

**Part of Speech Tagging for Bangla: A Study**  
Mridusmita Mitra,Pampa Bhattacharyya  
NLP Section,ICT&Services Group,CDAC Kolkata  
(mridusmita.mitra,pampa.bhattacharyya)@cdac.in

Abstract

Part of Speech (POS) Tagging is the process of tagging i.e marking or categorizing a particular word according to a particular part of speech or syntactic category. However the process does not simply refer to the marking of words as noun, adjective, adverb etc, it is little complicated than that. The reason being, in natural language, a word can represent more than one part of speech and often allotting a particular part of speech to a word within a sentence might become tricky. POS tagging is important in the context of machine translation and is the first stepping-stone in designing chunker and parser algorithms for natural languages. The authors started working on POS tagger as a part of the Bangla Analyzer work with the aim of designing an automated POS tagger, which would facilitate the process of building a Bangla to English Machine Translation system in future. The tagset has been designed after a detailed study of the available tagsets in English and Bangla. The tagset designed by CDAC Kolkata tries to analyze words according to three levels; clausal, phrasal and word level so as to extract maximum information possible from a particular sentence. Although the tagset chiefly stresses on syntactic category, in some cases contextual information extraction also becomes necessary. For this reason the tagset has been designed in a way so that it can provide syntactic as well as some lexical information at the same time. Some Bangla sentences have been manually tagged using this tagset. The result is being studied in order to critically review the tagset and to make it better equipped for tagging Bangla sentences. After providing some background information about POS tagging and existing POS tagsets , this paper aims to present some important issues related to POS tagging . The tagset has been discussed in detail in the paper with appropriate examples. The paper also attempts to highlight some problems associated with tagging Bangla sentences.