Parsing Sanskrit texts: Some relation specific issues

Amba Kulkarni[1] and K V Ramakrishnamacharyulu[2]

[1] Department of Sanskrit Studies,
University of Hyderabad,
Hyderabad
apksh@uohyd.ernet.in
[2] Department of Vyakarana,
Rashtriya Sanskrit Vidyapeetha,
Tirupati
kvrkus@gmail.com

Abstract. Building a sentential parser following a dependency framework needs a well defined set of relations. In case of Sanskrit, various texts on Śābdabodha theories discuss such relations. These relations are critically examined from the point of view of feasibility of building a rule based parser. We propose an intermediate parse with nice computational properties of a tree structure and then propose another layer to make this tree structure more useful for information retrieval and user understanding.

Key Words: Sanskrit, Parser, Kāraka relations

1 Introduction

Parsing unfolds a linear string of words into a structure which shows explicitly the relations between words. The parse of positional languages such as English are well expressed by constituency structure while languages like Sanskrit which are morphologically rich and to a large extent free word order are better represented by a dependency tree. A possible parse of

(1) rājā viprāya gāṁ dadāti.
gloss: King brahmin{dat.} cow{acc.} gives.
Eng: A king gives a cow to a brahmin.

is shown in Fig 1 as a dependency tree.

Texts dealing with the Śābdabodha theories describe an understanding arising from the sentences in terms of the relations between various constituents. For example the Vaiyākaraṇa's śābdabodha of (1) is

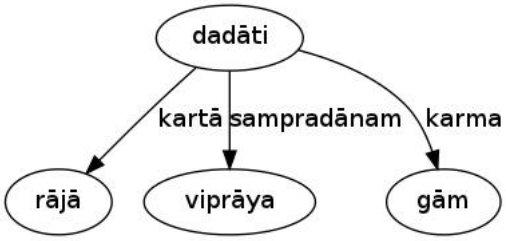

Fig. 1.

rājakartṛka gokarmaka viprasampradānaka dānānukūla vyāpāraḥ.

Thus according to Vaiyākaraṇas when a person hears sentence (1) then the understanding that he gets is of an activity of dāna whose kartā(agent) is a king, whose karma(goal) is a cow, and whose sampradāna(recipient) is a Brahmin.

For a Naiyāyika, however, the structure of the verbal cognition resulting from this utterance is different. The Naiyāyika's śābdabodha is

gokarmaka viprasampradānaka dānānukūla kṛtimān rājā.

So according to a Naiyāyika, sentence (1) results in the cognition of a king who is the agent of an activity of giving, whose goal is a cow and whose recipient is a Brahmin.

Thus the chief qualificand is different in both the cases, however the relations between the padārthas 'referent of words' are the same. Various relations described in the traditional grammar books have been compiled and classified by Krishnamacharyulu(2009) under the two broad headings viz. inter sentential and intra sentential relations. This work provided a starting point for developing guidelines for annotation of Sanskrit texts at kāraka level and also for the development of an automatic parser 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 (Marneffe and Manning, 2008) etc. for defining the set of relations, we are fortunate to have a well defined, time tested tagset for Sanskrit. This tagset can be put to use for two tasks – a) to develop an annotated corpus, b) to develop a parser that produces a parse tree of a given sentence. One question we would like to ask before putting it to actual use is whether the granularity of this tagset is suitable for the above two tasks?

The suitability of a tagset for manual annotation can be judged on the basis of the following parameters:

- The inter annotator agreement for various tags, and
- The grey / fussy tags which lead to errors in annotation.

A statistical parser that uses manually annotated data will also have these as the main concerns. A rule based parser, on the other hand, will have different considerations. A rule based parser performs better the less it depends on extra linguistic information. For example, consider a manually annotated text where kartā is sub-classified further into anubhavī kartā(experiencer), karaṇakartṛ, karmakartṛ, etc. If the tagged data is sufficiently large, it is possible that machine learns the distinction between a kartā and an anubhavī kartā from the tagged examples. On the other hand, to decide whether something is an anubhavī kartā or not, one needs to appeal to the semantics of the verb involved. Deciding whether the padārtha 'the referent of a word' is a kartā or not is relatively easier as it involves only the syntactic and morphological information. So with a goal to build a rule based parser, we critically examined all the tags in Krishnamacharyulu(2009). The basic principles we followed during this critical examination were

- The relations should be binary.
- information / cue for extracting any relation should be coded in the language string.

In the next section, we describe the notation for representing the relations and in section 3, we discuss various cues for extracting the relations. Section 4 discusses various criterion used for the choice of a relation, and section 5 discusses the issues of granularity. In section 6 we describe the post processing for making the parse more useful.

2 Convention for marking the relations

Let the *padārtha*s associated with the two words[3] 'a' and 'b' be related by relation 'R'. To be precise, 'b' has the property of 'R_ness' conditioned / determined by 'a'. We represent this graphically as in Fig 2.

Note the arrowhead at 'b'. This is a directed labelled graph where 'a' and 'b' represent the nodes and are joined by a labelled arrow 'R'. For example, in a sentence *rāmaḥ gacchati*, *rāma* is the kartā of the going activity. Fig 3 shows the corresponding graph.

For the ease of annotation, instead of annotating the sentences in graphical mode, we represent them as a text with three fields separated by a tab. The first

[3] Henceforth we shorten the phrase 'the padārthas associated with the word' by simply 'the word'.

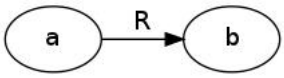

Fig. 2.

Fig. 3.

field contains the word number, the second field the word and the third field the relation. The first line of a sample annotation below then means: The word *rāmaḥ* at position 1 is the kartā of the word at the third position viz. *gacchati*.

```
1.  rAmaH  kartA,3
2.  vanam  karma,3
3.  gacchati
```

3 Clues for extracting the relations[4]

Sanskrit being inflectionally rich, the inflectional suffixes mark the relation between words. Similarly certain indeclinables mark some grammatical relations. Agreement between the words also indicate certain grammatical relations. We discuss below these clues for extracting relations.

1. Abhihitatva (property of being expressed)
   The Pāṇinian sūtra 'anabhihite' (2.3.1) (if not already expressed) is an important sūtra that governs the vibhakti assignment to the nominals. The vārttika[5] on this sūtra explains abhihita as the one which is expressed either by tiṅ (a finite verbal suffix), kṛt (a non-finite verbal suffix), taddhita (derivational nominal suffix) or samāsa (compound). E.g. in the sentence

   (2) *rāmaḥ vanaṁ gacchati.*

   the verb being in the active voice (kartari prayogaḥ), the verbal suffix 'ti' expresses the kartā, while in the following sentence in passive voice (karmaṇi prayogaḥ)

[4] These have been discussed in Kulkarni(2010). For the sake of completeness, we repeat the relevant portion here.

[5] tiṅkṛttaddhitasamāsaiḥ parisaṁkhyānam (ma. bhā. 2.3.1. vā.)

(3) *rāmeṇa vanaṁ gamyate.*

the karma is expressed by the verbal suffix. As such, in both cases, the one which is expressed (kartā and karma respectively) is in the nominative case and shows number and person agreement with the verb form.

Unlike the tiṅ suffix which is inflectional, kṛt, taddhita and samāsa mark the derivation process, and in the process the derivation generates a new lexical head from the old one. For example, in

(4) *dhāvan aśvaḥ.*

the kṛt suffix in 'dhāvan' expresses the relation of kartā (kartari kṛt (3.4.68)) and indicates the one which performs the action of dhāv 'running'.

2. Vibhakti
   The verbal as well as nominal suffixes in Sanskrit are termed vibhaktis. We have already seen that verbal suffixes (*tiṅ*), through abhihitatva, mark the relations between words. Now we consider the nominal suffixes. They fall under three broad categories.

   (a) vibhakti indicating a kāraka relation
   This marks a relation between a noun and a verb known as a kāraka relation. Sanskrit uses seven case suffixes to mark six kāraka relations viz. *kartā*, *karma*, *karaṇa*, *sampradāna*, *apādāna* and *adhikaraṇa*. The genitive suffix, in addition to marking a kāraka relation[6], is predominantly used to mark a noun-noun relation.

   (b) upapada vibhakti
   In addition to the noun-noun relations expressed by the sixth case, there are certain words, most of them indeclinables called upapadas, which also mark a special kind of noun-noun relation. These indeclinable, mark a relation of a noun with another noun, and in turn demand a special case suffix for the preceding noun. For example, the upapada 'saha' demands a third case suffix for the preceding noun as in:

   (5) *rāmeṇa saha sītā vanaṁ gacchati.*

   (c) special vibhakti
   Vibhaktis are also used to indicate various other relations such as
   - atyantasaṁyoga 'intimate and total contact' (as in *māsamadhitaḥ* 'he studied for a month without any break'),
   - kriyaaviśeṣaṇa 'adverbial usage' as in *vegena dhāvati* 'runs fast',
   - aṅgavikāra 'defect in a body-organ' as in *akṣṇā kāṇaḥ* 'blind with an eye',

[6] kartṛkarmaṇoḥ kṛti (2.3.65)

- nirdhāraṇa ‘specifying one out of many’ as in *nareṣu śreṣṭaḥ* ‘best among the men’ ,
- vibhakta ‘distinct / different’ as in *gopālāta śyāmaḥ avaraḥ* ‘Syāma is better than Gopāla’.

3. Indeclinables (avyaya)
   The indeclinables mark various kinds of relations such as negation, adverbial (manner adverbs only), co-ordination, etc. Sometimes they also provide information about interrogation, emphasis, etc. We distinguish the upapadas from the avyayas, mainly because, though most of the upapadas (which are termed karmapravacanīya by Pāṇini) are also indeclinables, they demand a special case suffix on the preceding word, whereas it is not so with indeclinables.

   For example, the relation of ‘na’ with ‘gacchati’ in the sentence

   (6) *rāmaḥ gṛhaṁ na gacchati.*

   is that of ‘negation (niṣedha)’. Similarly, the relation of ‘mandam’ with ‘calati’ in the sentence

   (7) *rāmaḥ mandaṁ calati.*

   is that of ‘adverbial (kriyāviśeṣaṇa)’. The relation of ‘eva’ with ‘rāma’ in the sentence

   (8) *rāmaḥ eva tatra upaviṣati.*

   is that of ‘emphasis (avadhāraṇa)’.

4. Samānādhikaraṇa
   Agreement in gender, number and case suffix marks *samānādhikaraṇa* (having the same locus), or the modifier-modified relation between two nouns as in

   (9) *śvetaḥ aśvaḥ dhāvati.*
   (10) *aśvaḥ śvetaḥ asti.*

   In (9) as well as (10), the words *aśvaḥ* and *śvetaḥ* have the same gender, number and vibhakti indicating samānādhikaraṇa. However, there is a subtle difference between the information being conveyed. In (10), the word *śvetaḥ* is a predicative adjective (vidheya viśeṣaṇa), while in (9) it is an attributive adjective.

4 Choice of relations and their representation

These cues now lead us to the following questions.

- Should the inflectional suffixes and derivational suffixes be treated at par?
- How to treat the function words? Should they be treated as a node in a tree or an edge?
- How to represent the inter sentential relations?
- Should anaphoric resolution be part of this annotation?

The basic principles we follow in arriving at the decision and the rationale behind them are

1. Preserve one-one mapping between the nodes of a tree and the words in a sentence.
   [This is more of a topological requirement than a linguistic one. If this condition is relaxed, the parse ceases to be a tree, loosing its nice computational properties[7].]
2. In case of derived nouns, consider only the inflectional suffix for establishing the relations.
   [This also is a topological requirement with the same reasoning as above.]
3. In case of derived indeclinables, use the derived suffix to mark the relations.
   [Since there is a deletion (lopa) of the inflectional suffix in case of derived indeclinables, the information encoded by the derived suffix is considered for the marking of relations.]
4. In case of indeclinables(other than the derived kṛdantas), the words themselves mark the relations.

5. A suffix or a word can represent one and only one relation.
   [Any meaningful unit in an interpretation under consideration expresses only one meaning.]

We present below various cases and explain the rationale behind these principles. These principles themselves provide answers to the questions raised above.

1. *abhihita* The verbal inflection in addition to marking various features associated with the verbal form also shows an agreement with the noun – a *kartā* in kartari prayogaḥ(active voice) and a *karma* in karmaṇi prayogaḥ (passive voice). So the question is, in case of an active voice, what is the head of the kartā relation – a verb or a noun? If we refer back to our convention for naming the relations, if 'b' has R_ness conditioned by 'a', then we mark the relation 'R' as a relation of 'a' in 'b'. In a sentence '*rāmaḥ gacchati*', the nominal case suffix in rāma does not indicate any kāraka relation[8]. And hence on the basis of agreement, we mark the relation of kartā in *rāma*. If kartṛ pada is missing in the sentence as in 'gacchāmi', we

[7] Amba Kulkarni thanks Gérard Huet for useful discussions.
[8] prātipadikārthaliṅparimāṇavacanamātre prathamā(2.3.46)

do not mark it.

Now consider another example:

(11) dhāvantam aśvaṁ paśya.

In this sentence, there are two verbs *dhāv* 'run' and *paśya(dṛś)* 'see'. *Dhāv* has a kṛt suffix (śatṛ) which is in the sense of kartā. *Aśva* is the karma of *paśya*, and is also the kartā of *dhāv*. Further *dhāvantaṁ* and *aśvaṁ* have same vibhakti and hence there is a viśeṣya-viśeṣaṇa bhāva (modifier-modified relation) between them. Now the question is which relation to mark for the word *aśva* - a *karma* of *paśya*, a *kartā* of *dhāv* or a *viśeṣya* of *dhāvat* or all of them? Marking all these three relations may generate a loop or multiple inheritance for the word *aśvaṁ* as shown in Fig 4.

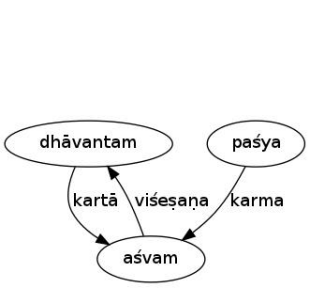

Fig 4

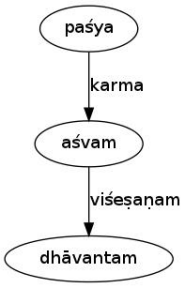

Fig 5

A loop destroys the nice tree structure of a parse, and hence we decide to mark only relations indicated through the inflectional suffixes and not through the derivational suffixes. Hence we propose to represent the relations as in Fig 5.

Now one may raise an objection that this representation is not faithful to the information content in the given string. The answer to this objection is: there is no loss of information in this representation. The information that *dhāvantam* marks the *kartā* of the verb *dhāv* is still available in the derivational suffix śatṛ, which we can still use during the post-processing to add the missing information. The advantage of postponing the marking of this information is that the resulting parse is a tree and we can use the existing computational tools for extracting a tree from the graph.

2. Treating indeclinables:
Should an indeclinable be treated as a function word or a content word? The indeclinables fall into three categories viz. kṛdanta, upapada and the rest. The treatment of each of them is discussed below with an example.

– kṛdanta avyayas:
Consider the following example:

(12) *rāmaḥ dugdham pītvā śālāṁ gacchati.*

Here the word *pītvā* is a *kṛdanta-avyaya* derived from the verbal root *pib* by adding a *kṛt* suffix *ktvā*. This *ktvā* which is a derived suffix marks the relation of precedence (*pūrvakālaḥ*) with reference to the main verb[9].

Here we mark the relation of precedence, though it is denoted by the derivational suffix. Since this is an indeclinable, there is an elision of inflectional suffix, and hence we mark the information encoded by the derivational suffix as a relation.

The sūtra *samānakartṛkayoḥ pūrvakāle(3.4.21)* states that the action denoted by the verb with *ktvā* suffix preceding another action, shares the *kartā* with it. Thus in the above example, *pib* precedes the action of going denoted by *gam* in *gacchati*. *Rāmaḥ* is the *kartā* of *gacchati* and is also *kartā* of the drinking action denoted by *pib*. So the graph showing the relations will be as shown in Fig 6.

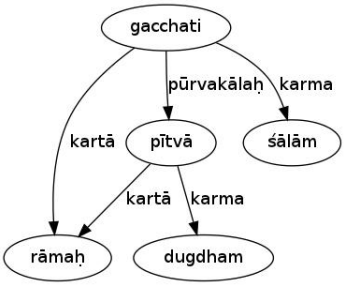

Fig 6

Rāmaḥ is the kartā of both an action of going as well as drinking. However, the nominal suffix in *rāmaḥ* can express only one relation. Further, because of two incoming arrows into a single node, the graph results in nodes having multiple inheritances which prohibits this parse from being a tree. Hence we do not mark the relation of *kartā* between *rāmaḥ* and *pītvā*, since it is not expressed by any suffix. This relation will be restored at the post-processing stage.

– upapada avyayas
The upapada avyayas such as *saha* demand a specific vibhakti for the noun with which it is connected. For example Pāṇini's sūtra *sahayukte'apradhāne (2.3.19)* assigns a third case to the noun to which

[9] samānakartṛkayoḥ pūrvakāle (3.4.21)

saha is attached as in *rāmeṇa saha*. Consider the sentences:

(13) *sītā rāmeṇa saha vanaṁ gacchati.*
(14) *sītā dugdhena saha roṭikāṁ khādati.*

In the first sentence, *rāma* is the *saha-kartā* while in the second sentence *dugdha* is a *saha-karma*. Thus to decide what kāraka role the noun with *saha* will have, extra-linguistic information is needed in the absence of which machine will end up producing two possible parses (one correct and the other wrong) for sentence (13) as in Fig 7 and 8 below.

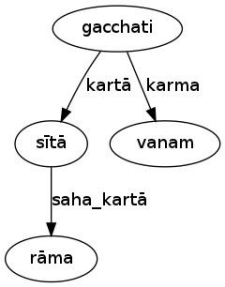

Fig 7

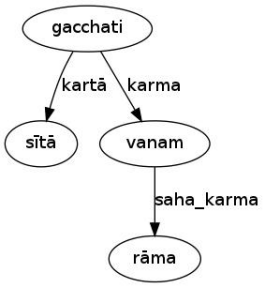

Fig 8

To arrive at the correct parse one needs to check the meaning-compatibility (*yogyatā*) of associated words. Further, in order to reduce the number of relations, we mark this relation as *saha-sambandha*, following the tradition, ensuring that there is no loss of information in doing so.

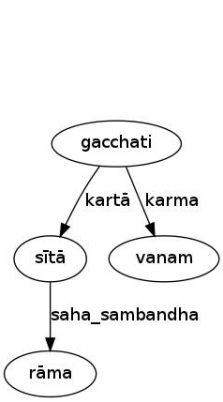

Fig 9

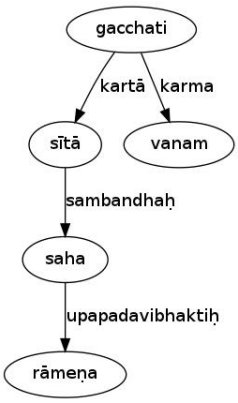

Fig 10

Now the next question is whether to treat this *upapada* as a content word or as a function word? In other words which parse to prefer – the one represented in Fig 9 or Fig 10?

The *upapada* acts more like a function word(*dyotaka*) than a content word(*vācaka*). So it is desirable to group the *upapada* together with the preceding content word and mark the relation with the content word as in Fig 9.

Though this solution is desirable, it creates a mismatch between the number of words and the nodes in the graph. To avoid this mismatch, we propose to generate a graph as in Fig 10 and then we collapse the intermediate node to generate the graph in Fig 9 mechanically later.

– rest of the indeclinables
Certain indeclinables such as *eva*, *iva*, *na* also mark relations such as *avadhāraṇā* 'emphasis', *sādharmya* 'similarity', *niṣedhya* 'negation' etc. These are all more like function words than content words. However, in order to preserve the one-one relation between the number of words in a sentence and the nodes in a graph, we treat these words as content words and mark the relations as in Fig 11 and 12.

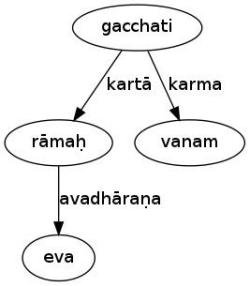

Fig 11

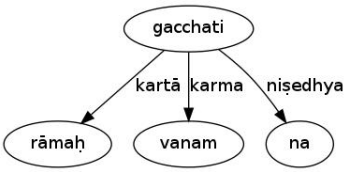

Fig 12

3. Treatment of Inter-sentential connectives
The inter-sentential connectives connect two sentences. For example, consider:

(15) *yadi tvam icchasi tarhi aham bhavataḥ gṛhaṁ āgacchāmi.*

Here *tvam icchasi* and *aham bhavataḥ gṛhaṁ āgacchāmi* are two independent sentences and the words *yadi-tarhi* connect them. Now this connection is through the main verbs *icchasi* and *āgacchāmi*. Fig. 13 shows one proposed parse.

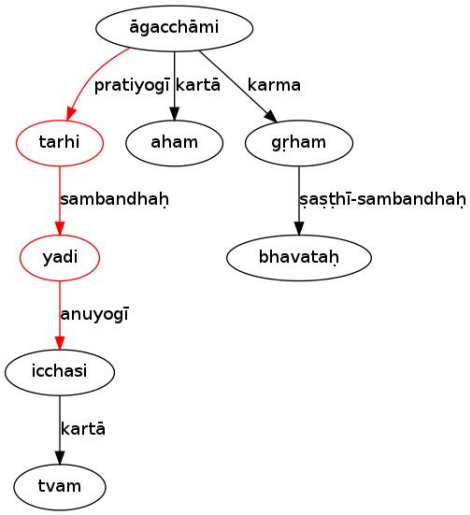

Fig 13

Following the Navya-nyāya convention, the relations are named *anuyogi* and *pratiyogi* (roughly relata 1 and relata 2). The words *yadi-tarhi* might be grouped together to form a node, but since this will create a mismatch between the number of words and the nodes, we name the relation between *yadi* and *tarhi* as *sambandhaḥ*. This scheme then can be extended to handle cases of ellipsis where either *yadi* or *tarhi* is dropped as below.

(16) *tvam icchasi tarhi aham bhavataḥ gṛhaṁ āgacchāmi.*
(17) *yadi tvam icchasi aham bhavataḥ gṛhaṁ āgacchāmi.*

The corresponding graphs are shown in Fig. 14 and Fig 15 respectively.

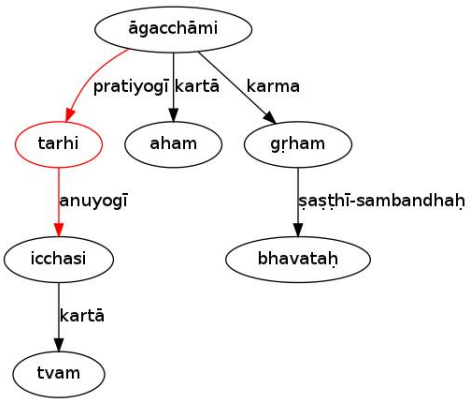

Fig 14

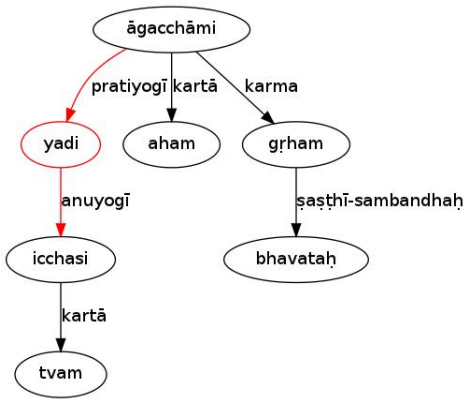

Fig 15

During the post-processing stage, we provide the missing words *yadi* / *tarhi* for proper interpretation of the graph.

4. Treatment of Anaphoras
   The convention for showing the anaphoric references is by co-indexing. Consider the sentence

   (18) *yatra nāryaḥ pūjyante ramante tatra devatāḥ.*

   The parse for this may be represented as in Fig 16.

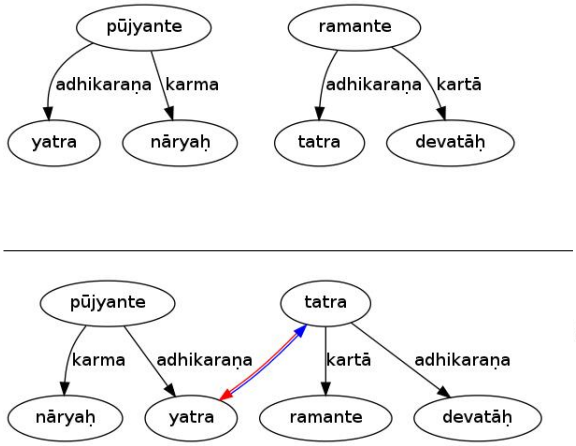

Fig 16, Fig 16a

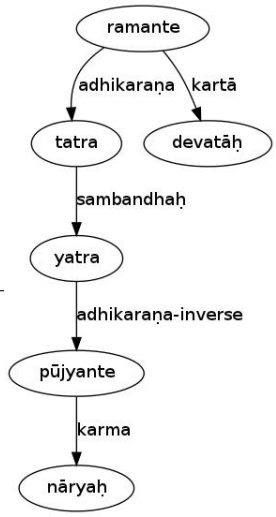

fig 17

This parse consists of two trees. But the given sentence is a single one because each of the part is incomplete without the other[10]. Then how do we account for a parse consisting of two trees?

To convert it into a single tree, words *yatra* and *tatra* would have to be joined together. One possible parse with single tree is as in Fig 17 making it a totally unintuitive parse!

Let us look at the information content again. The words *yatra* and *tatra* are in the seventh case with *yat* and *tat* as the nominal stems. The inflectional suffixes mark the relation of *adhikaraṇa* with the verbs. Now the anaphoric relation between them is due to the *nitya sambandhaḥ* between the pronominal stems, and not because of any suffixes. The relations we are marking are due to the suffixes, and therefore, we do not mark the relation

[10] arthaikatvāt ekaṃ vākyaṃ sākāṅkṡaṁ ced vibhāge syāt.

between the prātipadikas viz. yat and tat, leaving the parse structure as a forest. The co-indexing (denoted by double arrow) as in Fig 16a will turn the forest into a tree.

5. Treatment of conjunctions and disjunctions
   The problem in the representation of conjunction and disjunction is deciding the head. Following a Naiyayika we mark the conjunctive or disjunctive particle as a head. Fig 18. shows the analysis of sentence (19).
   (19) *rāmaḥ sītā lakṣmaṇaḥ ca vanaṁ gacchanti.*

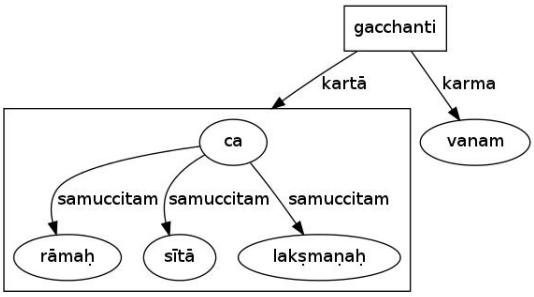

Fig 18

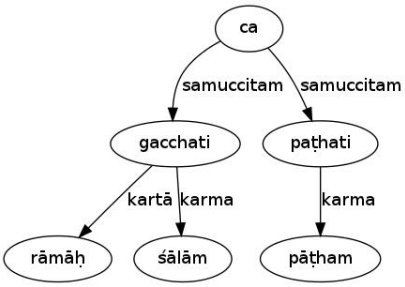

Fig 19

This is represented in Fig 18. Thus the kartṛtva is in *rāma, sītā and lakṣmaṇa* together and not in individuals separately. This is represented by the box which in turn contains all the individuals, joined by the conjunctive particle *ca*. In case of two sentences joined by the conjunctive particle, the later sentence is marked as a head(see Fig 19) as in

(20) *rāmaḥ śālāṃ gacchati pāṭhaṃ ca paṭhati.*

Note here that *rāmaḥ* which is the *kartā* for the verb *paṭhati* as well is not explicitly marked in this parse. The reason being, this information is not explicitly coded by any morpheme but is inferred through the property of the conjunctive particle. Such an information resulting due to inference will be shown in the post processed parse structure.

5 Granularity

The criterion for deciding the granularity is simple. If one can tell one relation from the other purely on the basis of syntax or morphology, then the two relations may be treated as distinct. We illustrate this with an example. Krishnamacharyulu(2009) sub-classifies the relation of *kartā* into the following subcategories.

– *anubhavī kartā*
Ex: **ghaṭo** naśyati

– *amūrtaḥ kartā*
Ex: **krodhaḥ** āgacchati

– *prayojaka kartā*
Ex: **devadattaḥ** viṣṇumitreṇa pācayati.

– *prayojya kartā*
Ex: devadattaḥ **viṣṇumitreṇa** pācayati.

– *madhyastha kartā*
Ex: devadattaḥ **yajñadattena** viṣṇumitreṇa pācayati.

– *abhipreraka / utpreraka kartā*
Ex: **modakaḥ** rocate.

– *karma-kartṛ*
Ex: **kāṣṭhaḥ** svayameva bhidyate.

– *karaṇa-kartṛ*
Ex: **asiḥ** chinatti.

– *ṣaṣṭhī kartā*
Ex: **ācāryasya** anuśāsanaṁ

Morphology and syntax are necessary to mark the *prayojaka kartā*, *prayojya kartā* and *ṣaṣṭhī kartā* mechanically. But these are not sufficient. The sufficiency comes in the form of yogyatā 'compatibility'. For example, in
devadattena annam pācayati.
devadatta is a *prayojya kartā*, and in
devadattaḥ agninā annam pācayati.
agni is karaṇa. The third case suffix in *devadatta* and *agni* are only the eligibility criterion for *devadatta* and *agni* to be either prayojya kartā or karaṇa. The sufficiency comes from their referents. Similarly the genitive case marks the necessity for a relation to be either kartā or karma as in *ācāryasya anuśāsanam* and *kāṣṭhasya jalanam.* The yogyatā between the referents decide the precise relation. Thus in case of these relations, morphology and syntax provide the necessary conditions. But this is not so with other relations. For example, only on the basis of morphology and syntax one can not claim that *asiḥ* is a karaṇakartṛ for *chinatti.* It is karaṇakartṛ because the referent of *asiḥ* also happens to be the karaṇa of the verb *chinn.* Thus in the case of karaṇakartṛ and karmakartṛ, only if the referent is a kartā as well as karaṇa or karma of the action indicated by the verb concerned, one can assign such relations. After

examining all the 103 tags proposed by Krishnamacharyulu(2009), we arrived at a set of only 31 relations(see appendix A) for which only morphology and syntax play as a necessary criterion.

6 Towards a more useful parse

Though the principles described above are good from the computational point of view, from a user's perspective some of these constraints are not good. Even from the point of view of information extraction, these constraints pose limitations. An information retrieval system or even an ordinary user interested in understanding the Sanskrit texts would like to

- mark the relations involving *upapada*s treating them as function words rather than content words,
- mark the relation between two sentences by semantic labels such as cause-effect (kāryakāraṇa) or reason(hetu-hetumadbhāva) etc. rather than just marking them by too general terms such as anuyogi, pratiyogi or sambandhaḥ,
- indicate the sharing of kārakas in case of conjunctive and disjunctive particles,
- mark the relations indicated by the derivational suffix in case of kṛdanta nouns, and
- show the co-indexing for anaphora resolution.

The post-processing module caters to this need. We show the relations added after this post-processing step by 'dotted arrows' so as to distinguish them from the ones which are marked before. Figs 20 and 21 show the graphs of sentences (21) and (22) after post-processing.

(21) rāmaḥ dugdham pītvā mohanena saha śālāṁ gacchati.
(22) rāmaḥ śālāṁ gacchati pāṭhaṁ ca paṭhati.

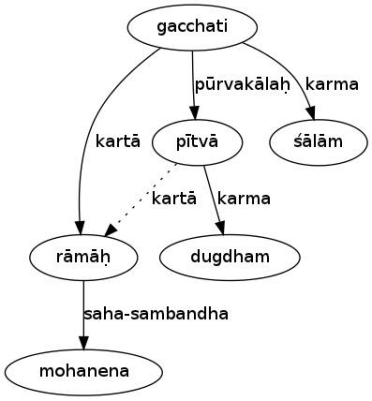

Fig 20

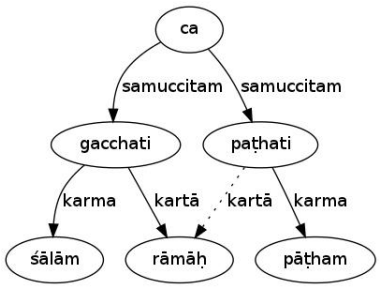

Fig 21

7 Conclusion

In this paper we discussed various issues in deciding the tagset of relations for parsing Sanskrit sentences. We followed a two stage procedure to account for the topological requirement of one-one mapping between the words and the nodes, no multiple inheritance, and no loops. In order to be faithful to the coded information, and also from the point of view of information retrieval, we process the parsed tree further, and a) collapse nodes corresponding to function words, b) make the information related to sharing explicit, and c) show the co-indexing.

This two stage parser has been implemented and is available as a part of Sanskrit-Hindi Machine Translation system at http://sanskrit.uohyd.ernet.in/scl/SHMT/shmt.html.

8 Acknowledgement

Authors thank the anonymous reviewers for useful suggestions. This work is a part of the Sanskrit Consortium project entitled 'Development of Sanskrit computational tools and Sanskrit-Hindi Machine Translation system' sponsored by TDIL Programme, DIT, Government of India.

A Table showing the recommended relations

kartā	prayojyakartā	prayojakakartā	karma
karaṇam	sampradānam	apādānam	ṣaṣṭhīsambandhaḥ
adhikaraṇam	sambodhanasūcakam	sambodhyaḥ	
hetuḥ	prayojanam	tādarthya	niṣedhyaḥ
kriyāviśeṣaṇam	viśeṣaṇam	śeṣasambandhaḥ	nirdhāraṇam
upapadasambandhaḥ	sambandhaḥ	pratiyogī	anuyogī
samuccitam	anyataraḥ	kartṛsamānādhikaraṇam	karmasamānādhikaraṇam
samānakālaḥ	anantarakālaḥ	pūrvakālaḥ	vīpsā

References

1. Akshar Bharati and Rajeev Sangal. *Parsing Free Word Order Languages in the Paninian Framework*, In Proc. of ACL: 93.
2. Akshar Bharati and Vineet Chaitanya and Rajeev Sangal. *Natural Language Processing: A Paninian Perspective* Prentice-Hall, New Delhi, 1995.
3. John Carroll, Guido Minnen, and Ted Briscoe. *Corpus annotation for parser evaluation.* In Proceedings of the EACL, 1999
4. Euguen Charniak and Mark Johnson. *Coarse- to-fine n-best parsing and MaxEnt discriminative re-ranking* In Proceedings of the 43rd annual meeting of the ACL, pp. 173-180. 2005.

5. Ann Copestake and Dan Flickinger. *An open-source grammar development environment and broad-coverage English grammar using HPSG* In Proceedings of the Second conference on Language Resources and Evaluation (LREC-2000), Athens, Greece.
6. Gadādhara Bhaṭṭāchārya *Vyutpattivādaḥ*, Chaukhamba Sanskrit Sansthanam, Varanasi, 1988.
7. Marie-Catherine de Marneffe and Christopher D. Manning. *The Stanford typed dependencies representation.* In COLING Workshop on Cross-framework and Cross-domain Parser Evaluation, 2008. URL http://nlp.stanford.edu/pubs/dependencies-coling08.pdf.
8. Tracy H. King, Richard Crouch, Stefan Riezler, Mary Dalrymple, and Ronald Kaplan. *The PARC 700 dependency bank.* In 4th International Workshop on Linguistically Interpreted Corpora (LINC-03).
9. K V R Krishnamacharyulu. *Annotating the Sanskrit Texts based on the śābdabodha systems* In 3rd International Sanskrit Computational Symposium, Ed. Amba Kulkarni and Gérard Huet, LNAI Springer Verlag, 2009
10. Amba Kulkarni, Sheetal Pokar and Devanad Shukl. *Designing a constraint based parser for Sanskrit* In 4th International Sanskrit Computational Symposium, Ed. G N Jha, LNAI Springer Verlag, 2010
11. Amba Kulkarni, Devanad Shukl and Sheetal Pokar. *Mathematical modelling of Akāṅkshā and Sannidhi for parsing Sanskrit* In 15th World Sanskrit Conference, 2012, Delhi.
12. I. Schröder. *Natural Language Parsing with Graded Constraints.* PhD thesis, Hamburg Univ., 2002
13. N S R Tatachārya. Śābdabodhamīmāṃsā: The sentence and its significance Part-I, Institute of Pondicherry and Rashtriya Sanskrit Sansthan, New Delhi, 2005