েছে   cf. তোমার মতো/ON সেও গিয়েছে
your like he-too gone-3P   your like he-too gone-3P
"He has gone too after saw you going"   "Like you, he has gone too"

প্রয়োজন : It is a noun and will be tagged as noun (NN) as well.

Ex: এই কাজটি করার কোন/DM প্রয়োজন/NN নেই/VB
    this work-the doing's any need not-exist
    "There is no need to do this work"

আমার/PRP+SF$ প্রয়োজন/NN আছে বলে আমি যাব
my need being that's-why I will-go-1P
"I will go since I have a purpose"

Also note:
Ex: প্রয়োজনে/NN+SFON আমি যাব
need-at I will-go-1P
"I will go if needed"

And it can be used as a predicative noun to express necessity of something. Bangla doesn't require the linking verb 'হয়' (to be) in Subject-Verb-Complement type sentences. In such sentences, the verb 'to be' is inherent.

Ex: এখন শুধু কাজ করা প্রয়োজন/NN   cf. এখন শুধু কাজ করাটা (হচ্ছে) প্রয়োজন/NN
now only work doing need   now only work doing-the (being-3P) need
"Now it is needed to do work only"   "Now it is being needed to do work only"

চাকরিটা আমার/PRP+SF$ প্রয়োজন/NN
job-the my need
"I need the job"

Similarly, it can take a 'subject complement'—
Ex: এখন প্রয়োজন/NN শুধু কাজ করা   cf. এখন প্রয়োজন/NN (হচ্ছে) শুধু কাজ করা
now need only work doing   now need (being-3P) only work doing
"Now it is needed to do work only"   "Now it is being needed to do work only"

বলে : Although it's a verb, it can also used as a postposition (RP) when follows a nominal phrase.
Ex: আমি/PRP বলে/ON নয়, এই ঘটনা সবাই জানে
I only not, this incident all know-3P
"Not only me, everybody knows about this incident"

আপনাকে কী/QDM বলে/ON ডাকব?
you-to what by will-call-1P?
"How would I call you?"

ওটা পাখি/NN বলে/ON মনে হচ্ছে
that bird likely-to-be mind-at being-3P
"That seems to be a bird"

And it is a subordinating conjunction (CS) when follows a clause.
Ex: পাখিটা/NN উড়ছে/VB বলে/CS মনে হচ্ছে
bird-the flying likely-to-be mind-at being-3P
"The bird seems to be flying"

But it is a perfective verb (VBT) when takes a clause as its complement which is an 'utterance' of the speaker.
Ex: ব্যাপারটা নিয়ে এখনো আলোচনা করা হয়নি বলে/VBT তিনি জানান
issue-the about yet discussion doing not-been-3P after-telling he informed-3P

"He informed that no discussion was held yet about this issue"

Also note the following contrast that 'বলে' can be tagged as either VBT or CS, depends on the actual interpretation of the sentence according to context:

Ex: ব্যাপারটা নিয়ে এখনো আলোচনা করা হয়নি বলে/VBT তিনি চলে গেলেন
    issue-the about yet discussion doing not-been-3P after-telling he after-leaving gone-3P
    "He has left after telling that no discussion was held yet about this issue"
    But,
    ব্যাপারটা নিয়ে এখনো আলোচনা করা হয়নি বলে/CS তিনি চলে গেলেন
    issue-the about yet discussion doing not-been-3P that's-why he after-leaving gone-3P
    "He has left because no discussion was held yet about this issue"

And, it can be adverb (RB) also: একটু বসুন, সে এলো বলে/RB
                                a-little-while sit-2P, he coming-2P shortly
                                "Take a sit for a little while, he is coming shortly"

বটে : It is a verb in 3rd person present form to mean 'to exist in the present'
    Ex: সে কে বটে/VB ?
        he who is (idiomatic use) ?
        "Who is he?"

        তাই বটে/VB
        that is (idiomatic use)
        "That's true"

It is a particle when bears the stress at clausal final position
    Ex: কী বলছ বটে/RP ?
        what talking-2P PARTICLE ?
        "What are you talking about actually?"

        সে সাহসী বটে/RP !
        he brave PARTICLE !
        "He is no doubt a brave fellow!"

মধ্যে : It is generally tagged as postposition to mean 'between' or 'among'.
    Ex: দু দলের মধ্যে/ON খেলা চলছে
        two team's between game continuing-3P
        "A game is being continued between two teams"

        পাঁচটি বইয়ের মধ্যে/ON তিনটি ব্যাকরণ বই
        five-the book's among three grammar book
        "Three out of five books are grammar books"

But it should be locative noun (NNL) when refers a location or a place.
    Ex: প্যাকেটটার মধ্যে/NNL কিছু কাগজ আছে
        packet-the-'s inside-at some papers is
        "There are some papers in that packet"

63

তোমার মধ্যে/NNL অনেক সম্ভাবনা আছে

your's inside-at many potentiality is

"You have many potentialities"

যাতে : It is tagged as QDM+SFON when it is referred by a postpositional phrase such as 'এ জন্য' (for this), 'সে জন্য' (for that) etc.

Ex: সে যাতে/QDM+SFON আসতে পারে এ/DM জন্য/ON আমি দরজা খুলে দিয়েছি

he that-by come-to can-3P this reason I door after-opening gave-1P

"I have opened the door so that he could come"

It will be tagged as eternal joining conjunction (CET) when co-occur with un-separated words 'এজন্য' (for this), 'সেজন্য' (for that) etc. Because when a word is split down into two words separated by a space, it becomes a phrase and should be considered to have different syntactic category.

Ex: সে যাতে/CET আসতে পারে এজন্য/CET আমি দরজা খুলে দিয়েছি

he that-by to-come can-3P this reason I door after-opening gave-1P

"I have opened the door so that he could come"

যে : When 'যে' introduces relative clauses, it is a question pronoun (QPR).

Ex: মানুষটা যে/QPR আমাকে চিনে

man-the who me knows-3P

"The man who knows me"

When 'যে' acts as complementizer, it is a subordinating conjunction (CS).

Ex: এটা দাবি করা হয় যে/CS পরীদের ডানা আছে

it claim doing be that angels' wing is

"Its being claimed that angels' have wings"

Note one thing that, although it has been included into the category subordinating conjunction (CS), in case of complementary distribution it is an initial complementizer. It is not possible for the word 'যে' to have same distribution as other complementizers which are final complementizers [Bayer 1999].

Ex: ছেলেটা ওর বাবা আসবে বলে/CS শুনেছে

boy-the his father will-come-3P that heard-3P

"The boy has heard that his father will come"

Here, 'বলে' is a final complementizer takes the clause 'ওর বাবা আসবে' as a complement positioned to the left. It will be ungrammatical if we use the initial complementizer 'যে' here:

*ছেলেটা ওর বাবা আসবে যে শুনেছে[2] because an initial complementizer takes the complement to the right, not at left: ছেলেটা শুনেছে যে/CS ওর বাবা আসবে

boy-the heard-3P that his father will-come-3P

"The boy heard that his father will come"

And it can also be used as a particle.

Ex: লোকটা যে/RP চলে গেল

_____

[2] However the sentence itself is not completely ungrammatical. Here 'যে' can be tagged as particle (RP), not as a subordinating conjunction (CS).

man-the PARTICLE after-going gone-3P

"It is not acceptable that the man has gone away"

যেমন : When used as a relative adjective, it is tagged as QJJ.

Ex: আমি সেটিই কিনেছি যেমন/QJJ পোশাক আমার পছন্দ

I exactly-that-one bought-1P what-type dress my choice

"I bought exactly that dress which one I like"

And it is tagged as coordinating conjunction (CC) when perform as a joiner of two clauses—

Ex: প্রথাগত ব্যাকরণে বাক্যে প্রধাণত দুটি অংশ থাকে, যেমন/CC উদ্দেশ্য এবং বিধেয়

traditional grammar-at sentence-at mainly two part stay-3P, such-as subject and predicate

"In traditional grammar a sentence has mainly two parts, such as subject and predicate"

সময় : Generally a common noun. But when contains the occurring time of an event or action, it is a temporal noun (NNT).

Ex: সময়/NN করে/VBT এসো

time after-doing come-2P

"Come when you will be abled to managed time"

But,

পড়ার/VBM+SF$ সময়/NNT কথা বলবে না

study's time word will-tell-3P not

"Don't talk at the time of study"

ওই/DM সময়/NNT একজন লোক এসেছিলো

that time one-SUFF person came-3P

"At that time a man came"

সঙ্গে : Generally it is a postposition (ON).

Ex: লোকটি কি তোমার সঙ্গে/ON এসেছে ?

man-the QUESTION-PARTICLE your with come-3P ?

"Has the man come with you?"

এই সঙ্গে/ON শেষ হবে আজকের অনুষ্ঠান

this with end will-be today's program

"Today's program will be concluded with this"

But should be tagged as a noun with postpositional suffix (NN+SFON) when modified by adjectives or possessive modifiers which may be elliptical (not mentioned, but understood from the context).

Ex: আমরা একই/JJ সঙ্গে/NN+SFON যাব

we same company-at will-go-1P

"We all will go together"

অসৎ/JJ সঙ্গে/NN+SFON সর্বনাশ

bad company-at ruin

"Bad association leads to destruction"

(তোমার) সঙ্গে/NN+SFON কে কে ছিল?

(your) company-at who who were-3P?
"Who were with (you)?"

Also note that, the word 'সঙ্গে' inflected with postpositional suffixes (সঙ্গ+এ), can be followed by another postposition (ON).
Ex: ওকে সঙ্গে/NN+SFON করে/ON নিয়ে এসো
    him company-at after-doing after-bringing come-2P
    "Bring him with you"

সাথে : Same as 'সঙ্গে'

এ কারণে : It will be tagged as এ/CSC কারণে/CS only when joins a subordinate clause to main clause, otherwise it will be tagged as এ/DM কারণে/NN+SFON.
Ex: আমি ছিলাম না, এ/CSC কারণে/CS সে আসেনি
    I was-1P not, this reason-at he came-not-3P
    "I wasn't available, that's why he didn't come"
    But,
    আমি ছিলাম না। এ/DM কারণে/NN+SFON সে আসেনি
    I was-1P not. This reason-at he came-not
    "I wasn't available. That's why he didn't come"

মাঝে মাঝে : These repetitive words will be tagged as locative nouns (NNL) when denotes a place, or in between of something. In these cases, it is generally preceded by possessive nominal.
Ex: এই বইটার/NN+SF$ মাঝে/NNL মাঝে/NNL অনেক পৃষ্ঠা নেই
    this book-the-'s between-at between-at many page not-exist
    "There are many pages missing in between of this book"

And it is temporal nouns (NNT) when denotes time.
Ex: আমরা মাঝে/NNT মাঝে/NNT এ ব্যাপারটা নিয়ে ভাবছি
    we sometimes sometimes this fact-the about thinking-1P
    "Sometimes we are thinking about this fact"

সঙ্গে সঙ্গে : Tag both words as postpositions (ON) when follow possessive nominal to act as postposition.
Ex: বিস্ফোরণের/NN+SF$ সঙ্গে/ON সঙ্গে/ON আগুন ধরে গেল
    explosion's with with fire after-catching gone-3P
    "It has been caught by fire as soon as explosion occurred"

But they should be tagged as adverbs (RB) when act as 'adverbs of manner'.
Ex: আমি সঙ্গে/RB সঙ্গে/RB বইটি কিনে আনলাম
    I with with book-the after-buying brought-1P
    "I have bought the book instantly"

Remember not to tag them as postpositions when not preceded by possessive nouns. Because sometimes that possessive noun is elliptical (not expressed, but understood from context). In this case, they are still considered as adverbs.
Ex: সে (আমার) সঙ্গে/RB সঙ্গে/RB এসেছে
    he (my) with with has-come-3P
    "He has come along with (me)"

সাথে সাথে : Same as 'সঙ্গে সঙ্গে'

5. Conclusion

Going through this annotation scheme developed for Bangla may reveal that the scope of this POS tagging guideline can productively be extended with many more examples and corresponding linguistic analysis for disambiguation process. Ambiguities are descriptively incomplete phenomenon. However, since the ambiguity is the phenomenon of lack of information in the scenario of pure syntactic distributions of words, the primary goal of this paper is to represent an annotator a basic shape about how to distinguish when an uncertainty arises between two or more alternative descriptions. And in addition, we need to mention that the tagset itself is under a development phase. In '**Fineness vs. Coarseness**' issue the tags are always chosen to be kept 'coarse' [Bharati et al. 2006; SPSAL 2007], because as number of tags are less as higher the accuracy of manual tagging. Since the tagset presented here is much finer in distinction, it can be simply perceptible that certain modifications to this design are required. Keeping in mind the enhancement of both coarseness and linguistic aspect, it has been decided that the tagset should be modified further according to following three issues at the current point of this stage:

1. Locative Nouns (NNL) and Temporal Nouns (NNT) have similar syntactic distribution. Thus they can be subsumed under a single category which is Spatial Noun (NST) as suggested in AnnCorra [Bharati et al. 2006].

2. Intensifiers such as খুব, বেশী, অনেক, have been categorized as adverbs. But it has been observed that their syntactic distribution is not as same as adverbs, because intensifiers are always tend to appear before adjectives or adverbs that they intensify, where adverbs are allowed to appear anywhere in a sentence. Thus, intensifiers should have different categorization (INTF) [Bharati et al. 2006; SPSAL 2007].

3. When the same word is written twice for various purposes such as indicating emphasis, deriving a category from another category. For example, আস্তে আস্তে (slowly slowly), ছোট ছোট (small small), লাল লাল (red red) have been tagged using the same tag for both words such as আস্তে/RB আস্তে/RB. But reduplication is a highly productive process and it has been proposed in AnnCorra [Bharati et al. 2006] to include a new tag **RDP** for annotating reduplicatives. In our current assessment, it has been resoluted to adapt this technique where the first word will be tagged by its respective syntactic category and the second word will be tagged as RDP, as for example আস্তে/RB আস্তে/RDP, to indicate that it is a case of reduplication distinguishing it from a normal sequence.

Finally, it should be stated that the descriptions and examples given here can be made more explicit, and can be validated by scrutinizing deeper linguistic phenomenon. It shouldn't be assumed that this document has restricted the scope of supplementary debate, by asserting all of the investigations pointed out here to be sufficient. Rather, it is left open for further discussions and suggestions to promote monographic studies of different syntactic phenomena of Bangla without being bound to some traditional and specific syntactic theories.

References:

[1] Azad, H. Bakyatattva (Syntax), Second Edition. Published by the University of Dhaka, Dhaka, Bangladesh, December 1994.

[2] Bayer, J. Two Grammars in One: Sentential Complements and Complementizers in Bengali and other South-Asian Languages. A paper presented at the International Symposium on 'Language Contact, Convergence and Typology in South Asia', ILCAA, Tokyo, December 6-9, 1999.

[3] Bharati, A., Sharma, D. M., Bai, L. and Sangal, R. AnnCorra: Annotation Corpora for POS and Chunk Annotation for Indian Languages. Language Technologies Research Centre, IIIT, Hyderabad, December 15, 2006.
http://ltrc.iiit.ac.in/MachineTrans/publications/technicalReports/tr031/posguidelines.pdf.

[4] Biswas, P., Dandapat, S., Bali, K., Choudhury, M. A Corpus-based Study of করে *(kare)* in Bangla: Theoretical and Computational Perspectives. In proceedings of 6th international conference on natural language processing (ICON-08), December 20-22, 2008.

[5] Dandapat, S., Sarkar, S., Basu, A. A Hybrid Model for Part-of-Speech Tagging and its Application to Bengali. Transactions on engineering, computing and technology VI ISSN 1305-5313. December 2004.

[6] Dandapat, S., Sarkar, S., Basu, A. Automatic Part-of-Speech Tagging for Bengali: An Approach for Morphologically Rich Languages in a Poor Resource Scenario. Proceedings of the ACL 2007 Demo and Poster Sessions, June 2007, pages 221-224.

[7] Ekbal, A., Haque, R., Bandyopadhyay, S. Maximum Entropy Based Bengali Part of Speech Tagging. Advances in natural language processing and applications research in computing science 33, 2008, pp. 67-78.

[8] Hasan, F. M., UzZaman, N., Khan, M. Comparison of Unigram, Bigram, HMM and Brill's POS Tagging Approaches for some South Asian Languages. In proceedings of conference on language and technology (CLT07), August 7-11, 2007.

[9] Paul, S. An HPSG Account of Bangla Compound Verbs with LKB Implementation. A Ph.D. dissertation submitted to the University of Hyderabad, April, 2004.

[10] Saha, G. K., Saha, A. B., Debnath, S. Computer Assisted Bangla Words POS Tagging. In proceedings of symposium on machine translation NLP & TSS (ISTRANS), 2004.

[11] Santorini, B. Part-of-speech Tagging Guidelines for the Penn Treebank Project (3rd Revision, 2nd printing), June 1990.

[12] Séaghdha, D. Ó. Annotation Guidelines for Compound Noun Semantics. May 29, 2008.
http://www.cl.cam.ac.uk/~do242/guidelines.pdf.

[13] A Part of Speech Tagger for Indian Languages (POS Tagger). Workshop on Shallow Parsing in South Asian Languages (SPSAL), Twentieth International Joint Conference on Artificial Intelligence, 2007.