

[CICLING 2010: 11th International Conference on Intelligent Text Processing and Computational Linguistics, March 21-27, 2010, Iași, Romania]

Abstract

Sobha Lalitha Devi, Vijay Sundar Ram R, Bagyavathi T and Praveen Pralayankar.
Syntactic Structure Transfer in a Tamil -Hindi MT System - A Hybrid Approach

We describe the Syntactic Structure Transfer (SST), a central design question in machine translation, between two languages Tamil (source) and Hindi (target), belonging to two different language families, Dravidian and Indo-Aryan respectively. Tamil and Hindi differ extensively at the clausal construction level and transferring the structure is difficult. The SST described here is a hybrid approach where we use CRFs for identifying the clause boundaries in the source language, Transformation Based learning(TBL) for extracting the rules and use semantic classification of postpositions for choosing the correct structure in constructions where there are one to many mapping in the target language. We have evaluated the system using web data and the results are encouraging.