

Discourse Analysis of Sanskrit texts

Amba Kulkarni and Monali Das
Department of Sanskrit Studies, University of Hyderabad
apksh@uohyd.ernet.in, monactc.85@gmail.com

ABSTRACT

The last decade has seen rigorous activities in the field of Sanskrit computational linguistics pertaining to word level and sentence level analysis. In this paper we point out the need of special treatment for Sanskrit at discourse level owing to specific trends in Sanskrit in the production of its literature ranging over two millennia. We present a tagset for inter-sentential analysis followed by a brief account of discourse level relations accounting the sub-topic and topic level analysis, as discussed in the Indian literature illustrating an application of these to a text in the domain of vyākaraṇa (grammar).

KEYWORDS: Discourse analysis, Sanskrit, Saṅgati.

1 Introduction

Sanskrit, the classical language of India, has a huge collection of literature in almost all branches of knowledge – astronomy, mathematics, logic, philosophy, medicine, technology, dramatics, literature, poetics – to name a few. It was the medium of communications for all serious discourses and scholarly communications till recent times. This resulted in continuous production of literature in Sanskrit in various branches of knowledge systems of human endeavour for almost over 2 millennia. The total corpus of Sanskrit is almost 100 times those in Greek and Latin put together. However, the picture changed completely in the last two centuries. The traditional learning methods are replaced by the Western learning systems. As a consequence the knowledge in Sanskrit texts is inaccessible to the modern Indian scholars.

The main reasons behind the difficulty in accessing Sanskrit texts are:

- Sanskrit is influenced by the oral tradition, and hence the Sanskrit texts are continuous strings of characters without any punctuation marks or word or sentence boundaries. The characters at the juncture of boundary undergo euphonic changes making it difficult to 'guess' the boundaries.
- Sanskrit is very rich in morphology and is inflectional. This also makes it difficult to remember various inflections of a word, which differ with the last character of the word and its gender.
- Though a substantial vocabulary in modern Indian languages is from Sanskrit, there have been cases of meaning shifts, meaning expansion and meaning reduction. This makes it difficult for an Indian to understand the Sanskrit texts faithfully, unless he knows the original meaning of the words.
- Another important aspect is the presentation of texts. There are various trends in the Sanskrit literature. One of them is that of nested commentaries. The original text which is in a cryptic sūtra form, is commented upon by later scholars for more clarification. In order to clarify a content in this commentary another commentary would follow, and this continues leading to nested commentaries (see Appendix A for an example). Since the modern scholars are trained in modern learning methodologies they find it difficult to get familiar with the structure and decide the boundaries of various topics and sub topics, and thereby understand the texts.

1.1 Discourse Analysis in Indian Grammatical Tradition

The rich tradition of linguistics in India is more than two millennia old. Pāṇini's (around 500 B.C.) contribution to the grammar is as important a milestone in the development as that of Euclid in case of development of geometry in Europe (Staal, 1965). The discussions on the problem of meaning and the process of understanding the texts by philosophers like Bhartṛhari, Gaṅgeśa, and Kumārilabhaṭṭa resulted into three distinct schools of thought. With an aim to understanding the Vedas these schools developed the theories of verbal cognition - *Śābdabodha*. These schools differ mainly in the chief qualificand of the cognition, however more or less they agree on various other relations at gross level. These three schools are *Vyākaraṇa* (Grammar), *Nyāya* (Logic) and *Mīmāṃsā* (Exegesis). Grammarians deal with the syntactic analysis to a considerable depth. Logicians and Mīmāṃsakas discuss various constraints such as *Akāṅkṣā* (expectancy), *Yogyatā* (mutual compatibility) and *Sannidhi* (proximity) to filter out nonsensical

analysis. In his seminal attempt to describe the relation between word and its meaning, a logician Gadādhara (Śāstri, 1929) has provided the meanings of various pronouns and rules for tracing their antecedents. The Mīmāṃsakas further discuss various types of discourse relations called *Saṅgatis* for checking the consistency and coherence of the text. The coherence is tested at various levels viz.

- (a) Śāstra saṅgati : The coherence at the level of the subject.
- (b) Adhyāya saṅgati : The coherence at the level of a chapter or a book.
- (c) Pāda saṅgati : The coherence at the level of a section.
- (d) Adhikaraṇa saṅgati : The coherence at the level of a topic.

Each topic can further have one or more sub-topics, each sub-topic can have one or more paragraphs and a paragraph may consist of one or more sentences. Thus the topic level analysis involves following steps:

- (i) Sentential analysis : Establishing relations among words in a sentence.
- (ii) Paragraph level analysis : Identifying inter-sentential relations based on either explicit or implicit connectives.
- (iii) Sub-topic level analysis : Establishing relations between the successive paragraphs showing the consistency of the argument leading to a sub-topic.
- (iv) Topic level analysis : Topic level analysis shows the relevance of each sub-topic towards the goal of the main topic and thus the coherence.

1.2 Computational Discourse Analysis

With the emergence of computational linguistics, it is now possible to build tools which can assist a scholar in accessing Sanskrit texts, reducing his learning time. The Computational Linguistic tools are centered around the Western Linguistic theories and hence remain suited for English and other European Languages. Sanskrit is morphologically rich and is dominated by oral tradition. This results in Sanskrit text as a continuous string of characters, merging not only the word boundaries but sometimes the sentence boundaries as well. This therefore poses a big challenge to the computational processing of Sanskrit texts, requiring new innovative methods to handle segmentation taking into account euphonic changes effectively. As a result we see that much of the Sanskrit computational work is still at the level of word analysis and segmentation (Huet, 2009; Hellwig, 2009; Kumar et al., 2010). The rich inflectional morphology further makes the constituency parsers inappropriate for syntactic analysis of a sentence. While for positional languages such as English, the information of the relation between words is coded in positions and hence the constituent structures makes sense, for inflectionally rich languages like Sanskrit, the information of the relation is in the inflectional suffixes, which in turn allows for flexible word order, and thereby the dependency structure is more appropriate to represent the semantics expressed through the suffixes. A full fledged constraint parser using the concepts of Ākāṅkṣā (expectancy) and Sannidhi (proximity) has been developed by (Kulkarni et al., 2010). This parser handles some inter-sentential relations as well, and the work on anaphora resolution has just begun. Thus the work on discourse analysis for Sanskrit is yet in its infancy.

On the other hand we see major efforts at the level of discourse analysis in English and other European languages. Halliday and Hasan (1976) articulated the discourse theory and discussed about cohesion in discourse. Two main discourse structures were proposed viz. tree structure (Mann and Thompson, 1987) and graph (Wolf and Gibson, 2005). The prominent discourse theories are Rhetorical Structure Theory (RST), Linguistics Discourse Model (LDM), Discourse GraphBank (DG), and Discourse-Lexicalized Tree Adjoining Grammar (D-LTAG).

RST (Mann and Thompson, 1987) associates discourse relation with discourse structure. Here discourse units relate two adjacent units by discourse relations. In RST the proposed structure is a tree. Discourse structure is modelled by schemas where leaves are elementary discourse units – non-overlapping text spans and discourse relation holds between daughters of the same non-terminal node.

LDM (Polanyi, 1988) deals with discourse structure in the form of a tree. It differs from RST in distinguishing discourse structure from discourse interpretation. The discourse structure comes from the context free rules i.e. parent is interpreted as the interpretation of its children and the relationship between them.

In DG (Wolf and Gibson, 2005) discourse units are related to both adjacent and non-adjacent units. It was observed that crossing dependencies and nodes with multiple parents appear in texts vastly while RST does not allow these. In order to overcome these problems, graph representation was proposed by DG.

D-LTAG (Webber et al., 2001) builds on the observation that discourse connectives have both the syntactic as well as semantic function in the discourse. It considers discourse relations triggered by lexical elements. In D-LTAG, the predicates (verbs) are discourse connectives.

Webber and Joshi further proposed a tagset (Webber and Joshi, 2012) for annotating a corpus for discourse. This tagset is used to annotate the Penn Discourse Treebank. This tagset is neutral and does not make any assumptions about the form of the overall discourse structure of text. In addition to marking the arguments for both explicit as well as implicit connectives, it also marks senses and attribution of each discourse connective.

In the recent years there have also been efforts to deal with the coherence at the level of topic (Webber, 2006; Webber and Joshi, 2012).

All these computational models for discourse analysis are centered around English and other European languages. They are not appropriate to handle morphologically rich and more or less free word order language like Sanskrit with a special discourse structure of scientific and philosophical texts. Further, India has a strong grammatical tradition. So it is natural to look at this tradition for building computational models rather than trying to 'fit in' available models for Sanskrit.

In this paper we present a framework for discourse analysis in Sanskrit. The second section presents a brief report on the set of relations used for developing a Dependency Tree bank of Sanskrit corpus. The third section lists various inter-sentential relations for paragraph level analysis, discussed in Sanskrit literature. The fourth section provides a brief report on the *Saṅgatis* (relations) needed for analysing the inter-relations between paragraphs describing the same sub-topic. The fifth section lists the *Saṅgatis* used by the Indian logicians to establish the coherence and then we illustrate with an example how these *Saṅgatis* are useful for proper understanding of a text. Then we give a brief outline of three major trends in the production of scientific literature, and the current status of Sanskrit computational tools.

2 Sentence Level Analysis

In the traditional learning schools, the sentence level analysis is introduced at a tender age of 9 or 10 immediately after the students have memorized Śabdārūpa (noun-word forms), dhātupāṭha (verbal forms) and Amarakośa (a thesaurus). Then the students are taught one chapter of Raghuvamśa of Kālidāsa to imbibe in them the methodology of analysing the text. There are two prominent approaches viz Daṇḍānvaya (also known as anvayamukhī) and Khaṇḍānvaya (also known as kathambhūtinī). In the first approach the teacher arranges all the words in prose order. In the second approach, on the other hand, the teacher gives the basic skeleton of a sentence and fills in other details by asking questions.¹ These questions are centered around the heads seeking their various modifiers. This later method of analysis is more close to the modern dependency parsing credited to (Tesnière, 1959). The dependency relations in Sanskrit have been proposed and thoroughly examined by the generations of scholars over a period of more than 2 millennia. Thus we are fortunate to have a well defined, time tested tagset for Sanskrit, unlike other languages such as English where special efforts were put in as described in PARC (King et al., 2003), Stanford dependency manual (M. Marneffe and Manning, 2006) etc. for defining the set of relations. Various relations described in the traditional grammar books have been compiled and classified by (Ramakrishnamacharyulu, 2009) under the two broad headings viz. inter- sentential and intra-sentential relations. This work provided a starting point for developing guidelines (Ramakrishnamacharyulu et al., 2011) for annotation of Sanskrit texts at kāraka (syntactico-semantic relations) level and also for the development of an automatic parser for Sanskrit. This tagset was further examined for the appropriateness of the granularity (Kulkarni and Ramakrishnamacharyulu, 2013). And a set of 31 relations were selected from among the 90 relations proposed in the original proposal. The reduction in the number of rules was to avoid the fine-grain distinction involving extra-linguistic knowledge. A constraint based parser² is developed to parse the Sanskrit sentences using these relations. A dependency tree bank of around 30K words is also annotated using this scheme.

3 Paragraph Level Analysis

The relations in the tag-set proposed by (Ramakrishnamacharyulu, 2009) contain intra-sentential relations as well. Some of the connectives connecting two sentences are single while most of them are parallel connectives or pairs. Each of these connectives takes two arguments. The relations are binary in nature except those indicated by the conjunctive and disjunctive particles. We follow the naiyāyikas (Indian Logicians) canonical form to represent the relations. In a sentence 'Rama sleeps', Rama is the agent of an activity of sleeping. This is represented as in Figure 1.



Figure 1: Convention for labelling relations

Note the direction of the arrowhead. This is interpreted as 'Rama' has an agent-hood conditioned/determined by an activity of sleeping.

In case of inter-sentential connectives, the two arguments, following logicians convention again,

¹A very good illustration of these approaches is given in Tubb and Boose (2007).

²<http://sanskrit.uohyd.ernet.in/scl/SHMT/shmt.html>

are named by the general terms *anuyogika*³ (combining) and *pratiyogī* (having a counter part). So, if C is the connective connecting two sentences S1 and S2 then the general structure is represented as in Figure 2.

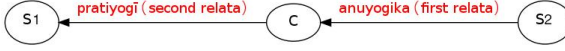


Figure 2: Discourse structure with single connective

When there are two parallel connectives C1 and C2 connecting S1 and S2 then the relation between them is represented as in Figure 3.

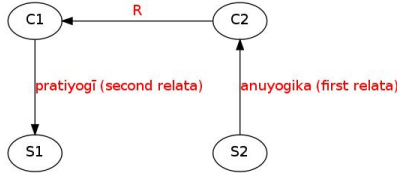


Figure 3: Discourse structure with paired connectives

Here R binds C1 and C2. The relation of the connectives with the sentence is through the main verbs. The sentences are further parsed as dependency trees. In case of paired connectives, we find instances of using either of the connectives or both. For example, in case of a paired connective ‘yadi-tarhi’ (if-then), we find instances of use of only ‘yadi’ (if), only ‘tarhi’ (then) and instances of both ‘yadi-tarhi’ (if-then). When only one of them is used in a sentence then the structure in Figure 3 collapses to that in Figure 2.

We present below various inter-sentential connectives in Sanskrit with an example for each. They are : yadi, tarhi, cet, tarhi-eva, yadyapi, tathāpi, athāpi, evamapi, yataḥ, tataḥ, yasmāt, tasmāt, ataḥ, atha, anantaram, api-ca, kim-ca, kintu, parantu. We illustrate below one example of each type.

1. Cet (If/provided) [See Figure 4] :

Sanskrit : Tvam icchasi cet aham bhavataḥ gṛham āgamiṣyāmi.
 Gloss : You desire provided I your house will_come.
 English : Provided you desire I will come to your house.

2. Yadi Tarhi (If-then) [See Figure 5] :

Sanskrit : yadi bhavān icchati tarhi aham bhavataḥ gṛham āgamiṣyāmi.
 Gloss : If you wish then I your house will_come.
 English : If you wish then I will come to your house.

It is possible that this sentence may be written with either of the connectives viz. only *yadi* or only *tarhi*. In that case the parse structure will be similar to the one in figure 4.

For the remaining examples, only if the relations differ we present a diagram.

³S2 is the anuyogi. So if the arrowhead is pointing towards S2 the name of the relation would have been anuyogi. In this diagram, the arrowhead is pointing towards C, and hence the name of the relation is inverse of anuyogi, i.e. anuyogika.

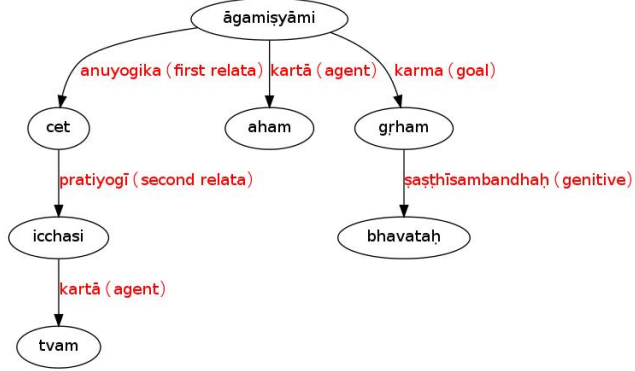


Figure 4: Cet

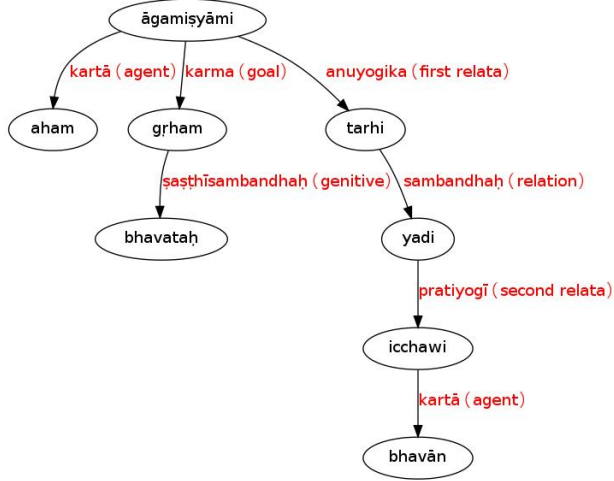


Figure 5: Yadi Tarhi

3. Yadyapi tathāpi (Even though, still):

Sanskrit : *yadyapi* ayaṁ bahu prayāsam kṛtavān *tathāpi* pariḥṣāyām tu anuttīrṇaḥ.

Gloss : Even-though he lot tried still examination failed.

English : Even-though he tried very hard, still he failed in the examination.

4. Athāpi (Hence) :

Sanskrit : pariḥṣāyām aham anuttīrṇaḥ *athāpi* punaḥ likhiṣye.

Gloss : in examination I failed hence again will write.

English : I failed in the exam, hence I will attempt again.

5. Yataḥ, Tataḥ (Because-hence) :

Sanskrit : yataḥ ayam samaye na āgataḥ tataḥ parīkṣāyām na anumataḥ.
Gloss: because he in_time not came hence in_exam not permitted.
English: Because he did not arrive in time, he was not permitted to write the exam.

6. Ataḥ (Therefore) :

Sanskrit : ayam samaye na āgataḥ ataḥ parīkṣāyām na anumataḥ.
Gloss : He in_time not came therefore in_exam not permitted.
English : He did not arrive in time therefore he was not permitted to write the exam.

7. Atha (Then) :

Sanskrit : prathamam ahaṁ śṛṇomi atha likhāmi.
Gloss : First I listen then write.
English : First I will listen and then will write.

8. Apica (And also) :

Sanskrit : bhikṣām aṭa apica gām ānaya.
Gloss : alms ask and also cow bring.
English : Seek for alms and also bring cows.

9. Kintu/Parantu (But) :

Sanskrit : gajendraḥ tīvram prayatnam_akarot kintu nakra-
grahāt na muktaḥ.
Gloss : gajendra lot tried but from_crocodile_jaw not es-
cape.
English : Gajendra tried a lot but could not escape from the jaw of the crocodile.

10. Pūrvakālikatvam (Preceding action):[see Figure 6]

Sanskrit uses a non-finite verb to indicate preceding action.

Sanskrit : rāmaḥ dugdham pītvā śālām gacchati.
Gloss : rama milk after_drinking school goes.
English : Ram goes to school after drinking milk.

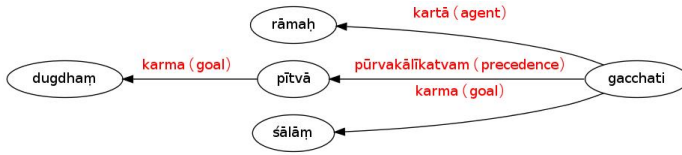


Figure 6: Pūrvakālikatvam

11. Prayojanam (Purpose of the main activity) :[see Figure 7]

Sanskrit : ahaṁ bhavantaṁ mama gṛhe bhoktum āhvayāmi.
Gloss : I you my in_house to_have_food invite.
English : I invite you to my house for lunch/dinner.

12. Samānakālikatvam (Simultaneity) :[see Figure 8]

Sanskrit : bālakaḥ jalam piban gacchati.
Gloss : boy water drinking goes.
English : The boy drinks water while going.

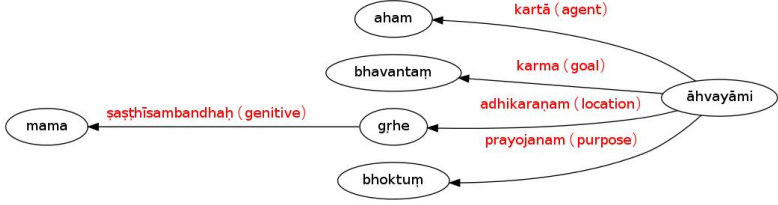


Figure 7: Prayojanam

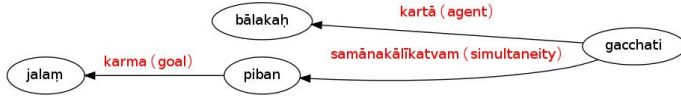


Figure 8: Samānakālikatvam

In this tagging scheme we have neither deciphered the sense of the connectives nor did we decipher the relations expressed by the two arguments. In Ramakrishnamacharyulu 2009, these relations are classified further into 9 sub-headings as below.

1. Hetuhetumadbhāvaḥ (cause effect relationship) : yataḥ, tataḥ, yasmāt, tasmāt, ataḥ.
2. Asāphalyam (failure) : kintu.
3. Anantarakālinatvam (following action) : atha.
4. Kāraṇasatve'api kāryābhāvaḥ / kāraṇābhāve'api kāryotpattiḥ (non-productive effort or product without cause) : yadyapi, tathāpi, athāpi.
5. Pratibandhaḥ (conditional) : yadi, tarhi, cet, tarhyeva.
6. Samuccayaḥ (conjunction) : ca, apica, kiñca.
7. Pūrvakālikatvam : The non-finite verb form ending with suffix *ktvā* 'adverbial participial'.
8. Prayojanam (Purpose of the main activity) : The non-finite verb form ending with suffix *tumun* 'to-infinitive'.
9. Samānakālikatvam (Simultaneity) : The non-finite verb form ending with suffix *Śatṛ* and *Śānac* 'present participle'.

In addition there are cases where the anaphora is used to indicate the simultaneity of events and the relation between events taking place in the same locus.

The analysis till this level is driven more by syntax and lexicon. The semantics is involved only to rule out incompatible parses.

4 Sub-Topic Level Analysis

Within each of the sub-topics, various paragraphs (each consisting of one or more sub-paragraphs) are connected by certain relations. The Mīmāṃsakas (exegetists) discuss 6 inter-paragraph relations in the text *Jaiminīya Nyāyamālā Vistara by Mādhavācārya*. These relations are as follows.

1. Ākṣepa (Objection)
2. Dṛṣṭānta (Example)
3. Pratyudāharāṇa (Counter-example)
4. Prāsaṅgika (Corollary)
5. Upodghāta (Pre-requisite)
6. Apavāda (Exception)

These relations differ for different types of texts. For example, a commentary on Pāṇini's *Aṣṭādhyāyī* by Patañjali has a different structure. The dominant structure, as observed in the commentary on a sūtra 2.1.1.1 'samarthaḥ padavidhiḥ⁴' consists of the following relations.

1. Praśna – question
2. Ākṣepa – objection
3. Samādhāna – justification
4. Uttara – answer
5. Vyākhyā – elaboration

Appendix B gives a small snapshot of these relations. To a certain extent some of these relations such as Praśna, Ākṣepa and Samādhāna are identifiable with the lexical cues (Tātāchārya, 2005; Tubb and Boose, 2007). Since these relations are different for different sets of texts, it is necessary to compile these various sets before we develop any discourse analysis tagset.

5 Topic Level Analysis

Six relations among topics, called *Saṅgatis* are proposed in Indian tradition. They are (Śāstri, 1916):

1. Prasaṅga - Corollary.
2. Upodghāta - Pre-requisite.
3. Hetutā - Causal dependence.
4. Avasara - Provide an opportunity for further inquiry.
5. Nirvāhakaikya - The adjacent sections have a common end.
6. Kāryaika - The adjacent sections are joint causal factors of a common effect.

⁴A compound is formed between the words which are mutually meaning-compatible.

5.1 Structure of Commentary on P2.1.1

Here we apply these *Saṅgatis* to reveal the underlying structure of a text in Grammar. The text selected is a commentary by patañjali on the sūtra *Samarthaḥ padavidhiḥ* (2.1.1) from Pāṇini's *Aṣṭādhyāyī*. The commentary consists of 213 paragraphs grouped into 14 topics as listed below.

- (1) The meaning of the words in the sūtra explaining the derivational morphology.
Here only one word *vidhi* is discussed. The commentator did not find it necessary to comment on the other words.
- (2) Type of sūtra.
The sūtras in Pāṇini's *Aṣṭādhyāyī* are classified into 6 types. Since it is not obvious from the sūtra to what type it belongs to, the commentator comments on its type and reasons thereof.
- (3) Purpose of this rule with determined type.
These three steps have the common goal of explaining the sūtra at hand. After this, the commentator explains this sūtra systematically.
- (4) Different characteristics of semantic connection (*samartha*).
- (5) The first meaning of *samartha* viz. *ekārthībhāva* 'single integrated meaning' is examined.
- (6) Various properties of single integrated meaning are examined.
- (7) Meaning of *vṛtti* 'formation of new morphemes' giving single integrated meaning are dealt with.
- (8) Possibility of the second meaning *vyapekṣā* of the word 'samartha' are ruled out.
- (9) Definition of a sentence where *vyapekṣā* is prominent.
- (10) Role of *sāmarthya* 'compatibility' in compound formation.
- (11) Purpose behind the use of the second word *padavidhiḥ*.
- (12) Objection that the sūtra is meaningless is refuted.
- (13) Rules for compound formation following syntactic agreement are explained.
- (14) Rules for deciding the gender and number of a compound.

These 14 topics are related to each other by one of the above 6 *Saṅgatis*. Figure 9 shows the relations among the topics.

6 Adhyāya Level Analysis

Among the scientific literature in Sanskrit we find three distinct trends. One is sūtra - bhāṣya - ṭīkā - ṭippaṇi popularly known as *Bhāṣya paramparā*. Here the original text is in the form of sūtras (cryptic aphorisms). This is followed by a commentary explaining the sūtras, optionally followed by an explanation (ṭīkā), a note (ṭippaṇi) etc. The commentaries may be nested, i.e. there is a commentary on the original sūtras and then commentary on this commentary, and further commentary on the sub-commentaries and so on. At each stage the number

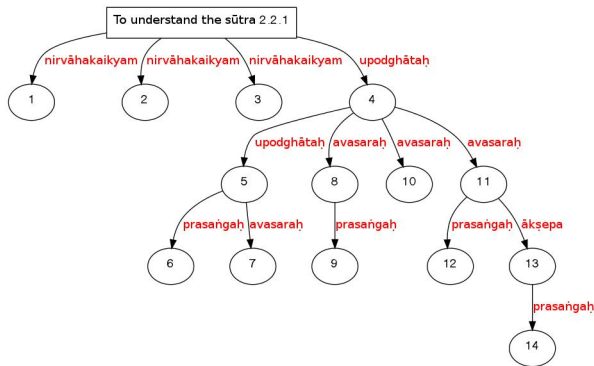


Figure 9: Discourse Structure of the commentary on "Samartha Padavidhih"

of commentaries may be more than one. The sūtras as well as the commentaries and sub-commentaries follow a certain discourse structure.

Another trend is where the original text establishes a theory, and the later scholars write criticisms on it attacking the original view and proposing a new view. This trend is known as *khaṇḍana-maṇḍana paramparā*. And there can be a series of such texts criticizing the previous theory in the series and proposing a new theory. The structure of these texts then leads to a tree structure, where the siblings indicate different criticisms of the same text leading to different view points.

The third trend is to write *prakaraṇa granthas* (books dealing with a specific important topic among several topics discussed in the texts in sūtra form). These books are thus related to the original sūtra texts, but also have their own nested commentaries.

The grammar of these discourse structures then necessarily differ.

7 Towards Computability

In this paper we have described various level of analysis the tradition is following in order to understand the Sanskrit texts. Based on the available literature, a tagging scheme for dependency analysis and a dependency parser are developed. This parser is further enhanced to handle the anaphora and inter-sentential relations as well. Sanskrit has an advantage of having a huge corpus in the form of printed texts, with important literary works well analysed at various levels through commentaries. These works should be useful for further identifying the cues for establishing various saṅgatis. It is well known that different interpretations of the same text have resulted in different Indian philosophical schools. For interpretations we need an objective analysis of the text. We also need to have all possible interpretations presented in a nut-shell. With the help of computational tools now it is possible to explore all the possible interpretations of a given text at various stages of analysis systematically and present it in a concise form leaving the task of interpretation to the user. For example an expression 'naisadharājagatyā' from the 'Nalacaritam' (biography of Nala) has 6 different interpretations as described in the commentaries. The current tools help a student of Sanskrit to understand these various interpretations (Varalakshmi, 2013) in a systematic way. With the availability

of a discourse level analysis, in future it should be then possible to understand how different interpretations emerge from the same text with different combinations of analysis at various stages.

References

- Ganeri, J. and Miri, M. (2010). Sanskrit Philosophical Commentary. *Journal of The Indian Council of Philosophical Research*, Vol. 27:187–207.
- Hellwig, O. (2009). Extracting Dependency Trees from Sanskrit Texts. In Kulkarni, A. and Huet, G., editors, *Proceedings, Third International Sanskrit Computational Linguistics Symposium*, pages 106–115. Springer-Verlag, LNAI 5406.
- Huet, G. (2009). Formal structure of Sanskrit text: Requirements analysis for a mechanical Sanskrit processor. In Huet, G., Kulkarni, A., and Scharf, P., editors, *Sanskrit Computational Linguistics 1 & 2*. Springer-Verlag, LNAI 5402.
- Joshi, S. D. (1968). *Patañjali's Vyākaraṇa Mahābhāṣya Samarthāhnikā (P 2.1.1) Edited with Translation and Explanatory Notes*. Center of Advanced Study in Sanskrit, Poona, first edition.
- King, T. H., Crouch, R., Riezler, S., Dalrymple, M., and Kaplan, R. (2003). The PARC 700 dependency bank. In *Proceedings of the Fourth International Workshop on Linguistically Interpreted Corpora (LINC-03)*.
- Kulkarni, A., Pokar, S., and Shukl, D. (2010). Designing a Constraint Based Parser for Sanskrit. In Jha, G. N., editor, *Proceedings, Fourth International Sanskrit Computational Linguistics Symposium*, pages 70–90. Springer-Verlag, LNAI 6465.
- Kulkarni, A. and Ramakrishnamacharyulu, K. V. (2013). Parsing Sanskrit Texts : Some Relation Specific Issues. In *Proceedings, Fifth Sanskrit Computational Linguistics Symposium*. D. K. Publisher.
- Kulkarni, A., Shukl, D., and Pokar, S. (2012). Mathematical Modeling of Ākāṅkṣā and Sannidhi for Parsing Sanskrit. 15th World Sanskrit Conference, Delhi.
- Kumar, A., Mittal, V., and Kulkarni, A. (2010). Sanskrit compound processor. In Jha, G. N., editor, *Proceedings, Fourth International Sanskrit Computational Linguistics Symposium*. Springer-Verlag, LNAI 6465.
- M. Marneffe, B. M. and Manning, C. D. (2006). Generating typed dependency parses from phrase structure parses. In *The fifth international conference on Language Resources and Evaluation, LREC 2006, Italy*.
- Mann, W. C. and Thompson, S. A. (1987). *Rhetorical Structure Theory : A Theory of Text Organization (Reprinted from The Structure of Discourse)*. ISI Reprint series 87–190.
- Polanyi, L. (1988). A formal model of discourse structure. *Journal of Pragmatics*, 12:601–638.
- Prasad, R., Miltsakaki, E., Dinesh, N., Lee, A., Joshi, A., Robaldo, L., and Webber, B. (2007). *The Penn Discourse Treebank 2.0 Annotation Manual*.
- Ramakrishnamacharyulu, K. V. (2009). Annotating Sanskrit Texts Based on Sābdabodha Systems. In Kulkarni, A. and Huet, G., editors, *Proceedings, Third International Sanskrit Computational Linguistics Symposium*, pages 26–39. Springer-Verlag, LNAI 5406.

Ramakrishnamacharyulu, K. V., Pokar, S., Shukl, D., and Kulkarni, A. (2011). *Annotation Scheme for Kaaraka level tagging and guidelines*.

Śāstri, A. (1916). *Kārikāvalī with the commentaries Muktvāvalī, Dinakarī, Rāmarudrī (edited with footnotes by Ananta Śāstri)*, page 286–287. Nirṇaya Sāgara Press, Bombay.

Śāstri, G. D. (1929). *Śaktivāda with Vivṛti commentary by Harinātha Tarka Siddhānta Bhaṭṭāchārya, edited with critical notes*. Chowkhamba Sanskrit Series Office, Benares.

Śrī Jīvananda Vidyāsāgar Bhaṭṭācārya (1989). *Śrī Mādhavācāryaviracita Jaiminiya Nyāyamālāvistarahaḥ*. Kṛṣṇadās Academy, Varanasi.

Staal, F. (1965). Euclid and Paṇini. *Philosophy of East and West*, 15:99–116.

Tātāchārya, N. S. R. (2005). *Śābdabodhamīmāṃsā: The sentence and its significance Part-I to V*. Institute of Pondicherry and Rashtriya Sanskrit Sansthan, New Delhi.

Tesnière, L., editor (1959). *Éléments de Syntaxe Structurale*. Klincksieck, Paris.

Tubb, G. A. and Boose, E. R. (2007). *Scholastic Sanskrit: A Manual for Students*. Columbia University, New York.

Varalakshmi, K. (2013). Śleṣa Alaṅkāra - A Challenge for Testing Sanskrit Analytical Tools. In *Proceedings, Fifth Sanskrit Computational Linguistics Symposium*. D. K. Publisher.

Webber, B. (2006). Accounting For Discourse Relations : Constituency and Dependency. *Festschrift for Ron Kaplan*.

Webber, B. and Joshi, A. (2012). Discourse Structure and Computation : Past, Present and Future. In *Proceedings of the ACL-2012 Special Workshop on Rediscovering 50 Years of Discoveries*.

Webber, B., Joshi, A., Sarkar, A., Prasad, R., Miltsakaki, E., and Forbes, K. (2001). D-LTAG System - Discourse Parsing with a Lexicalized Tree Adjoining Grammar. In Kruijff-Korbyová, I. and Steedman, M., editors, *Workshop on Information Structure, Discourse Structure and Discourse Semantics*.

Wolf, F. and Gibson, E. (2005). Representing discourse coherence: a corpus-based study. *Computational Linguistics*, 31:249–287.

A Nested Commentaries

This is an introductory section of a prakaraṇa grantha *Vaiyākaraṇa-siddhānta-laghu-mañjūṣā* with a commentary *Ratnaprabhā* by Sabhāpati Upādhyāya.

We show below the original text, followed by the commentary, as it appears in the printed text.

Original : tatra vākyasphoṭo mukhyo loke tasyaivārthabodhakatvāttenaivārthasamāptēśca,

gloss : There, the process of understanding the meaning of a sentence is primary. This process has the property of conveying the meaning and therefore it itself leads to the completion of meaning.

Commentary : tatreti nirdhāraṇe saptamī, tathā ca siddhāntagaṭakavākyasphoṭo mukhya ityarthah. *bodhakatvāditi – yadyapyāntarasphoṭasyaiva vācakatvasya siddhāntayaiṣyamāṇatayā

vāhyasya padasamūharūpavākyaśya na vācakatvaṃ tathāpi tattādātmyāpannatvena tasyāpi tattvaṃ bodhyam. *tenaiva – vākyenaiva. *arthasamāpteriti vākyaśyaiva nirākāṅkṣārthabodhakatvenārthasya pūrṇatvānnirākāṅkṣatvāditi yāvat.

If this piece is presented in this way, it is difficult to follow the commentary. We present below the original text segmented and commentary split into several footnotes placed at relevant places.

Segmented : tatra⁵ vākyaśpoṭaḥ mukhyaḥ loke tasya eva artha bodhakatvāt⁶ tena eva⁷ artha samāpteḥ ca,⁸

We observe that this makes it easy to read and understand the texts, since now we can 'see' the underlying structure. But we cannot use this technique further since nesting of footnotes after a certain limit becomes unwieldy. The current hyper text technology however makes it easy to present this text in the form of hyper text with links, allowing a smooth representation of the nested commentaries.

B Original text of Mahābhāṣya with relations among the paragraphs

This is the 4th sub-topic from the commentary by Patañjali on P2.1.1. The relations are marked manually, originally in the Nirṇaya sāgara edition of the mahābhāṣya which was further enhanced by Joshi in his edition (Joshi, 1968). The numbers indicate the serial number of paragraphs from the beginning of the commentary on P2.1.1.

Sub-Topic starts: *atha sāmāthyalakṣaṇabhedanirūpaṇādhikaraṇam*

(Now starts the section in which the different characteristic of semantic connection are examined.)

Relation: *praśnabhāṣyam* (question)

41. *atha kriyamāne'api samarthagrahane samarthamityucyate kiṃ samarthaṃ nāma* |

(Now, apart from the question whether (the word) *samartha* should be mentioned in P 2.1.1 (or not), (when) you say *samartha*, what do you really mean by *samartha*?)

Relation: *samādhānavārttikam* (justification)

Vārttika: pṛthagarthānāmekārthībhāvaḥ samarthavacanam || 1 ||

(The word *samartha* (means) single integrated meaning of words which (when un-compounded) have separate meanings (of their own).)

Relation: *vyākhyābhāṣyam* (elaboration)

42. *pṛthagarthānām padānāmekārthībhāvaḥ samarthamityucyate* |

((When) we say *samartha*, (it means) single integrated meaning of words which (when un-compounded) have separate meanings (of their own).)

⁵tatreti nirdhāraṇe saptamī, tathā ca siddhāntagaḥaṭakavākyaśpoṭo mukhya ityarthaḥ.

⁶bodhakatvāditi – yadyapyāntarasphotasyaiva vācakatvasya siddhāntayisyamānatayā vāhyasya padasamūharūpavākyaśya na vācakatvaṃ tathāpi tattādātmyāpannatvena tasyāpi tattvaṃ bodhyam.

⁷tenaiva – vākyenaiva.

⁸arthasamāpteriti vākyaśyaiva nirākāṅkṣārthabodhakatvenārthasya pūrṇatvānnirākāṅkṣatvāditi yāvat.

Relation: *praśnabhāṣyam* (question)

43. *kva punaḥ pṛthagarthāni kvaikārthāni |*

(But where (do words) have separate meanings (of their own), (and) where (do they) have a single meaning?)

Relation: *uttarabhāṣyam* (answer)

44. *vākye pṛthagarthāni|rājñāḥ puruṣa iti|samāse punarekārthāni rājapuruṣa iti |*

(In the uncompound word group (words) have separate meaning (of their own), like in *rājñāḥ puruṣaḥ*: 'king's man'. But in a compound, (words) have a single meaning, like *rājapuruṣaḥ*: 'king-man'.)

Relation: *ākṣepabhāṣyam* (objection)

45. *kimucyate pṛthagarthānīti yāvatā rājñāḥ puruṣa ānīyatāmityukte rājapuruṣa ānīyate rājapuruṣa iti ca sa eva |*

(What do you say: '(words) having separate meanings (of their own)'? Because when we say: 'let the king's man be brought', the king-man is brought. And (when we say): '(let) the king-man (be brought)', the same (man is brought).)

Relation: *samādhānabhāṣyam* (justification)

46. *nāpi brūmo'anyasyānayanam bhavatīti|*

(We do not say at all that a different person is brought.)

Sub-topic ends: *iti sāmāthyalakṣaṇabhedanirūpaṇādhikaraṇam*

(Here ends the section in which the different characteristics of semantic connection are examined.)