Computational Sanskrit & Digital Humanities

Selected papers presented
at
the 17^{th} World Sanskrit Conference

University of British Columbia, Vancouver
9–13 July 2018

D K Publishers Distributors Pvt. Ltd.

Edited by
Gérard Huet & Amba Kulkarni

First Published: 2018

ISBN: **978-93-87212-10-7**

Published by
D K Publishers Distributors Pvt. Ltd.
4224/1, Ansari Road, Daryaganj,
New Delhi-110002, India
Ph: 011-41562573-77 Fax: 011-41562578
E-mail: info@dkpd.com
www.dkpd.com

Printed at:
D K Fine Art Press P Ltd.
A-6, Nimri Colony, Ashok Vihar, Phase-IV
Delhi-110052, India

Contents

Yogyatā as an absence of non-congruity

Sanjeev Panchal *and* Amba Kulkarni

Abstract: Yogyatā or mutual congruity between the meanings of the related words is an important factor in the process of verbal cognition. In this paper, we present the computational modeling of yogyatā for automatic parsing of Sanskrit sentences. Among the several definitions of yogyatā we modeled it as an absence of non-congruity. We discuss the reasons behind our modeling.

Due to lack of any syntactic criterion for viśeṣaṇa (adjectives) in Sanskrit, parsing Sanskrit texts with adjectives resulted in high number of false positives. Hints from the vyākaraṇa texts helped us in the formulation of a criterion for viśeṣaṇa with syntactic and ontological constraints, which provided us a clue to decide the absence of non-congruity between two words with respect to the adjectival relation. A simple two way classification of nouns into dravya and guṇa with further sub-classification of guṇas into guṇavacanas was found to be necessary for handling adjectives. The same criterion was also necessary to handle the ambiguities between a *kāraka* and *non-kāraka* relations. These criteria together with modeling *yogyatā* as an absence of non-congruity resulted in 81% improvement in the precision.

1 Introduction

Three factors viz. ākāṅkṣā (expectancy), yogyatā (congruity) and sannidhi (proximity) play a crucial role in the process of śābdabodha (verbal cognition). These factors have been found to be useful in the development of a Sanskrit parser as well. The concept of subcategorisation of modern Linguistics comes close to the concept of ākāṅkṣā. Subcategorisation structures provide syntactic frames to capture different syntactic behaviours of verbs. Sanskrit being an inflectional language, the information of various relations is encoded in suffixes rather than in positions. These suffixes express the expectancy, termed as ākāṅkṣā in the Sanskrit literature. Kulkarni, Pokar,

and Shukl (2010) describe how the ākāṅkṣā was found to be useful in the proposition of possible relations between words. Sannidhi has been found to be equivalent to the weak non-projectivity principle (Kulkarni, P. Shukla, et al. 2013c). In this paper we will discuss the role of the third factor viz. yogyatā, in building a Sanskrit parser.

The concept of selection restriction is similar to the concept of yogyatā. The expectancy, or the ākāṅkṣā, proposes a possible relation between the words in a sentence. Such a relation would hold between two words only if they are meaning-wise compatible. It is the selection restriction or yogyatā which then comes into force to prune out incongruent relations, keeping only the congruent ones. Katz and Fodor (1963) proposed a model of selection restrictions as necessary and sufficient conditions for semantic acceptability of the arguments to a predicate. Identifying a selection restriction that is both necessary and sufficient is a very difficult task. Hence there were attempts to propose alternatives. One such alternative was proposed by Wilks (1975) who viewed these restrictions as preferences rather than necessary and sufficient conditions. After the development of WordNet, Resnik (1993) modeled the problem of induction of selectional preferences using the semantic class hierarchy of WordNet. Since then there is an upsurge in the field of computational models for the automated treatment of selectional preferences with a variety of statistical models and Machine learning techniques. In recent times, one of the ambitious projects to represent World Knowledge was taken up under the banner of Cyc. This knowledge base contains over five hundred thousand terms, including about seventeen thousand types of relations, and about seven million assertions relating these terms.[1] In spite of the availability of such a huge knowledge base, we rarely find Cyc being used in NLP applications.

The first attempt to use the concept of yogyatā in the field of Machine Translation was by the Akshar Bharati group (Bhanumati 1989) in the Telugu-Hindi Machine Translation system. Selectional restrictions were used in defining the `Kāraka Charts` that provided a subcategorisation frame as well as semantic constraints over the arguments of the verbs. On similar lines `Noun Lakṣaṇa Charts` and `Verb Lakṣaṇa Charts` were also used for disambiguation of noun and verb meanings. These charts expressed selectional restrictions using both ontological concepts as well as semantic

[1] `http://www.cyc.com/kb`, accessed on 30th August, 2017

properties. An example `Kāraka chart` for the Hindi verb `jānā` (to go) is given in table 1.

case relation	necessity	case marker	semantic constraint
apādānam (source)	desirable	se	not (upādhi:vehicle)
karaṇam (instrument)	desirable	se	(upādhi:vehicle)
karma(object)	mandatory	0/ko	-
kartā (agent)	mandatory	0	-

Table 1
Kāraka Chart for the verb jānā (to go)

Here upādhi is an imposed property. The first row in Table 1 states a constraint that a noun with case marker `se` has a kāraka role of apādānam (source) provided it is not a vehicle. The ontological classification was inspired from the ontology originated from the vaiśeṣika school of philosophy. The parsers for Indian languages were further improved. Bharati, Chaitanya, and Sangal (1995) mentions the importance of two semantic factors viz. animacy and humanity, in parsing, that remove the ambiguity among the kartā and karma(roughly subject and object). This hypothesis was further strengthened with experimental verification by Bharati, Husain, et al. (2008).

In the next section, we first state the importance of yogyatā in parsing, as a filter to prune out meaningless parses. Since yogyatā deals with the compatibility between meanings, and a word expresses meanings at different levels, we also discuss the mutual hierachy among these various meanings. In the third section, we look at various definitions of yogyatā offered in the tradition, and decide the one that is suitable for implementation. In the same section, we evolve strategies to disambiguate relations based on yogyatā. Finally the criteria evolved for disambiguation are evaluated. The evaluation results are discussed in section four, followed by the conclusion.

2 Yogyatā as a filter

Necessary condition for understanding a sentence is that a word having an expectancy for another word should become nirākāṅkṣa (having no further

expectancy) once a relation is established between them. Further, such related words should also have mutual compatibility from the point of view of the proposed relation. If they are not, then the expectancy of such words will not be put to rest and there would not be any verbal cognition. Therefore the role of yogyatā in verbal cognition is very important. The purpose of using yogyatā in parsing is not to make a computer 'understand' the text, but to rule out incompatible solutions from among the solutions that fulfill the ākāṁkṣās. For example, in the sentence

Skt: *yānam vanam gacchati.*
Gloss: vehicle{neut., sg., nom./acc.} forest{neut., sg., nom./acc.} go{present, 3rd per., sg.}

There are 6 possible analyses, based on the ākāṅkṣā. They are

1. `yānam` is the kartā and `vanam` is the karma of the verb gam,
2. `yānam` is the karma and `vanam` is the kartā of the verb gam,
3. `yānam` is the kartā of the verb gam and `vanam` is the viśeṣaṇa of `yānam`,
4. `yānam` is the karma of the verb gam and `vanam` is the viśeṣaṇa of `yānam`,
5. `yānam` is the viśeṣaṇa of `vanam` which is the kartā of the verb gam,
6. `yānam` is the viśeṣaṇa of `vanam` which is the karma of the verb gam.

If the machine knows that the kartā of an action of going should be movable, and that the designation of `yāna` is movable, but that of `vana` is not movable, then mechanically it can rule out the second analysis. The words `yānam` and `vanam` on account of the agreement between them have the potential to be viśeṣaṇas of each other. But the semantic incompatibility between the meanings of these words rules out the last four possibilities, leaving only the first correct analysis.

As another example, look at the sentence

Skt: *Rāmeṇa bāṇena Vālī hanyate.*
Gloss: Rama{ins.} arrow{ins.} Vali{nom.} is_killed.

`Rāma` and `bāṇa`, both being in instrumental case, can potentially be a kartā as well as a karaṇam of the verb `han` (to kill). If the machine knows that `bāṇa` can be used as an instrument in the act of killing, while `Rāma` being the name of a person, can not be a potential instrument in the act of

killing, it can then filter out the incompatible solution: `Rāma` as a karaṇam and `bāṇa` as a kartā.

Look at another sentence *payasā siñcati* (He wets with water). Here *payas* (water) is in instrumental case, and is a liquid, and hence is compatible with the action of `siñc` (to wet). But in the sentence *vahninā siñcati* (He wets with fire), `vahni` (fire) is not fit to be an instrument of the action of wetting, and as such it fails to satisfy the yogyatā. But now imagine a situation where a person is in a bad mood, and his friend without knowing it starts accusing him further for some fault of his, instead of uttering some soothing words of console. Third person watching this utters *kim vahninā siñcasi* (Why are you pouring fire?) - a perfect verbalisation of the situation. The words, here, are like a fire to the person who is already in a bad mood. This meaning of `vahni` is its extended meaning. Thus, even if a relation between primary meanings does not make sense, if the relation between extended meanings makes sense, we need to produce the parse. Therefore, in addition to the primary meanings, machine also, sometimes, needs an access to the secondary / extended meanings of the words.

2.1 Word and its Meanings

Every word has a significative power that denotes its meaning. In Indian theories of meaning, this significative power is classified into three types viz. abhidhā (the primary meaning), lakṣaṇā (the secondary or metaphoric meaning) and vyañjanā (the suggestive meaning). In order to use the concept of yogyatā in designing a parser, we should know what is the role of each of these meanings in the process of interpretation.

The secondary meaning comes into play when the primary meaning is incompatible with the meanings of other words in a sentence. The absence of yogyatā is the basic cause for this signification. Indian rhetoricians accept three conditions as necessary for a word to denote this extended or metaphoric sense. These three conditions are[2]

1. inapplicability / unsuitability of the primary meaning,
2. some relation between the primary meaning and the extended meaning, and

[2] *mukhyārthabādhe tadyoge rūḍhito'tha prayojanāt |*
anyo'rtho lakṣyate yat sā lakṣaṇāropitā kriyā ||
`(KP II 9)`

3. definite motive justifying the extension.

In addition to these two meanings, there is one more meaning, called vyañjanā or the suggestive meaning. This corresponds to the inner meaning of any text / speaker's intention. In order to understand this meaning, consider a sentence *gato'stam arkaḥ* which literally means 'the sun has set'. Every listener gets this meaning. In addition to this meaning, it may also convey different signals to different listeners. For a child playing in the ground, it may mean 'now it is getting dark and it is time to stop playing and go home', for a Brahmin, it may mean 'it is time to do the sandhyāvandana', and for a young man it may mean 'it is time to meet his lover'. This extra meaning co-exists with the primary meaning. It does not block the primary meaning. Therefore vyaṅgārtha (suggestive meaning) exists in parallel with the primary / secondary meaning.

Since the suggestive meaning is in addition to the primary / secondary meaning, and is optional, and also is different for different listeners, it involves subjectivity for processing. Hence it is not possible to objectively process this meaning for any utterance.[3] This also puts an upper limit on the meaning one can get from a linguistic utterance without the interference of subjective judgments. In summary, we observe that these three meanings are not in the same plane. Lakṣaṇā comes into play only when abhidhā fails to provide a suitable meaning for congruent interpretation. And the suggestive meaning can co-exist with the abhidhā as well as the lakṣaṇā, and as such, is outside the scope of automatic processing.

3 Modeling Yogyatā

Yogyatā is compatibility between the meanings of related words. This meaning, as we saw above, can be either a primary or a metaphoric one. Absence of any hindrance in the understanding of a sentence implies there is yogyatā or congruity among the meanings. There have been different views among scholars about what yogyatā is. According to one definition, yogyatā is *artha-abādhaḥ*[4] (that which is not an hindrance to meaning). It

[3] One of the reviewers commented that taking into account the advents in Big Data and Machine Learning techniques, it may even be possible to process such meanings by machines in future. However, we are of the opinion that machine would need semantically annotated corpus for learning, which does not yet exist.

[4] All the meanings we will be discussing below are found in NK p. 675.

is further elaborated as *bādhaka-pramā-virahaḥ* or *bādhaka-niścaya-abhāvaḥ* (absence of the decisive knowledge of incompatibility). There are other attempts to define it as an existing qualifying property. One such definition is *sambandha-arhatvam* (eligibility for mutual association), and the other one is *paraspara-anvaya-prayojaka-dharmavattvam* (a property of promoting mutual association). The first set of definitions presents yogyatā as an absence of incompatibility whereas the second set of definitions present it as the presence of compatibility between the meanings.

Let us see the implications of modeling yogyatā through these two lenses.

1. We establish a relation only if the two morphemes are mutually congruous.
   In this case we need to take care of not only the congruity between primary meanings but even between the metaphoric / secondary meanings.
2. We establish a relation if there is no incongruity between the two meanings.

The first possibility ensures that the precision is high and there is a less chance of Type-1 error, i.e. of allowing wrong solutions. The second possibility, on the other hand, ensures that the recall is high and there is less chance of Type-2 error, viz. the possibility of missing any correct solution. But there is a chance that we allow some unmeaningful solutions as well. If we decide to go for the first possibility, we need to handle both the primary as well as secondary meanings, and we need to state precisely under what conditions the meanings are congruous. And this means modeling congruity for each verb and for each relation. This is a gigantic task, and there is a possibility of missing correct solutions, if we do not take into account all the possible extensions of meanings. Therefore, we decided to go for the second choice allowing a machine to do some mistakes of choosing incongruous solutions but we did not want to throw away correct solutions even by mistake. This decision is in favour of our philosophy of sharing the load between man and machine. Our aim is to provide access to the original text by reducing the language learning load. So we can not afford to miss a possible solution. Thus at the risk of providing more solutions than the actual possible solutions, we decided to pass on some load to the reader of pruning out irrelevant solutions manually.

In the first step, we decided to use yogyatā only in those cases where a case marker is ambiguous between more than one relation. We noticed the following three cases of ambiguities with reference to the relations.

1. viśeṣya-viśeṣaṇa-bhāva (adjectival relation)
   Here both the viśeṣya and viśeṣaṇa agree in gender, number and case, and hence only on the basis of the word form, we can not tell which one is viśeṣya and which one is viśeṣaṇa.
2. a kāraka and a non-kāraka relation as in
   a. karaṇam (instrument) and hetu (cause), with an instrumental case marker,
   b. sampradānam (beneficiary), prayojanam (purpose) and tādarthya (being intended for), with a dative case marker,
   c. apādānam (source) and hetu (cause), with an ablative case marker.
3. śaṣṭhī sambandha (a genitive relation) and a viśeṣaṇa (an adjective)
   When two words are in the genitive case, it is not clear whether there is an adjectival relation between them, or a genitive relation.

We now discuss each of these three cases below.

3.1 Viśeṣya-viśeṣaṇa-bhāva (Adjectival relation)

We come across a term samānādhikaraṇa (co-reference) in Pāṇini to denote an adjective (Joshi and Roodbergen 1998: p. 6). One of the contexts in which the term samānādhikaraṇa is used is the context of an agreement between an adjective and a noun.[5] For example, *dhāvantaṁ mṛgaṁ* (a running deer), or *sundaraḥ aśvaḥ* (a beautiful horse). Pāṇini has not defined the term samānādhikaraṇa, either. The term samānādhikaraṇa (co-reference) literally means 'having the same locus'. Patañjali in the Samartha-āhnika discusses the term sāmānādhikaraṇya (co-referential) (literally a property of being in the same locus). In the example, *sundaraḥ aśvaḥ* (a beautiful horse), both the qualities of saundarya (beauty) and aśvatva (horse-ness) reside in an aśva (horse), which is the common locus. Similarly, in the case of *ācāryaḥ droṇaḥ*, or *agne gṛhapate* (O Agni! house-holder), both the words ācārya as well as droṇa refer to the same individual, so do agni

[5] sāmānādhikaraṇyam ekavibhaktitvam ca. dvayoścaitad bhavati. kayoḥ. Viśeṣaṇa-viśeṣyayoḥ vā sañjñā-sañjñinorvā (MBh 1.1.1)

and gṛhapati. This is true of various other relation-denoting terms such as `guru`, `śiṣya`, `pitā`, `putra`, etc. and upādhis (imposed / acquired properties) such as `rājā`, `mantrī`, `vaidya`, etc. From all this discussion, we may say sāmānādhikaraṇya (the property of having the same locus) is the semantic characterisation of a viśeṣaṇa.

In Sanskrit, there is no syntactic / morphological category as a viśeṣaṇa (an adjective). The gender, number and case of a viśeṣaṇa follows that of a viśeṣya (the head). From the point of view of analysis this provides a syntactic clue for a possible viśeṣya-viśeṣaṇa-bhāva between two words such as in *śuklaḥ paṭaḥ* (a white cloth). This agreement is just a necessary condition, and not sufficient. Because, a viśeṣaṇa, in addition to agreeing with the viśeṣya should also be semantically fit to be a qualifier of the viśeṣya. For example, there can be two words say `yānam` (a vehicle) and `vanam` (a forest), that match perfectly in gender, number and case, but we can not imagine a viśeṣya-viśeṣaṇa-bhāva between `yāna` and `vana`. Is it only the semantics that rules out such a relation or are there any clues, especially syntactic ones, that help us to rule out a viśeṣya-viśeṣaṇa-bhāva between such words?

In search of clues:

Pāṇini has not defined the terms viśeṣya and viśeṣaṇa. Patañjali uses two terms dravya (substance) and guṇa (quality) while commenting on the agreement between a viśeṣya and a viśeṣaṇa.

> *yad asau dravyaṁ śrito bhavati guṇaḥ tasya yat liṅgam vacanam ca tad guṇasya api bhavati.* `(MBh under A4.1.3 Vt VI.)`
> A quality assumes the gender and number of the substance in which it resides.

But then what is this guṇa?

We come across the description of guṇa by Kaiyyaṭa.

> *sattve niviśate apaiti pṛthag jātiṣu dṛśyate*
> *ādheyaḥ -ca-akriyājaḥ-ca saḥ asattva-prakṛti-guṇaḥ*
> `(MBh A4.1.44)`
> Guṇa is something which is found in things / substances *(sattve*

> *niviśate)*, which can cease to be there *(apaiti)*, which is found in different kinds of substances *(pṛthag jātiṣu)*, which is sometimes an effect of an action and sometimes not so *(ādheyaḥ-ca-akriyājaḥ-ca)*, and whose nature is not that of a substance *(asattva-prakṛti)*.

Thus guṇa is something which is not a substance, since it resides in other things. It is not a universal, since it is found in different kinds of substance. It is not an action, since guṇa is sometimes an effect of an action, as in the case of the color of a jar and sometimes not, as in the case of the magnitude of a substance. This characterisation of guṇais very close to the vaiśeṣika's concept of guṇa (Raja 1963).

Then, is this vaiśeṣika guṇa a viśeṣaṇa?

Patañjali commenting on the word guṇa under A2.2.11 provides an example contrasting two types of guṇas. While both `śukla` and `gandha` are qualities (guṇa) according to the vaiśeṣika ontology, the usage *śuklaḥ paṭaḥ* (a white cloth) is possible, while `gandham candanam` (fragrance sandal-wood) is not. Thus, only some of the vaiśeṣika guṇas have a potential to be a viśeṣaṇa, and not all.

If viśeṣaṇa is not a vaiśeṣika guṇa, what is it?

The characterisation of guṇa by Bhartṛhari in `Guṇa-samuddeśa` includes bhedakam as one of the characteristics of guṇa. But, in addition, guṇa, according to him, is also capable of expressing the degree of quality in a substance through a suffix. He defines guṇa as

> *saṁsargi bhedakaṁ yad yad savyāpāraṁ pratīyate*
> *guṇatvaṁ paratantratvāt tasya śāstra udāhṛtam* `VP III.5.1`
> Whatever rests on something else (saṁsargi), differentiates it (bhedaka), and is understood in that function (savyāpāra) is, being dependent, called `quality` in the śāstra. (Iyer 1971)

According to Bhartṛhari, apart from being a differentiator, a guṇa has another important characteristic, viz. that such a distinguishing quality can

also express the degree of excellence through some suffix (such as a comparative suffix `tarap`, or a superlative suffix `tamap`). This concept of guṇa of Bhartṛhari, thus is different from the concept of the guṇa of a vaiśeṣika. This definitely rules out the case of gandha, since we can not have gandhatara but we can have śuklatara to distinguish the white-ness between two white cloths.

Another clue from Pāṇini

We have another hint from Pāṇini through Patañjali. While in A4.1.3, Patañjali has used the terms dravya and guṇa in connection with agreement, in A1.2.52, he uses the term guṇavacana while describing a viśeṣaṇa

> *guṇavacanānāṁ śabdānām-āśrayataḥ liṅgavacanāni bhavanti-iti* (A1.2.52).
>
> The words which are guṇavacanas take the gender and number of the substance in which they reside.

The term guṇavacana is used for those words which designate a quality and then a substance in which this quality resides (Cardona 2009). In the example, *śuklaḥ paṭaḥ*, since śukla in addition to being a quality (white color), can also designate a substance, such as a paṭa (cloth), which is (white) in color, it is a guṇavacana word. But gandha (fragrance) designates only a quality, and can not be used to designate a substance that has a fragrance, and hence is not a guṇavacana.

Is guṇavacana necessary and sufficient to describe a viśeṣaṇa?

Let us look at the examples above. It definitely rules out yānaṁ and vanaṁ to be qualifiers of each other, since neither of them is a quality. But then what about dhāvan (the one who is running) in *dhāvan bālakaḥ* (a running boy)? Is dhāvan a guṇavacana?

Guṇavacana is a technical term, used by Pāṇini to define an operation of elision of `matup` suffix in certain quality denoting words such as 'sukla etc. So technically, a word such as `dhāvan`, though it designates a substance, is not a guṇavacana. This is clear from Patañjali's commentary on A1.4.1[6] where he

[6]The Vārtika *guṇvacanam ca* is followed by several other vārttikas, of which the following two are relevant. *samāsa-kṛt-taddhita-avyaya-sarvanāma-asarvaliṅgā jātiḥ* ||41 || *saṁkhyā ca* ||42 ||

states that compounds (samāsa), primary derivatives (kṛdantas), secondary derivatives (taddhitāntas), indeclinables (avyaya), pronouns (sarvanāma), words referring to universals (jāti), numerals (saṁkhyā) can not get the designation guṇavacana, since the latter saṁjñās (technical terms) supersede the previous ones.[7]

The very fact that Kātyāyana had to mention that words belonging to all the latter categories are not guṇavacana, indicates that all these category of words have a potential to get the guṇavacana designation, but Pāṇini did not intend to assign this sañjñā to these words. Whatever may be the reason, but this list of various categories, in fact, provides us a morphological clue for a word to be a viśeṣaṇa.

Here are some examples of viśeṣaṇas belonging to these different grammatical categories.

1. Samāsa (a compound)
   Bahuvrīhi (exo-centric) compounds refer to an object different from the components of the compound, and thus typically act as adjectives. For example, `pītāmbaraḥ` is made up of two components pīta (yellow) and ambara (cloth), but it refers to the 'one wearing a yellow-cloth' (and is conventionally restricted to Viṣṇu). An example of tat-puruṣ (endo-centric) compound as a viśeṣaṇa is *parama-udāraḥ* (extremely noble).
2. Kṛdanta (an adjectival participle)
   Nouns derived from verbs act as qualifiers of a noun. For example, in the expression *dhāvantam mṛgam* (a running deer), dhāvantam, a verbal noun, is a viśeṣaṇa. Only certain kṛdanta suffixes such as `śatṛ`, `śānac`, `kta`, etc. produce nouns that can be viśeṣaṇas, and not all.
3. Taddhita (a secondary derivative)
   Taddhitas with certain suffixes derive new nouns such as *bhāratīya* (Indian), *dhanavān* (wealthy), *guṇin* (possessing good qualities), etc. that denote a substance, as against certain other taddhita words such as *manuṣyatā* (humanity), *vārddhakya* (senility) etc. which derive new words designating qualities.
4. Sarvanāma (a pronoun)
   Pronouns also act as qualifiers. For example, in the expression *idam pustakam* (this book), *idam* is a viśeṣaṇa.

[7] *gaṇavacanasaṁjñāyāḥ ca etābhiḥ bādhanaṁ yathā syāt iti*

5. Jāti (a universal)
   In an expression *āmraḥ vṛkṣaḥ* (a mango tree), both the words āmraḥ and vṛkṣaḥ are common nouns. But one is a special and the other one is a general one. So the designation of āmra is a subset of the designation of vṛkṣa. Only in such cases, where there is a parājāti-aparājāti (hypernymy-hyponymy) relation, the one denoting an aparājāti (hyponymy) qualifies to be a viśeṣaṇa of the other one.
6. Saṁkhyā (a numeral)
   In an expression ekaḥ puruṣaḥ (a man), the word ekaḥ designates a number, which is a viśeṣaṇa of puruṣa.

There are still two more classes of words that are not covered in the above list, but which can be viśeṣaṇas. They are: words denoting an acquired property or an imposed property, and the relation-denoting terms. For example, ācāryaḥ in *ācāryaḥ droṇaḥ*, is an imposed property and putraḥ in *Daśarathasya putraḥ rāmaḥ* is a relation denoting term.

In summary, samastapada, certain kṛdantas, certain taddhitāntas, saṁkhyā, sarvanāma, ontological categories such as parā-aparā jātis, semantico-syntactic property such as guṇavacana and finally semantic properties such as relation denoting terms and upādhis, all these serve as characterisations of a viśeṣaṇa. This characterisation is only a necessary condition, and not sufficient, since it does not involve any mutual compatibility between the words. However, it brings in more precision in the necessary conditions for two words to be in viśeṣya-viśeṣaṇa-bhāva.

3.1.1 Deciding a Viśeṣya

Once we have identified the words that are mutually compatible with regard to an adjectival relation, the next thing is to decide the viśeṣya (head) among them. The commentary on A2.1.57 is useful in deciding the viśeṣya. This sūtra deals with the compound formation of two words that are in viśeṣya-viśeṣaṇa-bhāva. In Sanskrit compound formation, the one which is subordinate gets a designation of upasarjana. This provides us a clue about which word classes are subordinate to which ones. A noun may refer to a substance through an expression expressing the class character (jāti) such as utpalam (a flower), or through an action associated with it (kriyāvacana), as in dhāvan (running), or through a guṇavācaka such as nīlam. If there are two words designating common nouns, one denoting a special and the other one general, then the one which denotes a special type of common noun is

subordinate.[8] For example, in *āmraḥ vṛkṣaḥ*, āmra is a special kind of tree, and hence is a viśeṣaṇa and vṛkṣa is its viśeṣya. If one word designates a common noun and the other one either a guṇavacana or a kriyāvacana, then the word denoting the common noun becomes the viśeṣya.[9] Thus in *nīlam utpalam*, utpalam is the viśeṣya. In *pācakaḥ brāhmaṇaḥ* (cook Brahmin), brāhmaṇaḥ is the viśeṣya. When one of the words designate a guṇavacana and the other a kriyāvacana, or both the words designate either guṇavacanas or kriyāvacanas, then either of them can be a viśeṣya, as in khañjaḥ kubjaḥ (a hump-backed who is limping) or kubjaḥ khañjaḥ (a limping person with hump-back), similarly as in khañjaḥ pācakaḥ (a limping cook) or pācakaḥ khañjaḥ (a limping person who is a cook), etc.

On the basis of the above discussion, we have the following preferential order for the viśeṣya.

jātivācaka > {guṇvacana, kṛdanta}.

We saw earlier that a *viśeṣaṇa* can be any one of the following: a pronoun, a numeral, a kṛdanta, a taddhitānta, a samasta-pada, guṇavācaka, jāti, relation denoting terms, and an upādhi. So adding all these categories to the above preferential order, we get,

jātivācaka > upādhi > taddhitānta > guṇavacana > numeral > kṛdanta > pronoun.[10]

3.1.2 Flat or Hierarchical Structure?

After we identify all the words that have a samānādhikaraṇa relation between them, and mark the viśeṣya (the head) among them, the next task is to know whether a viśeṣaṇa is related to this viśeṣya directly, or through other viśeṣaṇas.

If there are n viśeṣaṇas, and all of them are related to the viśeṣya directly, then it results in a flat structure. But if a viśeṣaṇa is related to the viśeṣya

[8] *sāmānyajāti-viśeṣajātiśabdayoḥ samabhivyāhāre tu viṣeṣajātireva viśeṣaṇam. under A2.1.57, in BM*

[9] *jātiśabdo guṇakriyāśabdasamabhivyāhāre viśeṣyasamarpaka eva na tu viśeṣaṇa samarpakaḥ, svabhāvāt*, under A2.1.57, in BM

[10] This preferential order is purely based on some observations of the corpus, and needs further theoretical support, if there is any.

through other viśeṣaṇas, then there are exponentially large number of ways in which n viśeṣaṇas can relate to the viśeṣya. For example, if there are three words say `a`, `b` and `c`, of which `c` is the viśeṣya. Then computationally, there are three ways in which the other two words may relate to `c`.

1. Both `a` and `b` are the viśeṣaṇa of `c`. (This results in a flat structure.)
2. `a` is a viśeṣaṇa of `b` and `b` that of `c`.
3. `b` is a viśeṣaṇa of `a` and `a` that of `c`.

In positional languages like English, only the first two cases are possible. For example, consider the phrase 'light red car', which may either mean a car which is red in color and is light in weight, or a car which is light-red in color. In the second case, light-red is a compound.

Sanskrit being a free word order language, one can imagine, computationally, a possibility for the third type as well. The relation between the adjectival terms being that of sāmānādhikaraṇya (co-referential), semantically, only a flat structure is possible with adjectives. The other two cases of hierarchical structures result into compound formation in Sanskrit.

This is also supported by Jaimini's Mīmāṁsā sūtra

> *guṇānām ca parārthatvāt asambandhaḥ samatvāt syāt.* (MS 3.1.22)
> In as much as all subsidiaries are subservient to something else and are equal in that respect, there can be no connection among themselves.
> (Jha 1933)

Thus, a viśeṣaṇa is not connected to another viśeṣaṇa. The associated structure is a flat one, with all the viśeṣaṇas being connected to the viśeṣya.

3.2 Distinguishing a kāraka from a non-kāraka:

In Sanskrit, some case markers denote both a kāraka relation as well as a non-kāraka relation, as we saw earlier. In a sentence, if a verb denotes an action, then nouns denote the participants in such an action. These participants, which are classified into 6 types, viz. kartā, karma, karaṇam, sampradānam, apādānam, and adhikaraṇam are collectively called as kārakas. Other nouns in the sentence, which do not participate directly in the action,

express non-kāraka relations such as hetu (cause), prayojanam (purpose), etc. We get a clue to distinguish between the nouns which are related by a kāraka relation and those which are related by a non-kāraka one in the `Aruṇādhikāra` of the `Śābara bhāṣya`. There it is mentioned that

> *na ca amūrta-arthaḥ kriyātāḥ sādhanaṁ bhavatīti* (SB; p 654)
> No unsubstantial object can ever be the means of accomplishing an act.

Thus anything other than dravya can not be a kāraka. As we saw earlier, the guṇavacanas also can designate a dravya. And thus, all the dravyas and the guṇavacanas are qualified to be a kāraka. And the rest, i.e. nouns which denote either a guṇa which is not a guṇavacana or a kriyā (verbal nouns), may have a non-kāraka relation with a verb.

Let us see some examples.

Skt: *rāmaḥ daśarathasya ājñayā rathena vanam gacchati.*
Gloss: Rama {nom.} Dasharatha{gen.} order{ins.} ratha{ins.} forest{acc.} goes.
Eng: On Dasharatha's order, Rama goes to the forest by a chariot.

Skt: *rāmaḥ adhyayanena atra vasati.*
Gloss: Rama {nom.} study{ins.} here lives.
Eng: Rama lives here in order to study.

In the first sentence ājñā (order) is the cause for Rama's going to forest, ratha (chariot) is the instrument (or vehicle) for his going and in the second sentence adhyayana is the cause of Rāma's stay.

Since both hetu as well as karaṇam demand a 3^{rd} case suffix, ākāṅkṣā would establish a relation of karaṇam between `ājñayā` and `gacchati`,[11] between `rathena` and `gacchati` and also between `adhyayana` and `gacchati`. Now with the above definition of a kāraka, adhyayana, being a verbal noun (a kṛdanta) in the sense of bhāva, represents an abstract concept and therefore it does not designate a dravya (a substance). Hence it can not be a karaṇam. Similarly `ājñā`, which is a *guṇa* (according to *Vaiśeṣika* ontology, being a

[11] To be precise, the relation is between the meaning denoted by the nominal stem ājñā and the one denoted by the verbal root gam.

śabda), can not be a *karaṇa*. Thus use of congruity helps in pruning out impossible relations.

On the same grounds, establishment of apādānam and sampradānam relations between a non-dravya[12] denoting noun and a verb can also be prevented.

3.3 Congruous substantive for a Ṣaṣṭhī (genitive)

Pāṇini has not given any semantic criterion for the use of the genitive relation. His rule is *ṣaṣṭhī śeṣe* (A2.3.50) which means, in all other cases that are not covered so far, the genitive case suffix is to be used. The relation marked by the ṣaṣṭhī (genitive) case marker falls under the utthāpya (aroused) ākāṁkṣā. This is a case of uni-directional expectancy. Thus, there is no syntactic clue to which noun the word in genitive case would get attached. All other nouns in the sentence are potential candidates for a genitive relation to join with. The clue is, however, semantic. Patañjali in the `Mahābhāṣya` on A2.3.50 provides some semantic clues. He says there are hundreds of meanings of śaṣṭhī. Some of them are sva-svāmi-bhāva as in *rājñaḥ puruṣaḥ* (a king's man), avayava-avayavī-bhāva as in *vṛkṣasya śākhā* (branch of a tree) etc. So in order to establish a genitive relation, we need the semantic inputs. However, there are certain constraints. They are

1. A genitive connecting a verbal noun expressing bhāva such as lyuṭ etc. expresses a kāraka[13] relation and not the genitive one, as in *rāmasya gamanam*.
2. A genitive always connects with a viśeṣya, and never with a viśeṣaṇa, since there is a samānādhikaraṇa relation between the viśeṣya and viśeṣaṇa. For example, in the expression *rāmasya vīreṇa putreṇa*, the genitive relation of `rāmasya` is with `putreṇa` and not with `vīreṇa`.

Lexical resources such as Sanskrit WordNet[14] and Amarakośa[15] that are marked with the semantic information of part-whole relation, janya-janaka-bhāva, ājīvikā relation etc. help in identifying the genitive relations with confidence. When both the words refer to dravyas (substantives), then also there is a possibility of a genitive relation. So note that, while for other

[12] To be precise, a non-dravya and non-guṇavacana.
[13] *kartṛkarmaṇoḥ kṛti* (A2.3.65)
[14] `http://www.cfilt.iitb.ac.in/wordnet/webswn/english_version.php`
[15] `http://scl.samsaadhanii.in/amarakosha/index.html`

relations, we look for the absence of non-congruity for ruling out the relations, in the case of genitives, instead, we look for the presence of congruity, to prune out impossible relations. We took this decision, since we found it difficult to describe the non-congruity in the case of genitive relations.

Ambiguity between a genitive and an adjectival relation

Further we come across an ambiguity in the genitive relation, in the presence of adjectives. Look at the following two examples.

Skt: *vīrasya Rāmasya bāṇam*
Gloss: brave{gen.} Rama{gen.} arrow
Eng: An arrow of brave Rama

and

Skt: *Rāmasya putrasya pustakam*
Gloss: Rama{gen.} son{gen.} book
Eng: A book of Rama's son

In the first example, *vīra* being a guṇavacana, with the earlier characterisation of an adjective, vīra would be marked an adjective. while in the second one there is a kinship relation.

4 Evaluation

As stated earlier, ākāṅkṣā states the possibility of relations between two words. The mutual compatibility between the meanings further helps in pruning out the incompatible relations. We classified the content nouns into two classes: dravya and guṇa. Guṇas being further marked if they are guṇavacanas. We tested the mutual compatibility only when the suffix is ambiguous. To be precise, the yogyatā is used only to disambiguate between a kāraka versus non-kāraka relation, to establish the viśeṣya-viśeṣaṇa-bhāva, and to establish a genitive relation. This ensured that we do not miss on the metaphoric meanings. In the case of kāraka relations, if the noun denotes a guṇavacana, then the possible kāraka relation, on the basis of expectancy is

pruned out. Similarly, in the case of adjectival relations, the relations with a non-guṇavācaka guṇa is pruned out.

The performance of the system with and without yogyatā was measured to evaluate the impact of yogyatā. The corpus for evaluation of sentences consists of around 2300 sentences. It includes sentences with various grammatical constructions, a few passages from school text book, `Bhagavadgītā`, and a sample from Māgha's `Śiśupālavadham`. The ślokas in `Bhagvadgītā` as well as in `Śiśupālavadham` were converted to a canonical form.[16] The sentences with conjunction were not considered for the evaluation, since the nouns in conjunction conflict with the adjectives, and the criteria for handling conjunction are under development. The statistics showing the size of various texts, the average word length and the average sentence length is given in Table 2.

Type	Sents	Words	characters	avg sntlen	avg wrd len
Text books	260	1,295	9,591	4.98	7.40
Syntax	937	3,339	25,410	3.56	7.61
Māgha's SPV	66	623	5,851	9.40	9.39
Bhagvadgītā	940	5,698	42,251	6.06	7.41
Total	2,203	10,955	83,103	3.77	7.58

Table 2
Corpus Characteristics

All these sentences were run through a parser, first without using the conditions of yogyatā and second times using the conditions of yogyatā. In both the cases, the parser produced all possible parses. We also ensured that the correct parse is present among the produced solutions. Table 3 shows the statistics providing the number of solutions with and without using the filter of yogyatā. The number of parses produced was reduced drastically. This improved the precision by 63% in text book stories, by 67% in the grammatical constructs, and by 81% in case of the text from Bhagvadgītā and Māgha's kāvya. Better results in the case of these texts pertains to the fact that these texts have more usage of adjectives and non-kāraka relations as against the text book sentences, and artificial grammatical constructs.

[16] All the ślokas were presented in their anvita form, following the traditional Daṇḍānvaya method, where the verb typically is at the end, and viśeṣaṇas precede the viśeṣyas.

Corpus type	Sents	avg sols without yogyata	avg sols with yogyata	improvement in precision
Text books	260	39.76	14.56	63%
Syntax	937	19.5	6.33	67%
Literary	66	11,199	2,107	81%
BhG	940	2,557	478	81%
Total	2203	1439.54	268.85	81%

Table 3
Improvement

5 Conclusion

Yogyatā or mutual congruity between the meanings of the related words is an important factor in the process of verbal cognition. In this paper, we presented the computational modeling of yogyatā for automatic parsing of Sanskrit sentences. Among the several definitions of *yogyatā*, we modeled it as an absence of non-congruity.

Due to lack of any syntactic criterion for viśeṣaṇa (adjectives) in Sanskrit, parsing Sanskrit texts with adjectives resulted in high number of false positives. Hints from the vyākaraṇa texts helped us in the formulation of a criterion for viśeṣaṇa with syntactic and ontological constraints, which provided us a hint to decide the absence of non-congruity between two words with respect to the adjectival relation. A simple two way classification of nouns into dravya (substance) and guṇa (quality) with further classifications of guṇas into guṇavacanas was found to be necessary for handling adjectives. The same criterion was also found useful to handle the ambiguities between a *kāraka* and *non-kāraka* relations. These criteria together with modeling *yogyatā* as an absence of non-congruity resulted in 81% improvement in the precision.

Finally, the fact that there can not be an adjective of an adjective, having identified a viśeṣya, there is only one way all the viśeṣaṇas can connect with the viśeṣya. This theoretical input provided much relief from practical point of view, in the absence of which possible solutions would have been exponential.

6 Abbreviations

A: Pāṇini's Aṣṭādhyāyī, See Pande, 2004
Aa.b.c : adhyāya(chapter),pāda(quarter),sūtra number in Aṣṭādhyāyī
BM: Bālamanoramā, see Pande, 2012
MBh: Patañjali's Mahābhāṣya, see Mīmāṁsaka
KP: Kāvyaprakāśa, see Jhalakikar
MS: Mīmāṁsā sūtra, through SB
NK: Nyāyakośa, see Jhalkaikar
PM: Padamañjarī, see Mishra
SB: Śābara Bhāṣya, see Mīmāṁsaka, 1990
VP: Vākyapadīyam, see Sharma, 1974

References

Bhanumati, B. 1989. *An Approach to Machine Translation among Indian Languages*. Tech. rep. Dept. of CSE, IIT Kanpur.

Bharati, Akshar, Vineet Chaitanya, and Rajeev Sangal. 1995. *Natural Language Processing: A Paninian Perspective*. Prentice-Hall New Delhi.

Bharati, Akshar, Samar Husain, Bharat Ambati, Sambhav Jain, Dipti M Sharma, and Rajeev Sangal. 2008. "Two semantic features make all the difference in Parsing accuracy." In: *Proceedings of the 6th International Conference on Natural Language Processing (ICON-08)*. C-DAC, Pune.

Cardona, George. 2007. *Pāṇini and Pāṇinīyas on Śeṣa Relations*. Kunjunni Raja Academy of Indological Research Kochi.

—. 2009. "On the structure of Pāṇini's system." In: *Sanskrit Computational Linguistics 1 & 2*. Ed. by Gérard Huet, Amba Kulkarni, and Peter Scharf. Springer-Verlag LNAI 5402.

Devasthali, G V. 1959. *Mīmāṁsā: The vākya śāstra of Ancient India*. Booksellers' Publishing Co., Bombay.

Huet, Gérard, Amba Kulkarni, and Peter Scharf, eds. 2009. *Sanskrit Computational Linguistics 1 & 2*. Springer-Verlag LNAI 5402.

Iyer, K A Subramania. 1969. *Bhartṛhari: A study of Vākyapadīya in the light of Ancient comentaries*. Deccan College, Poona.

—. 1971. *The Vākyapadīya of Bhartṛhari, chapter III pt i, English Translation*. Deccan College, Poona.

Jha, Ganganatha. 1933. *Śābara Bhāṣya*. Oriental Institute Baroda.

Jhalakikar, V R. 1920; 7th edition. *Kāvyaprakāśa of Mammaṭa with the Bālabodhinī*. Bhandarkar Oriental Research Institute, Pune.

—. 1928. *Nyāyakośa*. Bombay Sanskrit and Prakrit Series, 49, Poona.

Jijñāsu, Brahmadatta. 1979. (In Hindi). *Aṣṭādhyāyī (Bhāṣya) Prathamāvṛtti*. Ramlal Kapoor Trust Bahalgadh, Sonepat, Haryana, India.

Joshi, S D. 1968. *Patañjali's Vyākaraṇa Mahābhāṣya Samarthāhnika (P 2.1.1) Edited with Translation and Explanatory Notes*. Center of Advanced Study in Sanskrit, University of Poona, Poona.

Joshi, S D and J.A.F. Roodbergen. 1975. *Patañjali's Vyākaraṇa Mahābhāṣya Kārakāhnikam (P 1.4.23–1.4.55)*. Pune: Center of Advanced Study in Sanskrit.

—. 1998. *The Aṣṭādhyāyī of Pāṇini with Translation and Explanatory Notes, Volume 7*. Sahitya Akadamy, New Delhi.

Katz, J J and J A Fodor. 1963. "The structure of a Semantic Theory." *Language* 39: 170–210.

Kiparsky, Paul. 2009. "On the Architecture of Panini's Grammar." In: *Sanskrit Computational Linguistics 1 & 2*. Ed. by Gérard Huet, Amba Kulkarni, and Peter Scharf. Springer-Verlag LNAI 5402, pp. 33–94.

Kulkarni, Amba. 2013b. "A Deterministic Dependency Parser with Dynamic Programming for Sanskrit." In: *Proceedings of the Second International Conference on Dependency Linguistics (DepLing 2013)*. Prague, Czech Republic: Charles University in Prague Matfyzpress Prague Czech Republic, pp. 157–66. URL: http://www.aclweb.org/anthology/W13-3718.

Kulkarni, Amba and Gérard Huet, eds. 2009. *Sanskrit Computational Linguistics 3*. Springer-Verlag LNAI 5406.

Kulkarni, Amba, Sheetal Pokar, and Devanand Shukl. 2010. "Designing a Constraint Based Parser for Sanskrit." In: *Fourth International Sanskrit Computational Linguistics Symposium*. Ed. by G N Jha. Springer-Verlag, LNAI 6465, pp. 70–90.

Kulkarni, Amba and K. V. Ramakrishnamacharyulu. 2013a. "Parsing Sanskrit texts: Some relation specific issues." In: *Proceedings of the 5th International Sanskrit Computational Linguistics Symposium*. Ed. by Malhar Kulkarni. D. K. Printworld(P) Ltd.

Kulkarni, Amba, Preeti Shukla, Pavankumar Satuluri, and Devanand Shukl. 2013c. "How 'Free' is the free word order in Sanskrit." In: *Sanskrit Syntax*. Ed. by Peter Scharf. Sanskrit Library, pp. 269–304.

Mīmāṃsakaḥ, Yudhiṣṭhira. 1990. *Mīmāmisā Śābara Bhāṣya*. Ramlal Kapoor Trust, Sonipat, Hariyana.

—. 1993. *Mahābhāṣyam, Patañjalimuniviracitam*. Ramlal Kapoor Trust, Sonipat, Hariyana.

Mishra, Sri Narayana. 1985. *Kāśikāvṛttiḥ along with commentaries Nyāsa of Jinendrabuddhi and Padamañjarī of Haradattamiśra*. Ratna Publications, Varanasi.

Pande, Gopaldatta. 2000, Reprint Edition. *Vaiyākaraṇa Siddhāntakaumudī of Bhaṭṭojidikṣita (Text only)*. Chowkhamba Vidyabhavan, Varanasi.

—. 2004. *Aṣṭādhyāyī of Pāṇini elaborated by M.M.Panditraj Dr. Gopal Shastri*. Chowkhamba Surabharati Prakashan, Varanasi.

Pande, Gopaldatta. 2012, Reprint Edition. *Vaiyākaraṇa Siddhāntakaumudī of Bhaṭṭojidikṣita containing Bālamanoramā of Śrī Vāsudevadīkṣita.* Chowkhamba Surabharati Prakashan, Varanasi.

Pataskar, Bhagyalata A. 2006. "Semantic Analysis of the technical terms in the 'Aṣṭādhyāyī' meaning 'Adjective'." *Annals of Bhandarkar Oriental Research Institute* 87: 59–70.

Raja, K Kunjunni. 1963. *Indian Theories of Meaning.* Adayar Library and Research Center, Madras.

Ramakrishnamacaryulu, K V. 2009. "Annotating Sanskrit Texts Based on Śābdabodha Systems." In: *Proceedings Third International Sanskrit Computational Linguistics Symposium.* Ed. by Amba Kulkarni and Gérard Huet. Hyderabad India: Springer-Verlag LNAI 5406, pp. 26–39.

Ramanujatatacharya, N S. 2005. *Śābdabodha Mīmāṁsā.* Institut Françis de Pondichérry.

Resnik, Phillip. 1993. "Semantic classes and syntactic ambiguity." In: *ARRPA Workshop on Human Language Technology.* Princeton.

Sharma, Pandit Shivadatta. 2007. *Vyākaraṇamahābhāṣyam.* Chaukhamba Sanskrit Paratishthan, Varanasi.

Sharma, Raghunath. 1974. *Vākyapadīyam Part III with commentary Prakāśa by Helaraja and Ambakartri.* Varanaseya Sanskrit Visvavidyalaya, Varanasi.

Shastri, Swami Dwarikadas and Pt. Kalika Prasad Shukla. 1965. *Kāśikāvṛttiḥ with the Nyāsa and Padamañjarī.* Varanasi: Chaukhamba Sanskrit Pratishthan.

Wilks, Yorick. 1975. "A preferential, pattern-seeking, semantics for Natural Language Interface." *Artificial Intelligence* 6: 53–74.