

MORPHOLOGICAL GENERATOR ENGINE (For Telugu)

Christopher Mala

M.S (Computational Linguistics)

IIT-Hyderabad.

christopher.mpg08@research.iit.ac.in

Abstract

Telugu is an Indian language spoken by over 70 million people in the country. It is rich in literature and is being studied by native and foreign linguists significantly. Due to recent progress in standardization of machine representation of text, applications such as machine translation and information retrieval are beginning to surface. Morphological analyzer and morphological generator are two essential and basic tools for building any language processing application for a natural language. Morphological analysis means to study the internal structure of the words of a language. A Morphological analyzer gives the morph analysis of a word i.e. for a given word a morphological analyzer will return its root word and word class along with other grammatical information depending upon its word class. For instance, it will provide gender, number, and case information for nouns and for verbs it will provide the information like tense, aspect, mood, etc. But the Morphological generator does exactly the reverse of it, i.e. for a given root word and its grammatical information it is capable of generating the word form of that root word. There is a need for a Morphological generator for Telugu that forms a vital integral part of applications like machine translation and universal dictionary. Here we present a tool that generates morphological forms of nouns, verbs, Adjectives, Adverbs and numbers of Telugu. Morphological Generator is developed based on the Morphological Analyzer which works on Word-Paradigm method and was developed by Prof. G. Uma Maheshwar Rao at CALTS (HCU). The present Morphological Generator is developed in Perl® and will be made available with IL-ILMT system. This tool can be used as an engine. It is tested with languages like Tamil, Hindi, Kannada. Input and output are in WX-notation.