

## Appendix 2.4

### Linguistic Standards for Indian Language (IL) Corpora Annotation POS Categories and Labels (September 29, 2010)

#### 2. Major Language Class

Bangla	: [BNG]	: E.g. lok <sub>[BNG]</sub>
English	: [ENG]	: E.g. myan <sub>[ENG]</sub>
Foreign	: [FRN]	: E.g. Admi <sub>[FRN]</sub>

#### 1. Major Lexical Classes in Bengali

Sl. No.	POS		Label	Example	Comments
1	Noun	বিশেষ্য	\NN\	balak, ghar, shahar, katha, manus	
2	Pronoun	সর্বনাম	\PR\	ami, tumi, se, tara, tui	
3	Demonstrative	নির্দেশক	\DM\	ye, ei, oi, tai,	
4	Verb	ক্রিয়া	\V\	Karchi, kratam, gelo, yabe	
5	Adjective	বিশেষণ	\JJ\	Bhala, manda, sundar, sada	
6	Adverb	ক্রিয়াবিশেষণ	\RB\	Kadachit, babd, karane, sahaje	
7	Postposition	অনুসর্গ	\PSP\	Pare, kache, age, niche	
8	Conjunction	সংযোজক	\CC\	Kintu, athaba, barang	
9	Particle	অব্যয়	\RP\	i, -o, -to na, ne, ni	
10	Quantifier	পরিমাণবাচক	\QT\	ek, dui, tin, pratham, payla, dvitiya	
11	Punctuation	যতিচিহ্ন	\PNC\	., : ;	
12	Residuals	অবশিষ্ট পদ	\RD\		

#### POS for Bengali

Sl. No	Category			Label	Annotation Convention**	Examples	Remarks
	Top level	Subtype (level 1)	Subtype (level 2)				
<b>1</b>	<b>Noun</b> (বিশেষ্য)			<b>NN</b>	<b>NN</b>		
1.1		Proper (ব্যক্তিবাচক)		\PPR\	\NN_PPR\	Mohan, ravi,	

1.2		Common (জাতিবাচক)		\COM\	\NN_COM\	kalam, ghar,	
1.3		Locative (স্থানবাচক)		\LOC\	\NN_LOC\	upare, niche,	
1.4		Collective (বাচক)		\COL\	\NN_COL\	sangha	
1.5		Abstract (বাচক)		\ABS\	\NN_ABS\	bhalobasa	
1.6		Material (বস্তুবাচক)		\MAT\	\NN_MAT\	Dhuti, bai,	
<b>2</b>	<b>Pronoun (সর্বনাম)</b>			<b>PR</b>	<b>PR</b>		
2.1		Personal (ব্যক্তিবাচক)		PRP	\PR__PRP\	se, tumi,	
2.2		Reflexive (আত্মবাচক)		PRF	\PR__PRF\	nijer	
2.3		Relative (সম্বন্ধবাচক)		PRL	\PR__PRL\	ye, yAra	
2.4		Reciprocal (ব্যতিহার)		PRC	\PR__PRC\	paraspara,	
2.5		Wh-word (প্রশ্নবাচক)		PRQ	\PR__PRQ\	ke, kara,	
<b>3</b>	<b>Demonstrative (নির্ণয়সূচক নির্দেশক)</b>			<b>DM</b>	<b>DM</b>	ei, oi, sei	
3.1		Deictic (প্রত্যক্ষ নির্দেশক)		DMD	\DM__DMD\	sei, oi, o, se	
3.2		Relative (সম্বন্ধবাচক)		DMR	\DM__DMR\	ye, yei	
3.3		Wh-word (প্রশ্নবাচক)		DMQ	\DM__DMQ\	kona,	
<b>4</b>	<b>Verb (ক্রিয়া)</b>			<b>V</b>	<b>V</b>		
4.1		Main (মুখ্যক্রিয়া)		VM	\V__VM\		
4.1.1			Finite (সমাপিকা)	VF	\V__VM__VF\	karalam	
4.1.2			Nonfinite (অসমাপিকা)	VNF	\V__VM__VNF\	kare, khey,	
4.1.3			Infinitive (অপূর্ণক্রিয়া)	VINF	\V__VM__VINF\	karate,	
4.1.4			Gerund (প্রযোজক ক্রিয়া)	VNG	\V__VM__VNG\	yoya, asa	
4.2		Auxiliary (গৌণ ক্রিয়া)		VAUX	\V__VAUX\	chila	
<b>5</b>	<b>Adjective (বিশেষণ)</b>			<b>JJ</b>	<b>JJ</b>	sundar	
<b>6</b>	<b>Adverb (ক্রিয়া বিশেষণ)</b>			<b>RB</b>	<b>RB</b>	aste, hathat	
<b>7</b>	<b>Postposition (অনুসঙ্গ)</b>			<b>PSP</b>	<b>PSP</b>	theke, diye	
<b>8</b>	<b>Conjunction (সংযোজক)</b>			<b>CC</b>	<b>CC</b>		
8.1		Coordinator (সমন্বয়ক)		CCD	\CC__CCD\	Ar, ebang,	
8.2		Subordinator (শর্তসংযোজক)		CCS	\CC__CCS\	kintu, noile,	
8.3		Quotative (উক্তিবাচক)		UT	\CC__CCS__UT\		NA

<b>9</b>	<b>Particles (অব্যয়)</b>			<b>RP</b>	<b>RP</b>		
9.1		Default (সাধারণ অব্যয়)		RPD	\RP__RPD\	to, ye,	
9.2		Classifier (বর্ণবাচক)		CL	\RP__CLF\	jan, khana	
9.3		Interjection (বিশয়বোধক)		INJ	\RP__INJ\	Are, ei,	
9.4		Intensifier (তীব্রতাবোধক)		INTF	\RP__INTF\	khuba,	
9.5		Negation (নঙ্করক)		NEG	\RP__NEG\	na, ne, ni	
9.6		Emphatic ( )		EMPH	\RP__EMPH\	-i, -o, -to	
<b>10</b>	<b>Quantifiers (পরিমাণবাচক)</b>			<b>QT</b>	<b>QT</b>		
10.1		General (সাধারণ)		QTF	\QT__QTF\	kichu, alpa,	
10.2		Cardinal (সংখ্যাবাচক)		QTC	\QT__QTC\	eka, dui, tina	
10.3		Ordinal (ক্রমবাচক)		QTO	\QT__QTO\	pratham, payla	
<b>11</b>	<b>Punctuation(যতিচিহ্ন)</b>			<b>PUNC</b>	<b>PUNC</b>	., : ;	
<b>12</b>	<b>Residuals (অবশিষ্ট শব্দ)</b>			<b>RD</b>	<b>RD</b>		
12.1		Foreign word (বিদেশী শব্দ)		FW	\RD__RDFW\	Word written in foreign script	
12.2		Symbol (প্রতীক)		SYM	\RD__SYM\	\$, &, *, (, )	
12.3		Echo-words ( অনুকার শব্দ)		ECH	\RD__ECH\	jala Tala,	
12.4		Unknown ( অজ্ঞাত শব্দ)		UNK	\RD__UNK\		

**\*\*** The annotation is to be done using the lowest level tag of the type hierarchy. Once the lower level tag is selected, the higher level tags should be stored automatically.

This POS tagset for Bangla Document was:

Prepared by: Kalika Bali, Microsoft Research India

Date: Aug, 2010

**Revised by: Niladri Sekhar Dash**

**Date: 29<sup>th</sup> September 2010.**