Hierarchal POS tagging for Tamil language using Machine learning approach

1Dr.V.Dhanalakshmi, HOD, Department of Tamil, FSH, SRM University, Kattankulathur. (dhanagiri@gmail.com).

2Mr.Anand Kumar, Assistant Professor, CIET, Coimbatore. (mailtoanandkumar@gmail.com)

Abstract:

This paper presents the intricacies involved in developing a hierarchal POS tagger generator using SVMTool for Tamil language. Tamil, a Dravidian language has a very rich morphological structure which is agglutinative. Tamil words are made up of lexical roots followed by one or more affixes, mostly suffixes. So tagging a word in a language like Tamil is very complex. We try to resolve this complexity by identifying the categorical ambiguities and developing three hierarchal tag sets at word grammatical category and grammatical feature level. These tag sets were used to annotate the corpora and trained using the SVMTool (An Open source tool available at http://www.lsi.upc.es/~nlp/SVMTool) to generate the POS tagger model. The results obtained in each level were encouraging.

References:


