The Knowledge Structure in Sanskrit Kośas

Amba Kulkarni
Department of Sanskrit Studies,
University of Hyderabad,
Hyderabad
apksh@uohyd.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, Vaijayantikośa etc. have a built in 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 Vaijayantīkośa. 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 benifitted.

Key Words : Amarakośa, Vaijayantīkośa, 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 (varṇamālā), he would memorise a few verses, subhāṣitas, 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 Vaijayantīkośa, Halāyudhakośa, Trikāṇḍśeṣa, Hārāvalī, 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 Vaijayantīkośa and Amarakośa, and compare the knowledge structure involved with that of the ConceptNet.

2 Amarakośa

Amarakośa primarily named as *Nāmaliṅgānuśāsana* (a work that deals with instructions related to the gender of nouns) is authored by Amarasiṃha - 4^{th} 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. (Patkar, 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

anuṣṭup 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*:**
  svargavargaḥ (heaven)
  vyomavargaḥ (sky)
  digvargaḥ (direction)
  kālavargaḥ (time)
  dhīvargaḥ (cognition)
  śabdādivargaḥ (sound)
  nāṭyavargaḥ (drama)
  pātālabhogivargaḥ (nether world)
  narakavargaḥ (hell)
  vārivargaḥ (water)

- ***Dvitīyakāṇḍam*:**
  bhūmivargaḥ (earth)
  puravargaḥ (towns or cities)
  śailavargaḥ (mountains)
  vanauṣadhivargaḥ (forests and medicines)
  siṃhādivargaḥ (lions and other animals)
  manuṣyavargaḥ (mankind)
  brahmavargaḥ (priest tribe)
  kṣatriyavargaḥ (military tribe)
  vaiśyavargaḥ (business tribe)
  śūdravargaḥ (mixed classes)

- ***Tṛtīyakāṇḍam*:**
  viśeṣyanighnavargaḥ (adjective)
  saṁkīrṇavargaḥ (miscellaneous)
  nānārthavargaḥ (polysemous)
  avyayavargaḥ (indeclinables)
  liṅgādisaṅgrahavargaḥ (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] Śloke ṣaṣṭaṃ guruṃ jñeyaṃ sarvatra laghu pañcamam |
Dvicatuḥpādayorhrasvaṃ saptamaṃ dīrghamanyayoḥ ||

[2] and as such is known as *Trikāṇḍī*

kāṇḍa	tokens	types
prathamakāṇḍam	2465	2300
dvitīyakāṇḍam	5827	5282
tṛtīyakāṇḍam	3288	2271

Table 1: Tokens and types in each kāṇḍas

3 Vaijayantīkośa

Vaijayantīkośa is a voluminous lexicon by Yādavaprakāśa. Approximately 18,000 tokens are there in Vaijayantīkośa.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), *antarīkṣa* (sky), *bhūmi* (earth), *pātāla* (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 Vaijayantīkośa has eight classes which are further sub-divided 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 diety)
  Lokapālādhyāyaḥ (guardian deities)
  Yakṣādhyāyaḥ (semi-divine beings)

- ***Antarikṣakakāṇḍaḥ (sky)***
  Jyotiradhyāyaḥ (light)
  Meghādhyāyaḥ (cloud)
  Khagādhyāyaḥ (bird)
  Śabdādhyāyaḥ (sound)

- ***Būmikāṇḍaḥ (earth)***
  Deśādhyāyaḥ (place)
  Śailādhyāyaḥ (hill)
  Vanādhyāyaḥ (forest)
  Paśusaṅgrahādhyāyaḥ (animals)
  Manuṣyādhyāyaḥ (mankind)
  Brāhmaṇādhyāyaḥ (priest tribe)
  Kṣatriyādhyāyaḥ (military tribe)
  Vaiśyādhyāyaḥ (bussiness tribe)
  Śūdrādhyāyaḥ (mixed class)

- ***Pātālakāṇḍaḥ (nether world)***
  Sarīsṛpādhyāyaḥ (serpent)
  Jalādhyāyaḥ (water)
  Purādhyāyaḥ (town & cities)
  Būtādhyāyaḥ (living being)

- ***Sāmānyakāṇḍaḥ (miscellaneous)***
  Gaṇādhyāyaḥ (multitude)
  Dharmakarmādhyāyaḥ (natural character)
  Guṇādhyāyaḥ (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".

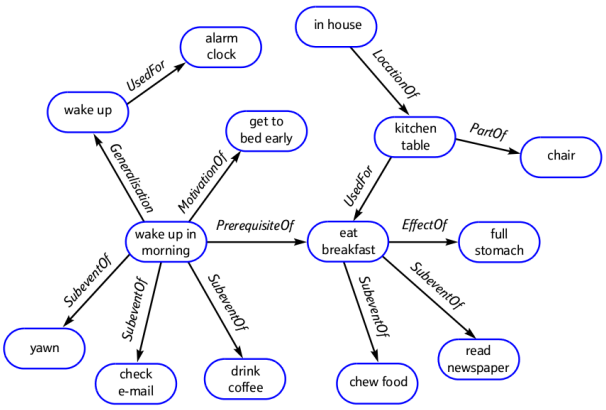

Figure 1: Knowledge representation in ConceptNet

5 Knowledge structure in Sanskrit Kośas

After examining the entries in Amarakośa and Vaijayantīkośa, we noticed that except the polysemous words (nānārthavarga), 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 hypernym indicating a more general term or the hyponym 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: Viṣṇuḥ 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).

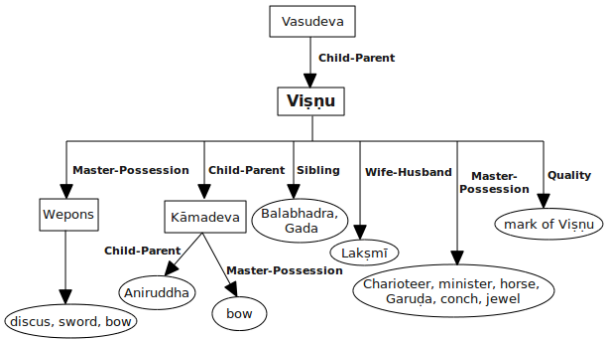

Figure 2: Relations of Viṣṇu from Amarakośa

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: Viṣṇuḥ from Vaijayantīkośa

In Vaijayantīkośa, 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 8^{th} incarnation of Viṣṇu, etc. (See Fig. 3).

[3] http://web.media.mit.edu/~hugo/conceptnet/

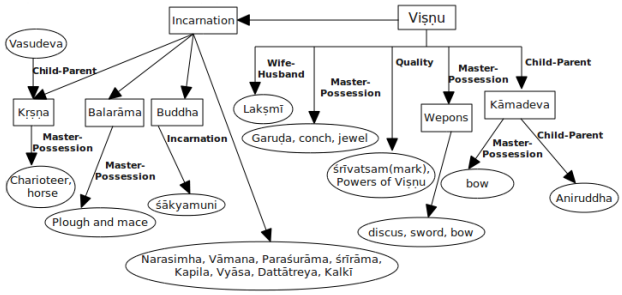

Figure 3: Relations of Viṣṇu from Vaijayantīkośa

5.3 Example 2: Samayaḥ from Amarakośa

The verses from 1.4.1 to 1.4.9 in Amarakośa deal with words related to time, units of measurement, special names of special days, etc.

Time (1.4.1)
Lunar day (1.4.1)
First lunar day (1.4.1)
{Day (1.4.2)
Morning (1.4.2 - 1.4.3)
Twilight (1.4.3)
Evening (1.4.3)
First four hours of a day (1.4.3)
Second four hours of a day (1.4.3)
Third four hours of a day (1.4.3)
Period of the day (1.4.3)
Night (1.4.3 - 1.4.4)
A dark night (1.4.5)
A moonlight night (1.4.5)
A night and two days (1.4.5)
First part of night (1.4.6)
Midnight (1.4.6)
Sequence of nights (1.4.6)
Space of three hours (1.4.6) }
Last day of the half month (1.4.7)
Precise moment of the full or the new moon (1.4.7)
Full moon day (1.4.7)
Full moon whole day(1.4.8)
Full Moon with a little gibbous on part of a day (1.4.8)
No moon day (1.4.8)
wanning crescent (1.4.9)
No moon whole day (1.4.9)

In this example, we see various concepts, in Indian tradition, associated with the concept of time. These may be broadly classified into the concepts associated with the apparent solar motion and those associated with the lunar motion.

5.4 Example 2: Samayaḥ from Vaijayantīkośa

In Vaijayantīkośa, we also see various other classifications such as pakṣa, 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 (*pakṣa*) (2.1.79)
first (white) *pakṣa* of a lunar month (2.1.79)
second (black) *pakṣa* of a lunar month (2.1.79)
pakṣa in which the *tithis* become shorter (2.1.79)
pakṣa in which the *tithis* become longer (2.1.79)
short fifteenth day of the second *(kṛṣṇa)pakṣa* (2.1.80)
short first day of a *pakṣa* (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 Amarakośa

Here is a group of verses from 2.8.1 to 2.8.10 belonging to the kṣatriyavarga, from the Amarakośa. 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 womens' appartments (2.8.8)
Outside guard of the womens' appartment (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 hight 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.

- *avayavāvayavī* (part-whole relation)
- *parāparājāti* (is a kind of relation)
- *janyajanaka* (child-parent relation)
- *patipatnī* (husband-wife relation)
- *svasvāmi* (master-possession relation)
- *ājīvikā* (livelihood)

There are a few other relations such as ādhāra-ādheya, vaṃśa-vaṃśīya, 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 behaviourial 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 benifitted.

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/amarakosha/home.html.`

References

Bharati Akshar, Kulkarni Amba and Nair S. Sivaja (2008) "Use of Amarakosha and Hindi WordNet in Building a Network of Sanskrit Words" International Conference On Natural Language Processing, C-DAC Pune.

Colebrooke H.T (1808) "Kosha or Dictionary of the sungskrita language by Umura singha with an English interpretations and annotations" Serampore.

Gustav, Oppert (1893) "The Vaijayantī of Yādavaprakāśa" Madras Sanskrit and vernacular text publication society, Madras.

Liu, H and Singh P (2004) “ConceptNet - a practical commonsense reasoning tool-kit”, BT Technology Journal Vol 22 No 4. October

Liu, H and Singh P (2004) “Commonsense Reasoning in and over Natural Language”, Proceedings of the 8th International Conference on Knowledge-Based Intelligent Information and Engineering Systems, Wellington, New Zealand.

Moosath, Parameshvaran T.C. (1914) “Paarameswarii : malayalam commentory of Amarakosha” National Book Stall, Kottayam.

Moosath, Parameshvaran T.C.(1956) “Triveni : malayalam commentory of Amarakosha” National Book Stall, Kottayam.

Nair S. Sivaja, Swain Pritilaxmi and Kulkarni Amba (2009) “Developing network of Sanskrit words across Part-Of-Speech categories”, CSATS’09, Rashtriya Sanskrit Vidhyapeeth, Tirupati.

Oka Krishnaji Govind (1981) “The Namalinganushasana : Amarakosa of Amarasimha with the commentary (Amarakoshodghaatana) of Ksheraswamin”, Upasana prakashan Delhi, Varanasi.

Patkar M.M (1981) “History of Sanskrit Lexicography” Munshiram Manoharlal Publishers Pvt. Ltd., Delhi.

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