

## **Introduction**

Morphological Analysis is an important area of NLP. A morphological analyzer takes a written or spoken text as input and analyzes it according to the rules of that language to provide the morphological analysis as output. Morphology is one of the constituting layers of linguistic structure of a language. It is concerned with the combination of morphemes to form new words. The importance of morphology and its analysis for a language vary according to the nature of that language. For most Indian languages, morphology is of much more importance because these languages contain most of the information in the words which show a rich inflectional and derivational tendency. That is why morphological analysis has been in the centre of NLP projects and activities in India. Various institutions are engaged in developing a variety of morphological tools such as analyzers, POS taggers, spell checkers, text processors and so on for Indian languages.

The present work is an attempt to make an analyzer for inflectional verb morphology of Sanskrit. Sanskrit acquires a unique place in Indian linguistic scenario. Apart from being one of the most ancient languages which also has a vast variety of literature, the language has scientifically defined structure. It has also left its impression on most of the languages of the subcontinent in morphology, syntax, grammar and many other linguistic aspects. Sanskrit follows a well defined process to derive and inflect words. The dominance of morphology in Sanskrit and other Indic languages leaves little scope for syntactic analysis. So, language almost follows free word-order and most of the information is contained in morphology. It is why morphological analysis has much more importance in developing any natural language understanding system for Sanskrit. The same is true for most of the other Indian languages.

The verb analyzer developed here recognizes verb forms in a given Sanskrit text. It further analyses these words to retrieve the related information. The core meaning of a Sanskrit verb is contained in the verb root or stem which becomes a base before verbal affixes are applied to it. The verbal suffixes which are called *tin* terminations signify tense, aspect, mood, person and number etc. The analyzer follows the reverse

Pāṇinian approach to identify and analyze these constituent parts in Sanskrit verb forms. However, the present work has limited itself only to the first class of Sanskrit verb root list, i.e. *bhvādigāṇa*. It attempts to provide an analysis strategy which is able to identify the *tiṅ* affix in the verb root and thus recognize a *tiṅanta* verb in given Sanskrit text. Further it analyses the string to retrieve the base and prefix (if any) from a verb form.

The dissertation titled “Computational Identification and Analysis of Sanskrit Verb-forms of *bhvādigāṇa*” has following chapters –

- The first chapter “Computational Morphology and Sanskrit” introduces computational morphology with its various implementations in different systems. Computational morphology is mainly concerned with morphological analysis and generation. Morphological analyzers and generators use rules and lexicons of a language to accomplish respective tasks. Efforts for developing NLP tools for Indian languages have also focussed on computational morphology a brief survey of which is given. Also discussed are some prominent features of Sanskrit morphology, particularly verb morphology.
- The second chapter “Sanskrit Verb Morphology” discusses Sanskrit verb morphology in detail. Sanskrit verb forms are produced by conjugation of verb roots (*dhātus*) in ten *lakāras*. *Dhātus*, the meaning-bearer unit of verb, may be primitive (around 2000, listed in *dhātupaṭha*) or derived (from verb root or nominal stem). These verb roots, also called *dhātus*, are divided in 10 classes and roots of a single class follow the same process in most of the cases. The rules of Pāṇinian *Aṣṭādhyāyī* apply on these *dhātus* to form various verb forms in different paradigms from these verb roots. There are 10 *lakāras* which denote tense, aspect and mood in Sanskrit. There are 18 basic verbal terminations
- Third chapter “Morphological Analysis of Sanskrit Verb Forms” focuses on the analysis strategy that is to be followed in this work. Various approaches have been developed for morphological analysis. The popular ones are cut and

paste technique and finite state technique. The strategy followed in this work is based on reversed Pāṇinian approach. Pāṇinian system follows a step-by-step method to derive full-fledged verb form from a given root. The analysis methodology tries to sort out various morphemes in the verb forms

- The partial implementation of the verb analysis methodology to computer program is discussed in the fourth chapter “Verb Analyzer for Sanskrit.” The analysis scheme is implemented to a computational program for identification and analysis of Sanskrit verb forms. The techniques used are Java programming language, Java Servlet Technology and JSP run on Apache Tomcat 4.0 Web-server. The system takes a running text of Sanskrit in Devanāgarī UTF-8 format as input. It first does the pre-processing and then tokenizes the text in order to separate all the words. Then it locates the *tiṅ* suffixes in *tiṅanta* forms and identifies them. In the next step, the system takes all the identified *tiṅanta* forms and separates the suffix in order to get the verbal base. In case, prefix is attached to it, it also searches for it.

The verb analyzer is a component of a larger system aimed at complete machine understanding of Sanskrit language. The system is available online at <http://sanskrit.jnu.ac.in> . The work done for this dissertation is a preliminary for complete verb analysis of Sanskrit and has immense scope of being developed as a comprehensive analyzer for Sanskrit verb forms.