

# Chapter 49

## Neural Network Model for Semantic Analysis of Sanskrit Text

**Smita Selot**

*SSTC, India*

**Neeta Tripathi**

*SSTC, India*

**A. S. Zadgaonkar**

*CV Raman University, India*

### ABSTRACT

*Semantic analysis is the process of extracting meaning of the sentence, from a given language. From the perspective of computer processing, challenge lies in making computer understand the meaning of the given sentence. Understandability depends upon the grammar, syntactic and semantic representation of the language and methods employed for extracting these parameters. Semantics interpretation methods of natural language varies from language to language, as grammatical structure and morphological representation of one language may be different from another. One ancient Indian language, Sanskrit, has its own unique way of embedding syntactic information within words of relevance in a sentence. Sanskrit grammar is defined in 4000 rules by Panini reveals the mechanism of adding suffixes to words according to its use in sentence. Through this article, a method of extracting meaningful information through suffixes and classifying the word into a defined semantic category is presented. The application of NN-based classification has improved the processing of text.*

### INTRODUCTION

Semantic, in basic form is the meaning of the sentence. In computer driven world of automation, it has become necessary for machine to understand the meaning of the given text for applications like automatic answer evaluation, summary generation, translation system etc. In linguistics, semantic analysis is the

DOI: 10.4018/978-1-7998-0951-7.ch049

process of relating syntactic structures, from words and phrases of a sentence to their language independent meaning. Given a sentence, one way to perform semantic analysis is to identify the relation of the words with action entity of the sentence. For example, *Rohit ate ice cream*, agent of action is Rohit, object on which action is performed is ice cream. This type of association creates predicate-arguments relation between the verb and its constituent. This association is achieved in Sanskrit language through *kAraka* analysis. Understanding of the language is guided by its semantic interpretation. Semantic analysis in Sanskrit language is guided by six basic semantic roles given by *pAninI* as *kAraka* values.

*pAninI*, an ancient Sanskrit grammarian, mentioned nearly 4000 rules called *sutra* in book called *asthadhyAyi*; meaning eight chapters. These rules describe transformational grammar, which transforms root word to number of dictionary words by adding proper suffix, prefix or both, to the root word. Suffix to be added depends on the category, gender, number of the word. Structured tables containing suffix are maintained for the purpose. These declension tables are designed in such a way that their position in the table are defined with respect to number, gender and *karka* value. Similar ending words follow the same declension, for example *rAma* is *a*-ending root word and words generated using *a*-ending declension table are *rAmH*, *rAmau* *rAmAH* by appending *H*, *au* and *AH* to *rAma*, respectively. Suffix based information of the word reveals not only syntactic but drives a way to find semantic based relation of words with verb using *kAraka* theory.

Next section describes Sanskrit language and *kAraka* theory, section three states the problem definition, followed by NN model for semantic analysis. Features extracted from corpus of pre-annotated text are supplied as input to system with objective of making system learn six *kAraka* defined by *pAninI*. This paper presents the concept of Neural Network, work done in the field of NN and Natural Language Processing, algorithm, annotated corpus and results obtained.

## **SANSKRIT AND *kAraka* THEORY**

Sanskrit language, with well-defined grammatical and morphological structure, not only presents relation of suffix-affix with the word, but also provides syntactic and semantic information the of words in a sentence. Due to its rich inflectional morphological structure; it is predicted to be suitable for computer processing. Work at NASA on Sanskrit language reported that triplets (role of the word, word, action) generated from this language are equivalent to semantic net representation (Briggs 1995).

Sanskrit grammar is developed by three main individuals over the years – *pAninI*, *Katyanan* and *Pa-tanjali*. *pAninI* was first one to identify nearly 4000 rules to define the language (Kak, 1987). Sanskrit is one of the 22 official languages of India and standardized dialect of old Indo-Aryan community. It is a rule-based language having flexibility and precision, both. It is an order free language – changing the position of the word, does not change the meaning of the sentence. Processing techniques for order free languages cannot be derived from rigid order language like English. *pAninI*'s idea of semantic interpretation was based on identification of relations of words with action in a sentence. as agent, object etc. This was called as *kAraka* theory. In Sanskrit, this relationship is developed by adding specific, pre-set syllables, known as case-endings or *vibhakti* to the basic noun form. Natural language processing concept in *pAnini* way describes the processing of language

Sanskrit is highly inflectional language, where words are generated from the root word. Structure of the word is given as <word>=<prefix><root\_word><suffix>. In *pAninI* system finite set of rules are enough to generate infinite sentences (Kak,1987). Extraction of prefix and suffix by morphologi-

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:  
[www.igi-global.com/chapter/neural-network-model-for-semantic-analysis-of-sanskrit-text/239977](http://www.igi-global.com/chapter/neural-network-model-for-semantic-analysis-of-sanskrit-text/239977)

## Related Content

---

### How Experiences with Words Supply All the Tools in the Toddler's Word-Learning Toolbox

Carmel Houston-Price and Beth Law (2014). *Computational Linguistics: Concepts, Methodologies, Tools, and Applications* (pp. 1345-1373).

[www.irma-international.org/chapter/how-experiences-with-words-supply-all-the-tools-in-the-toddlers-word-learning-toolbox/108781](http://www.irma-international.org/chapter/how-experiences-with-words-supply-all-the-tools-in-the-toddlers-word-learning-toolbox/108781)

### Formal Semantics for Metamodel-Based Domain Specific Languages

Paolo Arcaini, Angelo Gargantini, Elvinia Riccobene and Patrizia Scandurra (2014). *Computational Linguistics: Concepts, Methodologies, Tools, and Applications* (pp. 297-323).

[www.irma-international.org/chapter/formal-semantics-for-metamodel-based-domain-specific-languages/108727](http://www.irma-international.org/chapter/formal-semantics-for-metamodel-based-domain-specific-languages/108727)

### Error or Strength?: Competencies Developed in Adolescent Digitalk

Kristen Hawley Turner (2014). *Computational Linguistics: Concepts, Methodologies, Tools, and Applications* (pp. 70-88).

[www.irma-international.org/chapter/error-or-strength/108715](http://www.irma-international.org/chapter/error-or-strength/108715)

### Parsing Bangla Grammar Using Context Free Grammar

Al-Mahmud, Bishnu Sarker and K. M. Azharul Hasan (2013). *Technical Challenges and Design Issues in Bangla Language Processing* (pp. 137-154).

[www.irma-international.org/chapter/parsing-bangla-grammar-using-context/78474](http://www.irma-international.org/chapter/parsing-bangla-grammar-using-context/78474)

### Technology-Mediated Tasks in English for Specific Purposes (ESP): Design, Implementation and Learner Perception

Cédric Sarré (2014). *Computational Linguistics: Concepts, Methodologies, Tools, and Applications* (pp. 951-967).

[www.irma-international.org/chapter/technology-mediated-tasks-in-english-for-specific-purposes-esp/108759](http://www.irma-international.org/chapter/technology-mediated-tasks-in-english-for-specific-purposes-esp/108759)