Abstract

Tamil is a morphologically rich language with agglutinative nature. Being agglutinative language most of the word features are postpositionally affixed to the root word. The morphological generator takes lemma, POS category and morpho-lexical description as input and gives a word-form as output. It is a reverse process of morphological analyzer. In any natural language generation system, morphological generator is an essential component in post processing stage. Morphological generator system implemented here is based on a new algorithm, which is simple, efficient and does not require any rules and morpheme dictionary. A paradigm classification is done for noun and verb based on Dr.S.Rajendran's paradigm classification. Tamil verbs are classified into 32 paradigms with 1884 inflected forms. Like verbs, nouns are classified into 25 paradigms with 325 word forms. This approach requires only minimum amount of data. So this approach can be easily implemented to less resourced and morphologically rich languages.
A Novel Data Driven Algorithm for Tamil Morphological Generator

Reference

Index Terms

Computer Science
Text Mining

Key words
word-forms Paradigm Suffix table
Morpho-lexical information

Tamil morphological generator