Abstract

This paper describes the architecture of a Machine Translation System with source language as English and target language as Marathi. The basic approach used for the development of this system is Rule Based Machine Translation. The basic algorithm for obtaining the correct word order in the target language was developed based on specific traversals of the parse tree. One of the special features of the system is a Word Sense Disambiguation model. Presently only prepositions will be disambiguated and work is going on for verbs and nouns. The model is a generalized approach based on the categories/domains a word belongs to. Another feature is the target language generation module. The focus is on the grammar structure of the target language that will produce better and smoother translations. The architecture though developed specifically for English – Marathi language pair, may be extended to other language pairs with similar structure. The architecture is partially implemented in the form of Machine Translation system. A lexicon is built for morphological and semantic properties. The results, even at partial implementation stage, are really encouraging.
References

- Wren P. and Martin H. High School English Grammar and Composition. S Chand Publication
- Technology Development for Indian Languages, DIT, Government of India. Also available at: http://www.tdil-dc.in/index.php
- http://www.saakava.com

Index Terms

Computer Science

Text Processing
Keywords
Machine Translation  Word Sense Disambiguation  Parser  Transliteration
Marathi
Case-suffixes