Parsing Sanskrit Texts: Some relation specific issues

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

Department of Sanskrit Studies,
University of Hyderabad,
Hyderabad

Department of Vyakarana,
Rashtriya Sanskrit Vidyapeetha,
Tirupati

Sentence Level Analysis

Parsing: a linear string of words $->$ a structure showing the relations between words.

Structure: Constituency / Dependency

Constituency Structure: Eng Sentence

For positional languages such as English, constituency structure more or less leads to the dependency relations.

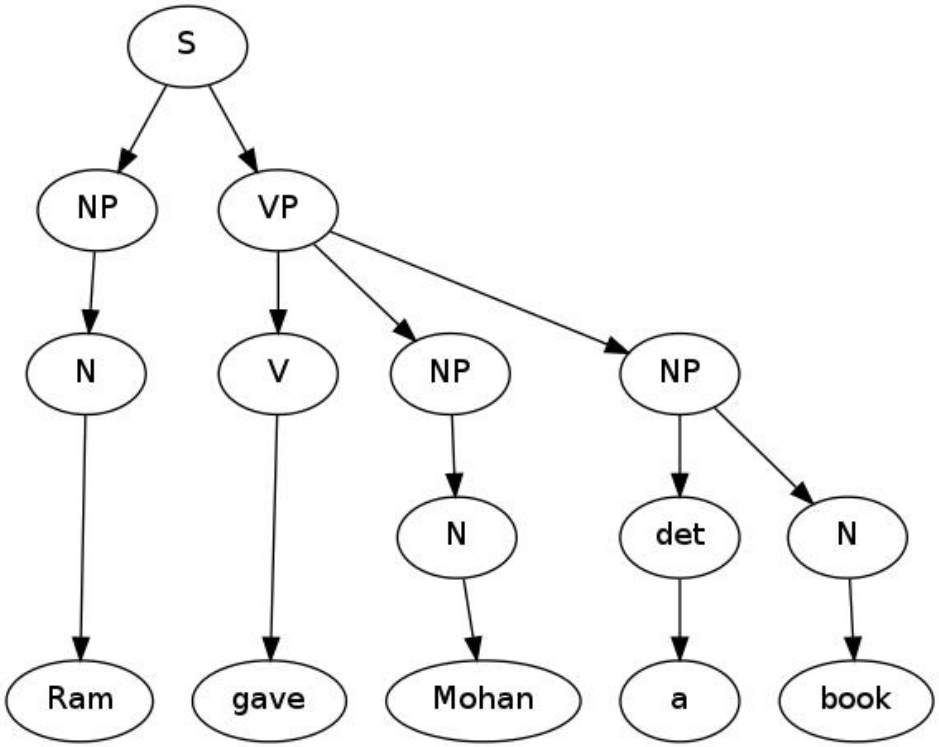


Figure: Constituency Parse of an English Sentence

Constituency Structure: Sanskrit Compound

In case of Sanskrit compounds, such a constitutional structure is very useful in determining the parse.

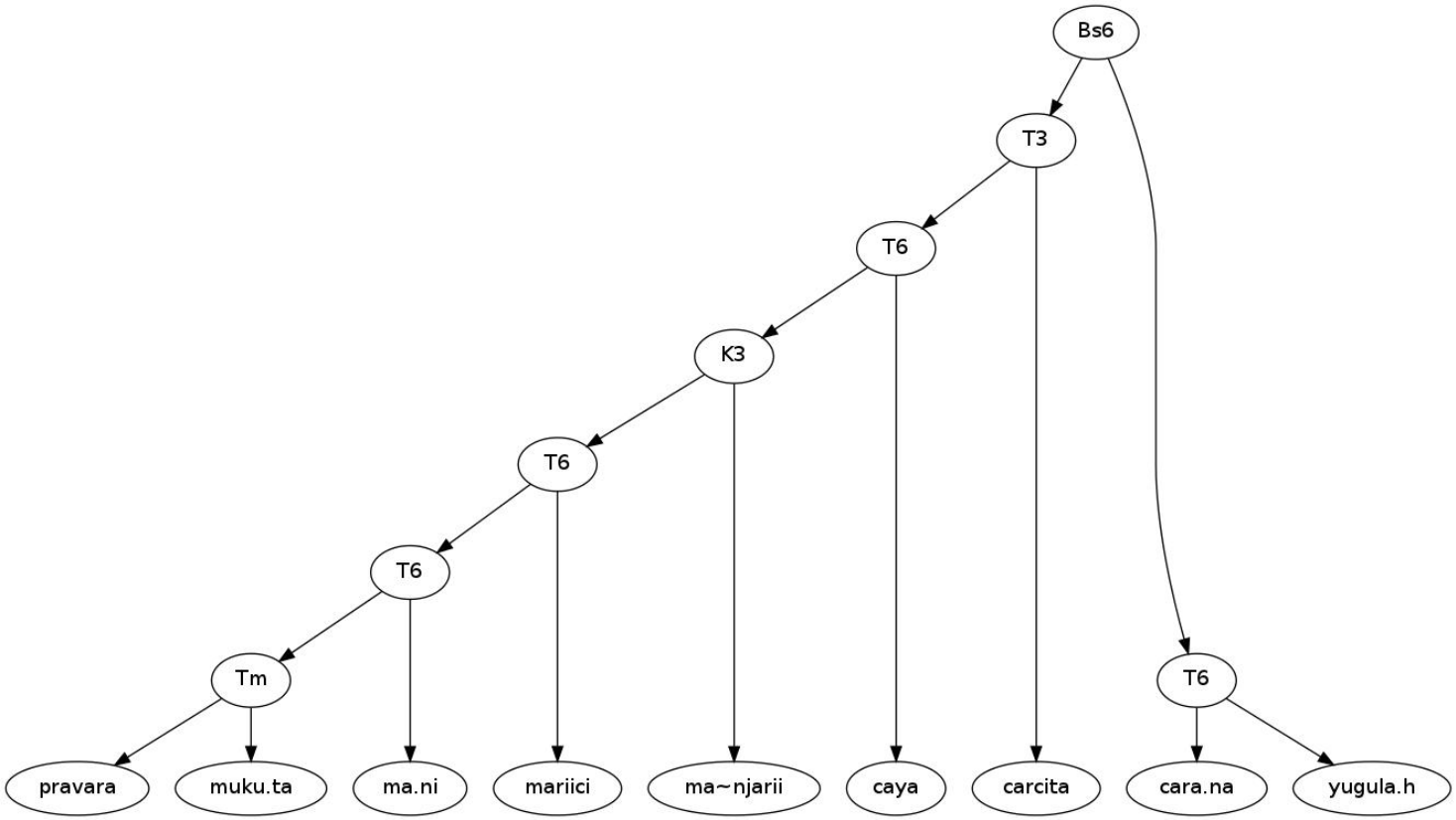


Figure: Constituency Structure of a Sanskrit Compound

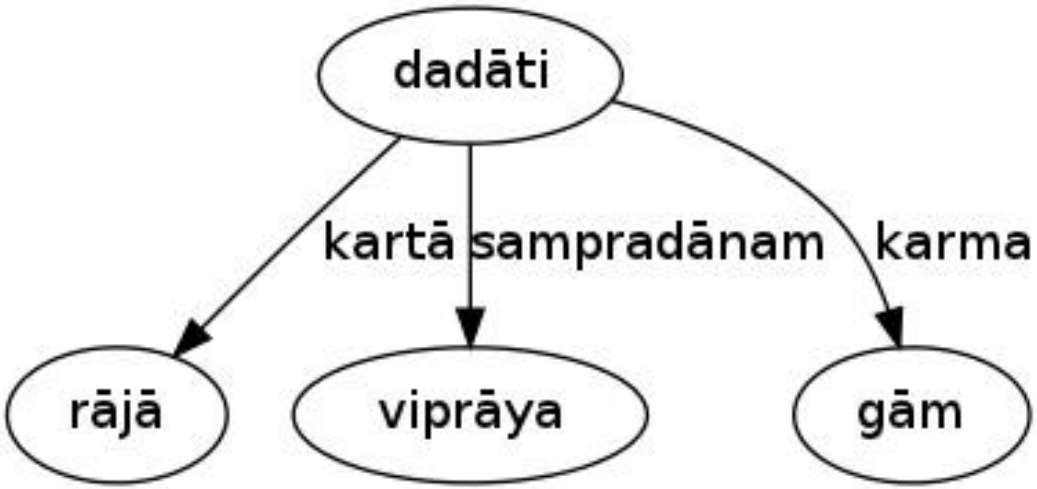
Structure of a Sanskrit Sentence

Sanskrit:

- Morphologically rich
- Free word order to a large extent

Dependency structure makes more sense than the constituency structure.

Dependency Structure: Sanskrit Sentence



Dependency Parse: *rājā viprāya gām dadāti*

Sentence Level Analysis

- Traditional Approaches:
  - Daṇḍānvyayaḥ [Anvyayamukhī]
  - Khaṇḍānvyayaḥ [Kathambhūtinī]

Sentence Level Analysis ...

- Modern Times
  - Dependency Parsing: credited to Tesniére [1959]

Tagset

Dependency Relations:

- Compilation of all relations: 90 [KVRK, 2009]
- Proposed a tagset

Tagset ...

Is this tagset suitable?

Tagset ...

Is this tagset suitable?

- For Manual Annotation
- For Statistical Parsing
- For Rule Based Parsing

Tagset ...

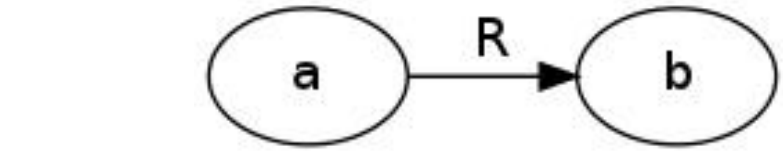
Concerns related to Manual Annotation / design of a Statistical Parser

- The inter annotator agreement
- The grey / fussy boundaries between semantics of tags lead to errors in annotation.

Concerns related to a design of a Rule based Parser

- Whether the relations can be decided purely on the basis of morphological and syntactic information?

Annotation Convention



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

Clues for extracting the relations

- Abhihitatva (property of being expressed)
- Vibhakti
- Indeclinables (avyaya)
- Sāmānādhikaraṇya (being in g-n-p agreement)
- Nitya sambandhaḥ

Clues: Abhitatva

- tiṅ
  - *rāmaḥ vanam gacchati.*
    marks the kartā relation
  - *rāmeṇa vanam gamyate.*
    marks the karma relation
- kṛt
  - *dhāvan aśvaḥ.*
    marks the kartā.
- taddhitaḥ
- samāsaḥ

Clues: Vibhakti

- kāraka vibhakti
  Marks noun-verb relations.
  e.g. kartā, karma, etc.
- upapada vibhakti
  Marks noun-noun relations through the upapadas.
  More of a morphological requirement.
  E.g. *rāmeṇa saha sītā vanam gacchati.*
- special vibhakti
  - kriyā-viśeṣaṇam
    vegena dhāvati
  - aṅgavikāraḥ
    akṣṇā kāṇaḥ
  - nirdhāraṇa
    nareṣu śreṣṭaḥ

Clues: Indeclinables

- negation
  rāmaḥ gṛham na gacchati.
- emphasis
  rāmaḥ eva tatra upaviṣati.

Clues: sāmānādhikaraṇya

- *śvetaḥ aśvaḥ dhāvati.*
- *aśvaḥ śvetaḥ asti.*

Clues: Nitya Sambandhaḥ

Relation between yat-tat pair

- *yadā - tadā*
- *yatra - tatra*

Choice of the relations

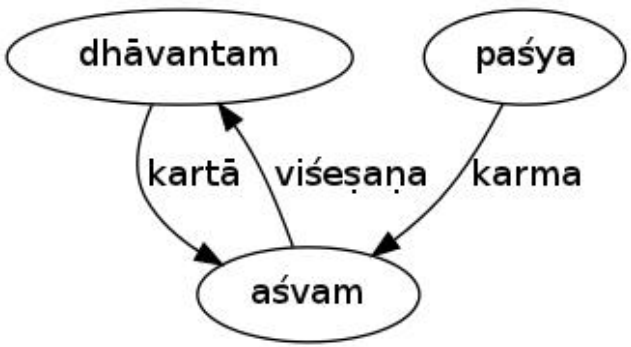
- Should the inflectional suffixes and derivational suffixes be treated at par?
- How to treat the function words? Should they be treated as a node in a tree or an edge?
- How to represent the inter sentential relations?
- Should anaphoric resolution be part of this annotation?

Basic Principles

1. Preserve one-one mapping between the nodes of a tree and the words in a sentence.
2. In case of derived nouns, consider only the inflectional suffix for establishing the relations.
3. In case of derived indeclinables, use the derived suffix to mark the relations.
4. A suffix or a word can represent one and only one relation.

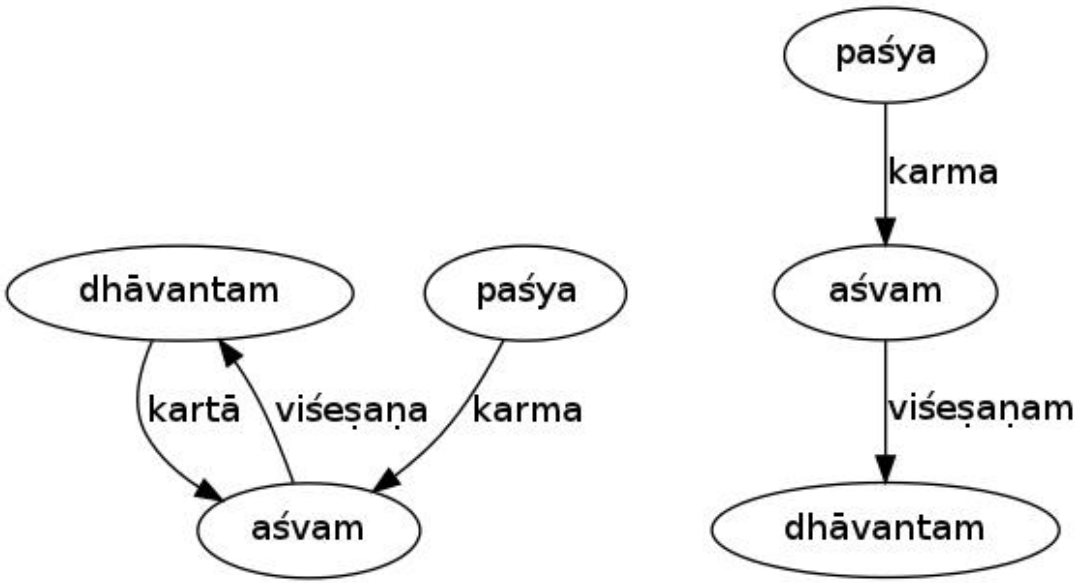
Which relation to mark?

Example: dhāvantam aśvam paśya.



Which relation to mark? Contd...

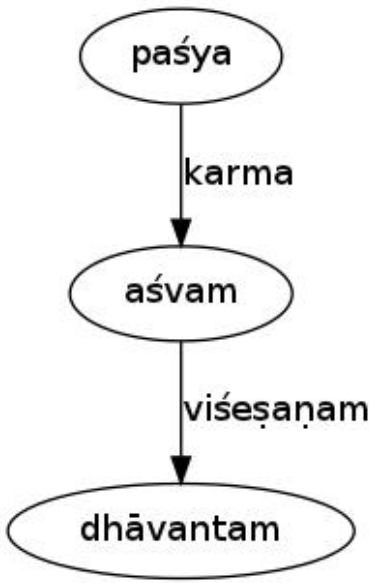
Example: dhāvantam aśvam paśya.



A loop destroys the nice tree structure of a parse. Hence mark only relations indicated through the inflectional suffixes and not the ones indicated through the derivational suffixes.

Which relation? .. contd

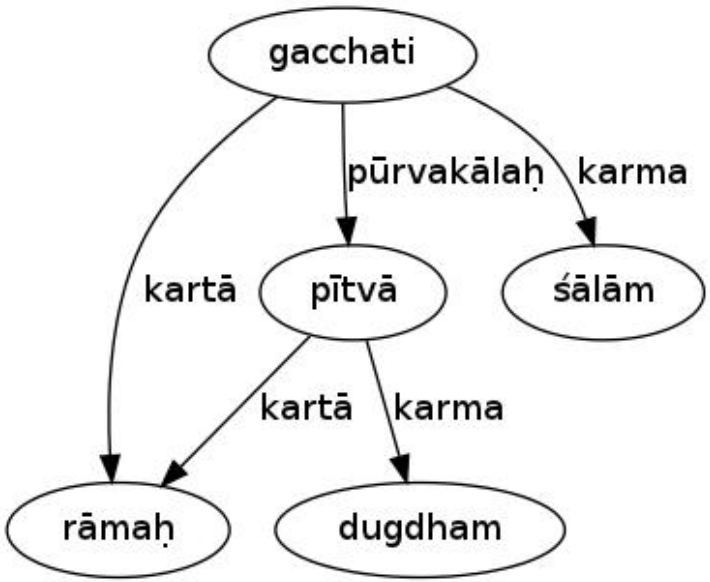
What about the faithfulness?



Information is available through the kṛt suffix, which is available for marking the relation of kartā at a later stage.

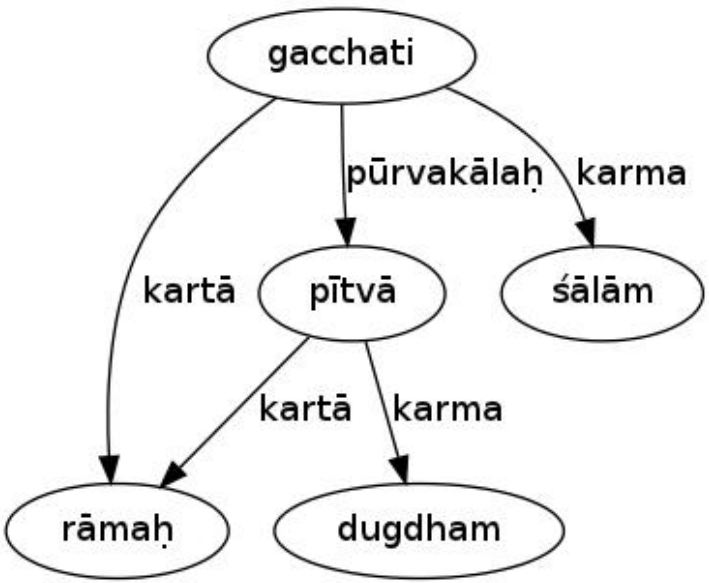
Treating indeclinables:

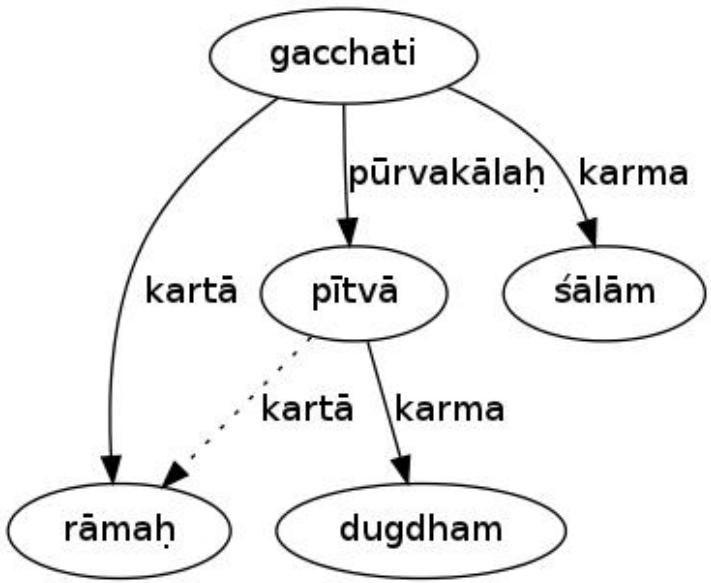
(1) kṛdanta avyayas:
rāmaḥ dugdham pītvā śālām gacchati.



Treating indeclinables:

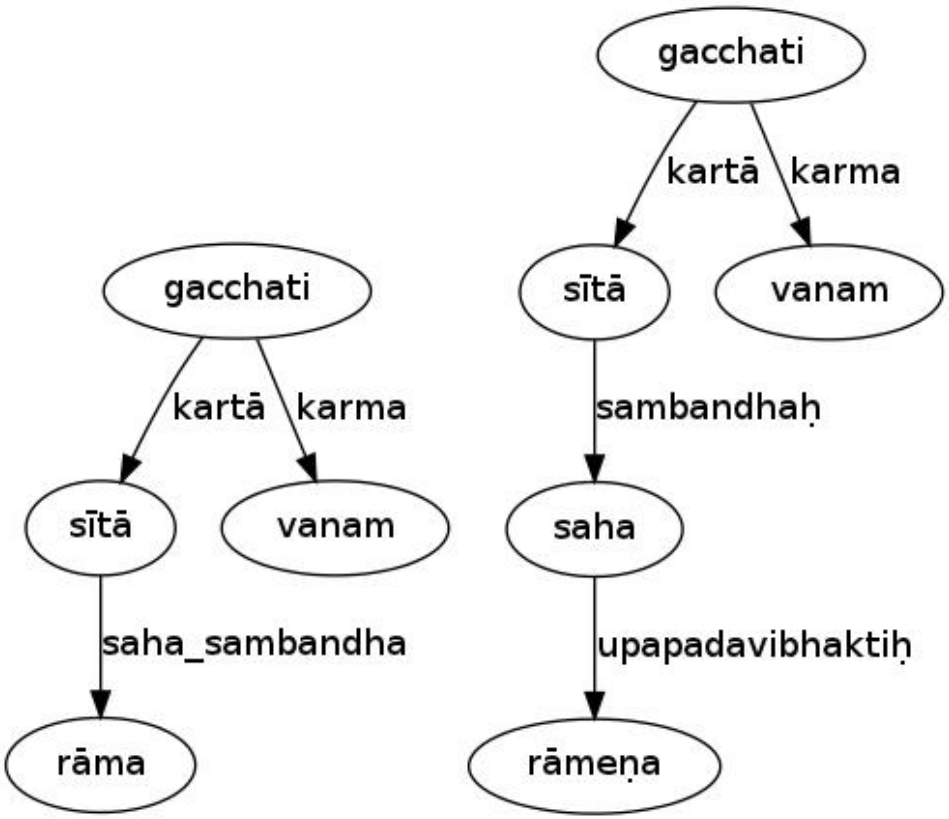
(1) kṛdanta avyayas:
rāmaḥ dugdham pītvā śālām gacchati.





Treating indeclinables:Content / function ?

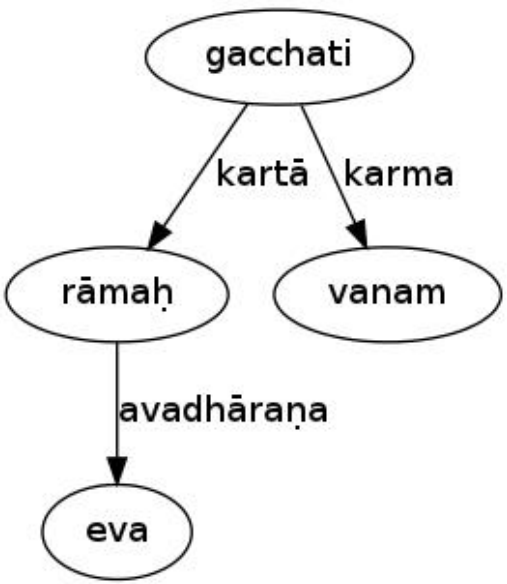
(2) upapada avyayas
sītā rāmeṇa saha vanam gacchati.

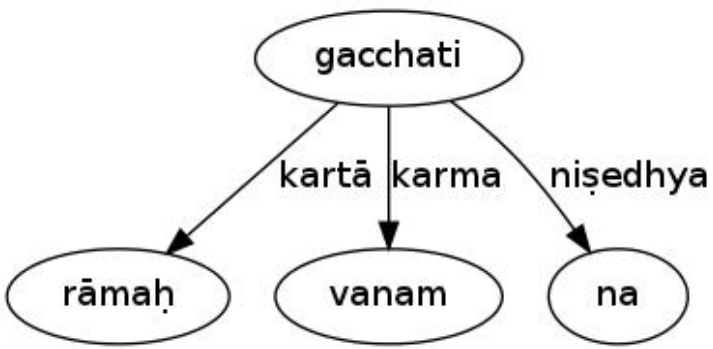


Treating indeclinables:Content / function ?

(3) Rest of the avyayas

- *eva – avadhāraṇā*
- *na – niṣedhya*



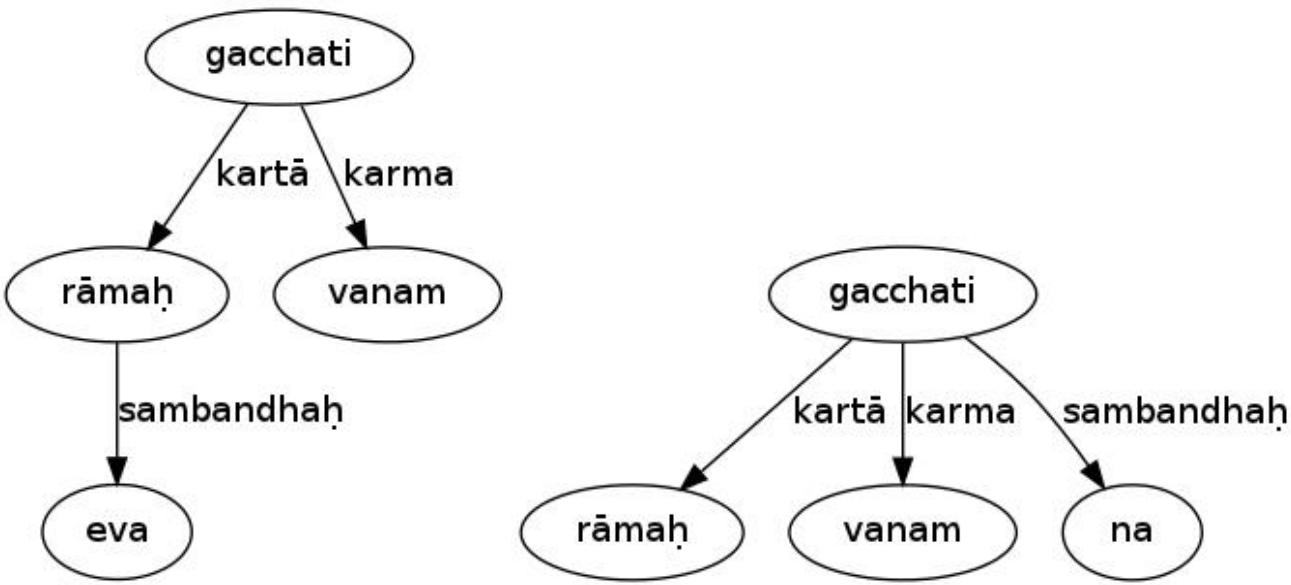


Number of relations explode.

Treating indeclinables:Content / function ?

(3) Rest of the avyayas

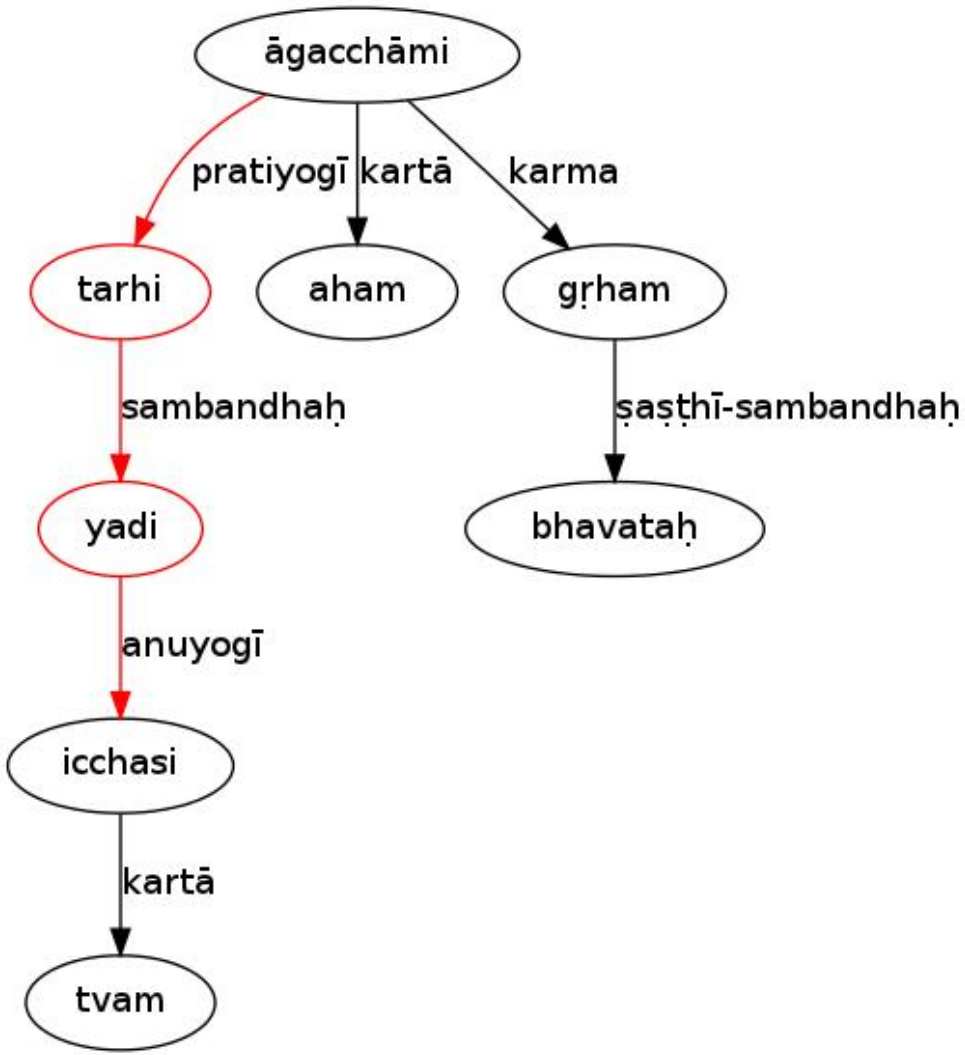
- *eva – avadhāraṇā*
- *na – niṣedhya*



Number of relations explode.

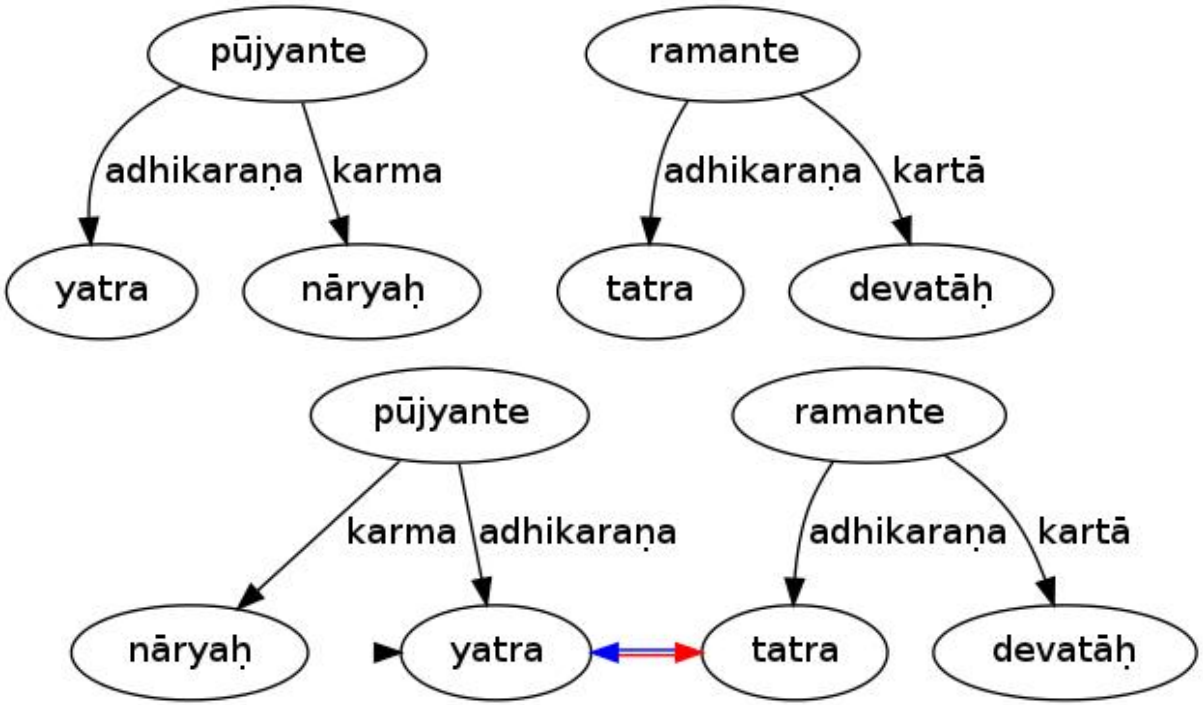
Inter Sentential Connectives

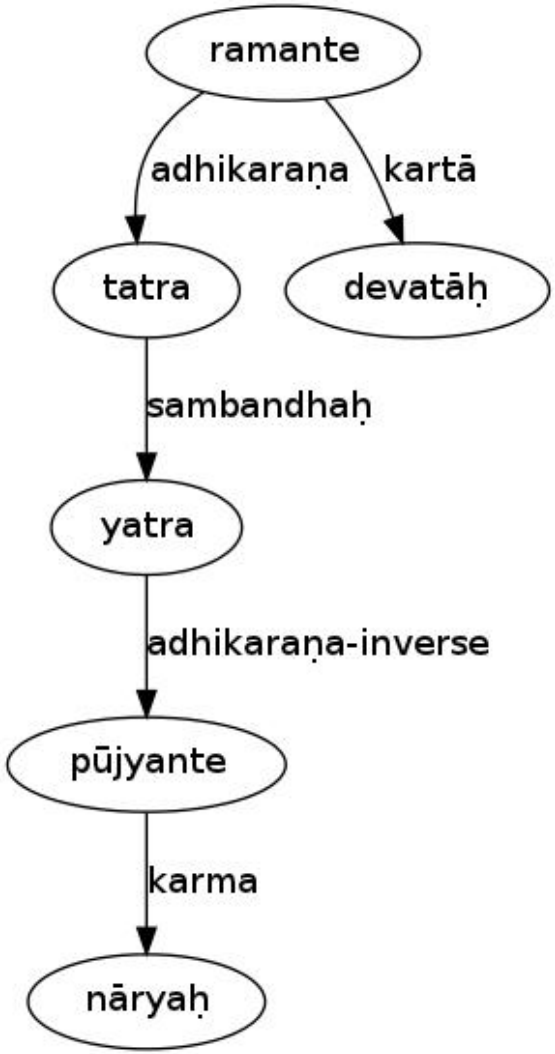
yadi tvam icchasi tarhi aham bhavataḥ gṛham āgacchāmi.



Treatment of Anaphoras

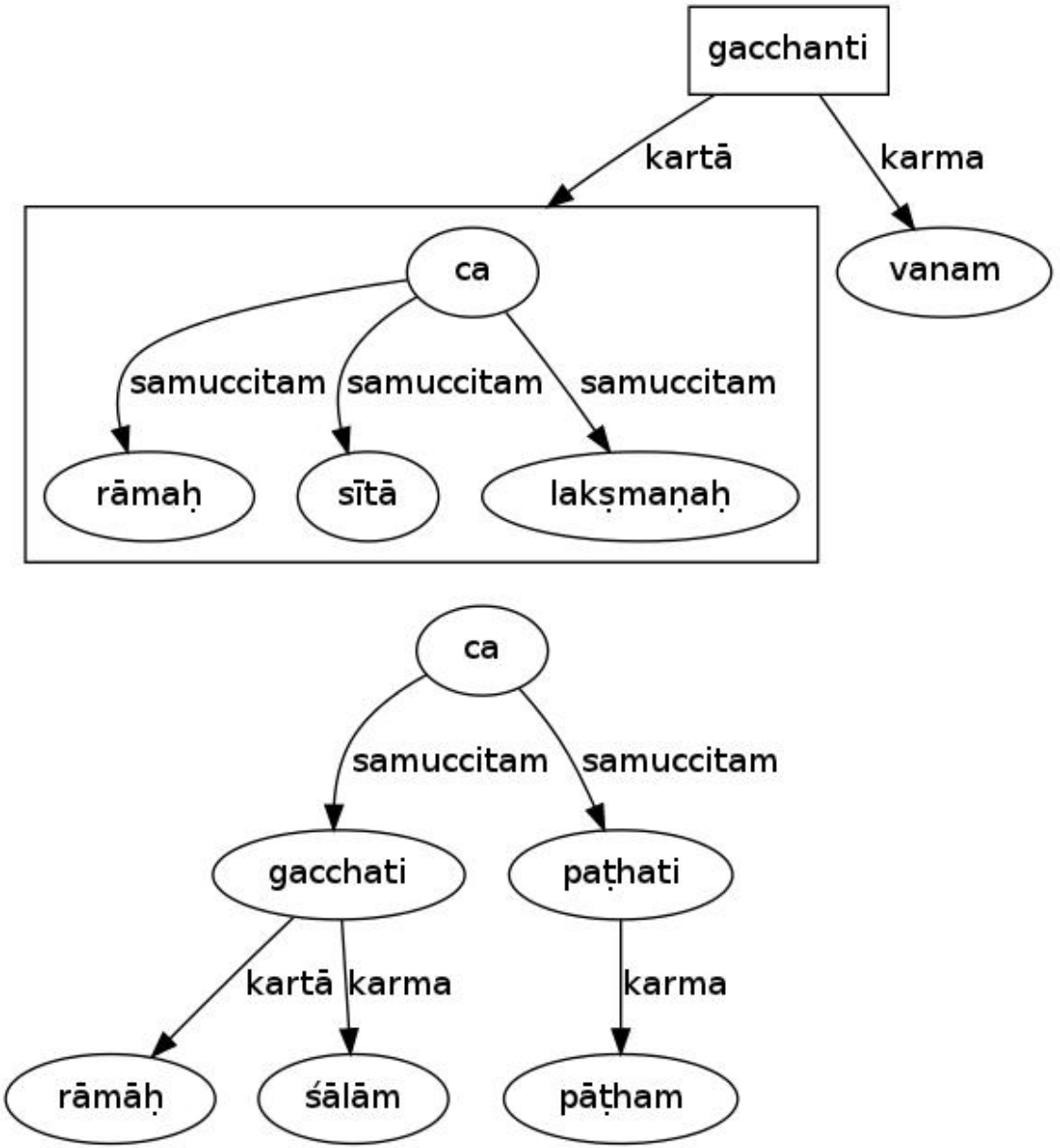
yatra nāryaḥ pūjyante ramante tatra devatāḥ




ramante
adhikaraṇa
kartā
tatra
devatāḥ
sambandhaḥ
yatra
adhikaraṇa-inverse
pūjyante
karma
nāryaḥ

Treatment of Conjunction and Disjunction

rāmaḥ sītā lakṣmaṇaḥ ca vanam gacchanti



Granularity

Criterion for Granularity:
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
Tradition clasifies *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.

Granularity ...

- *karma-kartṛ*
  Ex: **kāṣṭhaḥ** svayameva bhidyate.
- *karaṇa-kartṛ*
  Ex: **asiḥ** chinatti.
- *ṣaṣṭhī kartā*
  Ex: **ācāryasya** anuśāsanam

Necessary conditions for

- *prayojaka kartā*
- *prayojya kartā*
- *ṣaṣṭhī kartā*

are

- Morphology
- Syntax

- *devadattena annam pācayati.* (*prayojya kartā*)
- *devadattaḥ agninā annam pācayati.* (*karaṇa*)

asiḥ chinatti
Here *asiḥ* is *karaṇakartṛ* of *chinn* only because it is a *karaṇa* for *chinn*.

Semantics is involved in the decision.

31 relations: Necessary information for deciding the possibility involves only morphology and Syntactic information.

Set of Relations

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

Towards a more useful parse

- Treat *upapada*s as function words rather than content words
- Show the co-indexing for anaphora resolution
- Show the sharing of kārakas

DEMO