The Knowledge Structure in Sanskrit Kośas

Amba Kulkarni
Department of Sanskrit Studies, University of Hyderabad, Hyderabad
apksh@uchyd.ernet.in

Sivaja S. Nair
Department of Sanskrit Studies, University of Hyderabad, Hyderabad
sivaja.s.nair@gmail.com

Abstract

The Sanskrit kośas such as Amarakośa, Vaijayanṭikośa etc. have been based on knowledge structure of its own which apart from revealing the ontological classification, provides a holistic view of various concepts. Knowledge in these kośas concerns with many non-observational, culture specific facts. In this paper we present a few representative examples of the concept clusters from the two Sanskrit kośas; Amarakośa and Vaijayanṭikośa. There is a necessity to make these valuable resources available in suitable e-form so that the NLP community working in Indian Languages can be benefitted.

Key Words : Amarakośa, Vaijayanṭikośa, ConceptNet, Semantic relations

1 Introduction

The Indian tradition of transmitting knowledge orally is on the verge of vanishing. As the oral transmission demands, Indian traditional educational culture was organised to be formal and intensive as opposed to the modern culture which is more informal and extensive (Wood, 1985). In traditional circumstances, a child would receive his education largely by oral transmission, mainly through rote-learning. The method employed was through recitation and remembering. A child is taught the alphabet (varamālā), he would memorise a few verses, subhāṣītas, and then start reciting a dictionary of synonymous words – the Amarakośa – till it is memorised. It typically would take anywhere between 6 months to a year to memorise a list of approximately 10,000 Sanskrit words arranged as a list of synonyms. The close inspection of the structure of the Amarakośa gives much more insight into the way the words are organised. When a student memorises it, though in the beginning it appears as a linear list of words, as he starts understanding the meaning of the words, reads the commentaries on this text and starts using these words, the linear structure unfolds into a knowledge web with various links.

Many other kośas such as Vaijayanṭikośa, Halāyudhakośa, Trikāṇḍesā, Hārīvalī, Abhidhānacintāmaṇī etc. reveal similar structure when explored.

The modern efforts of building ConceptNet aimed at building a network around various concepts, have some parallels with the structure of these Sanskrit kośas.

In this paper, we present a few samples of knowledge structure from two important kośas viz Vaijayanṭikośa and Amarakośa, and compare the knowledge structure involved with that of the ConceptNet.

2 Amarakośa

Amarakośa primarily named as Nāmaṁgānūśāsana (a work that deals with instructions related to the gender of nouns) is authored by Amarasimha - 4th century A.D. (Oka, 1981) - and is the most celebrated and authoritative ancient thesaurus of Sanskrit with around 60 commentaries and translations into modern Indian as well as foreign languages such as Chinese, Tibetan, French, etc. (Pathak, 1981). It is considered as an essential requisite for a Sanskrit scholar and as such a child is asked to memorise it even before he starts his studies formally. It consists of 1608 verses composed in
anustup meter\(^1\) and are divided into 3 chapters called Kāṇḍas.\(^2\)

### 2.0.1 Classification

Each of the three Kāṇḍas is further subdivided into various vargas. The classification of three kāṇḍas into 25 vargas is as below.

- **Prathamakāṇḍam**: svargavargah (heaven) vyomavargah (sky) dīgavargah (direction) kālavargah (time) dhīvargah (cognition) sābdādvargah (sound) nāṭyavargah (drama) pīṭālabhogivargah (nether world) narakavargah (hell) vārīvargah (water)

- **Dvītyakāṇḍam**: bhūmivargah (earth) puravargah (towns or cities) sañcārvargah (mountains) vanusadāvargah (forests and medicines) simhādvargah (lions and other animals) manusyāvargah (mankind) brahmavargah (priest tribe) kṣatriyāvargah (military tribe) vaiśāyavargah (business tribe) sūdravargah (mixed classes)

- **Trīyakāṇḍam**: viśesyāṅghnavargah (adjective) samkārvargah (miscellaneous) nāmārthavargah (polysemous) avyayavargah (indeclinables) lingādisangrahavargah (gender)

Amarakośa contains 11,580 content words (tokens). Some of the tokens are repeated either within a kāṇḍa or across the kāṇḍas leading to only 9,031 types. The kāṇḍa-wise distribution of the tokens and types is shown in Table 1. The organisation of words is typically in the form of a set of synonymous words.

\(^1\)Sloko suṣṭam gurum jītyam vanavam laghu paścaṃam
\(^2\)Dvicaśāṃ padāyorhinnavam sapamam dhīgamamānyayoh
\(^\text{Table 1: Tokens and types in each kāṇḍas}\)

<table>
<thead>
<tr>
<th>kāṇḍa</th>
<th>tokens</th>
<th>types</th>
</tr>
</thead>
<tbody>
<tr>
<td>prathamakāṇḍam</td>
<td>2465</td>
<td>2300</td>
</tr>
<tr>
<td>dvītyakāṇḍam</td>
<td>5827</td>
<td>5282</td>
</tr>
<tr>
<td>trīyakāṇḍam</td>
<td>3288</td>
<td>2271</td>
</tr>
</tbody>
</table>

3 **Vaijayantikōsa**

Vaijayantikōsa is a voluminous lexicon by Yādavaprakāśa. Approximately 18,000 tokens are there in Vaijayantikōsa. The lexicon is divided into two broad divisions, viz. synonym sets and polysemous words. The synonym sets are further divided into five classes or kāṇḍa's, viz., svarga (heaven), antarikṣa (sky), bhūmi (earth), pīṭāla (nether world) and sāmānya (miscellaneous). The polysemous words are classified into three classes based on the number of syllables they contain, viz., two, three and more than three. Thus Vaijayantikōsa has eight classes which are further subdivided into two or more sub-sections called adhyāyās. There are total forty three in number. The classification of first five classes are shown in below.

- **Svargakāṇḍaḥ (heaven)**
  Ādidevādhyāyaḥ (supreme deity)
  Lokapālādhyāyaḥ (guardian deities)
  Yakṣādhyāyaḥ (semi-divine beings)

- **Antarikṣakāṇḍaḥ (sky)**
  Jyotirādhyāyaḥ (light)
  Meghādhyāyaḥ (cloud)
  Khagādhyāyaḥ (bird)
  Śabdādhyāyaḥ (sound)

- **Būmikāṇḍaḥ (earth)**
  Desādhyāyaḥ (place)
  Śaivaḥdhyāyaḥ (hill)
  Vanādhyāyaḥ (forest)
  Paśuṣangrhādhyāyaḥ (animals)
  Manusyādhyāyaḥ (mankind)
  Brāhmaṇādhyāyaḥ (priest tribe)
  Kṣatriyādhyāyaḥ (military tribe)
  Vaiśāyādhyāyaḥ (business tribe)
  Śūdrādhyāyaḥ (mixed class)
- **Pātālakāṇḍaḥ (nether world)**
  - Sarśrpādhyāyah (serpent)
  - Jalādhyāyah (water)
  - Purādhyāyah (town & cities)
  - Bātādhyāyah (living being)

- **Sāmānyakāṇḍaḥ (miscellaneous)**
  - Gāndhyāyah (multiplicity)
  - Dharmakarmādhyāyah (natural character)
  - Gūndhyāyah (quality)

4 ConceptNet

ConceptNet\(^3\) is a commonsense knowledgebase and natural-language-processing toolkit. It is a semantic network of commonsense knowledge. It aims to give computer an access to the commonsense knowledge - the kind of information that ordinary people know but usually leave unstated. ConceptNet is generated automatically from the English sentences of the Open Mind Common Sense (OMCS) corpus. Fig.1 shows the ConceptNet representation of the sentence “wake up in the morning and drink coffee”.

![Figure 1: Knowledge representation in ConceptNet](image)

5 Knowledge structure in Sanskrit Kośa

After examining the entries in Amarakośa and Vaijayanāthakośa, we noticed that except the polysemous words (nānāsthavarga), all other synsets in a class show some semantic relation to the class it belongs to and sometimes even to the preceding or following synsets. These semantic relations indicate various kinds of relations. They may be classified as hierarchical or associative. The hyponym indicating a more general term or the hypernym showing a more specific term are the examples of hierarchical relation. Similarly the holonym-meronym relation marking the whole-part relation is also a hierarchical relation. In addition various other relations are indicated by the adjacency of the synsets. These may be termed as associative relations, which indicate some kind of association of one synset with the other. This association may be the association among human beings, or the association of certain objects with certain other objects. We illustrate below some such relations with examples.

5.1 Example 1: Viṣṇu from Amarakośa

The verses from 1.1.18 to 1.1.29 describe various synsets representing Viṣṇu, and objects related to/associated with Viṣṇu. (See Figure 2).

![Figure 2: Relations of Viṣṇu from Amarakośa](image)

Thus we see here, the association of synsets indicate various kinship terms such as father, sibling, wife, son etc. Knowledge of such kinship terms is an essential data base for understanding various texts on our religion. Next we see various special instruments that are associated/used by Viṣṇu such as conch, discus, sword, jewel, bow, etc., then the charioteer, vehicle, ministers etc. Thus all the synsets together give a holistic view of the concept under consideration.

5.2 Example 1: Viṣṇu from Vaijayanāthakośa

In Vaijayanāthakośa, we see a few more links associated with the concept of Viṣṇu such as nine Viṣṇu Śaktis, incarnations of Viṣṇu, and also the instruments used by Balarāma - the brother of Kṛṣṇa - the 8th incarnation of Viṣṇu, etc. (See Fig. 3).

\(^3\)http://web.media.mit.edu/hugo/conceptnet/
5.4 Example 2: Samayaḥ from Vaijayantikaṣa

In Vaijayantikaṣa, we also see various other classifications such as pākṣa, fortnight before and after the new moon, and the associated month names from the verses 2.1.79 to 2.1.81.

- half of a lunar month (pākṣa) (2.1.79)
- first (white) pākṣa of a lunar month (2.1.79)
- second (black) pākṣa of a lunar month (2.1.79)
- pākṣa in which the tithis become shorter (2.1.79)
- pākṣa in which the tithis become longer (2.1.79)
- short fifteenth day of the second (kṛṣṇa)pākṣa (2.1.80)
- short first day of a pākṣa (2.1.80)
- lunar month (2.1.80)
- solar month (month in which the sun passes to another rāśi) (2.1.80)
- star month (2.1.81)
- month of 30 days (2.1.81)

Thus we see that these structures provide complete picture of the Indian calendar which is an essential part of Indian culture.

5.5 Example 3: Kṣatriyaḥ from Amarakoṣa

Here is a group of verses from 2.8.1 to 2.8.10 belonging to the kṣatriyavarga, from the Amarakoṣa. The words here refer to the king, military, ministers, various category of people engaged in the services of kings, etc.

- Man of the military tribe (2.8.1)
  - King (2.8.1)
    - Universal monarch (2.8.2)
    - An emperor (2.8.2)
    - King over a country (2.8.2)
    - Paramount sovereign (2.8.3)
    - Multitude of kings (2.8.3)
    - Multitude of military tribe (2.8.4)
    - Minister (2.8.4)
      - Deputy minister (2.8.4)
        - Priest (2.8.5)
        - Judge (2.8.5)
        - King’s companions (2.8.5)
        - Body guards of a king (2.8.6)
        - Warder (2.8.6)
Superintendent (2.8.6)
Village Superintendent (2.8.7)
Superintendent of many villages (2.8.7)
Superintendent of Gold (2.8.7)
Superintendent of Silver (2.8.7)
Superintendent of the women’s apartments (2.8.8)
Outside guard of the women’s apartment (2.8.8)
attendant of a king (2.8.9)
eunuch (2.8.9)
Prince whose territories lie on the frontiers of those of the enemy (2.8.9)
Neighboring prince (2.8.9)
Prince whose territories lie beyond those of the friend (2.8.10)
Enemy in the rear (2.8.10)

This again gives a good background of the military structure in ancient days, throwing light on the social structure in those days.

5.6 Implicit relations
These were three samples from three distinct topics involving totally different kind of relations. All these relations are semantic in nature. A more detailed study of such examples showed that following relations occur more frequently.

- **avayavana** (part-whole relation)
- **paraparajati** (is a kind of relation)
- **janyajanaka** (child-parent relation)
- **patipati** (husband-wife relation)
- **svasvami** (master-possession relation)
- **ajivika** (livelihood)

There are a few other relations such as ādāhāra-ādheya, vamsa-vamsiya, etc.

6 ConceptNet and the Sanskrit Kośas

There is a fundamental difference between the development of ConceptNet and the knowledge found in the Sanskrit Kośas. The ConceptNet is aimed at ‘capturing the common sense’. This common sense typically concerns with the behavioural observations, social norms etc. and also focuses typically on actions with which other concept nodes are related.

The Sanskrit kośas on the other hand contain words which are:
(a) culture specific (svarga, viṣṇu, etc.),
(b) reveal social or man made structures (kingdom, house, etc.),
(c) throw light on various social practices, in addition to the observational facts such as classification of animals, the flora and fauna etc.

This structure thus involves many non-observational facts, which are mostly part of our culture. In this sense the structure of these kośas is complementary to what the ConceptNet provides and is an essential part of the Indian culture.

There is a necessity to make these valuable resources available in suitable e-form so that the NLP community working in Indian Languages can be benefitted.

7 Computational Tools

The team at the Department of Sanskrit Studies, University of Hyderabad, has already taken a lead by starting a pilot study of Amarakośa. There is a need to explore and build e-structures with other Sanskrit Kośas mentioned above in order to make this knowledge available for NLP applications.

The e-version of Amarakośa is available with various structured queries at http://sanskrit.uohyd.ernet.in/ \~ anusaaraka/sanskrit/samsaadhanii/amarakosa/home.html.

References


Colebrooke H.T (1808) “Kośa or Dictionary of the sanskritia language by Umura singha with an English interpretations and annotations” Serampore.


Pandit, Sivadatta (1915) “Namalinganusasana (Amarakosha) With the commentary of (Vyakhyasudha or Ramasrami) of Bhanuji Dikshit” Tukaram javaji proprietor of the “Nirmaya Sagar” Press, Bombay.