Abstract

Function tagging is one of the essential steps in Myanmar to English machine translation system. In this paper we propose a set of function tags for Myanmar and address the question of assigning function tags to Myanmar words. A small functional annotated tagged corpus manually serves as the training data because the large scale Myanmar Corpus is unavailable at
present. Part of the challenge of statistical function tagging for Myanmar language comes from the fact that Myanmar has free phrase order and a complex morphological system. In the task of function tagging, we use the output of morphological analyzer which tags the function of Myanmar sentences with correct segmentation, POS (part-of-speech) tagging [1] and chunking information [2]. We use Naïve Bayesian statistics to disambiguate the possible function tags of each word in the sentence. Function tagging can be exploited by NLP applications such as syntactic and semantic analysis, information retrieval and machine translation. Experiments show that our analysis achieves a good result with simple sentences and complex sentences.

Reference

- Versteegen, L. 1999. The Simple Bayesian Classifier as a Classification Algorithm.

Index Terms

Computer Science
Natural Language
Key words

Function tagging  function tagset  Myanmar

language

functional annotated tagged corpus