

Developing network of Sanskrit words across Part-Of-Speech categories

Sivaja S. Nair, Pritilaxmi Swain, Amba Kulkarni
Department of Sanskrit Studies,
University of Hyderabad,
Hyderabad, India,
sivaja.s.nair@gmail.com, pritisans@yahoo.co.in, apksh@uohyd.ernet.in

January 11, 2009

1 Introduction

Words in any language are related to various other words in it. Some relations are lexical while some are semantic. WordNet [English WordNet] is an excellent tool that allows one to navigate through various relations a word has with others to get a holistic view of the meaning it conveys. The synsets convey the sense, antonyms give the words with opposite sense, meronymy and hyponymy help one in identifying the parts of the object and the objects of which the given word is a part of, hypernymy and hyponymy give an idea of ontological classification. In case of verbs the entailment relation helps in understanding the activity-subactivity relation.

Usefulness of English wordnet led of the development of WordNets for other languages and also the interlinking of various WordNets. Eversince the development of English WordNet[Fellbaum, 1999] the computational lexicography work has gained momentum and acquired a new direction. Several projects purely dedicated to building WordNets for different languages, linking the existing WordNets and building multilingual WordNets were taken up during the last decade[Vossen, 2002 and Sinha et. al, 2006]. Though the usefulness of WordNet for NLP is still to be established, there are several efforts to show its significance and relevance for the NLP related work[Agirre E. et. al, 1996]. EuroWordNet is a multilingual WordNet for several European languages (Dutch, Italian, Spanish, German, French,

Czech and Estonian).

In India, there have been efforts at several places all over the country to develop WordNets for Indian Languages (Tamil, Marathi, Hindi, Sanskrit)[Tamil WordNet, Marathi WordNet, Hindi WordNet and Sanskrit WordNet]. Though there were initiatives to start the work on Sanskrit WordNet[Mohanty et. al, 2002] nothing concrete has yet come out. In the recent developments, Kulkarni[Kulkarni et al, 2008b] presents the interlinking of amarakosha with the Hindi WordNet.

2 Richness of Sanskrit morphology

Now the question is, Is the present structure of wordnet capable enough to express the network of relations among different words in Sanskrit? To understand the complexity involved in representing the network of words in Sanskrit we first understand the complexity of word formation in Sanskrit.

Two important aspects of language study are its grammar and its lexicon. *Pāṇini's Aṣṭādhyāyī* and *Amarasimha's Nāmaliṅgānuśāsanam* popularly known as *Amarakośa* both belonging to roughly 5th century B.C. serve as monumental works in the area of grammar and lexicography respectively. Though lexicographic work such as *Nighaṇṭu* existed before *Amarakośa*, *Amarakośa* dealt with essentially non-vedic words and hence gained importance very soon.

Some languages build extensively while others to a limited extent only. Raghuvira(1981) in the introduction of his ambitious project of building English - Hindi dictionary of technical terms, where he borrows heavily from Sanskrit, describes the richness of word-formation in Sanskrit in the following words.

While every language builds to a certain extent, it is only a very small number that build constantly, and not only single stray words but whole systems. These are the three great classical languages of the world. ... are Sanskrit, Chinese and Latin (with Greek)[Raghuvira, 1981].

Figure 1 describes the rich word formation in Sanskrit through the Finite State Transducer(FST).

Thus, as is clear from figure1, the relation between words across Part of Speech(POS) also becomes very significant in case of Sanskrit. However English WordNet does not contain syntagmatic relations linking words from different syntactic categories except for a few such as legal-lawyer, big-size [Fellbaum, 1999]. To get an idea of the richness in building words in Sanskrit, we show in figure2 the compositionality in the meaning of nouns derived from verbs by adding non-finite suffixes(*kṛt*). Sanskrit has around 140 such *kṛt* suffixes, and the derivation is quite productive. As is evident from the figure2, such a network of Sanskrit words explaining the relationships among them is a valuable resource for any NLP work related to Sanskrit. The important role of verbs in building Sanskrit WordNet is also highlighted by Kulkarni[Kulkarni et. al, 2008].

Thus, there are two distinct tasks: one is to develop a network of words within a syntactic category which is more or less parallel to the concept of English WordNet, though in case of Sanskrit the ontological classification may be influenced by the *Vaiśeṣika* ontology. The development of WordNet for nouns extracting synsets from *Amarakośa* is described in Bharati[Bharati et. al, 2008]. Owing to the productive nature of Sanskrit in word building there is another important and unique task of developing a network between the words belonging to different syntactic categories but related semantically. In what follows we describe the approach we would like to build this network.

3 New features of networking of Sanskrit words

Verbal suffixes play a very significant role in Sanskrit. *Pāṇini* has given certain rules for deriving *kṛdanta* forms from the roots. He has provided nearly 130 *kṛt* suffixes in *Aṣṭādhyāyī*. Most of the suffixes occur with most of the roots. Some of the *kṛdanta*'s can take noun inflections while others are indeclinables. 15 high frequent *kṛt* suffixes namely *anīyar*, *artha*, *ghan̄*, *kta*, *ktavatu*, *ktivā*, *lyuṭ*, *ṇamul*, *ṇvul*, *ṇyat*, *śānac*, *śatr̄*, *tavyat*, *tṛc*, *tumun*, *yat* are selected here for experiment.

Semantically, one suffix may correspond to more than one meanings. For example the suffix *ghan̄* occurs in the sense of all 6 karakas as well as *bhāva* as shown in table1.

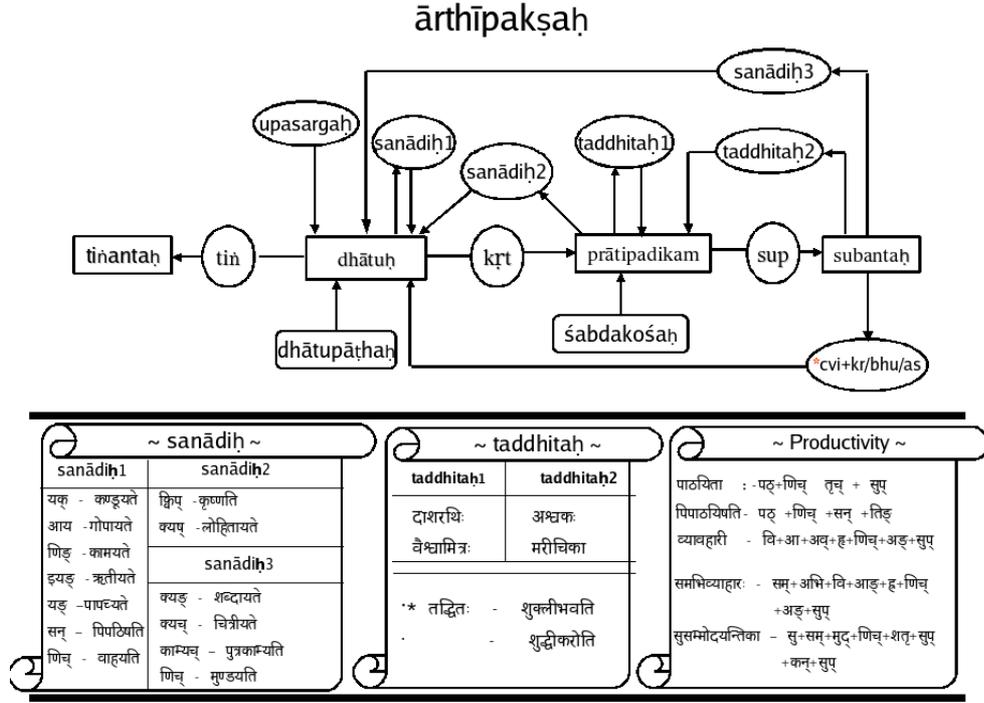


Figure 1: Word Formation in Sanskrit

Legends:

<i>dhātuḥ</i>	verbal root	<i>sup</i>	nominal suffix
<i>subantaḥ</i>	noun	<i>kṛt</i>	nonfinite verbal suffix
<i>tin</i>	finite verbal suffix	<i>śabdakośaḥ</i>	lexicon
<i>dhātupāṭhaḥ</i>	verbal root list	<i>taddhitaḥ</i>	derivational suffix
<i>sanādīḥ</i>	derivational suffixes	<i>upasargaḥ</i>	verbal prefix

At the same time it is possible to have different suffixes representing the same meaning. For example, multiple meanings of *ghañ* suffix as shown in table1.

<i>pratyaya</i>	<i>artha</i>	<i>udāharaṇa</i>
<i>ghañ</i>	<i>kartā</i>	<i>ātmarāmaḥ</i>
<i>ghañ</i>	<i>karma</i>	<i>abhravarṣam</i>
<i>ghañ</i>	<i>karaṇa</i>	<i>apamānam</i>
<i>ghañ</i>	<i>sampradāna</i>	<i>dāsaḥ</i>
<i>ghañ</i>	<i>apādāna</i>	<i>sphoṭaḥ</i>
<i>ghañ</i>	<i>adhikaraṇa</i>	<i>rāma</i>
<i>ghañ</i>	<i>bhāva</i>	<i>pākaḥ</i>

Table 1: Multiple meanings of *ghañ*

4 Proposal

Each of the *kṛt* suffixes represent a semantic relation between a verb root and its form. These relations may be classified into the following *kartā*, *karma*, *karaṇa*, *sampradāna*, *apādāna*, *adhikaraṇa*, *bhāva*, *bhūta*, *bhaviṣyat*, *ābhīksṇya*, *āvaśyaka*, *atisarga*, *nimitta*, *auciwyā*, *prāptakāla*, *praiṣya*, *sāmarthya*, *yogyatā* etc.

There is no one to one mapping between a *kṛt* suffix and the semantic relation. Each of these relations may correspond to more than one *kṛt* suffixes and a *kṛt* suffix may represent more than one semantic meanings. For example the suffix *ghañ* has more than one meanings as shown in table1. Similarly there may correspond more than one suffixes which express the semantic relation as shown in table3.

Though on the face of it the broad meaning is same in all these cases, however, we observe that these words are not replacable and hence do not form a synset. Therefore we represent this data as shown in the following table.

Since *kṛt* suffixes derive nouns from verbs, we now link these nouns with the synsets defined using *amarakośa*.

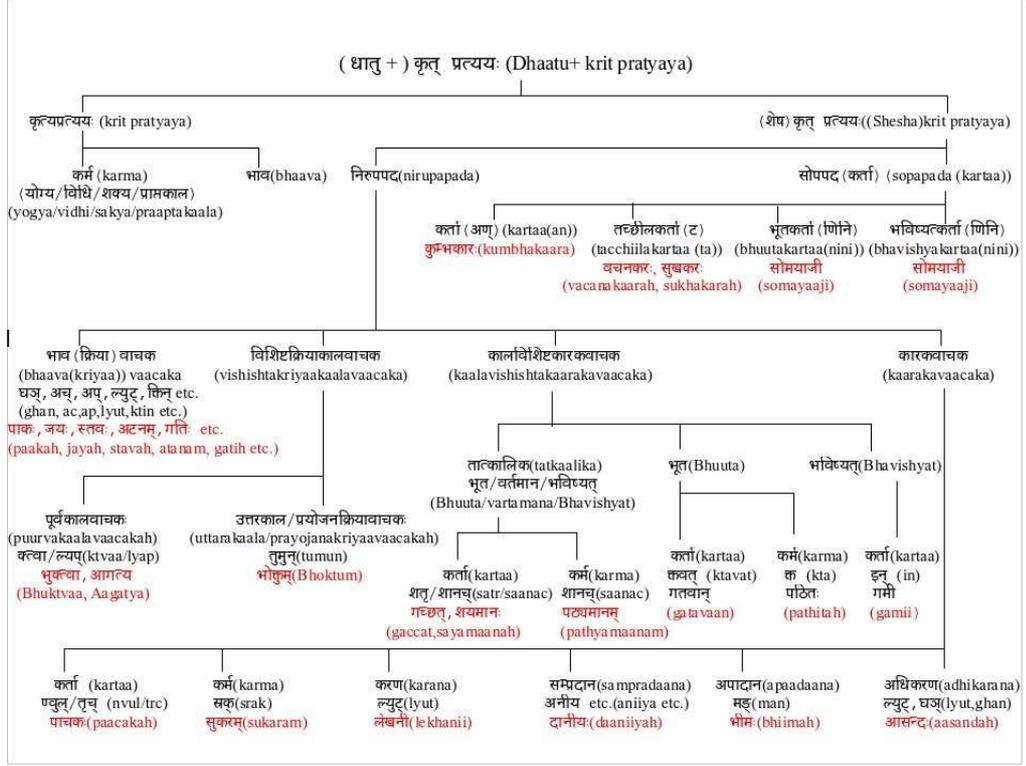


Figure 2:

<i>pac</i>	<i>kartari</i>	<i>pac + anīyar</i>	<i>pacanīya</i>
<i>pac</i>	<i>kartari</i>	<i>pac + ktavatu</i>	<i>paktavat</i>
<i>pac</i>	<i>kartari</i>	<i>pac + nvul</i>	<i>pācaka</i>
<i>pac</i>	<i>kartari</i>	<i>pac + nyat</i>	<i>pācya</i>
<i>pac</i>	<i>kartari</i>	<i>pac + śatṛ</i>	<i>pacat</i>
<i>pac</i>	<i>kartari</i>	<i>pac + tavyat</i>	<i>paktavya</i>
<i>pac</i>	<i>kartari</i>	<i>pac + tqc</i>	<i>paktṛ</i>

Table 2: *Kartari* meanings of *pac dhātu*.

<i>anīt dhātu</i>	<i>it dhātu</i>	<i>kṛt pratyaya</i>	<i>artha</i>	<i>rūpa</i>
<i>bhū1</i>	<i>bhū</i>	<i>anīyar</i>	<i>kartari</i>	<i>bhavanīya</i>
<i>bhū1</i>	<i>bhū</i>	<i>anīyar</i>	<i>bhāve</i>	<i>bhavanīya</i>
<i>bhū1</i>	<i>bhū</i>	<i>anīyar</i>	<i>yogyatā</i>	<i>bhavanīya</i>
<i>bhū1</i>	<i>bhū</i>	<i>anīyar</i>	<i>prāptakāle</i>	<i>bhavanīya</i>

Table 3: Sample entry in the database

A synset in *amarakośa* is the set of words that having same meaning spread over one or more *padas*, and *ślokas*. Typically this meaning is captured by the commentators of *amarakośa* by a single word which we call a head word. Example as shown in the following table4.

Word	Chapter-Varga-Verse-Line	Gender	Class	Synset-id-word
<i>amara</i>	1.1.7.1	puM.	<i>svargavarga</i>	<i>svarga</i>

Table 4: Sample entry in the *amarakośa* database

The information about to which class (*varga*) these words belong to is also unique and gives us a broad ontological classification. We link the *kṛt* words with *amarakośa* words, showing the synsets these *kṛt* words belong to figure2 consider the root *bhū* and its forms in the meaning of *bhāve*. There are eight words derived form *bhū* denoting the meaning *bhāve* as shown below. The automatic linking of the words with the *amarakośa* leads to figure2.

5 Conclusion

The synsets based on the *amarakośa* provide a basic WordNet for nouns. Since Hindi and Sanskrit share a lot, the ontology from Hindi has been borrowed by linking the *amarakośa* synsets to the Hindi WordNet. Further the linking of verbs with the derived nouns and further to the *amarakośa* words makes it easy to navigate across the POS categories.

The root is **भू** and the meaning is **भावे**

रूपम् = (धातु + प्रत्यय)	अमरकोशं synsets
भवनीय = (भू+अनीयर्)	
भाव = (भू+घञ्)	अर्थ = विद्वान् भाव[1.7.12.1, पु.] नाट्यवर्गः
भूत = (भू+क्)	अर्थ = देवयोनि अप्सरसा[1.1.11.1, स्त्री-वह.] भूत[1.1.11.2, नपु.] गन्धर्वा[1.1.11.1, पु.] गुरुका[1.1.11.2, पु.] किनर्[1.1.11.1, पु.] मिशाचा[1.1.11.2, पु.] रक्षा[1.1.11.1, पु.] सिद्धा[1.1.11.2, पु.] विद्याधर्[1.1.11.1, पु.] यक्षा[1.1.11.1, पु.] स्वर्गवर्गः अर्थ = मातम् आसादिन्[3.1.104.2, वि.] भावित्[3.1.104.2, वि.] भूत्[3.1.104.2, वि.] लब्ध[3.1.104.2, वि.] प्राप्त[3.1.104.2, वि.] विन्न[3.1.104.2, वि.] विशेष्यनिम्नवर्गः
भूतवत् = (भू+कवत्)	
भवन = (भू+ल्यट्)	अर्थ = गृहम् गृह[2.2.4.2, नपु.] उदलसित[2.2.4.2, नपु.] गृह[2.2.4.2, नपु.] निकेतन[2.2.4.2, नपु.] सङ्गना[2.2.4.2, नपु.] वेश्मना[2.2.4.2, नपु.] आगात्[2.2.5.1, नपु.] आलया[2.2.5.2, पु.] भवना[2.2.5.1, नपु.] गृह[2.2.5.2, पु.] वह्[2.2.5.1, पु.] मन्दिर[2.2.5.1, पु.] निकाय्या[2.2.5.2, पु.] निलया[2.2.5.2, पु.] निशान्त[2.2.5.1, पु.] सदन[2.2.5.1, नपु.] वस्त्या[2.2.5.1, नपु.] पुत्र्यः
भव्यमान = (भू+शानच्)	
भावितव्य = (भू+तव्यत्)	
भव्य = (भू+यत्)	अर्थ = शुभम् भद्रा[1.4.25.2, नपु.] कल्याण[1.4.25.2, नपु.] मङ्गला[1.4.25.2, नपु.] शिवा[1.4.25.2, नपु.] शुभा[1.4.25.2, नपु.] स्वश्रेयसा[1.4.25.2, नपु.] भविका[1.4.26.1, नपु.] भावुका[1.4.26.1, नपु.] भव्या[1.4.26.1, नपु.] क्षमा[1.4.26.1, पु-नपु.] कुथला[1.4.26.1, नपु.] शस्ता[1.4.26.2, नपु.] कालवर्गः

[Test another root 2](#)

Figure 3: Snapshot of the web display

6 Acknowledgement

Authors thanks K V RamaKrishnamacharyulu for useful discussions at various stages of the work and for beautifully classifying the *krt* suffixes in figure2.

References

- [1] English WordNet <http://wordnet.princeton.edu/>
- [2] Agirre E. and Rigau G. (1996) “Word Sense Disambiguation using Conceptual Density”, COLING, Denmark
- [3] Fellbaum, Christiane (1999) “WordNet An Electronic Lexical Database” MIT Press, Massachusetts
- [4] Hindi wordNet:<http://www.cfilt.iitb.ac.in/wordnet/webhwn>
- [5] Kulkarni, Malhar and Bhattacharya Pushpak (May 15th - 17th 2008) “Verbal roots in the Sanskrit WordNet”, 2nd international Sanskrit Computational Linguistics Symposium, Brown University.
- [6] Marathi WordNet:<http://www.cfilt.iitb.ac.in/wordnet/webmwn>
- [7] Mohanty,S. Dasadhikary, K. P., Santi P. K., Nayak, S.N. (2002) “Making of Sanskrit WordNet”, Proceedings of the Int. Conference on Universal Knowledge and Language, 25-29 November 2002, Goa, India.
- [8] Raghuvira (1981) “Comprehensive English - Hindi Dictionary of Governmental and Educational Words and Phrases.” International Academy of Indian Culture, New Delhi
- [9] Sinha, Manish and Mahesh Reddy and Pushpak Bhattacharyya (2006) “An Approach towards Construction and Application of Multilingual Indo-WordNet” In: Global WordNet Conference
- [10] Tamil WordNet <http://www.languageinindia.com/march2002/rajendran3.html>.
- [11] Vossen, Piek (2002) “EuroWordNet” In: EuroWordNet Project-report”.

- [12] Bharati Akshar, Kulkarni Amba, and Nair Sivaja, "Use of Amarakosha and Hindi WordNet in Building a Network of Sanskrit Words." (2008)
In: 6th International Conference on Natural Language Processing.