A Fine-Grained Tagset for Bengali Language

Article · January 2014

2 authors, including:

Mohammad Abdullah Al Mumin
Shahjalal University of Science and Technology
13 PUBLICATIONS   **54** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   Tagset View project

Project   Corpus Linguistics View project

A Fine-Grained Tagset for Bengali Language

Arun Krishna Paul[1] and Md. Abdullah Al Mumin[2]

[1, 2]*Department of Computer Science and Engineering, Shahjalal University of Science and Technology, Bangladesh*
E-mail: *aruncse2007007@gmail.com[1] and mumin-cse@sust.edu[2]*

Abstract

The lexical tags, called Tagsets, play a significant role providing the large amount of information about a word and its neighbors, telling us something about how the words are pronounced, being used in stemming for information retrieval[1]. So, a standard tagset is necessary for working with a language in any computational linguistic field. Two major kinds of tagsets for a language are fine-grained tagset, which uses a large number of tags, and coarse-grained tagset, which uses a small number of tags. The goal of this paper is to propose a fine-grained tagset, containing a total of 1070 tags, for tagging Bengali texts. Being a completely inflectional language, Bengali requires more tags for tagging a text than English or some other non-inflectional languages. A good example of this is the proper noun 'হাসান' (masculine gender)[*TABLE 1*], which takes 39 forms in Bengali.

Key Words: Bengali Tagset, Inflectional Language, Fine-Grained Tagset, Coarse-Grained Tagset.

1. Introduction

Parts-of-Speech (POS) tagging – the task of assigning appropriate POS tag to every word in a given text – is an important process used as a building block for several Natural Language Processing (NLP) tasks. A POS tagsetusually defines the list of morpho-syntactic categories that are applicable at the word-level to a specific language, though in some cases they may also include pure morphological or syntactic information [2]. Computer scientists and linguistics are working overnight for finding morphological or syntactic information from each word.

To understand human speech or to extract information or to classify data, firstly, a system has to understand what is the function of each word used in the speech or in the document. Based on the function of each word, we annotate them by some tags called Parts-of-Speech. So, identifying the appropriate tag for each word used in the text or speech plays a significant role in language processing.

Between the two types of tagsets, fine-grained tagset is the most appropriate way for finding the nuances of the words. For instance, consider the two examples: আমিকাজটিকরি। আমিকাজটিকরাই। In the above two examples, the verbs 'করি' and 'করাই' are expressing completely different meanings. In the first sentence, the subject is doing the task, but in the second sentence, the subject is making someone else do the work. But both of these words have come from the same root 'কর'. This root word 'কর' has 160 tags according to our tagset and each tag is containing significant amount of linguistics information. But the verb 'do' only has 5 forms (like do, did, done, does, doing) in English. If we would use a small number of tags (coarse-grained tagset) , we would have to compromise finding the nuances of the words. So, for working with Bengali, a fine-grained tagset is a must as it is a completely inflectional language.

Significant amount of improvement is shown in some languages like English, German, French, etc., we are yet to see any significant advancement in Bengali. However, Bengali is the fifth mostly spoken language in the world. So, research on the Bengali word classes is the demand of the time.Starting working on Bengali language processing is very hard, as there is no useful data source available for Bengali research. Coping up with all types of difficulties, we have proposed a fine-grained Parts-of-Speech Tagset for Bengalilanguage.

Traditional Bengali grammars generally provide five basic Parts-of-Speech [like – বিশেষ্য (Noun), বিশেষণ (Adjective), সর্বনাম (Pronoun), অব্যয় (Combination of Conjunction, Preposition, Indeclineables, etc.) and ক্রিয়া (verb)] different from English grammars which provide eight basic parts-of-Speech (like: Noun, Pronoun, Adjective, Verb, Adverb, Preposition, Conjunction and Interjection). However, most linguists want to capture finer grammatical

distinctions, leading to larger tag sets. For easing computational process using our tagset as well as for universality of our tagset, we have used English POS notation.

In the next section, we will briefly discuss the nature of Bengali language and the necessity of fine-grained tagset leading to the discussion in section 3 on previous research on tagsets. Section 4 discusses the design methodology for this tagset followed by a description of the design in section 5. The next section shows the tags for proper noun masculine gender according to our tagset. In section 6, we discuss the availability of the tagset and in section 7, we show an annotated text with our tagset followed by conclusion in the last section.

2. Nature of Bengali Language and the Necessity of Fine-Grained Tagset

At the beginning, Bengali language started with a few number of vocabularies, but with the passage of time, its vocabularies have increased significantly from a variety of languages. Bengali words have come from five broad categories, according to the Bengali grammarians. The categories are: words directly coming from Sanskrit; words coming from Aryan language; words (slightly changed) coming from Sanskrit; words coming from indigenous languages and words coming from foreign languages [3]. Besides these, there are some mixed words in Bengali like (পকেট-মার= English + Bengali), (ডাক্তার-খানা= English + Persian), etc.

But the maximum number of words in Bengali language had directly come from Sanskrit language. As Sanskrit is a completely inflectional language, a significant dominance of inflectional features is also observed in Bengali. But, there also exists some agglutinative features in Bengali language. Bengali language follows the SOV (Subject + Object + Verb) structure.

Bengali, being a completely inflectional language, requires a lot of linguistic knowledge to extract significant information form any text. Consider the sentence: I go to school. In Bengali, আমিবিদ্যালয়েযাই।Ifsomeone writes আমিবিদ্যালয়যযাই।Then, it does not make one's statement completely clear to us. So, the inflectional part '্যে'plays an important role in the understanding of the writer's view. The word 'বিদ্যালয়'takes 34 forms and each form signifies the word differently. So, for finding sufficient information from Bengali language, we need a significant number of tags.

3. Previous Research on Tagsets

The early efforts in POS tagsets were UPenn, Brown and C5 mainly for English using flat structure [2]. They were coarse-grained tagsets – Penn Treebank tagset used 45-tags, C5 tagset used 61 tags. The CLAWS2 introduced the hierarchical structure in parts of speech tagging for decomposing tags [2].

Though several tagsets have been developed for Bengali language like Bangla POS Tag Set (91 tags) by Goutam Kumar Saha, CDAC-Kolkata[4], 53 morpho-syntactic tags by BRAC University [5], and Bangla Tagset by ILPOSTS, Microsoft Research Private Limited [6]. Among them, first two are coarse-grained tagsets, strictly following Penn Treebank standard. The last one is hierarchical tagset following CLAWS2 standard.

53 morpho-syntactic tags by Brac University, Bangla POS Tagset (91 tags) by Goutam Kumar Saha, CDAC-Kolkata have compromised with the inflectional feature of this language. These tagsets are completely influenced by Penn TreeBankTagset designed for English.

The hierarchical Tagset produced by Microsoft Research India, Bangalore has considered the inflectional feature of Bengali language, but for its conformity with IL-POSTS frame, it has compromised some of the significant features of Bengali language – repeated words (আমি'স্বরস্বর'বোধকরছি), mixed verbs (দর্শনকরলাম), compound proper noun ('হাসানটাসানকে'দিয়েএকাজহবেনা).

So, we have attempted to design a fine grained Bengali tagsetof the Bengali language.

4. The Design Methodology for this Tagset

The design methodology of this tagset is based on the linguistics knowledge of Bengali language. Any Bengali word can be different based on any of the following characteristics like – number, gender, complementary foreign words, repeated words, compound words, agglutinative words, objective case, possessive case, locative case, definiteness, finiteness, emphatic feature, person, tense, mood, negativity, honorificity, insult, persuasion. Although, some of the linguistics features like number, person, tense, case-marker, aspect, mood, finiteness, distributive, definiteness, emphatic, negative, honorificity, numeral, realis are used in Annotation Guidelines for Bangla[6], Microsoft Research Labs Indian Pvt. Ltd., Bangalore, they have skipped the following significant features like gender, complementary foreign words (now used as like as bengali word – মেজরসমস্যা), repeated words expressing

another parts of speech (তারকবিকবিভাবআছে – here the words), agglutinative words (আবালবৃদ্ধবণিতা), persuasion (আমিতাকেপড়াচ্ছি). Below, we are discussing our word classifications for this fine-grained tagset.

4.1 Number
Bengali nouns fall in two number categories. They are singular number and plural number such as - 'হাসান'and 'হাসানেরা'. The same word 'হাসান'is inflected with some inflections to make the word 'হাসানেরা'.

4.2 Gender
Gender plays an important role in information extraction. Consider the sentence: বাঘতারবাচ্চাকেদুধখাওয়াচ্ছে।This sentence is not semantically correct, as a masculine gender cannot do this task. There are four types of genders in Bengali. They are masculine, feminine, common and neuter. We have classified words based on gender, so that we can easily find gender specific information from the tag.

4.3 Complementary Foreign Words
We have tagged foreign words based on the syntactic roles the word plays in the sentence. For example: সেঅনেকমেজরসমস্যাসৃষ্টিকরেছে।Here, the word 'মেজর' is a foreign word, but here it is used as a circumstantial adjective. So, we have tagged it as a circumstantial adjective.

4.4 Repeated Words
When a single word is used next to next, its syntactic category may change. For example, 'তারকবিকবিভাবআরগেলনা'. Here, the words 'কবিকবি' are no longer common noun, they are treated as adjective in our tagset.

4.5 Compound words
Compound words are treated as Common Noun, Pronoun etc., based on the syntax. For example: রহিম-করিমভালছেলে।Here, the compound word 'রহিম-করিম' is tagged as compound form of proper noun. On the other hand, মাবাবাসবাইচলেগেছে।Here, 'মা-বাবা' is treated as compound form of common noun.

4.6 Agglutinative words
Agglutinative words also take different tags based on syntax.

4.7 Objective Case
Bengali objective cases take different forms unlike English. Consider the two sentences: আমিকরিমকেএকটিবইদেই।I give Karim a book. In English the object does not change its form, but in Bengali, it changes, 'করিম'becomes 'করিমকে' in Bengali objective case.

4.8 Possessive Case
Possessive cases provide significant information about word such as – করিমেরঘড়িকাল।Here, 'ঘড়ি'is the subject of the sentence. 'করিমের'expresses who is the owner of the 'ঘড়ি'।We have signified possessive cases as different tags.

4.9 Locative Case
Locative cases perform like English adverbial in Bengali such as – গাছেপাখিবসেছে।We know 'গাছ'is a common noun in bengali. But the inflection 'ে'added with 'গাছ'has given it locative property. We have tagged it differently than as a usual common noun.

4.10 Definiteness
Bengali nouns may end with a definite article like – আমটিমিষ্টি।Here, the base noun is common noun ('আম')and 'টি'is used here to specify the mango.

4.11 Finiteness
Finiteness is used with verbs. Verbs may be of two types – finite and non-finite. Consider the two examples: সেপড়েখেলতেযাবেand সেপড়ে।The first 'পড়ে'is a non-finite verb and the second 'পড়ে'is a finite verb. We tag verbs based on the syntax.

4.12 Emphatic Feature
Emphatic features may be expressed by inflection in Bengali such as – Hasan, too, goes to school. In Bengali, it is expressed as 'হাসানওস্কুলেযায়'।

4.13 Person
English verbs do not change with person, except third person singular number. But, Bengali verbs change form with the change of person form. For example, I go and you go. আমিযাইand তুমিযাও। Here, as the verbs have changed its forms, we have treated persons separately in Bengali for their linguistics importance.

4.14 Tense
Bengali verbs are inflectionally sensitive with the change of tense. For instance, the verb 'do' have five unique forms in English – do, does, did, doing, done. But, in Bengali, it can take more than 200 forms based on tenses, person, honorificity, etc. We have to separately tag them for more fine grained information.

4.15 Mood
Mood expresses speakers' attitudes towards the listener. For example, 'দয়াকরেভেতরেআসুন'।Here, the speaker is requesting someone to come in. We have taken into consideration of different types of moods used while expressing views and tagged them separately.

4.16 Negativity
Negative forms may be used in two ways in Bengali. They may be used with a verb word like 'যায়না'or they may be used as a separate word like 'যায়না'।When, they are used with the verb word, we tag them as verb and add 'N' at the end of the verb as negative tag. When it is used separately, we tag it as a Negative indeclinable.

4.17 Honorificity
This information is used in person. To indicate a respected person, we use 'তিনি' instead of 'সে'in Bengali. We also change in verb forms based on honorificity.

4.18 Insultory
It is treated same as honorificity in Bengali.

4.19 Persuasion
Like imperative, we can add persuasive sense in Bengali verbs to make someone do a work for me. For example: সেবাচ্চাটাকেখেলাচ্ছে।Here, 'খেলাচ্ছে'verb is the persuasive form. It is tagged differently than the verb 'খেলে'।

4.20 Some Peculiarities
There are some peculiar forms in Bengali language. For example, two words can act as a single part of speech in Bengali – 'উদযাপনকরা'or 'celebrate' in English. In Bengali, they are called mixed verbs. Mixed verbs in Bengali can be formed in three ways like verbal noun + finite verb = mixed verb, verbal adjective + finite verb = mixed verb and indeclinable + finite verb = mixed verb. We treat them separately as verbal noun, verbal adjective and indeclinable and if any finite verb is found following them, we treat these two words together as a mixed verb.

Above mentioned, more than one feature can be present in a single Bengali word, then we have tagged it separately. 'হাসানকেও'is tagged as Proper Noun Objective Emphatic Masculine (NPOEM) according to our tagset.

5. A Description of our Tagset for Proper Noun Masculine Gender

Our tagset is based on morpho-syntactic information of Bengali language. Our full tagset contains 1070 tags. Our full tagset can be found in *http://www.sust.edu/cse/nlp/finegrainedtagset.*There is asample of complete tagging of Bengali proper noun (masculine gender) based on our tagset below:

TABLE 1: Tags for Proper Noun Masculine Gender

Number	Name	Masculine (Tag)
1	Proper Noun Singular	NPM (হাসান)
2	Proper Noun Plural	NPSM (হাসানেরা)
3	Proper Noun Singular Possessive	NP$M (হাসানের)
4	Proper Noun Plural Possessive	NPS$M (হাসানদের)
5	Proper Noun Singular Objective	NPOM (হাসানকে)
6	Pronoun Noun Plural Possessive Objective	NPS$OM (হাসানদেরকে)
7	Proper PSoun Singular State	NPSTM (হাসানে)
8	Proper Noun Singular Definitive	NPDM (হাসানটা)
9	Proper Noun Singular Possessive Definitive	NP$DM (হাসানেরটা)
10	Proper Noun Plural Possessive Definitive	NPS$DM (হাসানদেরটা)
11	Proper Noun Singular Definitive Objective	NPDOM (হাসানটাকে)
12	Proper Noun Plural Possessive Definitive Objective	NPS$DOM (হাসানদেরটাকে)
13	Proper Noun Singular Definitive Possessive	NPD$M (হাসানটার)
14	Proper Noun Plural Possessive Definitive Possessive	NPS$DSM (হাসানদেরটার)
15	Proper Noun Singular Definitive Emphatic	NPDEM (হাসানটাও)
16	Proper Noun Singular Possessive Definitive Emphatic	NP$DEM (হাসানেরটাও)
17	Proper Noun Plural Possessive Definitive	NPS$DEM (হাসানদেরটাও)
18	Proper Noun Singular Definitive Objective Emphatic	NPDOEM (হাসানটাকেও)
19	Proper Noun Plural Possessive Definitive Objective Emphatic	NPS$DOEM (হাসানদেরটাকেও)
20	Proper Noun Singular Possessive Definitive Objective Emphatic	NP$DOEM (হাসানেরটাকেও)
21	Proper Noun Singular Emphatic	NPEM (হাসানও)
22	Proper Noun Plural Emphatic	NPSEM (হাসানেরাও)
23	Proper Noun Singular Possessive Emphatic	NP$EM (হাসানেরও)
24	Proper Noun Plural Possessive Emphatic	NPS$EM (হাসানদেরও)
25	Proper Noun Singular Objective Emphatic	NPOEM (হাসানকেও)
26	Proper Noun Plural Possessive Objective Emphatic	NPS$OEM (হাসানদেরকেও)
27	Proper Noun Singular State Emphatic	NPSTEM (হাসানেও)
28	Proper Noun Singular Definitive Possessive Emphatic	NPD$EM (হাসানটারও)
29	Proper Noun Singular Possessive Definitive Possessive Emphatic	NPDEM (হাসানেরটারও)
30	Proper Noun Plural Possessive Definitive Possessive Emphatic	NPSDEM (হাসানদেরটারও)
31	Proper Noun Singular Possessive Definitive State	NP$DSTM (হাসানেরটায়)
32	Proper Noun Plural Possessive Definitive State	NPS$DSTM (হাসানদেরটায়)
33	Proper Noun Singular Additive Inflection	NPADM (হাসানসহ)
34	Proper Noun Singular Objective Additive Inflection	NPOADM (হাসানকেসহ)
35	Proper Noun Plural Possessive Objective Additive Inflection	NPS$OADM (হাসানদেরকেসহ)
36	Proper Noun Singular Definitive Objective Additive Inflection	NPDOADM (হাসানটাকেসহ)
37	Proper Noun Plural Possessive Definitive Objective Additive Inflection	NPS$DOADM (হাসানদেরটাকেসহ)
38	Compound Proper Noun	NPPM (হাসান-টাসান)
39	Proper Noun Singular Possessive Definitive Possessive	NPDM (হাসানেরটার)

The calculation of our total tagsets is shown below:

TABLE 2: The Classification of the Total 1070 Tagset

Tag Descriptions	Number of Tags
Proper Noun Masculine Gender	39
Proper Noun Feminine Gender	39
Proper Noun Neuter Gender	25
Common Noun Masculine Gender	37
Common Noun Feminine Gender	37
Common Noun Common Gender	40
Common Noun Neuter Gender	31
Compound Common Noun Neuter Gender	21
Agglutinative Common Noun Neuter Gender	22
Foreign Common Noun Neuter Gender	28
Common Noun from Other word-classes Neuter Gender	32
Material Noun Neuter Gender	37
Foreign Material Noun Neuter Gender	23
Compound Material Noun Neuter Gender	16
Collective Noun Neuter Gender	25
Compound Collective Noun Neuter Gender	13
Foreign Collective Noun Neuter Gender	19
Verbal Noun Neuter Gender	21
Compound Verbal Noun Neuter Gender	11
Abstract Noun Neuter Gender	27
Compound Abstract Noun Neuter Gender	19
Adjective (Bengali words)	18
Adjective (Foreign words)	18
Emphatic Adjective (Bengali words)	15
Emphatic Adjective (Foreign words)	15
Repetitive Adjective (Bengali words)	15
Repetitive Adjective (Foreign words)	15
1st Person Singular Pronoun	13
1st Person Plural Pronoun	12
2nd Person Plural Pronoun	13
2nd Person Singular Pronoun	13
3rd Person Singular Pronoun	13
3rd Person Plural Pronoun	13
2nd Person Singular Pronoun (Insultory)	13
2nd Person Plural Pronoun (Insultory)	13
3rd Person Singular Pronoun (Honorificity)	13
3rd Person Plural Pronoun (Honorificity)	13
Reflexive Pronoun Singular	12
Reflexive Pronoun Plural	12
Collective Pronoun Singular	12
Question Pronoun Singular	11
Question Pronoun Plural	11
Indefinite Pronoun Singular	10
Additive Pronoun Singular	11
Additive Pronoun Plural	11
Reciprocal Pronoun Singular	14
Reciprocal Pronoun Plural	14
Indeclineables	12
First Person Verb	32
Second Person Verb	32
Third Person Verb	32
Second Person Insultory Verb	32
Third Person Honorificity Verb	32
Verb	6
Total	**1070**

6. Availability

The full tagset is available free of charge for educational and research purposes, however, the license allows making short citations of this paper. the tagset will be distributed through the Computer Science and Engineering (CSE) department of Shahjalal University of Science and Technology (SUST). [1]

7. Sample of an Annotated Document with this Tagset

Below, there are some examples of tagging a corpus using our tagset.

কীর্তনখোলার/NP$F ভাঙনে/NVST বিলীন/SA বেলতলা/NPN ফেরিঘাট/NCPN ঃ/PUNC ণটি/NAC ফেরিঘাট/NCPN নিয়ে/PI বিপাকে/NCOST সড়ক/NCN বিভাগ/NCN

আবারও/SNA কীর্তনখোলার/NP$P রাক্ষুসী/NCF খাবার/NCO$ কবলে/NCOST ফেরীঘাট/NCPN ।/PUNC

নদীভাঙনের/NCP$ তীব্রতায়/NCOST ৭/NA দিন/NCN আগে/PI বিলীন/SA হয়েছে/VBFTPPP নগরীর/NC$N বেলতলা/NPN ফেরিঘাট/NCPN ।/PUNC

ফলে/SNA বন্ধ/NCO রয়েছে/VBTPPP এখানকার/NCO$ ফেরি/NPN চলাচল/NV ।/PUNC

এর/AP আগে/PI নদী/NCN ভাঙনে/NVST বন্ধ/NV ছিল/VBTPPI কীর্তনখোলা/VPF চরকাউয়া/NPN ফেরিটি/NCD ।/PUNC

জোড়াতালি/NCP দিয়ে/PI সেটি/APD চালু/NV করা/VAUX হলেও/VBTPPASTE বন্ধ/VB হলা/VBTPPAI নতুন/IA করে/PI বেলতলা/NPN ফেরি/NCN ।/PUNC

8. Conclusion

As natural language processing is a new field and a small amount of work is done on Bengali language, we are yet to get a standard tapset for Bengali language. For a lack of standard common tagset, the research in Bengali language is not up to our mark. To provide a common platform for every research group, we have proposed a tagset that has analyzed Bengali language from the morpho-syntactic perspective. As Bengali is a highly inflectional language, we have to consider inflectional feature of every word class of this language. For this reason, we have developed a tagset of 1070 tags. We have presented a fine-grained tagset for Bengali language. Each tag of this tagset bears significant amount of information about a Bengali word and its syntactic structure**.**

We hope that the Tagset will initiate all the researches for NLP applications for Bengali that was notdone in common criteria for a lack of a standard tagset in Bengali language.

References

[1]   Jurafsky, D. and Martin, J. (2008). *Speech and language processing.* Prentice-Hall.

[2]   Sankaran, B., Bali, K., Choudhury, M., Bhattacharya, T., Bhattacharyya, P., Jha, G. N., Rajendran, S., Saravanan, K., Sobha, L. and Subbarao, K., V., (2008), *A Common Parts-of-Speech Tagset Framework for Indian Languages.*, in 'LREC' , European Language Resources Association, .

[3]   Chowdhury, M. (2009), বাংলাভাষাব্যাকরণ

[4]   Saha, G.K., *Bangla POS Tag Set*, CDAC-Kolkata.

[5]   Mahmud, A. & Khan, M., *Research Report on Bangla Tagset*, Center for Research on Bangla Language Processing, BRAC University, Dhaka, Bangladesh.

[6]   *Annotation Guidelines for Bangla*. Microsoft Research Labs India Pvt. Ltd., Bangalore.

[7]   Ali, M., *An Unsupervised Parts-of-Speech Tagger for the Bangla language,*Department of Computer Science, University of British Columbia. 2010.

[8]   Azad, H., *Bakyatatta*, The University of Dhaka, 1994.

[9]   Dandapat, S., Sarkar, S., and Basu, A., *Automatic Part-of-Speech Tagging for Bengali: An Approach for Morphologically Rich Languages in a Poor Resource Scenario*, Proceedings of the Association for Computational Linguistic, pp 221-224, 2007

[10]   Dandapat, S., and Sarkar, S., *Part of Speech Tagging for Bengali with Hidden Markov Model*, In Proceeding of the NLPAI Machine Learning Competition, 2006.

[11]   Dasgupta, S., and Khan, M., *Feature Unification for Morphological Parsing in Bangla*, Proc. 7th ICCIT, Dhaka, Bangladesh, (2004).

---

[1] http://www.sust.edu/cse/nlp/finegrainedtagset

[12] Dashgupta, S., Khan, M., Pavel, D.S.H., Sarkar, A.I., and Khan, M., (2005), '*Morphological Analysis of Inflecting Compound words in Bangla'*, Proceeding of the International Conference on Computer, and Communication Engineering (ICCIT), Dhaka, Pp. 110-117.

[13] Dash, N., S., *The morphodynamics of the Bengali compounds: decomposing from lexical processing,* Language in India, pp. 1-25 (2006).

[14] Ekbal, A., and Bandyopadhyay, S., *Lexicon Development and POS tagging using a Tagged Bengali News Corpus*, In Proc. of FLAIRS-2007, Florida, 261-263, 2007.

[15] Ekbal, A., Haque, R., and Bandyopadhyay, S*., "Maximum Entropy based Bengali Part of Speech Tagging"*, In A. Gelbukh (Ed.), Advances in Natural Language Processing and Applications, Research in Computing Science (RCS) Journal, Vol. (33), 67-78, 2008.

[16] Hasan, F.M., UzZaman, N. and Khan, M., *Comparison of Unigram, Bigram, HMM and Brill's POS Tagging Approaches for some South Asian Languages*, Proc. Conference on Language and Technology (CLT07), Pakistan, 2007.

[17] Hoque, M., M., and Ali, M., M. *A Parsing Methodology for Bangla Natural Language Sentences*, Proceedings of International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, pp. 277-282, 2003.

[18] Jones, D., A., Berwick, R., C., Cho, F., Khan, Z., Kohl, K., T., Nomura, N., Radhakrishnan, A., Sauerland, U., and Ulicny, B., *Verb Classes and Alternations in Bangla, German, English and Korean,* Massachusetts Institute of Technology, Center for Biological and Computational Learning and the Artificial Intelligence Laboratory (1994).

[20] Mehedy, L., Arifin, S., M., N., and Kaykobad, M., *"Bangla Syntax Analysis: A Comprehensive Approach"*. ICCIT, pp.287-293., vol-1 (2003).

[21] Nath, J., Monojit, C., Mukherjee, A., Biemann, C. and Ganguly, N. *Unsupervised Parts-of-Speech Induction for Bengali,* International Conference on Language Resources and Evaluation (2008)

[23] Sarker, M., Z., H., Rahman, S. and Mottalib, M., A.,*Design & Implementation of Bottom-Up-Context-Sensitive Algorithms for Bengali Parser in Natural Language Processing,* International Journal of Knowledge-based and Intelligent Engineering Systems, pp.79-89 Vol. 11 (2007)

[24] Seddiqui, H., M., Rana, M., S., Mahmud, A., A. and Sayed, T*., Morphological Analysis of Bangla Words*, Proceedings of International Conference on Computer and Information Technology (ICCIT), pp.313-316 Vol. 6 (2003)

[25] Seddiqui, H., M., Rana, M., S., Mahmud, A., A. and Sayed, T.,*Parts of Speech tagging using Morphological Analysis in Bangla,* Proceedings of International Conference on Computer and Information Technology (ICCIT), pp.374-376 Vol. 6 (2003)