Sentential Parsing of Upamā

Amruta Barbadikar and **Amba Kulkarni**
Department of Sanskrit Studies
University of Hyderabad
amruta.barbadikar@gmail.com, ambakulkarni@uohyd.ac.in

Abstract

This study examines the computational treatment of *Upamā alaṅkāra*, drawing upon *Mammaṭa*'s grammatical classification as its foundation. It introduces a systematic tagging scheme for all twenty-five subcategories of *Upamā*, ensuring fine-grained annotation of poetic comparisons. The paper further addresses non-standard and irregular usages of *Upamā* by proposing suitable representational model. In addition, it considers annotation strategies for other similarity-based *alaṅkāras*, thereby situating *Upamā* within the broader framework of figurative expressions in Sanskrit.

1 Introduction

An important category of *alaṅkāra* is *arthālaṅkāra*, or the ornamentation of meaning. *Arthālaṅkāras* represent various ways in which meaning can be expressed with artistic elegance. Their use elevates a simple sentence into something more impactful and refined. These include comparisons as in *upamā, utprekṣā, rūpaka, ananvaya*, cause-and-effect relations as in *vibhāvanā, viśeṣokti*, inversions as in *nidarśanā*, and other conceptual structures. Literary language, or *kāvya*, is not just used to convey information, it expresses ideas in a rich and multi-layered way, full of depth and subtlety. The careful selection of words, their arrangement, and the implied meanings distinguish poetic language from ordinary speech.

Contemporary natural language processing tools, however, are primarily trained to focus on grammatical aspects such as syntax and morphology. When applied to poetic texts, especially those rich in *alaṅkāras*, often fail to capture the semantic depth and aesthetic dimensions inherent in such language. They can offer structural analysis but fall short of interpreting the poetic significance.

Among the various *arthālaṅkāras*, *Upamā*, holds a central place. It expresses similarity between two entities and typically involves four components:

1. **Upameya**: The subject of comparison.

2. **Upamāna**: The object of comparison.

3. **Sādhāraṇa-dharma**: The shared quality in the form of an Event or a property or a State.

4. **Upamā-dyotaka**: Comparative marker or comparator, such as '*iva*', 'yathā', 'tulya', etc.

A sentence containing *Upamā* conveys additional information by highlighting the similarity between the *upameya* and the *upamāna*, marked by a comparative expression and grounded in a shared property. Such sentences carry greater informational and aesthetic depth than ordinary statements. When a sentential parser lacks the capacity to process these figurative elements and their interrelations, essential semantic nuances may be lost. Owing to its structural clarity and frequent occurrence in Sanskrit poetry and poetics, *Upamā* stands out as both theoretically significant and practically approachable.

Some attempts have been made towards processing of *Upamā*. Jadhav et al. (2023) provided graphs for human cognition of *upamā* sentences and proposed a representation schema for machines to process the limited range of 6 subtypes of *pūrṇopamā* category. Jadhav et al. (2025) explored the potential of Large Language Model in identification and classification of *Upamā*. Kulkarni and Kulkarni (2025) has proceeded towards the refined manual modelling with a limited instances of *Upamā* occurring in the second canto of *Raghuvaṃśa*, primarily consisting of *Samāsagā-upamā* denoted by *iva*. However, no comprehensive guidelines covering all 25 subtypes of *Upamā* have been identified. Their systematic implementation for parsing sentences containing *Upamā Alaṅkāra* also remains unexplored.

While Anusārakam[1], a sentential analyser based on Pāṇinian *kāraka* and *śabdabodha* theories, primarily focuses on syntactic and semantic analysis based on grammatical roles, it does not account for the semantic ornamentation introduced by figures of speech (*alaṅkāras*). To address this, our aim is to extend the parser with a specialised module designed to identify and analyse *Upamā alaṅkāra* in Sanskrit sentences. In order to extend it, we propose the tagging guidelines for the 25 types of *Upamā* constructions.

In this paper, first we describe the subtypes of *Upamā* proposed by *Mammaṭa* in his treatise *Kavyaprakāśa*. This is followed by the grammatical classification and tagging scheme for annotating *Upamā*. We present the performance and error analysis of the system. The real life examples from literature pose more challenges and we present a few of them. Finally, we discuss example sentences with '*iva*' where it is used to denote other *Alaṅkāras* such as *Utprekṣā* etc. and suggest solutions for their correct automatic tagging.

2 Upamā Alaṅkāra

Arthālaṅkāras are commonly classified into various types based on similarity, dissimilarity, cause and effect, association, and so on. Among these, the category grounded in similarity stands out both in terms of conceptual prominence and the number of individual *alaṅkāras* it contains. *Upamā* is the foremost figure in this group, occupying a foundational place in Sanskrit poetics. Instances of *Upamā* appear even in the Vedic hymns, indicating its deep roots in the earliest Sanskrit literature. Several literary theorists have regarded *Upamā* as the source or model for other *arthālaṅkāras*.[2]

Bharata initiated the theoretical treatment of *alaṅkāras* in his *Nāṭyaśāstra*, identifying *Upamā* as one of just four figures of speech. The essence of *Upamā* proposed by Bharata was adopted by later scholars with only minor changes. However, a variety of classifications emerged among theoreticians, shaped by their differing interpretive frameworks.

Bharata and *Daṇḍin* focused on the nature of the subject matter when classifying *Upamā*. **Bharata** mentions five types of *Upamā* [3]: *Praśaṃsā* (praise), *Nindā* (censure), *Kalpitā* (conceit), *Sadṛśī* (likeness), and *Kiñcitsadṛśī* (partial resemblance). *Daṇḍin*, in contrast, expands the scope significantly by identifying 32 distinct types of comparisons classified under *Upamā*.[4]

Vāmana, recognises two types—*Kalpitā* and *Laukikī*—depending on whether the comparison is imaginative or grounded in common worldly perception. *Vāmana*, approaching from a different angle, contributes to a more detailed classification based on the syntactic structure. As discussed earlier, an *Upamā* typically consists of four elements: the subject of comparison (*Upameya*), the object of comparison (*Upamāna*), a shared quality (*Sādhāraṇa-dharma*), and a comparative marker (*Upamā-dyotaka*), such as '*iva*', '*yathā*', '*tulya*', etc. When all the four components are

---

[1] https://sanskrit.uohyd.ac.in/scl → Anusarakam

[2] upamāprapañcaśca sarvo'laṅkāra iti vidvadbhiḥ pratipannameva | 17th Chapter, *Abhinavabhāratī*

[3] praśaṃsā caiva nindā ca kalpitā sadṛśī tathā |
yā kiñcitsadṛśī jñeyā sopamā pañcadhā budhaiḥ|| 17.47, *Nāṭyaśāstra*

[4] *dharmopamā, vastūpamā, viparyāsopamā, anyonyopamā, niyamopamā, aniyamopamā, samuccayopamā, atiśayopamā, utprekṣitopamā, adbhutopamā, mohopamā, saṃśayopamā, nirṇayopamā, śleṣopamā, samānopamā, nindopamā, praśaṃsopamā, ācikhyāsopamā, virodhopamā, pratiṣedhopamā, caṭūpamā, tattvākhyānopamā, asādhāraṇopamā, abhūtopamā, asambhāvitopamā, bahūpamā, vikriyopamā, mālopamā, vākyārthopamā, prativastūpamā, tulyayogopamā and hetūpamā | Kāvyādarśa, 2.15 - 2.50*

explicitly present, *Vāmana* terms it a *Pūrṇopamā* (complete simile).[5] If one or more of these elements are omitted, the simile is called *Luptopamā* (elliptical).[6] Additionally, *Vāmana* draws a distinction between similes expressed through a single word and those constructed with more than one word or a paraphrase.

Udbhaṭa, in his seminal contributions, builds upon *Vāmana*'s syntactic model by introducing grammatical categories into the classification. He includes forms such as compound words (*Samāsa*) and derivatives with *Taddhita* suffixes that inherently convey comparison. This attention to grammatical structure sets a precedent followed by later scholars such as *Mammaṭa* and *Viśvanātha*. A classification based on grammatical cues offers more precision and can be effectively translated into computational analysis. In contrast, the earlier subject-based classifications, while insightful, pose challenges for machine interpretation and pattern recognition.

We follow the classification provided by *Mammaṭa* in *Kāvyaprakāśa*, which is based on *Udbhaṭa* with precise definitions and examples. In the following section, we explain different types of *Upamā* with an example and suggest a tagging scheme representing the syntax faithfully. *Upamā* being a figure of speech, deals more with semantics rather than syntax. Therefore, an attempt is also made to provide a semantic interpretation from the syntactic representation.

3   Types of *Upamā*

In *Kāvyaprakāśa*, the classification of *Upamā* rests primarily on syntactic considerations. The distinctions are drawn according to how the four essential components of *Upamā*—the subject of comparison (*upameya*), the object of comparison (*upamāna*), the shared property (*sādhāraṇa-dharma*), and the comparative marker (*upamā-vācaka*)—are expressed in a construction. Broadly, two principal types are recognised: *Pūrṇopamā* (complete) and *Luptopamā* (elliptical).

In *Pūrṇopamā*, all the four components are explicitly present, while in *Luptopamā* one or more these are absent. Another basis of classification is the nature of similarity indicated by the comparative marker: *Śrautī*, where the resemblance is grasped immediately upon hearing the word, and *Ārthī*, where the listener needs to process the meaning to discern the implied similarity. A further division concerns the syntactic setting of the marker: *Vākyagā* (appearing as stand-alone words), *Samāsagā* (occurring as a component of a compound), and *Pratyayagā* (expressed by affix denoting similarity).

Thus, *Pūrṇopamā* comprises of six subtypes arising from the intersection of the two semantic kinds (*Śrautī* and *Ārthī*) with the three syntactic settings (*Vākyagā*, *Samāsagā*, *Pratyayagā*).

Table 1: Classification of *Pūrṇopamā*

Meaning indicated	Syntactic environment
Śrautī (implicit)	Vākyagā
	Samāsagā
	Pratyayagā
Ārthī (explicit)	Vākyagā
	Samāsagā
	Pratyayagā

In contrast, *Luptopamā* is divided by the number of omitted elements: *Ekāluptā* (one element omitted), *Dviluptā* (two omitted), and *Triluptā* (three omitted), producing a total of nineteen recognised varieties. *Ekāluptā* comprises omissions of either the shared property (*Dharmaluptā*), the object of comparison (*Upamānaluptā*), or the comparative marker (*Vācakaluptā*). *Dviluptā* involves the loss of two components, yielding *Dharma-vācaka-luptā*, *Dharma-upamāna-luptā*, and *Vācaka-upameya-luptā*. Finally, *Triluptā* denotes the omission of three—*sādhāraṇa-dharma*, *upamāna*, and *upamā-vācaka*, simultaneously. Refer to Table 2 for details.

---

[5] guṇadyotakopamānopameyaśabdānāṃ sāmagnye pūrṇā| 2.5, *Kāvyālaṅkārasūtravṛtti*
[6] lope luptā| 4.6, *Kāvyālaṅkārasūtravṛtti*

Table 2: Classification of *Luptopamā*

Category	Subtypes and further divisions
Ekāluptā	• *Dharmaluptā* (5): *Śrautī Vākyagā, Śrautī Samāsagā, Ārthī Vākyagā, Ārthī Samāsagā, Ārthī Pratyayagā* • *Upamānaluptā* (2): *Ārthī Vākyagā, Ārthī Samāsagā* • *Vācakaluptā* (6): *Samāsagā*; *Pratyayagā—kyac* (karma), *kyac* (ādhāra), *kyaṅ* (kartā), ṇamul (karma), ṇamul (kartā)
Dviluptā	• *Dharma-vācaka-luptā* (2): *Samāsagā, Pratyayagā—kvip affix* • *Dharma-upamāna-luptā* (2): *Vākyagā, Samāsagā* • *Vācaka-upameya-luptā* (1): *Pratyayagā—kyac affix*
Triluptā	*Dharma-upamāna-vācaka-luptā* (1): *Samāsagā*

In the 10th *ullāsa* of *Kāvyaprakāśa*, examples are provided for each of the *Upamā* types. This classification preserves the subtle distinctions of classical Sanskrit poetics while offering a clear and systematic framework, particularly useful for the computational analysis of *Upamā*.

4   Tagging guidelines for *Upamā*

Prior to the development of a parser - whether it is developed as an expert system following the grammar rules or as ML (Machine Learning) system using annotated data, we need guidelines for the annotation. In this section, we present the tag scheme and the guidelines devised for the annotation of the four essential elements of *Upamā* and describe the annotation of different constructions enumerated in Section 3, including subtle variations found in literary texts.

4.1   About the Notation and Convention

We explain the tagging scheme with the following example (see figure 1):

Skt: *sundaram   mukham   paśya,*
yathā   kamalam.
Gloss: beautiful{sg,nom,neut}   face{sg,nom,neut}   see{imper,2p,sg}
like   lotus{sg,nom,neut}
Eng: See the beautiful face,
like a lotus.

Here, the main sentence is *sundaraṃ mukhaṃ paśya* where *sundara* (beautiful) is an adjective (*viśeṣṇa*) of *mukham* (face) which is a *karma* of an activity of seeing *(paśya)*. Now, there is a comparison of *mukha* with *kamala* (lotus) as an object of comparison, *mukha* being a subject of comparison and the shared property being beauty *(saundarya)*. Now, in dependency parsing, we capture the semantics that can be extracted from the syntax and the morphology of the words in the sentence. In this example, the marker *yathā* marks *kamalam* as an *upamānam* of *mukham*. Hence, we mark only the *upamānam* and *upamā-dyotakaḥ*. The *upameya* is understood to be *mukham*.
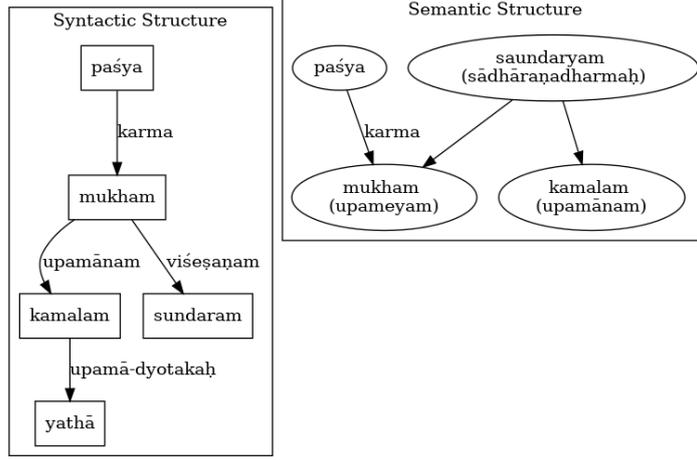
Figure 1: Example parse tree with *Upamā* specific tags

Since, in the dependency parser, we mark only the relations that are expressed through position or through markers, the semantic property of being a *sādhāraṇa-dharma* is not shown in the dependency diagram. Thus, *sundara* is marked only as a *viśeṣaṇa* of mukham and not shown explicitly as the *sādhāraṇa-dharma.* This decision is a very conscious decision. Only the syntactic information is marked. In addition, we decided to name the relation as *'upamāna'* following the convention we followed in dependency tagging, rather than *'upamāna-upameya-bhavaḥ'*, to keep it compact. The system followed in the marking the relations is as follows.

If the word 'Y' has a relation of 'R' with respect to 'X' then we mark it as: X → Y [R] (see Figure 2)
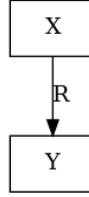
Figure 2: Representation of X → Y

For example, *rāmaḥ gacchati (Rāma goes.) rāmaḥ → gacchati [kartā]*, refer to figure 3,
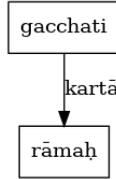
Figure 3: Parse Tree for *'rāmaḥ gacchati'*

On similar lines, we decided to mark the relation between *mukham* and *kamalam* as *'upamāna'*. We have also shown the semantic structure of *Upamā* in the sentence in figure 1. This semantic structure is arrived at after discarding the *'dyotaka'* elements. We also retain the same word form as in the sentence, and not changing it to its stem, for easy understanding of the semantic graph.

4.2 Annotating *Upamā*: Guidelines

In this part, we present the guidelines for annotations in the form of parse trees for each of the types of *Upamā* described earlier. We present examples according to their grammatical category

discussed in the previous section, viz. *Vākyagā, Samāsagā* and *Pratyayagā*.

- **Vākyagā Upamā**: In this category, the *Upamā alaṅkāra* is expressed using explicit words. We get all four elements of *Upamā* as separate words. In *Vākyagā Śrautī Upamā* the *upamā_dyotaka*s are indeclinables *(avyayas)*.
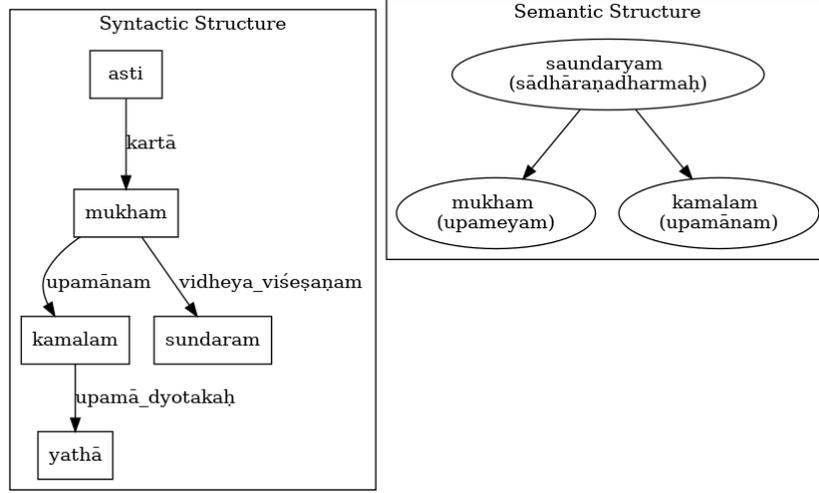
  1. **Pūrṇopamā Śrautī Vākyagā**:

Figure 4: *Vākyagā Śrautī Pūrṇopamā*

Skt: *mukham   kamalam   yathā   sundaram   asti.*
Gloss: face{sg,nom,neut}   lotus{sg,nom,neut}   like{indecl}   beautiful{sg,nom,neut}   is{pres,3p,sg}
Eng: The face is beautiful like a lotus.
Upameyam: *mukham,* Upamānam: *kamalam,*
Upamā-dyotakaḥ: *yathā,* Samāna-dharma: *saundarya.*

Above are the four components mentioned in this sentence. Figure 4 shows the relations marked for the sentence, in syntactic as well as semantic level. *Mukham* being the *upameya* has a primary role of *kartā* in the sentence. At semantic level, due to the *upamāna* relation with *kamalam,* we understand that the *mukham* is the *upameya.* Here, we have to note that both *mukham* and *kamalam* having same number, gender and case, making them compatible to be marked as *kartā.* But the comparator is always positioned after the *upamāna* in *śrautī upamā.* Hence, we can deduce that *kamalam* is the *upamāna* and *mukham* is the *upameya.*

For humans, differentiating between these factors is accomplished due to their access to world knowledge and the contextual cues provided in discourse. For machines, which lack such inherent world knowledge and contextual understanding, this task becomes significantly more challenging. Therefore, lexical clues provided by classical literary analysis prove to be of great help in the automatic processing of such figurative expressions.

Similarly, in *Śrautī Upamā,* the comparative marker is always associated with and positioned next to the *Upamāna.*[7] For instance, we observe that *yathā* is marked as the *Upamā-dyotaka* along with *kamalam.*

The adjective *sundaram* indicates the shared property of beauty (*saundarya*). In this sentence, it is the only adjective present. Sometimes, multiple adjectives may appear

---

[7]yathevavādiśabdā yatparāstasyaivopamānatāpratītiriti yadyapyumānaviśeṣaṇānyete tathāpi śabdaśaktimahimnā śrutyaiva ṣaṣṭhīvat saṃbandhaṃ pratipādayantīti tatsadbhāve śrutī upamā| *Kāvyaprakāśa*,87, 1o[th] Ullasa||

to represent the shared qualities. At times, adjectives may qualify either the *Upamāna* or the *Upameya*, or in some cases, the action itself may be the common element of comparison.

2. ***Vākyagā Śrautī Dharmaluptā*** :

   Skt: *mukham kamalam yathā asti.*

   Gloss: face{sg,nom,neut} lotus{sg,nom,neut} like{indecl} is{pres,3p,sg}
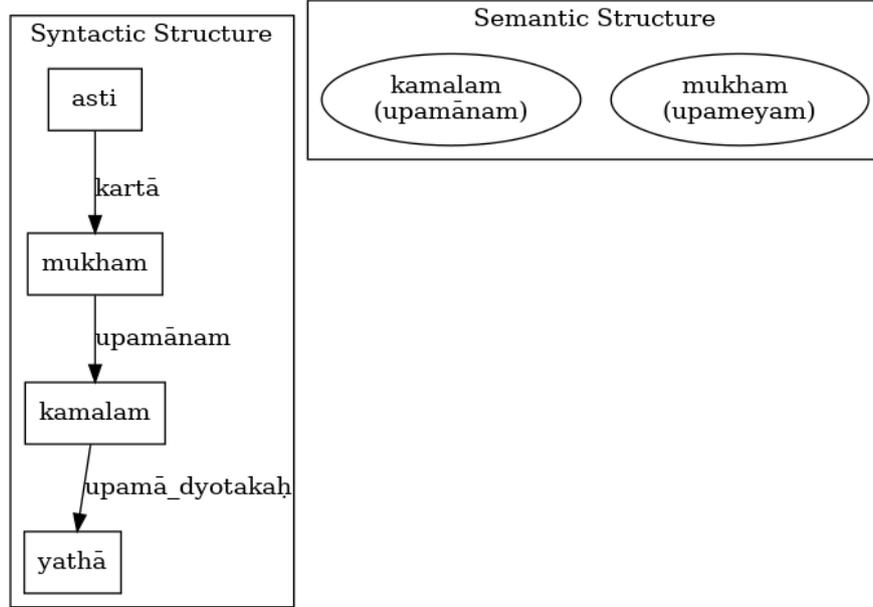
   Eng: The face is like a lotus.

Figure 5: *Vākyagā Śrautī Dharmaluptā*

In this sentence, the shared property is not mentioned. However, the parse tree will be similar to *Pūrṇopamā Śrautī Vākyagā* without the word '*sundaram*'. See figure 5 for the syntactic parse structure and semantic representation. Note the absence of common property (Sādhāraṇa-dharma) in the semantic structure.

3. ***Vākyagā Ārthī Pūrṇopamā*** :

   Skt: *mukham kamalena tulyam sundaram asti.*

   Gloss: face{sg,nom,neut} lotus{sg,instr,neut} comparable{sg,nom,neut} beautiful{sg,nom,neut} is{pres,3p,sg}

   Eng: The face is beautiful, comparable to a lotus.

   Here, *tulyam* is the comparative marker which is an independent word. In *Ārthī Upamā*, the structure of the sentence will be a bit different than *śrautī upamā*. *Tulyam*, the marker, has agreement with *mukham* which is the *upameya* unike *śrautī upamā*. But *tulya* has the third case expectancy which acts as a *upamāna*. Figure 6 shows the parse tree for the given sentence with the semantic representation of *Upamā*.

   There are three ways the comparative marker *tulya* can have an agreement with the other elements.

   (a) **Type I:** The marker can be related to *upmeya* as in the given example.

   (b) **Type II:** It can be related to *upamāna*.

    For example, '*mukhasya tulyam sundaram kamalam asti.*' (See figure 7)

   (c) **Type III:** It also can be related to both *upameya* and *upamāna*, as in, '*mukham ca kamalam ca tulyam sundaram asti.* (See figure 8)

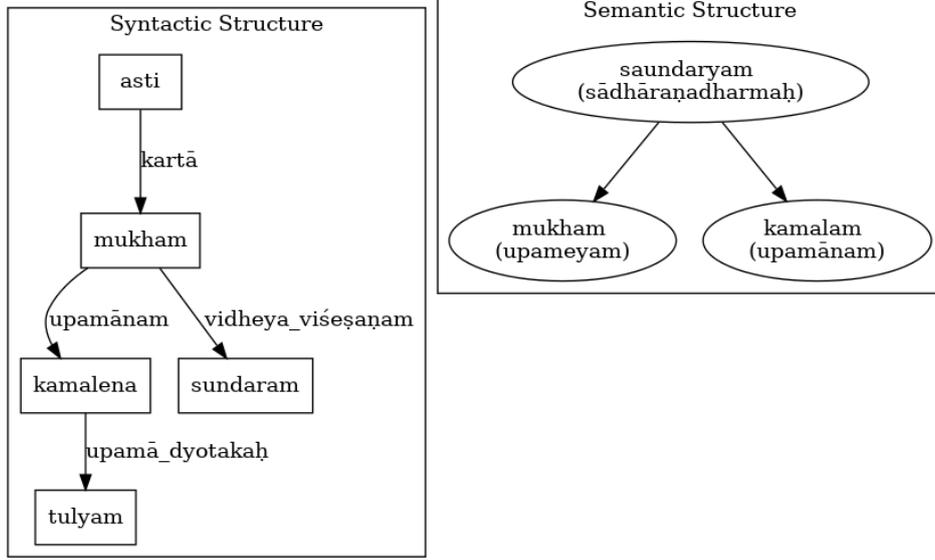   It makes it challenging for the parser to identify the *upamāna* and *upameya* when only
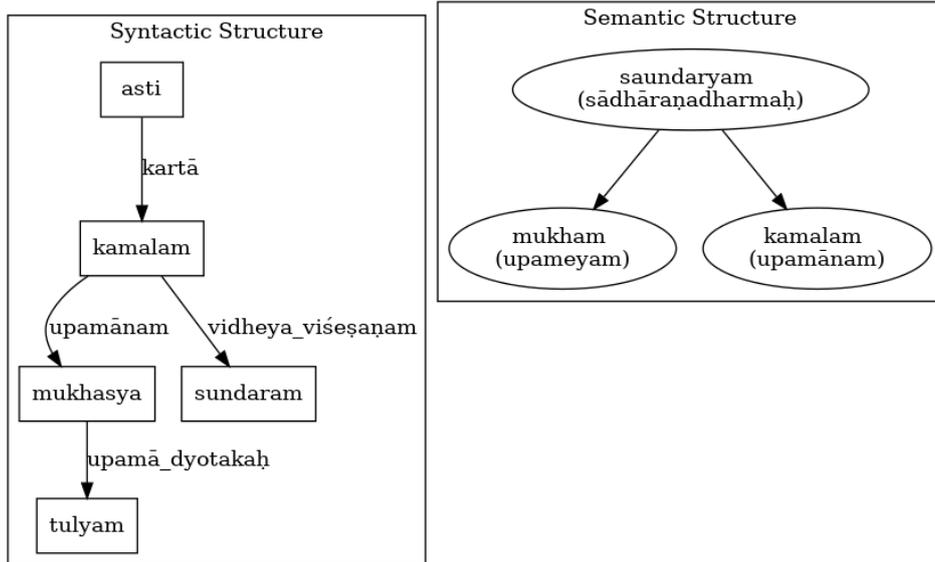
Figure 6: *Vākyagā Ārthī Pūrṇopamā: type I*
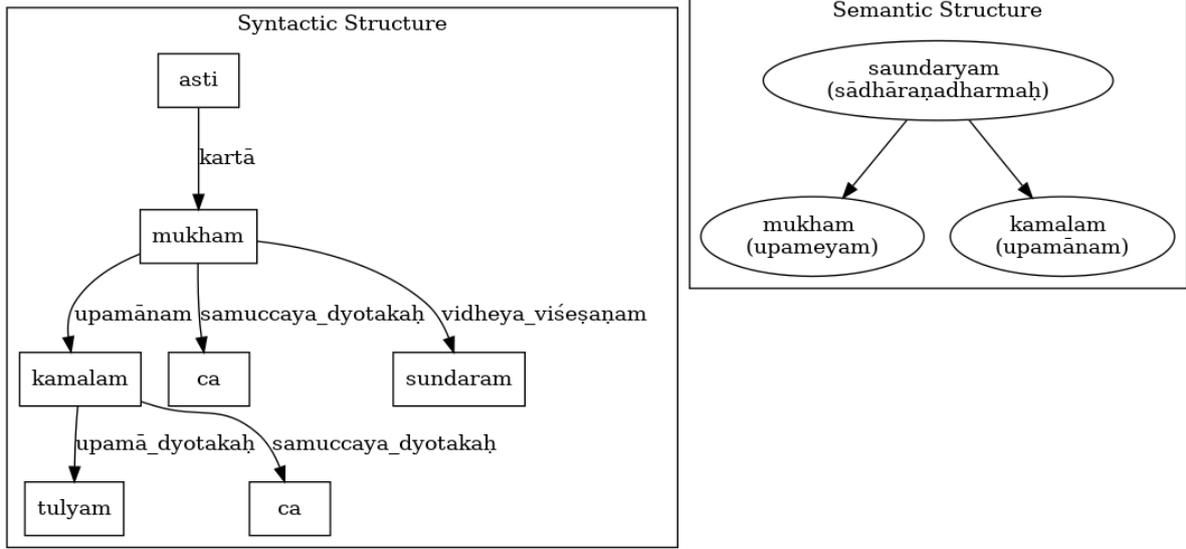

Figure 7: *Vākyagā Ārthī Pūrṇopamā: type II*

Figure 8: *Vākyagā Ārthī Pūrṇopamā:type III*

the case, number and gender agreement provides the hint for identification. These words can also bear the *kāraka* roles as per the meaning of the sentence.

4. **Dharmaluptā Ārthī Vākyagā**

   Skt: *mukham   kamalena   tulyam   asti.*

   Gloss: face{sg,nom,neut}   lotus{sg,instr,neut}   comparable{sg,nom,neut} is{pres,3p,sg}

   Eng: The face is comparable to a lotus.

   The analysis of this example will be as *Pūrṇopamā Ārthī Vākyagā* without the mention of common property, that is *sundaram*. Figure 9 provides the parse tree for this sentence along with the semantic representation of the components of *Upamā*.
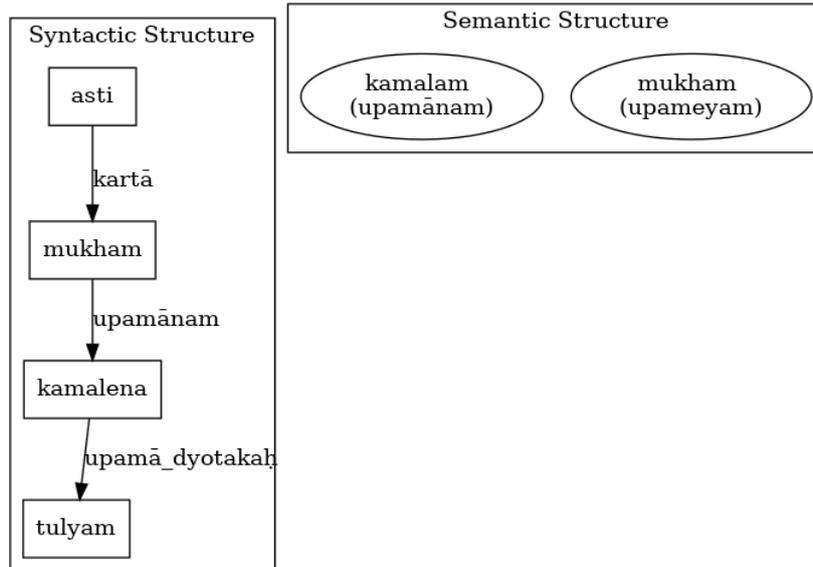
Figure 9: *Vākyagā Ārthī Dharmaluptā*

5. **Vākyagā Ārthī Upamānaluptā**

   Skt: *mukhasya   sadṛśaṃ   sundaraṃ   kim   api   na   asti.*

   Gloss: face{sg,gen,neut}   similar{sg,nom,neut}   beautiful{sg,nom,neut}

what{sg,nom,neut}   any   not   is{pres,3p,sg}

Eng: There is no beauty comparable to the face.

In this sentence, 'sadṛśam' is the *dyotaka* which is an independent word. *Upameya* and the *samāna-dharma* are as before. There is no entity found to be compared with the face, therefore this expression does not have a *upamāna*. Here, the phrase '*kim api na*' triggers the *upamānalupta* type. The syntactic parse structure and semantic representation is shown in figure 10.
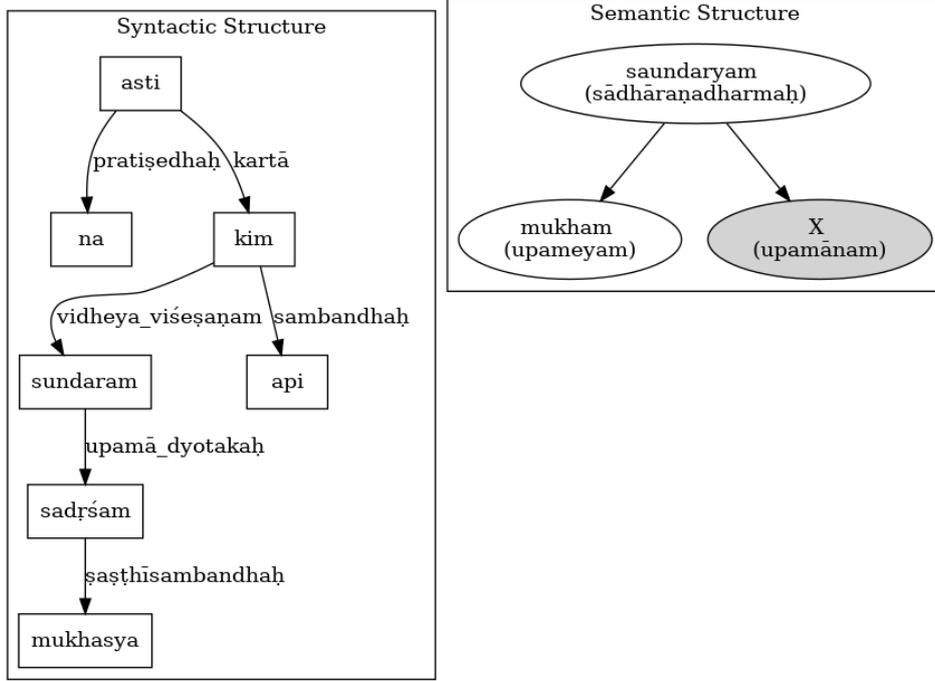

Figure 10: *Vākyagā Ārthī Upamānaluptā*

6. ***Vākyagā Ārthī Dharma-upamānaluptā***

   Skt: *mukhasya   sadṛśam   kim   api   na   asti.*

   Gloss: face{sg,gen,neut}   similar{sg,nom,neut}   what{sg,nom,neut}   any{indecl}   not   is{pres,3p,sg}

   Eng: Nothing similar to the face exists.

   The syntactic and semantic annotation for this example (see figure 11) will be similar to the previous type (*Upamānaluptā Ārthī Vākyagā*) without the mention of the common property '*sundaram*'.

- ***Samāsagā Upamā***

  The *Upamā* expression using compounds is considered in *samāsagā Upamā*. There are a total of 9 types of *samāsagā Upamā*. In this category the *dyotaka*s are either elided or compounded with the other element. Thus, we mark the compound itself as *upamānam*. In some types, the compound which is having the elements of *Upamā* carries the *kāraka* role in the sentence. In such cases, we have to understand the pesence of *Upamā* from the compound type.

  1. ***Samāsagā Śrautī Pūrṇopamā***

     Skt: *mukham   kamalam-iva   sundaram   asti.*

     Gloss: face{sg,nom,neut}   lotus{sg,nom,neut}-like{indecl}   beautiful{sg,nom,neut}   is{pres,3p,sg}

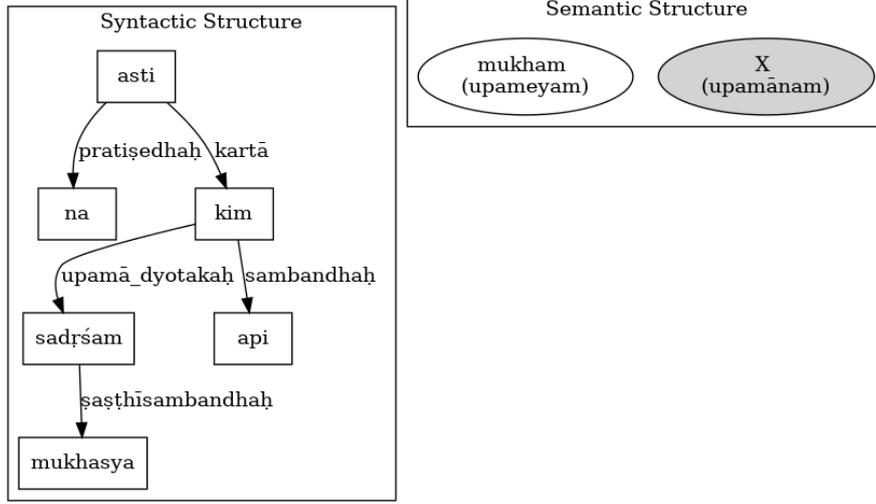   **Eng:** The face is beautiful like a lotus.

Figure 11: *Vākyagā Ārthī Dharma-upamānaluptā*

In this sentence, except for the *Upamā-dyotakaḥ* that is *iva*, all other components are the same as *Pūrṇopamā Śrautī Vākyagā*. According to *'ivena samāso vibhaktyalopaśca'*, a *vārtika* under *Aṣṭādyāyī* 2.1.4 and 2.4.71, *iva* is a particle (*avyaya*) which always forms a compound '*kevala-samāsaḥ*' with the *upamāna* without deletion of the case marker of the previous word. The syntactic parse tree and the semantic structure is as shown is figure 12.
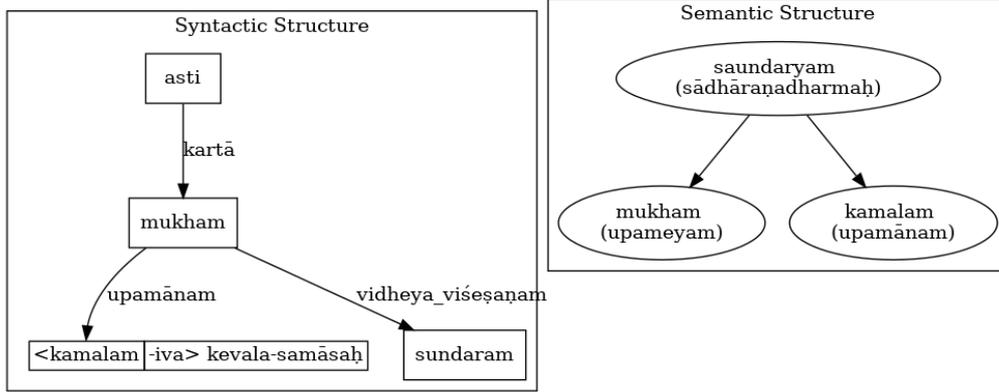
Figure 12: *Samāsagā Śrautī Pūrṇopamā*

2. **Dharmaluptā Śrautī Samāsagā**
   Skt: *mukham   kamalam-iva   asti.*
   Gloss: face{sg,nom,neut}   lotus{sg,nom,neut}-like{indecl}   is{pres,3p,sg}
   Eng: The face is like a lotus.
   For this sentence, the analysis will be the same as *Pūrṇopamā Śrautī Samāsagā* without the entry of the common property *sundaram*. Figure 13 demonstrates the syntactic parse tree and the semantic representation of the given sentence.

3. **Samāsagā Śrautī Vācakaluptā**:
   Skt: *mukham   kamala-sundaram   asti.*
   Gloss: face{sg,nom,neut}   lotus(like)-beautiful{sg,nom,neut}   is{pres,3p,sg}
   Eng: The face is beautiful like a lotus.
   Here, the *upameya* and the *samāna-dharma* joined as a compound to give the meaning 'beautiful like a lotus' ('*kamala iva sundaram*') by '*upamānāni sāmānyavacanaiḥ*'
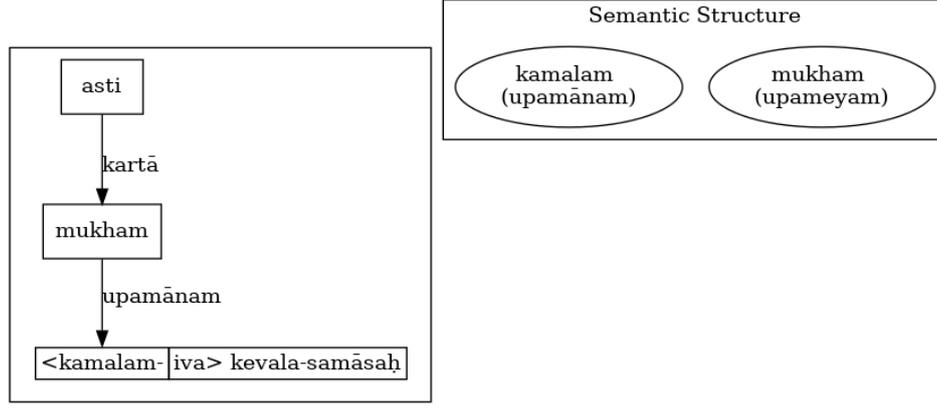
Figure 13: *Samāsagā Śrautī Dharmaluptā*

(A2.1.55). In this compound, the comparative marker gets elided. It is a kind of '*karmadhāraya-samāsa*'. Through this information, we get that the first component of the compound '*kamala*' is a *upamāna* and the second component, i.e. '*sundaram*', is a quality which is shared by the *upamāna* and the *upameya*. Figure 14 shows the syntactic structure and the semantic representation of this example.
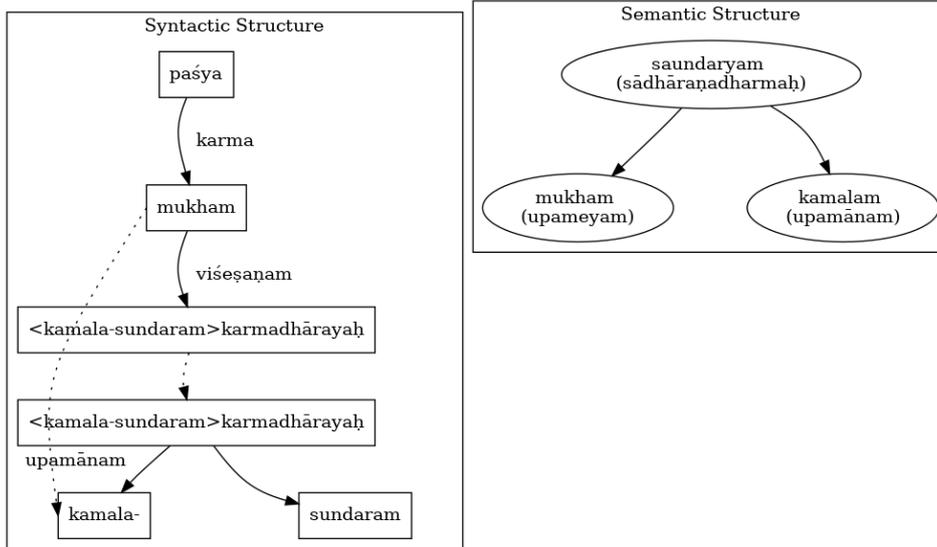

Figure 14: *Samāsagā Śrautī Vācakaluptā*

4. ***Samāsagā Śrautī Dharma-vācakaluptā***:
   Skt: *asau   rājakuñjaraḥ   rājate.*
   Gloss: that{sg,nom,masc}   king-elephant(like){sg,nom,masc}   shines{pres,3p,sg}
   Eng: That, elephant like king, shines.
   In this example, the word *rājakuñjaraḥ* is a compound, meaning '*rāja kuñjaraḥ iva*', expressing the *dyotaka* and the *dharma* with *samāsa*. This is an instance of *karmadhāraya* compound, which is formed by the *sūtra* -'*upamitaṃ vyāghrādibhiḥ sāmānyā'prayoge*' *(A2.1.56)*, where the first component (here, *rājā*) is the *upameya* and the second component (here, *kuñjara*) is a *upamāna*. The parse tree is as in figure 15.

5. ***Samāsagā Ārthī Pūrṇopamā*** :
   Skt: *mukham   kamala-tulyam   sundaram   asti.*
   Gloss: face{sg,nom,neut}   lotus-similar{sg,nom,neut}   beautiful{sg,nom,neut}   is{pres,3p,sg}
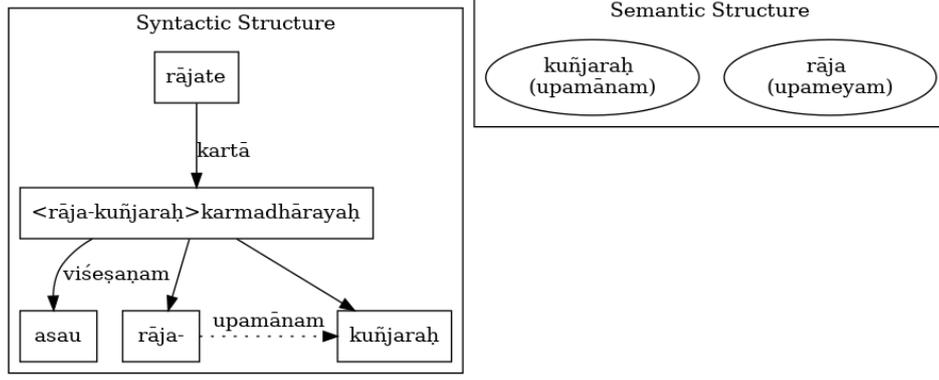
Figure 15: *Samāsagā Śrautī Dharma-vācakaluptā*

Eng: The face is beautiful, comparable to a lotus.

Here, the marker, *tulyam* is compounded with the *upamāna - kamalam*. Being a compound, it is treated as a single word. Based on the type of the copound and the lexicon involved, we mark the *upamāna* and the *dyotaka*. The parse tree along with the semantic structure for the given instance is as shown in the figure 16.
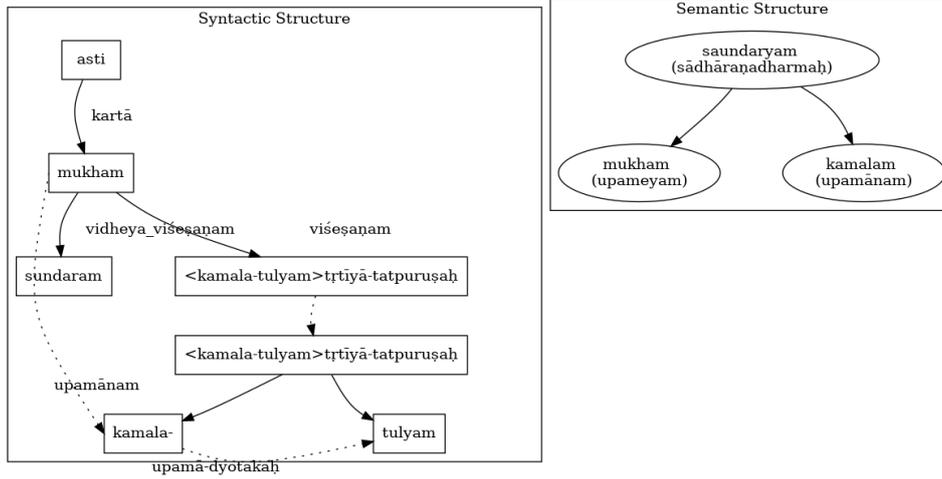

Figure 16: *Samāsagā Ārthī Pūrṇopamā*

6. ***Samāsagā Ārthī Dharmaluptā*** :

    Skt: *mukham   kamala-tulyam   asti.*

    Gloss: face{sg,nom,neut}   lotus-comparable{sg,nom,neut}   is{pres,3p,sg}

    Eng: The face is comparable to a lotus.

    The parse of the given example will be similar to the *Pūrṇopamā Ārthī Samāsagā* graph without the shared property. Figure 17 shows the syntactic parse structure and semantic representation of the given sentence.

7. ***Samśagā Ārthī Upamānaluptā***

    Skt: *mukha-sadṛśam   sundaram   kim   api   na asti.*

    Gloss: face-similar{sg,nom,neut}   beautiful{sg,nom,neut}   what{sg,nom,neut} any{indecl}   not{indecl} is{pres,3p,sg}

    Eng: Nothing beautiful, similar to the face, exists.

    Here, *sadṛśam* is compounded with *mukham* with *Ṣaṣṭhītatpuruṣa samāsa*, therefore it is *samāsagā Upamā*. The figure 18 shows the parse tree and semantic structure.

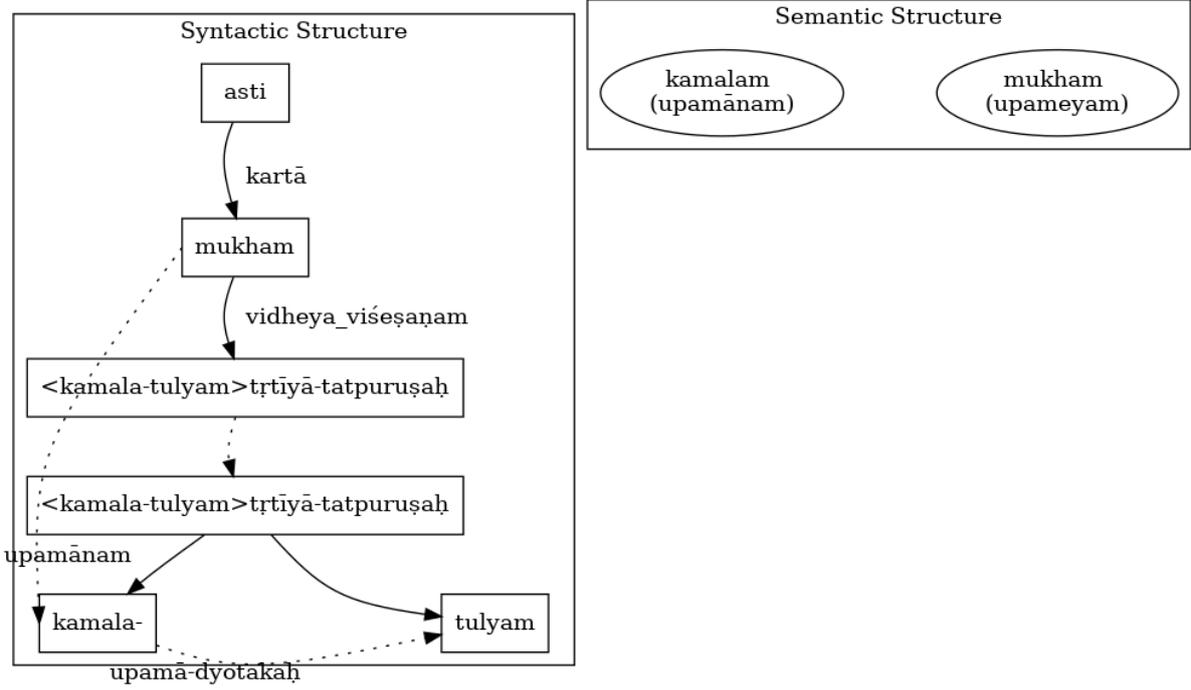8. ***Samāsagā Ārthī Dharma-upamānaluptā***

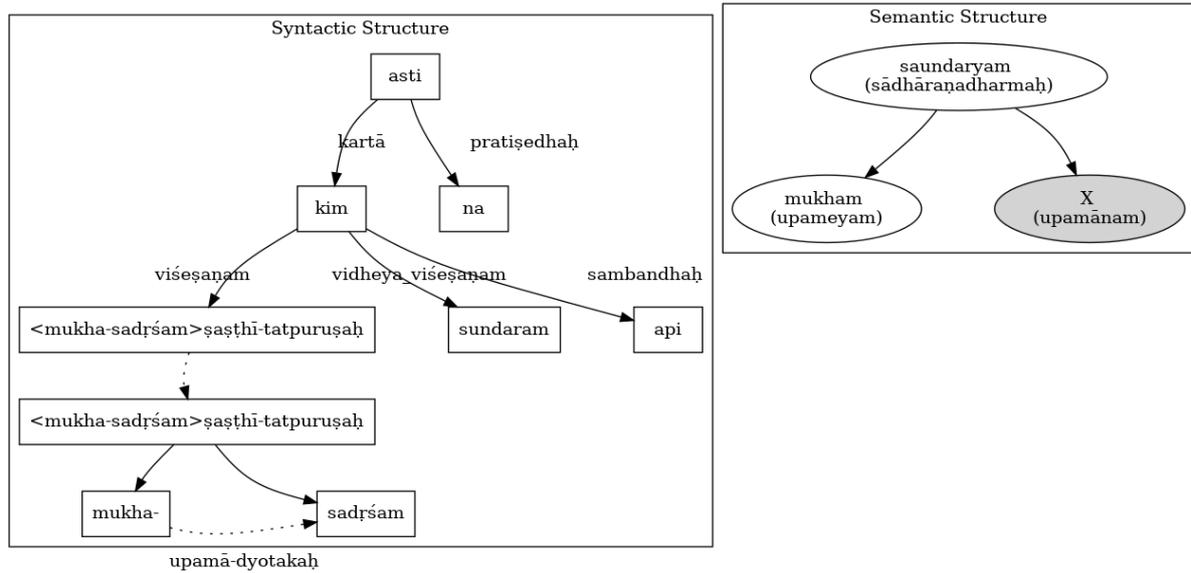Figure 17: Samāsagā Ārthī Dharmaluptā

Figure 18: Samāsagā Ārthī Upamānaluptā

Skt: *mukha-sadṛśam   kim   api   na   asti.*

Gloss: face-similar{sg,nom,neut}   what{sg,nom,neut}   any{indecl}   not{indecl}
is{pres,3p,sg}

Eng: Nothing similar to the face exists.

The analysis for this example will be similar to the previous *Upamānaluptā Ārthī Sam̐sagā* type without the common property. This example is demonstrated with the parse tree and semantic structure in figure 19.
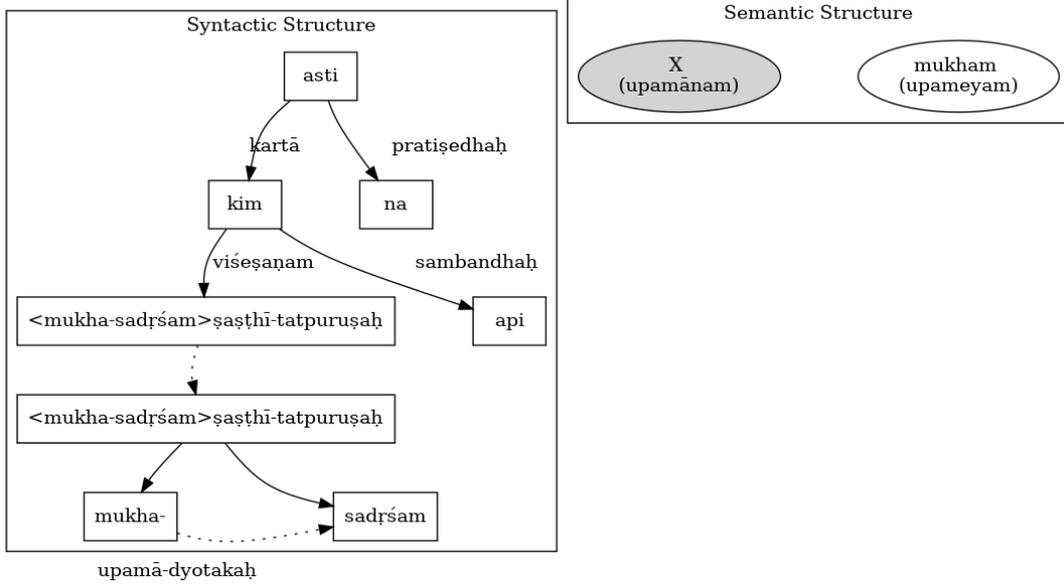

Figure 19: Samāsagā Ārthī Dharma-upamānaluptā

9. **Triluptā**: ***Samāsagā Ārthī Dharma-upamāna-vācaka-luptā***

   Skt: *mṛganayanā   manaḥ   harate.*

   Gloss: deer-eyed{sg,nom,fem}   mind{sg,acc,neut}   steals{pres,3p,sg}

   Eng: The doe-eyed (woman) steals the mind.

   mṛganayanā is a *bahuvrīhi compound (upamānapūrvapada-vyadhikaraṇa-bahuvrīhi)* under the *sūtra anekamanyapadārthe (A2.2.24)* and the *vārtika - 'saptamyupamānapūrvapadasyottarapadalopaśca'*. *Upamāna-pūrvapada-vyadhikaraṇa-bahuvrīhi* is a kind of *bahuvrīhi* compound, where the preceding component is *upamāna*. This compound has the meaning '*mṛga-locane iva cañcale nayane yasyāḥ sā*'. Here, *mṛga-locane* is *upamāna* out of which only a part *mṛga* is explicit in the sentence, and the latar component-*nayane* is *upameya*. The parse tree, along with the semantic representation of the given instance, is shown in the figure 20.

- *Pratyayagā Upamā*:

  The comparison in *upamā* expressed through affixes is considered as *Pratyayagā Upamā*. *Taddhita* affixes are attached to a noun to form a new *prātipadika* (nominal stem). Similarly, some *sanādi (kyac, kyaṅ, kvip)* and *kṛt (ṇamul)* affixes are attached to a noun to form a new verb. There are 10 types of affixes which are used in *Upamā* to denote similarity.

  1. ***Pratyayagā Śrautī Pūrṇopamā***

     Skt: *mukhasya   saundaryam   kamalavat   asti.*

     Gloss: face{sg,gen,neut}   beauty{sg,nom,neut}   lotus-like{indecl}
     is{pres,3p,sg}

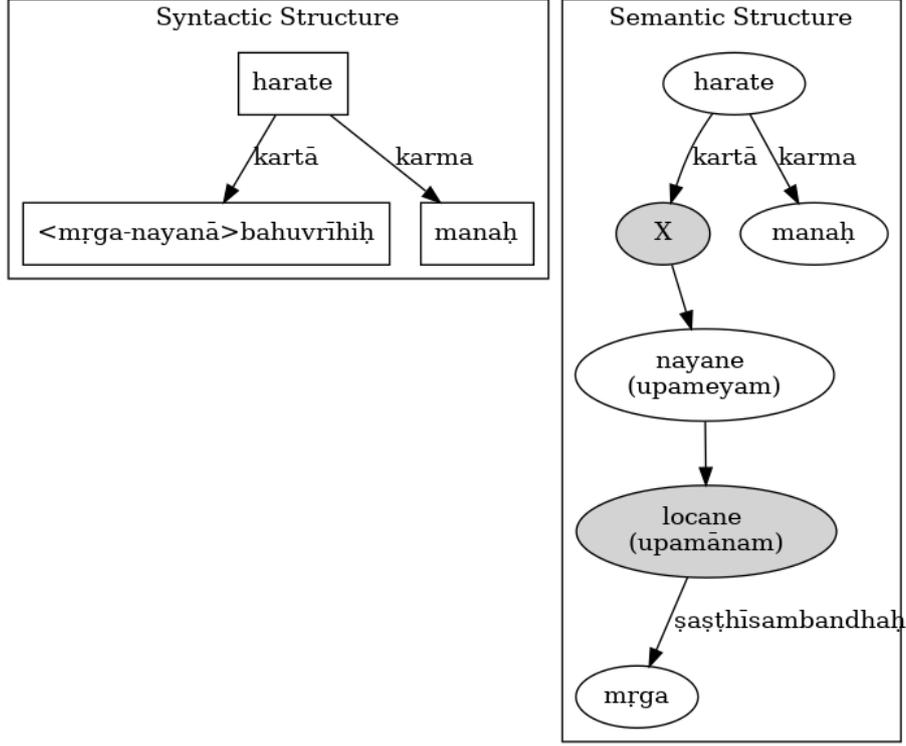   **Eng:** The beauty of the face is like a lotus.

Figure 20: *Samāsagā Ārthī Dharma-upamāna-vācaka-luptā*

Here, the *upamāna* along with the *dyotaka* are presented by the word *kamalavat*. By '*tatra tasyeva*'(A5.1.116), to depict the similarity, the '*vati*' *tadhhita* affix is added to the *upamāna*. Figure 21 demonstrates the parse tree and the semantic structure for the sentence.

2. ***Pratyayagā Ārthī Pūrṇopamā***

   Skt: *mukham   kamalavat   sundaram   asti.*

   Gloss: face{sg,nom,neut}   lotus-like{indecl}   beautiful{sg,nom,neut} is{pres,3p,sg}

   Eng: The face is beautiful like a lotus.

   In this sentence, *vati*, a *taddhita* affix is added in the meaning of similarity by '*tena tulyam kriyā cedvatiḥ*'(A5.1.116). Insteadof *kamalena tulyam* the taddhita ending word *kamalavat* is used. Figure 22 demonstrates the syntactic and semantic representation of the example.

   The following verse has the examples of *kyac* and *kyaṅ* affixes.

   > *pauraṃ **sutīyati** janaṃ samarāntare'sāv**antaḥpurīyati** vicitracaritracuñcuḥ/*
   > ***nārīyate** samārasīmni kṛpāṇapāṇerālokya tasya caritāni sapatnasenā//*

3. ***Pratyayagā Śrautī Vācakaluptā karma - kyac***:

   Skt: *asau   pauraṃ   sutīyati.*

   Gloss: that{sg,nom,masc}   citizen{sg,acc,neut}   treats-like-children{pres,3p,sg}

   Eng: That one treats citizens as children.

   The word '*sutīyati*' has the *kyac* affix in the meaning of '*sutam iva ācarati*' by the *sūtra* - *upamānādācāre* (A3.1.10). Refer to figure 23 for the semantic and syntactic structure of the sentence.

4. ***Pratyayagā Śrautī Vācakaluptā ādhāra - kyac***:
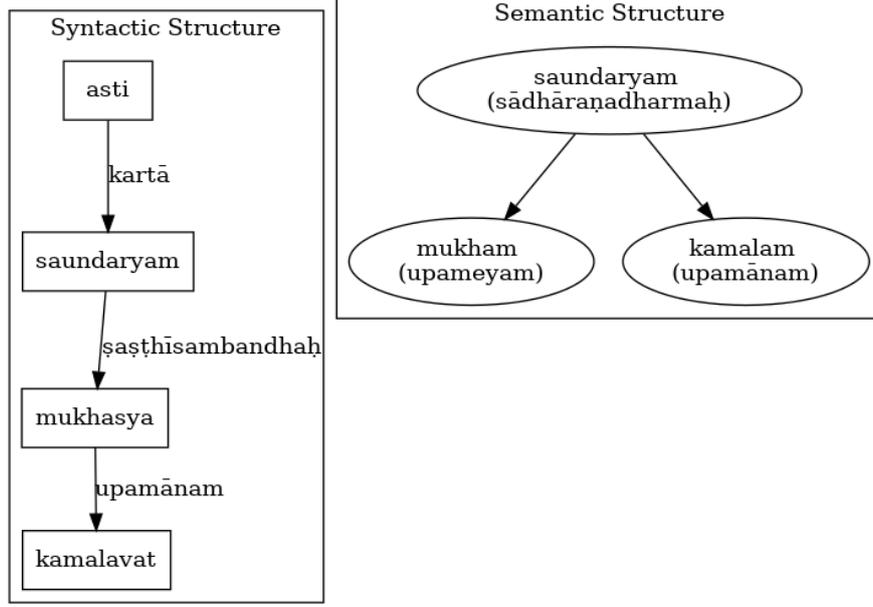
   Skt: *asau   samare   antaḥpurīyati.*

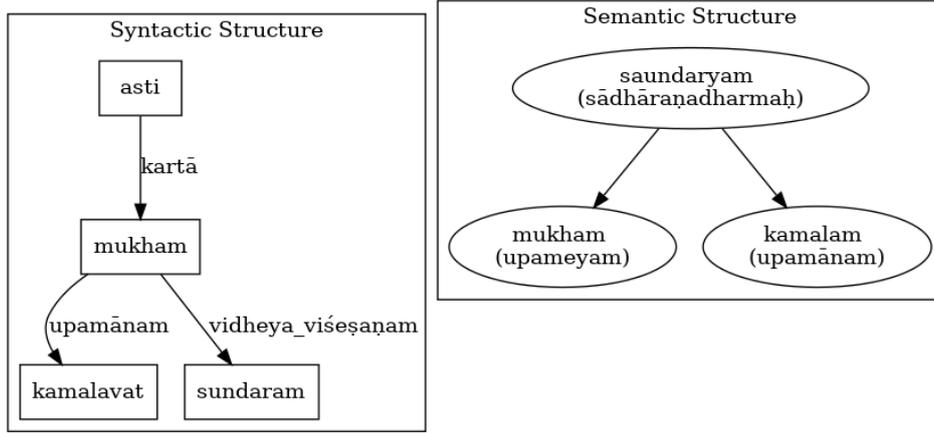Figure 21: *Pratyayagā Śrautī Pūrṇopamā*
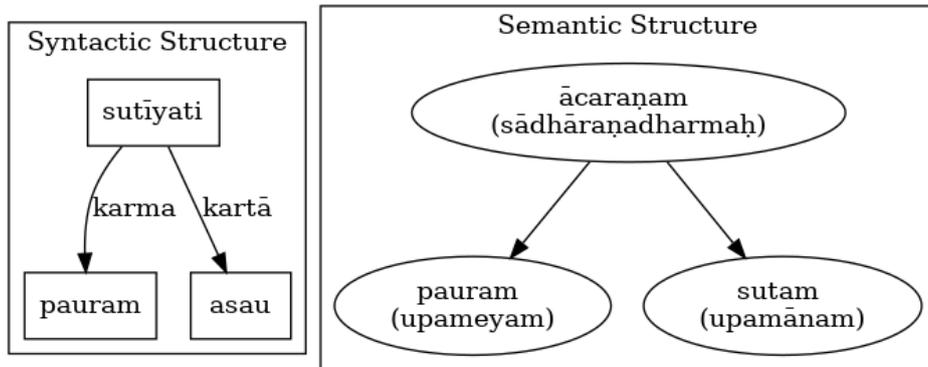

Figure 22: *Pratyayagā Ārthī Pūrṇopamā*

Figure 23: *Pratyayagā Śrautī Vācakaluptā karma - kyac*

Gloss: that{sg,nom,masc}   battle{sg,loc,neut}   behaves-as-in-apartment-{pres,3p,sg}

Eng: That one behaves in battle, as he is in his home.

The word '*antaḥpurīyati*' having *kyac* suffux gives the meaning of '*antaḥpure iva ācarati*' by the *vārtika* '*adhikaraṇācceti vaktavyam*' on the *sūtra upamānādācāre* (A3.1.10) . Figure 25 shows the annotation parse tree and syntactic representation of the given example.
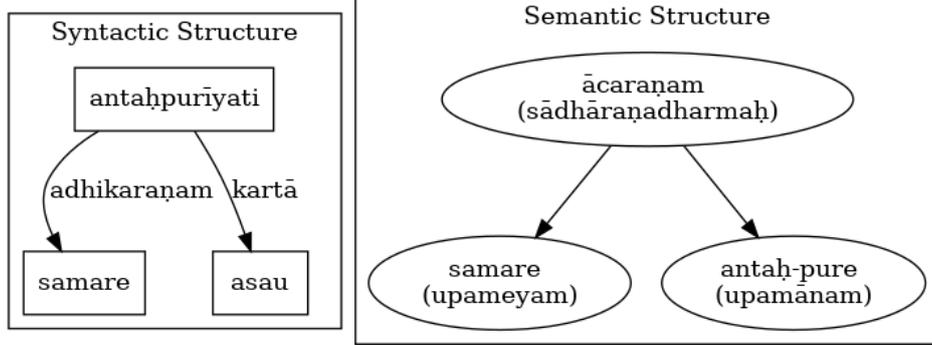

Figure 24: *Pratyayagā Śrautī Vācakaluptā ādhāra - kyac*

5. ***Pratyayagā Śrautī Vācakaluptā kartā - kyaṅ***:

Skt: *senā   nārīyate.*

Gloss: army{sg,nom,fem}   behaves-like-woman{pres,3p,sg}

Eng: The army behaves like a woman.

The word from the above given verse, *nārīyate* has *kyaṅ* affix in the meaning of *kartā* for *nārī iva ācarati* by the *sūtra* '*kartuḥ kyaṅ salopaśca*' (A3.1.11). Figure 28 shows the parse tree with the semantic structure for the same.
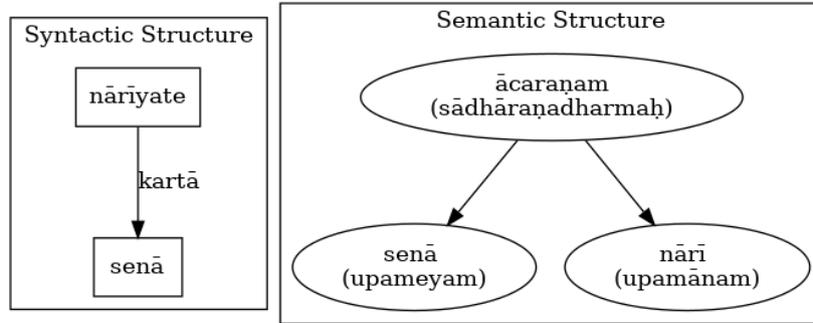

Figure 25: *Pratyayagā Śrautī Vācakaluptā kartā - kyaṅ*

6. **Pratyayagā Śrautī Vācakaluptā karma - ṇamul**:

Following verse demonstrates the affix *ṇamul* in two different meanings. The first line has the affix in *karma* meaning and the second line has it in *kartā* meaning.

> *mṛdhe **nidāghagharmāṃśudarśaṃ** paśyanti taṃ pare |*
> *sa punaḥ **pārthasaṃcāraṃ** saṃcaratyavanīpatiḥ ||*

Skt: *taṃ   gharmāṃśudarśaṃ   paśyanti.*

Gloss: he{sg,acc,masc}   Sun-see{sg,acc,masc}   see{pres,3p,pl}

Eng: He was looked upon as the Sun.

Here, '*nidāghagharmāṃśudarśaṃ*' (for *nidāghagharmāṃśum iva paśyati*') has the *ṇamul* affix in the meaning of *karma* by the *sūtra* '*upamāne karmaṇi ca' (A3.4.45)*. Refer to the figure 26 for the detailed parse tree and semantic structure of the sentence.
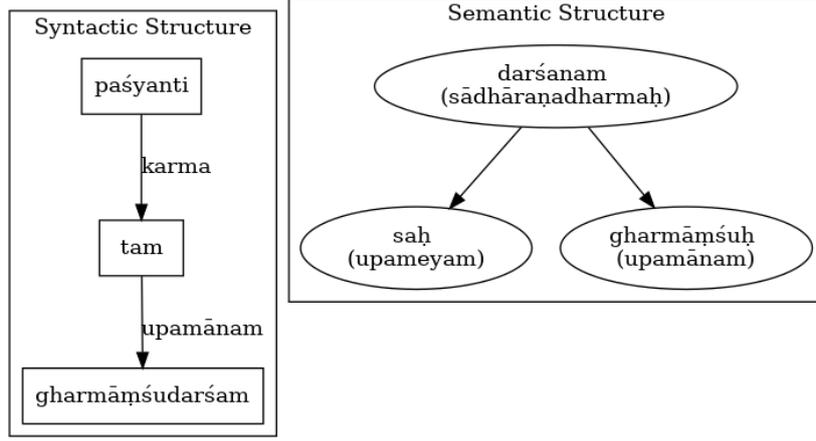
Figure 26: *Pratyayagā Śrautī Vācakaluptā karma - ṇamul*

7. **Pratyayagā Śrautī Vācakaluptā kartā - ṇamul**:

   Skt: *saḥ   pārthasañcāraṃ   sañcarati.*

   Gloss: he{sg,nom,masc}   movement-of-Arjuna{sg,acc,masc}   move{pres,3p,sg}

   Eng: He moves like Arjuna.

   '*pārthasaṃcāraṃ*' (for pārthaḥ iva sañcaraṇam) has the '*ṇamul*' affix in the meaning of *kartā* by the same *sūtra* as above. Figure 27 shows the syntactic and semantic structure for the given example.
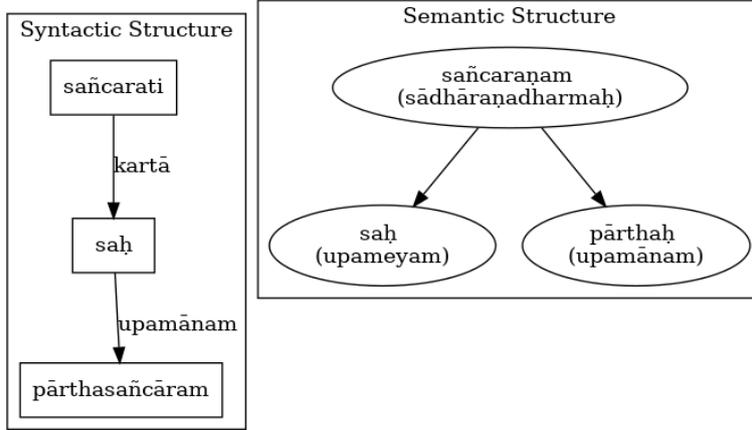

Figure 27: *Pratyayagā Śrautī Vācakaluptā kartā - ṇamul*

8. **Pratyayagā Śrautī Vācaka-upameyaluptā : kyac**

   Skt: *rājā   sahasrāyudhīyati*

   Gloss: king{sg,nom,masc}   behaves-like-having-thousand-weapons{pres,3p,sg}

   Eng: The king acts as if he is the 'kārtavīrya' having thousand arms.

   Here, '*sahasrāyudhīyati*' has *kyac* affix giving the meaning - '*sahasrāyudham iva ācarati ātmānam*'. Figure 28 provides the parse tree and the semantic representation of the stated instance.

9. **Pratyayagā Śrautī Dharma-vācakaluptā - kvip** :

   Skt: *yāminyaḥ   dinanti*

   Gloss: night{pl,nom,fem}   acts-like-day{pres,3p,pl}

   Eng: The nights feels like days.

   Here, '*yāminyaḥ*' (nights) is the *upameya*. The *kvip* affixed word '*dinanti*' having the
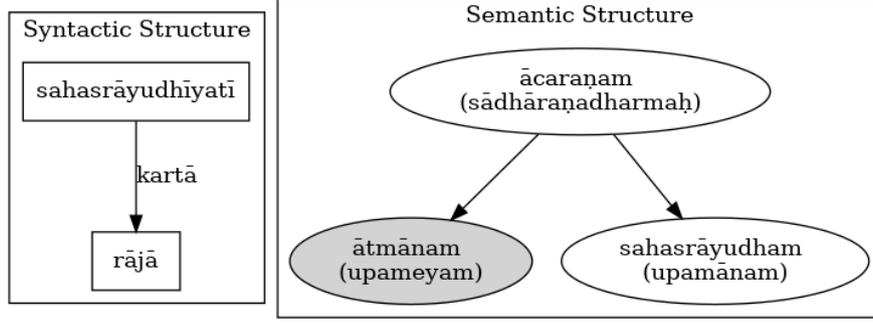
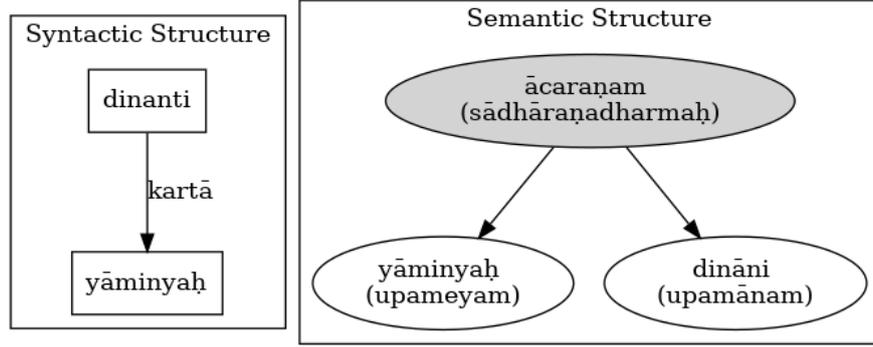Figure 28: *Pratyayagā Śrautī Vācaka-upameyaluptā : kyac*

Figure 29: *Pratyayagā Śrautī Dharma-vācakaluptā - kvip*

meaning '*dinam iva ācarati*' expressing *upamānam*(*dinam*), while implicitly provides the *dyotaka* and the *dharma*. The affix '*kvip*' gets elided due to the *sūtra veraprktasya A6.1.67).* Therefore, the common property 'ācāraḥ', which is expressed through the suffix, is also considered as omitted.[8] Figure 29 shows the parse tree and semantic representation.

10. **Pratyayagā Ārthī Dharmaluptā**

    Skt: *duṣṭasya   manaḥ   **viṣakalpaṃ**   asti*

    Gloss: wicked{sg,gen,masc}   mind{sg,nom,neut}   poison-like{sg,nom/acc,neut} is{pres,3p,sg}

    Eng: The mind of a wicked person is like a poison.

    In this example, the word *viṣakalpaṃ* has the *taddhita* affix '*kalpap*' by the *sūtra* '*īṣadasamāptau kalpabdeśyadeśīyaraḥ (A5.3.67)*'. Figure 30 shows the parse tree along with the semantic structure for the given example.

We observe that out of 25 subtypes of *Upamā*, the parser can analyse *Vākyagā Upamā*s of 6 types and *Samāsagā* of 9 types, making a total of 15. The 10 *Upamā*s involving *Taddhita*, *sanādi* and *kṛt* suffixes cannot be handled by the current system due to the non-availability of morphological analysers to analyse these type of words. In the analysis, the major cue to capture the *Upamā* is the comparative marker. In the types *Vākyagā*, *Samāsagā*, where the *dyotaka* is explicit, the parser is able to recognise it as *Upamā* and assign appropriate tags, without any difficulty. Due to the addition of the 'Compound-type Identification module' (Krishnan et al., 2025) the 9 *samāsagā upamā* sentences can also be processed by the parser.

---

[8]*vivaraṇakārāstu kippratyaya eva tulyācārarūpasādhāraṇadharmavācakaḥ/ tasya "veraprktasya"-iti sūtreṇa nityaṃ lupta-tvāddharmalopa ityāhuḥ/ 'Balabodhinī' commentary on 10.89, Kāvyaprakāśa*
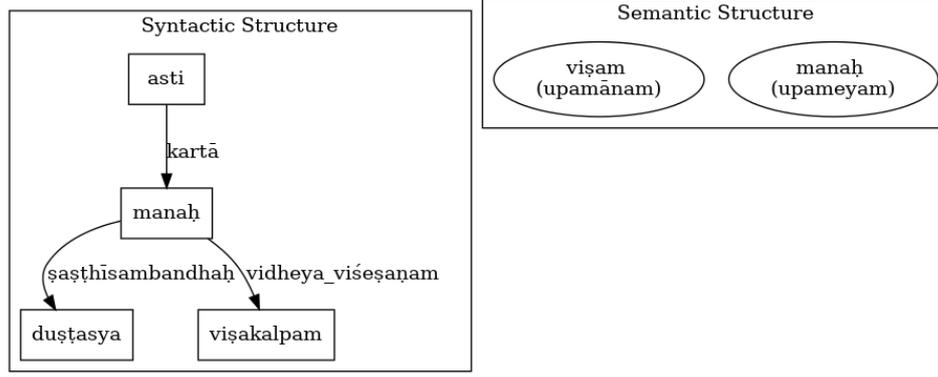
Figure 30: *Pratyayagā Ārthī Dharmaluptā*

4.3 Handling of special nuances from literature

In *Kāvyaprakāśa*, *Mammaṭa* systematically addresses the major grammatical categories and illustrates each of them with corresponding examples. These examples typically employ a single *upamāna–upameya* pair and conform to the standard construction of *Upamā*. However, in actual poetic practice, the use of *Upamā* often departs from this standard form. A poet, exercising creative freedom, may construct expressions involving more than one *upamāna–upameya* pair. In some instances, a single *upameya* may be compared with multiple *upamānas*. Thus, poetic constructions frequently display structural variations beyond the canonical type.

When designing guidelines for parse tree representation, it is therefore necessary to accommodate such creative possibilities alongside the standard forms. Furthermore, certain other *alaṅkāras* exhibit structures similar to *Upamā*. To avoid ambiguity, these must be represented in a manner that distinguishes them from *Upamā*. Accordingly, the guidelines are extended to incorporate these nuances, ensuring that both the creativity of poetic expression and the structural particularities of related *alaṅkāras* are adequately captured.

- **More than one *Upamāna*s and *upameya*s with one comparative marker**:

  Skt: *sītā api anugatā rāmaṃ śaśinaṃ rohiṇī yathā*

  Gloss: Sītā{sg,nom,fem} also{indecl} followed{ptcp,past,nom,sg,fem} Rāma{sg,acc,masc} moon{sg,acc,masc} Rohiṇī{sg,nom,fem} like{indecl}

  Eng: Sītā also followed Rāma, as Rohiṇī follows the moon.

  This sentence is taken from *Rāmāyaṇam* 1.1.28. In this sentence, there are two *upamāna*s and their respective *upameya*s. But only one comparative marker is given. In this case, we can mark the relation of *yathā* with *rohiṇī* after which *upamā-dyotaka* is present in the sentence. We mark the relation of *upamānam* between *rāmam* and *śaśinam*; also in between *sītā* and *rohiṇī*. Figure 31 shows the parse tree for the given example.

- ***Iva* not connected with the *upamāna* in *samāsa*.**
  Iva forms a compound with *upamāna* without the deletion of the case marker even when it is a *samāsa*.[9] The samāsa which is formed with *iva* should have the *upamāna* as the initial component. But, in some verses[10] we have observed that *iva* is attached to some other word than the *upamāna*. In such cases, we treat it as *Vākyagā Upamā*. We mark *iva* as an independant word with the relation *upamā-dyotakaḥ* with the *upamāna* and not with the word with which *iva* is attached.

---

[9]*ivena samāso vibhaktyalopaśca/ vārtikam under Aṣṭādyāyī 2.1.4 and 2.4.71*

[10]*udbāhuriva vāmanaḥ/ 1.3, Raghuvaṃśa*
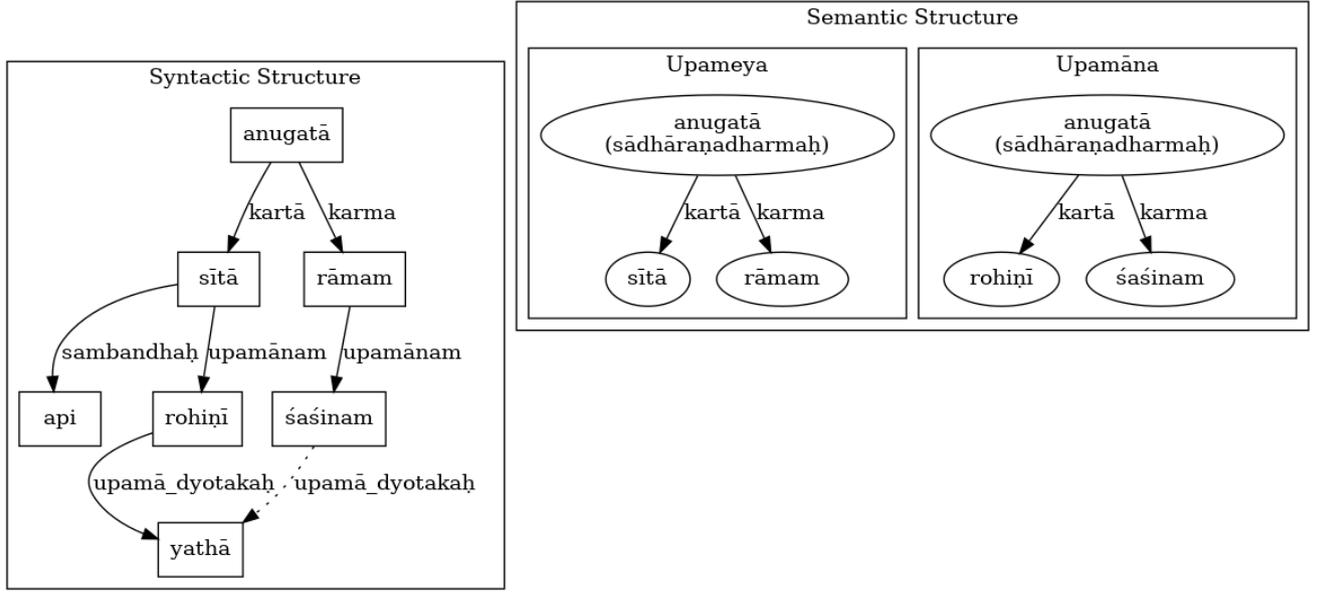*jugopa gorūpadharāmivormim/ 2.3, Raghuvaṃśa*

Figure 31: Representation of more than one *upamāna*s and *upameya*s

iti vismṛtānyakaraṇīyamātmanaḥ sacivāvalambidhuraṃ dharādhipam |
parivṛddharāgamanubandhasevayā mṛgayā jahāra catureva kāminī ||RV, 9.69||

Gloss: (*iti*) thus{indecl}   (*vismṛtānyakaraṇīyamātmanaḥ*) whose-other-duties.
forgotten-by-himself{sg,gen,masc}   (*sacivāvalambidhuraṃ*) minister-reliance.
burden{sg,acc,masc}
(*dharādhipam*) king{sg,acc,masc}   (*parivṛddharāgamanubandhasevayā*)
with-service-of-increasing-passion{sg,instr,fem}   (*mṛgayā*) hunt{sg,nom,fem}
(*jahāra*) carried-off{aor,3p,sg,act}   (*caturā*)cleverly   (*iva*) like{indecl}
(*kāminī*) beloved-woman{sg,nom,fem}

Eng: Thus, the hunt, with service bound up with growing passion, carried off the king
who, having forgotten other duties, bore the burden of reliance on his ministers just as a
clever woman seizes her lover.

This verse is taken from *Raghuvaṃśam*(RV). In this verse the *upamā-dyotaka* is *iva* which
is attached with an adjective of the *upamāna* that is *caturā* and not with the *upamāna -
kāminī*. To show the annotation for such sentence, here, we show the relevant words from
the the above mentioned verse having the components of *Upamā*. Figure 32 shows the
parse tree for the given example.
Example: adhipam mṛgayā jahāra catureva kāminī

- **Differentiating between *Upamā* and other *arthālaṅkāras* when *iva* is the marker.**

  In addition, several other *arthālaṅkāras* also employ a *dyotaka* (comparative marker) similar
  to that used in *Upamā*. For instance, *Utprekṣā* and *Ananvaya* exhibit surface structures
  that can be mistaken for *Upamā* by a computational parser. However, their semantics and
  poetic effect differ significantly. In *Utprekṣā*, the *Upamāna* is imaginary or hypothetical,
  whereas in *Ananvaya* the *Upamāna* and the *Upameya* refer to the same entity. On the same
  grounds, *Upameyopamā* is an *alaṅkāra* where the comparison is done reciprocally. Here, we
  show the representative parse trees for these *alaṅkāras*.

  - ***Ananvaya***
    The following example is taken from *Rāmāyaṇa* (6.107.52). In the source text the śloka
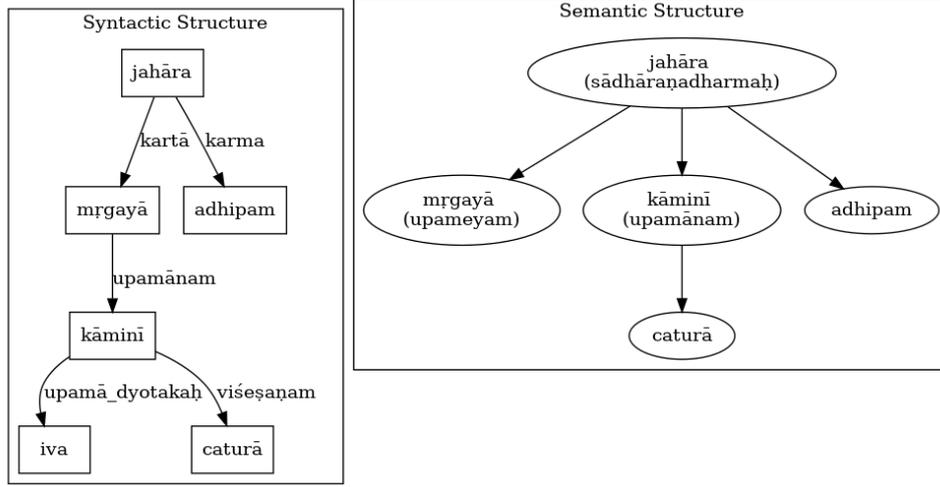
Figure 32: Example parse tree for *Iva* not connected with the *upamāna*

is as 'rāmarāvaṇayoryuddham rāmarāvaṇayoriva'. Here, we have to supply '*yuddham*' as the *upamāna*.

Skt: *rāma-rāvaṇayoḥ   yuddham   rāma-rāvaṇayoḥ   yuddham   iva   asti*

Gloss: Rāma-Rāvaṇa{du,gen,masc}   battle{sg,nom/acc,neut}
Rāma-Rāvaṇa{du,gen,masc}   battle{sg,nom/acc,neut}   as.if{indecl}
be{pres,3p,sg}

Eng: The battle of Rāma and Rāvaṇa is only like the battle of Rāma and Rāvaṇa.

In *ananvayā*, the *upameya* is depicted as there is no other entity to compare with it. In other words, the *sādhāraṇadharma* which is being compared cannot be seen in any other object. Hence, the *Upameya* itself is presented as the *upamāna* as in the given example. The war between *Rāma* and *Rāvaṇa* is so great that it can be compared only with itself. The parse tree representation is shown is figure 33.
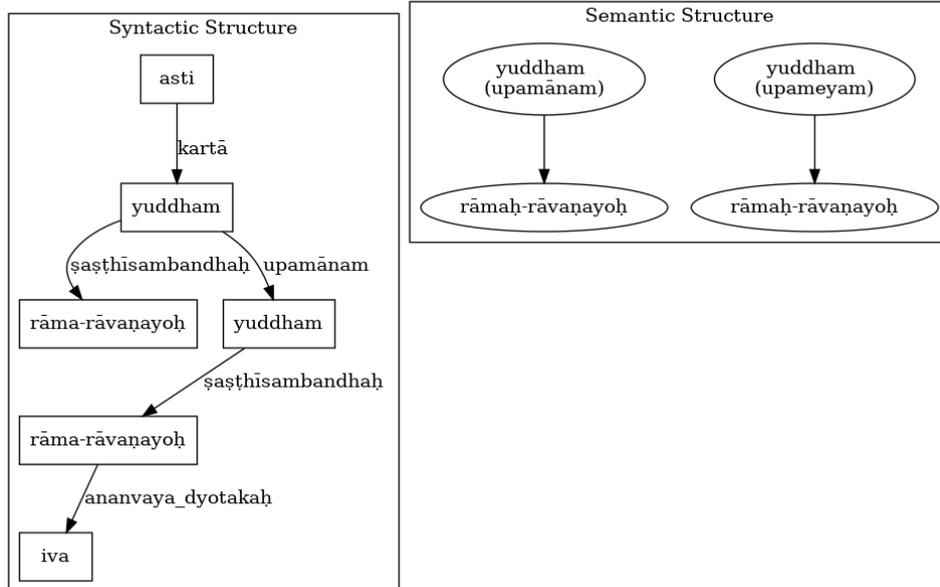

Figure 33: Representation of *Ananvayā*

– **Upameyopamā**
   Skt: *dharaṇī   iva   tasya   dhṛtiḥ   asti/ tasya   dhṛtiḥ   iva   dharaṇī   asti/*

Gloss: earth{sg,nom,fem}   like{indecl}   his{sg,gen,masc}   firmness{sg,nom,fem} be{pres,3p,sg}   his{sg,gen,masc}   firmness{sg,nom,fem}   like{indecl} earth{sg,nom,fem}   be{pres,3p,sg}

Eng: His firmness is like the Earth; and the Earth is like his firmness.

In this *alaṅkāra*, to show the degree of similarity between the *upamāna* and the *upameya*, both are used interchangeably. The example 'his fortitude' (tasya dhṛtiḥ) is compared with 'the earth' (dharaṇī) and vice versa as demonstrated in figure 34.
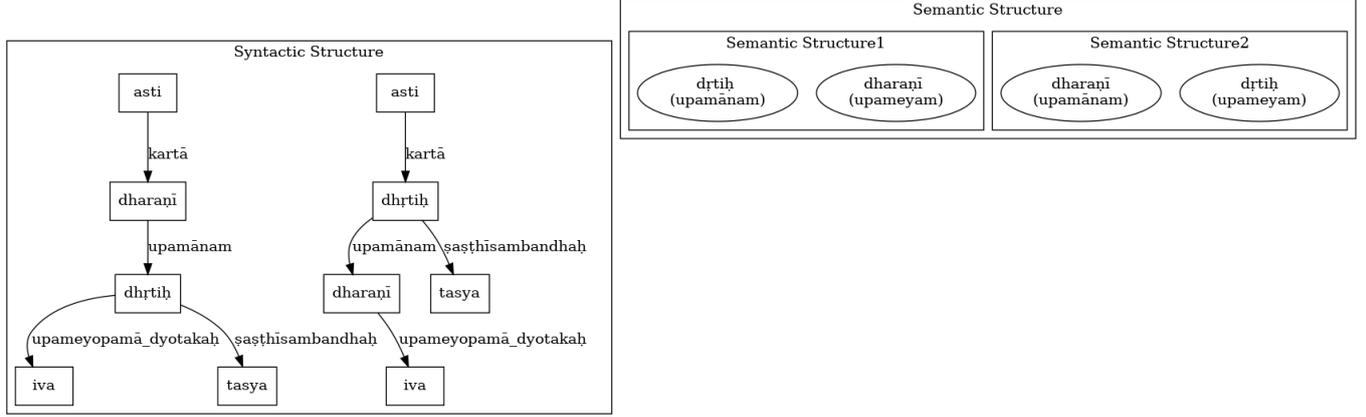

Figure 34: Representation of *Upameyopamā*

– **Utprekṣā**:

The figures explained before have the common property as the basis of the comparison. In *Utprekṣā* the common property is the subject of suggestive meaning (*vyaṅgyasādharmya*). In this *alaṅkāra*, the object is imagined or forced with the said common property to be compared as *upamāna*. Hence, we do not show the suggested *sādhāraṇadharma* in the semantic representation as it may vary from person to person.

Skt: *limpati   iva   tamaḥ   aṅgāni*

Gloss: cling{pres,3p,sg}   as.if{indecl}   darkness{sg,nom,neut} limbs{pl,acc,neut}

Eng: It is as if darkness clings to the limbs.

In this example, the marker *iva* is attached to the action *limpati*, which is a clue for identifying *utprekṣā*. Unlike *Upamā*, there is no *upamāna-upameya-bhāva* present in the sentence. The action of smearing (*limpati*) is imposed by the poet. Figure 35 shows the parse tree for the same.
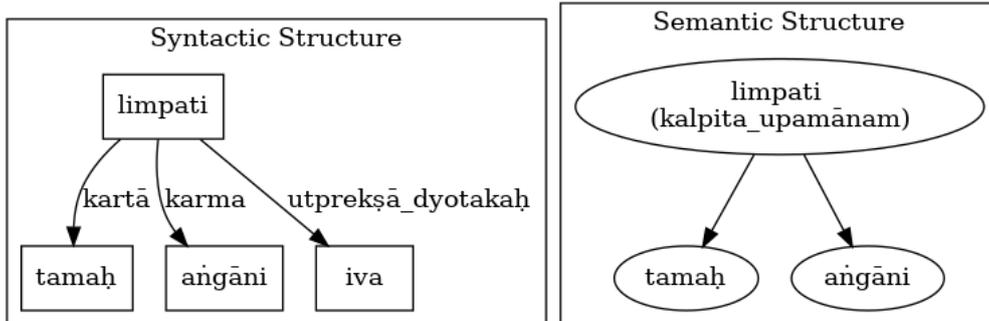

Figure 35: Representation of *Utprekṣā*

5 Conclusion

Upamā is primarily found in all litarary texts to enhance the richness of the poetry. To analyse sentences carrying *Upamā*, the parser must be designed to recognise and appropriately tag these constructions.

We presented a tagging scheme to annotate the *Upamā* construction with different grammatical categories along with the special literary constructions and similar other *alaṅkāra*s. However, there are further challenges to make the parser capable of processing such complex constructions. Presented below is a list of challenges that are particularly difficult for a machine to tag at the initial stage of processing.

To improve the parser's coverage across a broader range of *Upamā* constructions, the following four areas of enhancement are proposed:

1. **Integration of a *taddhita* and *sanādi* Module:** A dedicated module to analyze *taddhita* and *sanādi* affixes and their semantic functions in conveying comparison.

2. **Enhancement of the *Samāsa* Module:** A robust *samāsa* analysis system to identify embedded relationships within compounds where explicit comparative markers are absent.

3. **Expansion of the Comparative Marker Lexicon:** An enriched list of comparative markers (*dyotaka*s) drawn from classical literature to cover a wider spectrum of linguistic realisations.

4. **Formulation of Special Case Rules:** Custom rules to address nuanced literary examples not fitting standard classifications, such as verses with only one comparative marker, or multiple *upamāna–upameya* pairs, so that the parser can handle such deviations gracefully.

By addressing the listed areas, the system can be better equipped to analyze the diverse and complex instantiations of *Upamā* found in Sanskrit literature.
.

References

Ācārya Viśveśvara. 2017. *Kāvyaprakāśa*. jñānamaṇḍala Limited, Varanasi.

Manomohan Ghosh. 1951. *The Nāṭyaśāstra*, volume 1. Asiatic Society of Bengal, Calcutta.

Bhakti Jadhav, Amruta Barbadikar, Amba Kulkarni, and Malhar Kulkarni. 2023. Issues in the computational processing of upamāalaṅkāra. In Jyoti D. Pawar and Sobha Lalitha Devi, editors, *Proceedings of the 20th International Conference on Natural Language Processing (ICON)*, pages 180–190, Goa University, Goa, India, December. NLP Association of India (NLPAI).

Bhakti Jadhav, Himanshu Dutta, Shruti Kanitkar, Malhar Kulkarni, and Pushpak Bhattacharyya. 2025. An introduction to computational identification and classification of upamā alaṅkāra. In Amba Kulkarni and Oliver Hellwig, editors, *Computational Sanskrit and Digital Humanities - World Sanskrit Conference 2025*, pages 1–14, Kathmandu, Nepal, June. Association for Computational Lingustics.

Ganganath Jha. 1928. *Kāvyālaṅkārasūtravṛttiḥ*. Oriental Book Agency, Poona.

Ganganath Jha. 1967. *Kāvyaprakāśa of Mammaṭa with English Translation*. Bharatiya Vidya Prakashan, Varanasi.

Sriram Krishnan, Pavankumar Satuluri, Amruta Barbadikar, T S Prasanna Venkatesh, and Amba Kulkarni. 2025. Compound type identification in Sanskrit. In Amba Kulkarni and Oliver Hellwig, editors, *Computational Sanskrit and Digital Humanities - World Sanskrit Conference 2025*, pages 90–108, Kathmandu, Nepal, June. Association for Computational Lingustics.

Revati Milind Kulkarni and Malhar A. Kulkarni. 2025. Morph-dependency model for upamā. *Language in India*, 25(9):146–156.

Amba Kulkarni. 2021. Sanskrit parsing following indian theories of verbal cognition. volume 20, New York, NY, USA, April. Association for Computing Machinery.