

# Converting Constituency Structures to Dependency Structures for Sanskrit: Linguistic Issues

Pawan Goyal<sup>1</sup> and Amba Kulkarni<sup>2</sup>

<sup>1</sup>*Indian Institute of Technology, Kharagpur.*

<sup>2</sup>*Department of Sanskrit Studies, University of Hyderabad*

## Abstract

According to the annotation scheme utilized, the treebanks are generally of two types,

- 1) Phrase structure (also called constituency structure) Treebanks such as English Penn Treebank and
- 2) Dependency Treebanks such as the Prague Dependency Treebank.

In this work, we explore the relationship between constituency structures and dependency structures for Sanskrit language, with main emphasis on the conversion from constituency to dependency structures. This work aims not only at designing an algorithm to convert Treebanks from one type of representation to the other, but also to bring out the difference in the representational coverage of these two approaches.

The dataset that we are using in this work has its origin in Apte's student guide. The work was initiated in 1986 by Brendan Gillon, who assigned a syntactic parse to each sentence, giving each sentence an English translation and annotating each sentence for miscellaneous syntactic and semantic facts. In 2004, Gérard Huet re-engineered the document in order to parse it mechanically, and he verified its correct syntactic structure after typographical corrections. He devised an abstract syntax to formalize this constituency structure.

In this talk, we will describe our algorithm for converting a constituent tree into a dependency tree. We will also focus on the comparison of the information presented in the constituency tree to that in a dependency tree. Using some concrete examples, we would also like to open the discussion as to which of the two structures, dependency or constituency, preserve most of the syntactico-semantic information in the sentence and what are the advantages of using one representation over another, for annotating Sanskrit corpus.