Classification of Verbs –Towards Developing a Bengali Verb Subcategorization Lexicon

Professor Sivaji Bandyopadhyay
Computer Science and Engineering Department
Jadavpur University, Kolkata, INDIA.

Abstract

Subcategorization lexicon, especially for verbs is an important resource for any language. VerbNet is the largest online English verb lexicon with explicitly stated syntactic and semantic information based on Levin’s verb classification. We are developing the verb subcategorization lexicon in Bengali with the help of English VerbNet, Bengali chunker and a Bengali – English bilingual lexicon. The main hypothesis considered is that the subcategorization frames for a Bengali verb are same with the subcategorization frames for its equivalent English verb with an identical sense tag. In the Bengali-English bilingual lexicon, the equivalent English verbs for a Bengali verb are grouped together that share the same sense. Another important challenge in the task is that while English is a fixed phrase order language Bengali is a free phrase order language. The identified subcategorization frames for Bengali verbs are being used to create Bengali verb classes. The verb subcategorization frame acquisition system has demonstrated precision, recall and F-measure values of 74.11%, 70.83% and 72.44% respectively on a test set of 120 sentences. The flexible ordering of different phrases, additional attachment of optional phrases in Bengali sentences makes this frames acquisition task challenging. We are using English VerbNet as well as the information of Bengali verb subcategorization lexicon for identifying emotion holders in Sentiment Analysis.

Related publications