

# Lexical Resources for Recognition, Analysis and Word Formation Process for Sanskrit Morphology

Subhash Chandra, Bhupendra Kumar, Vivek Kumar and Sakshi

Computational Linguistics, Department of Sanskrit

Department of Sanskrit, Faculty of Arts, University of Delhi, Delhi, India

E-mail: subhash.jnu@gmail.com, bhupendrakmr87@gmail.com, vishnorapatyam@gmail.com, sakshi911003@gmail.com

## Abstract

Sanskrit is morphologically very rich language. Major work on grammatical tradition for Sanskrit is done by Pāṇini in his Aṣṭādhyāyī (AD) which particularly contains about 3,959 rules of Sanskrit morphology, syntax and semantics. Sanskrit word formation process is taught in all major Indian Universities offering Sanskrit courses at Undergraduate (UG) and post graduate (PG) level. This paper introduces a web based word formation process tools for students and teachers of Sanskrit with the aim of teaching and learning Sanskrit morphological inflectional process based on Pāṇini rules and *prakriyā granthas* of AD. The system is developed by combining rule and example based approaches used by Pāṇini. There are three components entitled Recognizer, Analyzer and Word Formation Process (WFP) Generator in the system that generate word formation process for *subanta* (nominal), primary verb (*tinanta*) and secondary verb (*sanādyanta*). Currently this system covers *subanta* and *tinanta* only and it is being used by the Sanskrit students and teachers for learning and teaching Sanskrit Grammar.

**Keywords:** Language Resources, Sanskrit morphology, E-Learning tools for Sanskrit, Sanskrit rupa siddhi, Morphology, Morphological Analyzer, Sanskrit Morphological System, Morphological Analysis Methods, Morphological Recognizer and Analyser for Sanskrit.

## 1. Background

There are two types of Sanskrit morphology Nominal (*subanta*) (Chandra, 2006 and Chandra, 2012) and Verbal (*tinanta*) (Chandra, 2006; Chandra, 2012; Jha et al, 2009; Jha et al, 2006 and Agrawal, 2007). Nominal may primary (*kṛidanta*) (Singh, 2008; Shailaja, 2014 and Murali et al, 2014), secondary (*taddhita*), compound (*samāsānta*) [11] and feminine (*strīpartayayānta*) (Bhadra, 2007) it derives with the addition of 21 morphological suffixes in eight *vibhaktis* and three numbers according to the end character of the base words called *prātipadika* (Chandra, 2006; Chandra, 2012; Jha et al, 2009 and Jha et al, 2006). Therefore a single word generates 24 forms. Sanskrit verbal system is very complex with verbs inflecting for different combinations of tense, aspect, mood, number, and person. There are approximately 2000 verb roots listed in Pāṇini's dhātupāṭha (DP). There are two broad ways of classifying Sanskrit verbal roots. First *parasmaipadī*, *ātmanepadī* and *ubhayapadī* (that derived in *parasmaipadī* and *ātmanepadī* both) and second verb roots are divided into 10 classes (*gaṇas*) according to the structure of the verb forms. Exponents used in verb conjugation include prefixes, suffixes, infixes, and reduplication. Sanskrit verbs are two types. The forms which derived with the addition of 18 suffixes (9 for *ātmanepadī* and 9 for *parasmaipadī*) in three persons, three numbers and with the addition of multiple prefixes that are called primary verb. There are 12 secondary suffixes that add with specific verb roots and nominal bases and create new verb roots and again derive with 18 verbal suffixes. These verbs called secondary verb

forms. Each Sanskrit verbs are derived into 10 *lakārs* (Chandra, 2006; Chandra, 2012; Jha et al, 2009; Jha et al, 2006 and Bhadra et al, 2009). Therefore, a single verb root may generate a lots of forms and creates very complication to detection, analysis and word formation process.

## 2. Materials and Methods

Detection and analysis of the Sanskrit morphology is very essential and challenging task. Rule of Sanskrit grammar by Pāṇini (Sharma, 2003), Hindi meaning, explanation by siddhāntakaumudī (Shastri, 1994) and DP are used as material for this work. Combining rules base and example based (hybrid) methods are used for detection and analysis of Sanskrit morphology. Computational Linguistic (Jurafsky & Martin, 2008) and Software engineering methods are also used for computational analysis. Sample of the rules for recognition is shown in table 1. System generates complete word formation process with the help of Pāṇini's rules based on the detection and analysis. Computational rules have been developed for detection and analysis and a database is also created for rules, Hindi meaning and explanation of Panini's AD rules for the word formation process. This system accepts Devanagari Unicode texts through web based user interface and generates results in same format in Hindi language only. Methodology can be understood with the flowchart shown in Figure: 1.

## 3. Component of the System

There are three components of the system for recognition, analysis and Word formation process generation.

### 3.1 Recognizer

This component recognize the input words with the help of Recognition rule and example database. Sample of the database is shown in the Table 1.

Sr	EndStr Ln	EndChr	Example
1	7	ाभ्याम्	रामाभ्याम्
2	6	भ्याम्	हरिभ्याम्
3	6	स्मात्	सर्वस्मात्
4	6	स्मिन्	सर्वस्मिन्
5	5	ेभ्यः	रामेभ्यः
6	5	ाणाम्	रामाणाम्
7	5	ानाम्	रामानाम्
8	5	ेषाम्	सर्वेषाम्
9	4	स्मै	सर्वस्मै
10	3	ान्	रामान्
11	3	ात्	रामात्
12	3	ाद्	रामाद्
13	3	स्य	रामस्य
14	3	योः	रामयोः
15	3	ेषु	रामेषु
16	3	ेसु	रामेषु
17	3	ीन्	हरीन्
18	3	ीन्	हरीन्
19	2	ाः	रामाः
20	2	म्	रामम्
21	1	ौ	रामौ

Table 0: Sample of rules for recognition

### 3.2 Analyzer

After Recognition this component analyzes the input

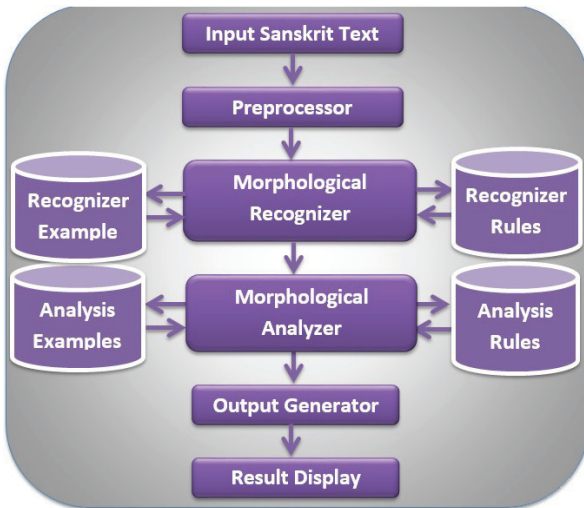


Figure 1: Methodology

words with the help of Analysis rule and example database. Sample of the database is shown in the Table 2.

### 3.3 Word Formation Process Generator

This component is the main component of this system which generates the word formation process. WFP is done with the help of WFP database and Example database.

## 4. Result and Discussions

System accepts Devanagari Sanskrit text in Unicode format as input and does detection and morphological analysis of input text. Based on detection and analysis, system generates complete WFP in tabular format with all essential information (as shown in Figure 2). All rules shown for WFP are linked with over mouse function for meaning of the rule and hyperlinked for the explanation of the rule. User can see the meaning of the rules used in WFP through keeping mouse over on any rule. Explanation may be seen after clicking on the particular rules. A snapshot is shown in Figure 2. Few limitations of the system are also seen in detection, analysis and word formation process. The recognizer is depend on the dataset and rules for recognition, second component analyzer is depend on the first component and analysis rules and third component is depend on second component.

Sr	addr	CaseNum	End	Gen	suffi x
1		3.2/4.2/5.2	अ	M	भ्याम्
2		3.2/4.2/5.2	इ	M	भ्याम्
3		5.1	सर्वादि	M	ङ्स्
4		7.1	सर्वादि	M	ङि
5		4.3/5.3	अ	M	भ्यस्
6		6.3	अ	M	आम्
7		6.3	अ	M	आम्
8		6.3	सर्वादि	M	आम्
9		4.1	सर्वादि	M	ङे
10		2.3	अ	M	शस्
11		5.1	अ	M	ङ्स्
12		5.1	अ	M	ङ्स्
13		6.1	अ	M	ङसि
14		6.2/7.2	अ	M	ओस्
15		7.3	अ	M	सुप्
16		7.3	अ	M	सुप्
17	ि	2.3	इ	M	शस्
18	ि	2.3	इ	M	शस्
19		1.3	अ	M	जस्
20		2.1	अ	M	अम्
21		1.2/2.2	अ	M	औ/औद्

Table 0: Sample of rules for Analysis

It mean if first component fail or does wrong detection then analyzer may also do the wrong analysis and if

analyzer does wrong analysis then WFP generator also generates wrong WFP. The system is being used by students and improvement is being done as per feedback from the users.

## 5. Conclusions

The system detects, analyze and generates WFP based on Pāṇinian formulation. The system is very useful for students and teachers for learning and teaching. Various language resources such as database for AD rules with Hindi meaning and Explanation and other relevant information, Computation rules for identification and analysis, database for Pāṇini's DP, various small dataset for Sanskrit grammar are also developed. This system is the part of E-learning system (under development) for Sanskrit. In future our aim to develop multilingual system for Sanskrit morphology Word Formation Generation.

## 6. Acknowledgements

This research was funded by Research Council, University of Delhi, Delhi, India under Research and Development Grants (R&D) 2014-15.

## 7. References

- Agrawal, Muktanand. (2007). *Computational identification and analysis of Sanskrit verb-forms of bhvaadigana*, Diss. Special Center for Sanskrit Studies, Jawaharlal Nehru University, New Delhi.
- Bhadra, Manji, Singh, Surjit Kumar, Kumar, Sachin, Chandra, Subhash, Agrawal, Muktanand, Chandrashekar, R, Mishra, Sudhir Kumar & Jha, Girish Nath. (2009). *Sanskrit Analysis System (SAS)*. Sanskrit Computational Linguistics Lecture Notes in Computer Science by Springer Berlin Heidelberg, pp. 116-133.
- Bhadra, Manji. (2007). *Computational analysis of gender in Sanskrit noun phrases for Machine Translation*. Diss. Special Center for Sanskrit Studies, Jawaharlal Nehru University, New Delhi.
- Chandra Subhash & Jha Girish Nath. (2011). *Computer Processing of Sanskrit Nominal Inflections: Methods and Implementation*. Cambridge Scholars Publishing (CSP), UK.
- Chandra, Subhash. (2006). *Machine Recognition and Morphological Analysis of Subanta padas*. Diss. Special Center for Sanskrit Studies, Jawaharlal Nehru University, New Delhi.
- Chandra, Subhash. (2012). Restructuring of Paninian Morphological Rules for Computer processing of Sanskrit Nominal Inflections. In *Proceedings of the Workshop on Indian Language Data: Resources and Evaluation*, Lütfi Kırdar Istanbul Exhibition and Congress Centre, Turkey.
- Jha, Girish Nath, Agrawal, Muktanand, Chandra, Subhash, Mishra, Sudhir Kumar, Mani, Diwakar, Mishra, Diwakar, Bhadra, Manji & Singh, Surjit Kumar. (2009). *Inflectional Morphology Analyzer for Sanskrit*, Sanskrit Computational Linguistics Lecture Notes in Computer Science by Springer Berlin Heidelberg, pp. 219-238.
- Jha, Girish Nath, Bhowmik, Priti, Mishra, Sudhir Kumar, Chandrashekar, R, Chandra, Subhash, Mendiratta, Sachin & Agrawal, Muktanand. (2006). *Towards a Computational analysis system for Sanskrit*. In *Proceedings of first National Symposium on Modeling and Shallow parsing of Indian Languages* at Indian Institute of Technology Bombay pp 25-34 on 2nd to 4th April 2006.
- Jurafsky, Daniel & James, H. Martin. (2008). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Dorling Kindersley, India.
- Murali, N, Ramasree, R. J. and Acharyulu, K.V. R. K. (2014). Kridanta Analysis for Sanskrit. *International Journal on Natural Language Computing (IJNLC)* Vol.3, No.3, June.
- N, Shailaja. (2014). *Comparison of Paninian Dhātuvṛttis*. Thesis Department of Sanskrit Studies University of Hyderabad.
- Satuluri, Pavankumar. (2015). *Sanskrit Compound Generation: With a Focus on the Order of Operations*. Thesis Department of Sanskrit Studies School of Humanities University of Hyderabad.
- Sharma, Rama Nath. (2003). *The Ashtaadhyaayi of Paanini*. Munshiram Manoharlal Publishers Pvt. Ltd., Delhi.
- Shastri, Vasudev Lakshman Panashikar (Ed.). (1994). *Siddhaantakaumudi*. Chaukhamba Sanskrit Pratisthan, New Delhi.
- Singh, Surjit Kumar. (2008). *Kridanata Recognition and processing for Sanskrit*, Diss. Special Center for Sanskrit Studies, Jawaharlal Nehru University, New Delhi.

## सुबन्त रूपसिद्धि निर्मापक Word Formation Generation for Sanskrit Nominal Inflections

The "Expert System for Sanskrit Grammar for E-learning (संस्कृत व्याकरण ई-शिक्षक)" is a result of the research project carried out by [Dr. Subhash Chandra](#), Assistant Professor, Computational Linguistics, Department of Sanskrit, University of Delhi, Delhi under R&D Grant, University of Delhi, 2014-2015 for the development of E-learning tools for Sanskrit. The coding for the application was done by [Dr. Subhash Chandra](#). Data set and rules were prepared by Ph.D. Research Scholars ([Mr. Bhupendra Kumar](#), [Mr. Madhav Prasad Wostj](#), [Mr. Vivek Kumar](#), [Ms. Sakshi](#)) under supervision of [Dr. Subhash Chandra](#).

सुबन्त रूपसिद्धि प्रक्रिया के लिए कृपया यूनीकोड में पद लिखें ।  
(Enter Word/s in Unicode for Sup Generation Process)

रामः

रूपसिद्धि के लिए क्लिक करें

### Result:

शब्दरूप/पद = रामः	
पद पहचान/लिङ्ग विभक्ति एवं वचन = राम पुल्लिङ्ग प्रथमा एकवचन	
Recognition Code = श_M_1.1	
<u>अर्थवदधातुरप्रत्ययः प्रातिपदिकम्</u> सूत्र से राम की अव्यत्यय पक्ष में प्रातिपदिक सञ्जा	राम
<u>प्रत्ययः</u>	राम
<u>परञ्च</u>	राम
<u>झाप्धातिपदिकात्</u> सूत्र के अधिकार में	राम
<u>स्वोवसप्तौट्छष्टाभ्याम्भिस्ङेभ्याम्भ्यस्ङसोसंझोत्सुप्</u> सूत्र से ड्यन्त/आवन्त/प्रातिपदिक में परे सुँ, औ, जस् आदि इङ्गीम प्रत्यय	राम + सुँ/औ/जस्.....
<u>द्वकयोद्विवचनैकवचने</u> सूत्र से एकत्व की विवक्षा में एकवचन / द्वित्व की विवक्षा में द्विवचन का प्रत्यय होने पर	राम + सु
<u>उपदेशेऽजनुनासिक इत्</u> सूत्र से उपदेशावस्था में अनुनासिक अच् (स्वर) की इत्सञ्जा ।	राम + सु
<u>तस्य लोपः</u> सूत्र से इत्सञ्जाक वर्ण का लोप होने पर	राम + सु
<u>सुसिङन्तं पदम्</u> सूत्र से सुबन्त / तिङन्त की पद सञ्जा	राम + सु
<u>सप्ततुषो रुः</u> सूत्र से सकार / सप्तुआन्द के षकार के स्थान पर रु आदेश	राम + रु
<u>उपदेशेऽजनुनासिक इत्</u> सूत्र से उपदेशावस्था में अनुनासिक अच् (स्वर) की इत्सञ्जा ।	राम + रु
<u>तस्य लोपः</u> सूत्र से इत्सञ्जाक वर्ण का लोप होने पर	राम + रु
<u>विरामोऽवसानम्</u> सूत्र से वर्णों के अभाव की अवसान सञ्जा	राम + र्
<u>बरवसानयोर्विसर्जनीयः</u> सूत्र से रफ को विसर्ग आदेश होकर	राम + ः
वर्ण सम्मेलन करने पर रामः रूप सिद्ध होता है ।	रामः

Figure 0: Screen Shot Web based System for Sanskrit Grammar with Result details