Abstract

Parsed Corpus is essential for various Natural Language Processing (NLP) activities, but its availability cannot be guaranteed for most of the languages. Furthermore, development of parsed corpus or parser is not an easy task. This is because success of these approaches almost invariably demands a huge amount of computational resources that are not typically available for most of the natural languages. Aim of the present research is to develop parsing schemes for natural language sentences. Parsing approaches viz. rule based or statistical, suffer from an inherent drawback that they typically lack a significant volume of computational resources. Languages lacking such computational resources are called “resource deficient” languages. Marathi language is one of the official languages in India, morphologically rich and free word order in nature. In this thesis, we propose a way of formulation of Marathi language grammar in Link Grammar formulation and parsing of such formulated grammar by using proposed Link Grammar Parsing Algorithm for Marathi. In this work we have pursued following two tasks:

Formulating the Marathi Link Grammar. We have formulated Marathi Link Grammar first, by introducing link types on studying various phenomenon of Marathi grammar. We present detail study of phrases such as Noun Phrase, Verb Phrase and agreement between subject/object to verbs. The study also includes sentence structures of Simple, Complex and Compound types with respect to computational purpose. We ensured to get good functionality of this link types so that parsing analysis can be effortless. This study leads us to propose functional links for Phrasal approach and Karaka approach. On studying, Marathi sentence structures, sentential links for Complex and Compound sentence types are also proposed.

Parsing of the formulated Marathi Link Grammar. This task, is divided again in two subtasks. First is designing an effective parsing algorithm for Marathi Link Grammar and secondly implementing and testing the parsing results. We designed the Parsing algorithm which may call subroutines designed to parse complex or compound type of sentence. Valid linkage shall be built on successful parsing. To acquire links automatically while parsing, Marathi WordNet and NER approaches are also proposed. The parsing algorithm developed in this work have been implemented and tested on a reasonably sized database. We have been able to demonstrate the efficacy of it. We feel that our research will pave the way for development of parsing other Indian languages too.

Keywords : Parsing, Link Grammar, Link Grammar Parsing Algorithm, Marathi Link Grammar