Anuprāsa Identifier and Classifier: A computational tool to analyze Sanskrit figure of sound

Amruta Barbadikar
Department of Sanskrit Studies,
University of Hyderabad
amruta.barbadikar@gmail.com

Amba Kulkarni
Department of Sanskrit Studies,
University of Hyderabad
ambakulkarni@uohyd.ac.in

Abstract

Anuprāsa is a *śabdālaṅkāra* (figure of sound), in which the poetry is embellished by the repetitive occurrence of letters.[1] The task of processing the decorative language consisting of such figures is a path not explored in the field of Sanskrit computational linguistics. This paper discusses a tool that identifies and classifies *anuprāsa alaṅkāra*. *Anuprāsa*, being a figure of sound, makes the least use of semantics. This tool is essentially developed upon the insights taken from the school of *alaṅkāra*s, especially from the treatise of ĀCĀRYA VIŚVANĀTHA from 14th century AD.

1 Introduction

There is a varied scope for research in the field of Sanskrit computational linguistics. Segmentation(Goyal and Huet, 2013), Morph-analysis and Generation (Kulkarni and Shukl, 2009), Compound type analysis (Kulkarni and Kumar, 2013; Kulkarni and Kumar, 2011) and Generation (Satuluri and Kulkarni, 2013), Sentence analysis (Goyal et al., 2009; Kulkarni et al., 2020) and Generation (Kulkarni and Pai, 2019), Discourse analysis (Kulkarni and Das, 2012), Translation (Agrawal and Madaan, 2020), etc. tasks are being carried out with extensive efforts. Most of the tasks are grammar oriented. Other works which attempt to process the rhetoric of the poetic language are very limited. Except for the Meter identification (Melnad et al., 2015; Rajagopalan, 2018; Neil, 2023; Terdalkar and Bhattacharya, 2023) and a tool to identify and classify the *yamaka alaṅkāra* (Barbadikar and Kulkarni, 2023), the Natural Language Processing (NLP) tools are far away from processing poetic beauty in Sanskrit.

Although, some examples in languages other than Sanskrit for processing figurative language can be found. Shutova (2011) presented a computational approach to process metaphor using statistical methods. Englard (2013) used the rhetorical analysis of text to predict the author. For Hindi, Audichya and Saini (2021), worked out the *alaṅkāra*s in Hindi to present the hierarchical structure with a taxonomical listing of *alaṅkāra*s. However, the computational implementation was not exercised. Naaz and Singh (2022) were able to contribute by presenting three different tools for Hindi. 'Text2Mātrā' produces the *laghu* and *guru mātras* for the input, 'RPaGen' detects the rhyming quality of the poem and 'FoSCal' generates a score according to the quantity of *anuprāsa* used over the poem. For Sanskrit, the automatic Meter Identification task has been worked out, from various applications perspectives, by different scholars. Melnad et al. (2015), Rajagopalan (2018), Neil (2023), Terdalkar and Bhattacharya (2023) are some of the notable contributors to the available state-of-the-art Sanskrit Meter Identification systems.

Highly complex language structures, use of intended and implicit meaning, multiple meanings of a word, multiple words having a similar meaning, and unavailability of useful state-of-the-art tools are the factors that discourage the processing of decorative language used in poetry. The tradition of *alaṅkāraśastra* (poetics) is developed over a long period ranging from the 1st century AD. The study of figures of speech is an important stream of this tradition. Figures of speech

[1]Here, the repetition of sounds is desirable. In Sanskrit, there is one to one mapping of sound with the denoting letter and we compute letters, not sounds. Hence we use the terms letters and sounds interchangeably.

are employed by the poets to enhance the beauty of the poetry. Even in the Vedic literature, the use of such devices can be traced.

ĀCĀRYA BHARATA from 1st century AD, known as the father of Indian poetics, in his treatise NĀṬYAŚĀSTRA, describes only 4 *alaṅkāra*s. Whereas in KUVALAYĀNANDA of APPAYA DĪKṢITA (16th century AD) 125 types of *alaṅkāra*s are enlisted. Sanskrit has a rich tradition of poetics that is 2000 years old starting from BHARATA'S NĀṬYAŚĀSTRA. There are six primary schools of poetics in Sanskrit viz. *rasa, alaṅkāra, rīti, dhvani, vakrokti* and *auchitya*. The school of *alaṅkāra* is one of the most cherished schools. It is the ornamentation of poetry through the specific arrangement of syllables or words or astonishing meanings to enhance the rhetorical effect. We aim to concentrate on the computational analysis of the provided text essentially with the *alaṅkāra* point of view.

According to the school of *alaṅkāra, alaṅkāra*s are mainly of two types viz. *śabdālaṅkāra* (figures of sound) and *arthālaṅkāra* (figures of speech). The combination of these two is called as *ubhayālaṅkāra*. The count of *alaṅkāra*s differs from scholar to scholar. Approximately the count exceeds the number of 50. Some *alaṅkāra*s that use the phonetic or structural beauty may be easier to identify, but others would be quite tricky to recognize even for the experts in this field because of the involvement of deeper semantics. As we aim at dealing with these *alaṅkāra*s from the computational point of view, it is feasible to identify different syntactic constructions without considering the semantics in *śabdālaṅkāra*s like *yamaka* and *anuprāsa*. This research is aimed at the identification and classification of *anuprāsa* without considering the meaning.

It is a non-trivial task to analyze highly semantic and aesthetically rich texts without the help of machine learning or any advanced techniques of NLP. Like the Indian grammatical tradition, the rhetoric tradition has provided a robust theory upon which a foolproof rule-based system can be built. Hence, we relied upon a rule-based approach to accomplish this task. As we are dealing with *śabdālaṅkāra*, it allowed us to ignore the sense of the poetry making this task easier. We employ a simple rule-based algorithm after extracting various syntactic clues from the school of *alaṅkāra*s. For classification purpose, we select the best and most convincing, inclusive scheme proposed by VIŚVANĀTHA, a prominent scholar in the tradition of *alaṅkāraśāstra*.

2 *Anuprāsa*

Anuprāsa is a *śabdālaṅkāra*. Essentially, it is the repetition of consonants. This repetition should be in proximity such that one should remember the prior instance.[2] The phenomenon of *anuprāsa* is similar to Alliteration.[3]

Like *Yamaka*, *anuprāsa* holds an important place in the *alaṅkāra*s. In the tradition of *alaṅkāraśāstra*, *anuprāsa* was originally introduced as a subtype of *yamaka* viz. *mālā yamaka*.[4] *Yamaka* is a repetition of the longer sequence patterns of syllables engaged in the poetry (Barbadikar and Kulkarni, 2023), especially in metrical verses, whereas in *mālā yamaka* repetition of consonants is considered, which is similar to *anuprāsa*. In *yamaka* where repetition of the longer patterns are engaged in the poetry especially the metrical verses when employed in a more complex way, this might create a hindrance in the process of experiencing *rasa*.[5] Because the meaning is different in each repetition, the listener might find it difficult to understand the meaning of the complete verse and lose interest. But *anuprāsa* is considered to be a contributor to the emergence of *rasa*. *Anuprāsa* can be traced in any type of text suggesting any kind of *rasa*.

[2] *pūrvānubhavasaṃskārabodhinī yadi adūratā* //1.55, KĀVYĀDARŚA

[3] In alliteration, consonant sounds in two or more neighbouring words or syllables are repeated. The repeated sounds are usually the first, or initial, sounds as in "seven sisters", but repetition of sounds in non-initial stressed, or accented, syllables is also common: "appear and report."-"Alliteration." `https://www.merriam-webster.com/dictionary/alliteration`. Merriam-Webster, 2023.

[4] *nānārūpaiḥ svarairyuktaṃ yatraikaṃ vyañjanaṃ bhavet /*
tanmālāyamakaṃ nāma vijñeyaṃ paṇḍitairyathā//16.84, NĀṬYAŚĀSTRA

[5] *tadetatkāvyāntargaḍubhītam/* in the vṛtti of 83rd kārikā, 9th chapter, KĀVYAPRAKĀŚA

3 The conceptual development of the types of *anuprāsa*

Anuprāsa is independent of the form of poetry, that is, it is used in prose format also, like in KĀDAMBARĪ of BĀṆABHAṬṬA. The classifications found are based on the variations in repetitions in terms of the categories of repeated consonants, number of repetitions, number of repeated consonants and mood emergence due to the combination of different consonants. Hence, we do not observe many variations in the definition and classification, but the number of subtypes considered varies.

In this section, we present a brief overview of various types of *anuprāsa* furnished by different scholars. BHARATA, known as the first scholar of the tradition, has enlisted only four *alaṅkāra*s namely *yamaka, upamā, rūpaka* and *dīpaka* among which *yamaka* was the only figure of sound. *Mālā yamaka*, a subtype of *yamaka* can be considered as the inspiration behind *anuprāsa*. Example of *mālā yamaka* given by BHARATA is,

> *asau hi rāmā rativigrahapriyā*
> *rahaḥpragalbhā ramaṇaṃ manogatam /*
> *ratena rātriṃ ramayet pareṇa vā*
> *na cedudeṣyattaruṇaḥ paro ripuḥ* //16.86, NĀṬYAŚĀSTRA

Other scholars after BHARATA considered *anuprāsa* as an individual *alaṅkāra* and classified it from different perspectives. BHĀMAHA provided only two types, whereas BHOJA extended the count to 6 types. Some scholars being excessively analytical tried to increase the count even more.

BHĀMAHA (6^{th} century AD) for the first time put forward *anuprāsa* as a separate *alaṅkāra* in his treatise KĀVYĀLAṄKĀRA. BHĀMAHA declares the arrangement of similar letters as *anuprāsa*.[6] Moreover, he provides two types of *anuprāsa*. One is *grāmyānuprāsa*. As the name suggests, the repetition of letters without any pattern or elegance is *grāmyānuprāsa*. Learned people assume it as an ordinary repetition.[7] For example,

> *sa lolamālānīlālikulākula galo balaḥ* / 2.6, KĀVYĀLAṄKĀRA

The another type is *lāṭānuprāsa*. It is a repetition of a complete *pada* (word). But the meaning of *pada* does not change. For example,

> *dṛṣṭiṃ dṛṣṭisukhāṃ dhehi candraścandramukhoditaḥ* / 2.8, KĀVYĀLAṄKĀRA

'Lāṭa' is a name of a geographical region. Poets belonging to the region 'Lāṭa' used to employ this kind of repetition in plenty. Most of the scholars in the tradition included *lāṭānuprāsa* in the classification of *anuprāsa*. *Lāṭānuprāsa* shows similarity with *yamaka alaṅkāra*. The only difference is that, in *yamaka* the repeated word or the sequence of sounds should possess different meanings in each repetition.

After BHĀMAHA, DAṆḌIN (8^{th} century AD) added the clause of proximity to *anuprāsa*. According to DAṆḌIN the repetition of letters such that the listener remembers the previous occurrence of the repeated letter is called *anuprāsa*.[8] He added that the repetition in *anuprāsa* nourishes the *rasa*.

UDBHAṬA (9^{th} century AD) introduced *chekānuprāsa* in his work KĀVYĀLAṄKĀRA-SĀRA-SAṄGRAHA (8^{th} AD). *Chekānuprāsa* is one repetition of two groups of consonants.[9] Here, the sequence may not be the same. The example given is as follows,

> *sa devo divasān ninye tasmin śailendrakandare /*
> *gariṣṭhagoṣṭhī-prathamaḥ pramathaiḥ paryupāsitaḥ* // 3.3, KĀVYĀLAṄKĀRA-SĀRA-SAṄGRAHA

[6] sarūpavarṇavinyāsamanuprāsaṃ pracakṣate| 2.5, KĀVYĀLAṄKĀRA
[7] grāmyānuprāsamanyttu manyante sudhiyo'pare| 2.6, KĀVYĀLAṄKĀRA
[8] pūrvānubhavasaṃskāra-bodhinī yadyadūrata| 1.55, KĀVYĀDARŚA
[9] chekānuprāsastu dvayordvayoḥ susadṛśoktikṛtau| 3.2, KĀVYĀLAṄKĀRA-SĀRA-SAṄGRAHA

In addition to this, *vṛttyanuprāsa* was also defined. Here, the combination of the repeated consonants is considered. According to it, three *vṛttis* are defined viz. *paruṣa* (harsh), *upanāgarikā* (soft) and *grāmyā* (other than the prior two). Again, the count and the definitions of *vṛttis* vary from scholar to scholar.

BHOJA from 11th century conducted a vast review on this *alaṅkāra*. In addition to *vṛttyanuprāsa* and *lāṭānuprāsa* he added 4 more types viz. *śrutyanuprāsa*, *varṇānuprāsa*, *padānuprāsa* and *nāma-dvirukti*. *Varṇānuprāsa* is similar to *chekānuprāsa*. Also, *padānuprāsa* and *nāma-dvirukti* can be included into *lāṭānuprāsa*.

JAYADEVA (12th century AD) in CANDRĀLOKA introduced two types of *anuprāsa* viz. *sphuṭānuprāsa*, which is the repetition of consonants within a *pāda* or a half of a *pāda* and *arthānuprāsa*, where the repetition of consonants occur in the two words which are connected semantically. For example,

candanaṃ khalu govinda-caraṇa-dvandva-vandanam/ 5.6, CANDRĀLOKA

Here, the two words '*candanam*' and '*govind-caraṇa-dvandva-vandanam*' are connected with the 'upamāna-upameya' relation and possess the repetition of consonants '*nd*'.

The criteria for classifying *anuprāsa* into different types is basically the number of repetitions, consideration of the order of the repeated letters and what letters are being repeated. Focusing on these points different classifications are framed. Due to such finite dimensions of classification, we observe a limited number of types. Also, similar kinds of types are explained in various other classification schemes.

For the tool development, we follow one comprehensive classification. VIŚVANĀTHA'S *anuprāsa* classification provided in his treatise SĀHITYADARPAṆA has five classes that cover the extract of all the other interpretations available. Categories proposed by the rhetoricians like DAṆḌIN, UDBHAṬA, VĀMANA, RUDRAṬA, MAMMAṬA, JAYADEVA, BHOJA, etc. are covered under the umbrella of the classification of VIŚVANĀTHA.

4 Viśvanātha's classification

The tenure of VIŚVANĀTHA (14th century AD) comes in the later part of the tradition of *alaṅkāraśāstra*. He provides a well-defined and comprehensive theory for the classification of *anuprāsa* which facilitates the clarity for implementation. VIŚVANĀTHA'S classification includes other prominent classifications. Moreover, it uses widely accepted nomenclature. According to him, the *anuprāsa* is classified into 5 sub-classes. The examples for these 5 types are taken from 10th *pariccheda* of SĀHITYADARPAṆA.

1. ***Chekānuprāsa***
   Chekānuprāsa is the double occurrence of consonants with the same sequence. In each repetition, vowel endings may vary. In the example given below, one repetition of '*n-d-h*', '*v-r*' and '*p-v-n*' is in the same order. That means the order of the repeated consonants is not changed. The repetition of '*v-r*' is not changed to '*r-v*' irrespective of the changing vowels in between.

   ādāya bakulagandhānandhīkurvan pade pade bhramarān |
   ayameti mandamandaṃ kāverīvāri-pāvanaḥ pavanaḥ||

2. ***Vṛttyanuprāsa***
   Vṛtti is the mood or emotion. It is defined as the arousal of a specific mood resulting from a certain combination of letters. Repetition of one or many consonants in any order to produce a specific mood (*vṛtti*) is called *vṛttyanuprāsa*. The emotional effect differs according to the repetitive sound pattern and the combination of the letters used. This effect should complement the actual *rasa* of the poetry.

unmīlanmadhugandhalubdhamadhupavyādhūtacūtāṅkura-
krīḍatkokilakākalīkalakalairudgīrṇakarṇajvarāḥ /
nīyante pathikaiḥ kathaṃ kathamapi dhyānāvadhānakṣaṇa
prāptaprāṇasamāsamāgamaraṣollāsairamī vāsarāḥ //

In this example, the first foot has multiple repetitions of the consonant '*dh*'. In the second foot, there is repetition of the consonants '*k*' and '*l*' in any order. The third foot has the repetition of '*dh*' only once. The last foot has the repetition of '*p*', '*r*', '*s*' and '*m*' in different orders.

3. ***Śrutyanuprāsa***
Śrutyanuprāsa is the repetition of a group of consonants with a similar manner of articulation. According to DAṆḌIN it also is beneficial to the *rasa*.[10] These are further sub-classified into five classes according to the place of articulation.

(a) *Kaṇṭhya* (Velar) - *{k, kh, g, gh, ṅ, h}*
(b) *Tālavya* (Palatal) - *{c, ch, j, jh, ñ, y}*
(c) *Mūrdhanya* (Retroflex) - *{ṭ, ṭh, ḍ, ḍh, ṇ, r, ṣ}*
(d) *Dantya* (Dental) - *{t, th, d, dh, n, l, s, v}*
(e) *Oṣṭhya* (Labiel) - *{p, ph, b, bh, m, v}*

For example, the following verse

dṛśā dagdhaṃ manasijaṃ jīvayanti dṛśaiva yāḥ /
virūpākṣasya jayinīstāḥ stumo vāmalocanāḥ //

has a repetition of Palatal varṇas '*j*' and '*y*'.

4. ***Antyānuprāsa***
Repetition of syllables at the end of the *padas* (words) or at the end of the foot. Specifically, after the penultimate vowel that is the last but one vowel of the *pāda* or *pada*. the For example,

keśaḥ kāśastabakavikāsaḥ kāyaḥ prakaṭitakarabhavilāsaḥ /
cakṣurdagdhavarāṭakakalpaṃ tyajati na cetaḥ kāmamanalpam //

In this example, the ends of the first and second feet match, similarly the ends of the third and fourth feet.

5. ***Lāṭānuprāsa***
Lāṭānuprāsa is not just the repetition of the consonants but the repetition of a word (*pada*) with similar meaning but different implications. This kind of *anuprāsa* is similar to *yamaka* as repetitions are considered for longer syllable sequences. Following is an example of *lāṭānuprāsa*.

smerarājīvanayane nayane kiṃ nimīlite /
paśya nirjitakandarpaṃ kandarpavaśagaṃ priyam //

In the above given example, the words '*nayane*' (meaning eyes) and '*kandarpam*' (meaning desire for love) are repeated with the same sense. The words are sometimes an independent word or a part of a compound word. According to the role of the word in the sentence, the implication changes. The first appearance of *nayane* in *smerarājīva*nayane (meaning - a woman with lotus like eyes) is in the form of an element of the compound, and contributes its meanings to form a meaningful compound. The other occurrence of '*nayane*' is an independent word to give the meaning as 'two eyes'. Similarly in '*kandarpam*', both repetitions possess the same meaning but the implication differs in each occurrence.

[10] *yayākayācicchrutyāyatsamānamanubhūyate/*
tadrūpāṃhipadāsattiḥsānuprāsārasāvahā// 1.52, KĀVYĀDARŚA

Type	Count	Order	Unit	Position in the input
lāṭa	≥ 2	same	word	next to each other or with a few interventions.
cheka	2	same	sequence of syllables without vowels	anywhere within a proximity of $8 + 2*$ length of syllable sequence.
vṛtti	a)> 2 b)≥ 2	any	a) sequence of syllables without vowels b) a consonant	anywhere within a proximity of $8 + 2*$ length of syllable sequence.
śruti	≥ 2	any	syllables from the same class	within a proximity of 8 syllables.
antya	≥ 2	same	syllables after the second last vowel.	end of feet and words.

Table 1: Differentiation from the implementation point of view

5 Implementation of Anuprāsa Identifier and Classifier

From the definitions of various *anuprāsa*s, we note that some types have more stringent conditions than others. Hence, the examples that satisfy more stringent conditions may also satisfy less stringent conditions and thus can be categorised under two different types of *anuprāsa*. For example, *lāṭānuprāsa* demands that the repetition is of words and not syllables. *Chekānuprāsa* demands the repetition of syllables in the same order. Thus, any example of *lāṭānuprāsa* is also potentially an example of *chekānuprāsa* as well. However, due to the stringent conditions of *lāṭānuprāsa*, it is appropriate to classify such an instance only under *lāṭānuprāsa*. A similar situation exists with other pairs as well. In order to decide the proper exclusive sequence in which these *anuprāsa*s should be identified, we look at the necessary conditions for each of them.

From the table 1, we understand that the natural order for identifying the *anuprāsa* type is *lāṭa, cheka* and *vṛtti*. The conditions of *śruti* and *antya* type of *anuprāsa* do not clash with any other classes and hence can be identified either in the beginning or at the end.

We use the frequencies of n-grams[11] for identifying *lāṭānuprasa* and frequencies of n-grams of the sequence of letters ignoring the vowels and their positions identifying the *chekānuprāsa* and the frequencies of n-grams of consonants for identifying *vṛttyanuprāsa* and only consonants having the same place of articulation and their positions in the input for identifying *śrutyanuprāsa*. The *pādāntyānuprāsa* is identified by looking at rhymes at the end of the *pādas*. For *padāntyānuprāsa*, we consider the rhyming in the space-separated word endings.

Unicode Devanagari is unsuitable for processing and identifying the n-gram and consonant frequencies since the basic units in Unicode are a mix of consonants with a vowel *'a'* inherent in them. Hence we convert the input internally into WX notation[12] and process it. In addition to Devanagari and IAST schemes, we also accept input in various other transliteration schemes such as Velthuis, SLP, Kyoto Harward, WX notation, etc.

Normalization of the input is an important step in processing. To analyse *antyānuprāsa*, the *daṇḍas* ('|') and spaces to mark the word and the foot boundary are preserved. For other types, the normalization of various elements is defined below.

- Spaces :
  In the oral tradition, the spaces between the words do not carry any significance. *Anuprāsa* deals with the sound patterns, and as such, we ignore the spaces between the words.

- *Anunāsikyas* :
  Since the *anuprāsa* is identified based on sound patterns, the variations in spelling need to be taken care of. Sanskrit allows some spelling variations concerning nasalization. All the homogenous nasal stops are converted into *anusvāras*. The *anusvāra* when followed by consonants, can be converted into homogenous nasal stop viz. *ṅ, ñ, ṇ, n* and *m*.

[11] The sequences of letters of length 'n' are called n-grams.
[12] `https://en.wikipedia.org/wiki/WX_notation`

For example, *'aṃbuja'* versus *'ambuja'*, *'aṃka'* versus *'aṅka'*. Similarly, the nasal stop *'m'* at the end of a word is written as an anusvāra when it is followed by a word starting with a consonant. We normalize all the nasal stops to *anusvāra*.

- Special characters:
  (1) A special character that needs special attention is the *avagraha*. The *avagraha* is a writing convention to indicate the elided *'a'* during the *sandhi* operation. Since for the purpose of *anuprāsa* identification, we look at the *sandhi*ed text only, we ignore the *avagraha* if it is present in the input text.
  (2)Similarly, the *daṇḍa* ('|') used to denote the sentence-end, or in the case of a verse, to denote the end of two *pādas*. Except for *pādāntyānuprāsa*, the *daṇḍa* is also ignored.

The broad algorithm is as follows.

- Read the sequence of letters.
- Convert it to WX notation.
- Check for *pādāntyānuprāsa* by dividing the input into 4 equal parts and comparing the sequence of letters after the second last vowel at the end of each part.
- Check for *padāntyanuprāsa* by comparing the word endings from the penultimate vowel to the end of the consecutive words. If the sequence of letters matches at least in two words mark the repeated sequence as *padāntyānuprāsa*.
- Get the n-grams ($n \geq 2$) along with their positions with and without vowels.
- Remove all the small n-grams that can be subsumed by the large n-grams with matching positions (index).
- If the frequency of n-gram with vowels is more than 1, mark it as *lāṭānuprāsa*.
- Else if the frequency of n-grams without vowels is 2, mark such sequences as *chekānuprāsa*.
- Else if the frequency of n-grams ($n \geq 2$) without vowels is greater than 2 and if the frequency of single consonants is greater than or equal to 2 the n-gram or the consonants are marked as *vṛttyanuprāsa*.
- If the frequency of consonants belonging to the same class is greater than 2, mark them as *śrutyanuprāsa* of the type to which these consonants belong.

The use of else if ensures that the classification prefers a type with a more stringent definition than the others.

As a general rule in *anuprāsa*, the repetitions should not be far away to make the reader forget the previous occurrence. If the distance is large, the instance will not be able to produce amusement for the reader. To strike out such cases we have added one function in which the distance is calculated through the indices of the repeated consonants. For a single letter repetition, the maximum distance is considered to be 4 to 5 *akṣara*s, that is 8 to 12 letters approximately, considering the frequent conjuncts in Sanskrit.

6 Interface

We have designed a user-friendly interface to access this tool. This is an integrated tool for both *yamaka* and *anuprāsa* (see figure 1). User can provide their input text in various available encodings. Figure 1 shows all five types of *anuprāsa* highlighted in red colour corresponding to the input given by the user. The highlighted sequence facilitates the user with a better comprehension of the *alaṅkāra* and helps the user understand the difference between each type of *anuprāsa* effectively and easily. The interface is available in the 'tools' section at `https://sanskrit.uohyd.ac.in/scl/`

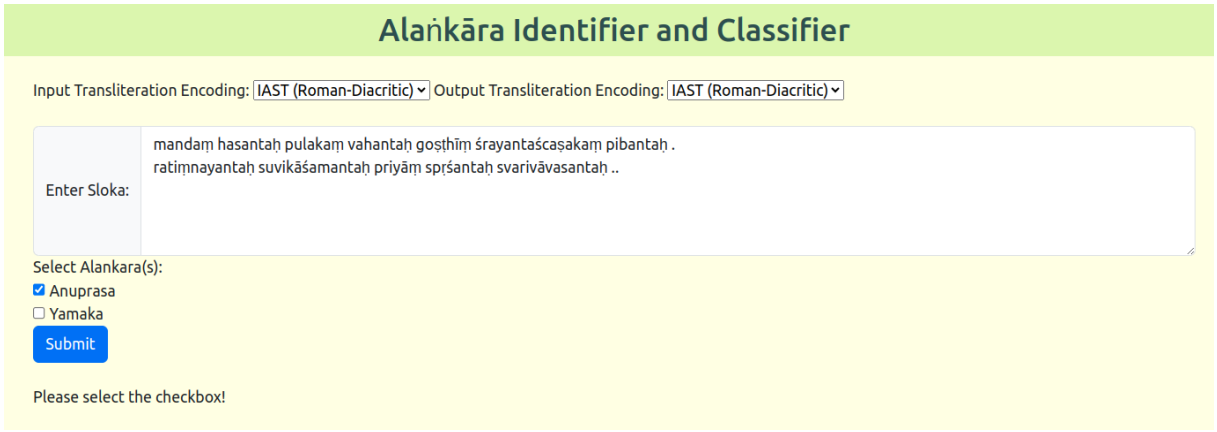

Figure 1: Alaṅkāra Identifier and Classifier: *Anuprāsa* input

Figure 2: Alaṅkāra Identifier and Classifier: *Anuprāsa* output

7 Evaluation

We tested our tool on a data set of 70 *ślokas* and 10 sample prose. The selected *ślokas* are primarily given as examples of *anuprāsa* in the *śāstr*ic texts and some from the RAGHUVAṂŚA of KĀLIDĀSA. The prose examples were passages consisting of 2 to 3 sentences from BĀṆABHAṬṬA'S KĀDAMBARĪ. Most of the examples contained more than one type of *anuprāsa*. This tool could successfully handle these *anuprāsa* instances. Since the three classes of *anuprāsa* viz. *lāṭānuprāsa*, *chekānuprāsa*, and *vṛttyanuprāsa* relax the conditions as we go from the first one to the third, the latter is a strict superset of the previous one. Hence, a pattern satisfying the conditions of *lāṭānuprāsa*, though it is an example of *chekānuprāsa* and *vṛttyanuprāsa*, is shown only under the *lāṭānuprāsa*. Similarly, the patterns satisfying the conditions for *chekānuprāsa* are not again displayed under *vṛttyanuprāsa*. Similarly, only those patterns that are not covered under *lāṭānuprāsa* and *chekānuprāsa*, are considered for *vṛttyanuprāsa*.

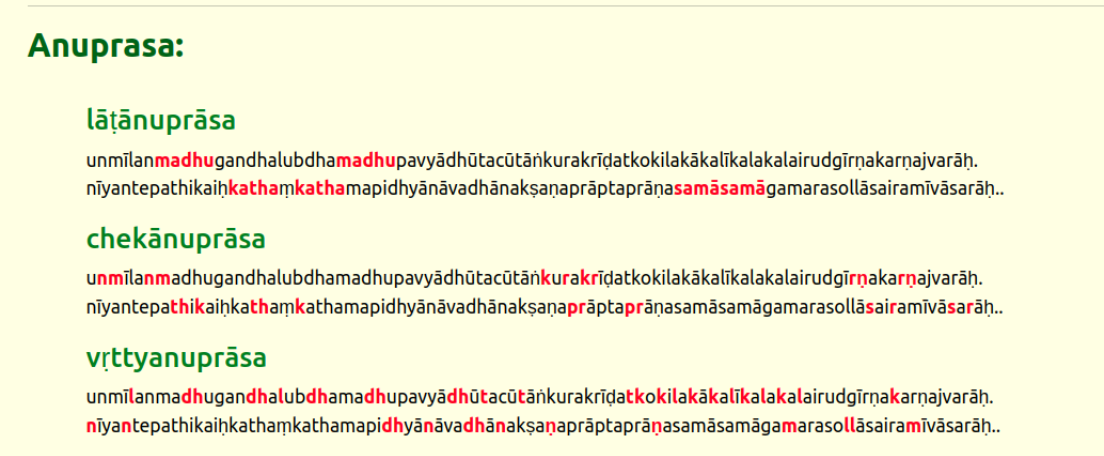

Figure 3: The cascade effect in *lāṭānuprāsa*, *chekānuprāsa* and *vṛttyanuprāsa*

Since this tool is not supported with word segmenter or meter identifier, for analysis of *antyānuprāsa* it completely relies on the spaces and the *daṇḍas* to mark the *pada* and *pāda* boundary. If the user has not provided the *daṇḍa* or spaces in the appropriate place, the tool is not able to identify *antyānuprāsa*.

8 Conclusion

We have discussed a tool useful for the identification and classification of Sanskrit poetry focusing on *anuprāsa*. This can be taken as a booster for figurative language processing in Sanskrit and other Indian languages as well. The concept of *anuprāsa* along with its classification is adopted by other Indian languages like Hindi, Marathi, Telugu, Kannada, etc. The same model with the necessary amendments can be deployed for the identification and classification of *anuprāsa* in modern Indian languages as well.

This module is extendable for other classifications presented in the tradition. The *vṛttyanuprāsa*, a type of *anuprāsa* can be researched extensively to identify *rasa* depending upon the repetition of clusters of consonants.

'Anuprāsa Identifier and Classifier' is useful for teaching this figure of sound by presenting a demonstration of different examples. While creating a masterpiece of poetry, a good poet does not deliberately enforce the figures in the poetry. Such upcoming masterpieces in Sanskrit can be tested with this tool.

References

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

Prateek Agrawal and Vishu Madaan. 2020. A Sanskrit to Hindi language machine translator using rule based approach. In *Proceedings of the 17th International Conference on Natural Language Processing (ICON): System Demonstrations*, pages 13–15, Patna, India, December. NLP Association of India (NLPAI).

Milind Kumar Audichya and Jatinderkumar R. Saini. 2021. Towards Natural Language Processing with figures of speech in Hindi poetry. In *International Journal of Advanced Computer Science and Applications*.

V. Balasubramanyam. 2017. *Citram - Poetry of Sound*, volume 1. Rashtriya Sanskrit Sansktan, New Delhi.

Amruta Barbadikar and Amba Kulkarni. 2023. Yamaka identifier and classifier: A computational tool for the analysis of Sanskrit figure of sound (upcoming).

Satyadev Chowdhary. 1965. *Kāvyālaṅkāra*. Vasudev Prakashan, Delhi.

Durgaprasad and Kashinath Parab. 1886. *Kāvyālaṅkāraḥ*. Nirnayasagar Press, Mumbai.

Benjamin Englard. 2013. A rhetorical analysis approach to Natural Language Processing. In *ArXiv*.

Edwin Gerow. 1971. *A Glossary of Indian Figures of Speech*. Mouton & Co. N. V. Publishers, Hague.

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

Pawan Goyal and Gérard Huet. 2013. Completeness analysis of a sanskrit reader. In *Proceedings, 5th International Symposium on Sanskrit Computational Linguistics. DK Printworld (P) Ltd.*

Pawan Goyal, Vipul Arora, and Laxmidhar Behera. 2009. Analysis of sanskrit text: Parsing and semantic relations. In Gérard Huet, Amba Kulkarni, and Peter Scharf, editors, *Sanskrit Computational Linguistics*, pages 200–218, Berlin, Heidelberg. Springer Berlin Heidelberg.

Mari Hattori. 1997. On the rhyme (yamaka) in Sanskrit poetics. *Annals of the Bhandarkar Oriental Research Institute*, 78(1/4):263–274.

Amba Kulkarni and Monali Das. 2012. Discourse analysis of Sanskrit texts. In *Proceedings of the Workshop on Advances in Discourse Analysis and its Computational Aspects*, pages 1–16, Mumbai, India, December. The COLING 2012 Organizing Committee.

Amba Kulkarni and Anil Kumar. 2011. Statistical constituency parser for Sanskrit compounds. In *ICON 2011*.

Amba Kulkarni and Anil Kumar. 2013. Clues from Aṣṭādhyāyī for compound type identification. In *5th international SCLS 2013*.

Amba Kulkarni and Madhusoodana Pai. 2019. Sanskrit sentence generator. In *Proceedings of the 6th International Sanskrit Computational Linguistics Symposium*, pages 1–13, IIT Kharagpur, India, October. Association for Computational Linguistics.

Amba Kulkarni and Devanand Shukl. 2009. Sanskrit morphological analyser: Some issues. In *the Festscrift volume of Bh. Krishnamoorty, Indian Linguistics*.

Amba Kulkarni, Pavankumar Satuluri, Sanjeev Panchal, Malay Maity, and Amruta Malvade. 2020. Dependency relations for Sanskrit parsing and treebank. In *Proceedings of the 19th International Workshop on Treebanks and Linguistic Theories*, pages 135–150, Düsseldorf, Germany, October. Association for Computational Linguistics.

Keshav Melnad, Peter Scharf, and Pawan Goyal. 2015. Meter identification of Sanskrit verse. In *Sanskrit Syntax: Selected Papers Presented at the Seminar on Sanskrit Syntax and Discourse Structures*.

Shriramachandra Mishra. 1996. *Kāvyādarśa*. Chowkhamba Vidyabhavan, Varanasi.

Komal Naaz and Niraj Kumar Singh. 2022. Design and development of computational tools for analyzing elements of Hindi poetry. In *IEEE Access*.

Tyler Neil. 2023. Skrutable: Another step toward effective Sanskrit meter identification. In *Proceedings of the Computational Sanskrit & Digital Humanities: Selected papers presented at the 18th World Sanskrit Conference*.

S. Rajagopalan. 2018. A user-friendly tool for metrical analysis of Sanskrit verse. In *Computational Sanskrit & Digital Humanities, Selected papers presented at the 17th World Sanskrit Conference*.

C. Shankara Rama Sastri. 1956. *Kāvyālaṅkāra of Bhāmaha*. The Sri Balamanorama Press, Mylapore, Madras.

Pavankumar Satuluri and Amba Kulkarni. 2013. Generation of sanskrit compounds. In *ICON 2013*.

Pandit Rangacharya Raddi Shastri. 1938. *Kāvyādarśa*. Bhandarkar Oriental Research Institute, Pune.

Shrikrishnamoori. 1909. *Kāvyālaṅkārasūtravṛttiḥ*. Sri Vani Vilas Press, Srirangam.

Ekaterina V. Shutova. 2011. *Computational approaches to figurative language*.

Krishnan Sriram, Amba Kulkarni, and Gérard Huet. 2023. Validation and normalization of DCS corpus and development of the Sanskrit heritage engine's segmenter. In *Proceedings of the Computational Sanskrit & Digital Humanities: Selected papers presented at the 18th World Sanskrit Conference*, pages 38–58, Canberra, Australia (Online mode), January. Association for Computational Linguistics.

Renate Söhnen. 1995. On the concept and presentation of "yamaka" in early indian poetic theory. *Bulletin of the School of Oriental and African Studies*, 58(3):495–520.

Hrishikesh Terdalkar and Arnab Bhattacharya. 2023. Chandojnanam: A Sanskrit meter identification and utilization system. In *Proceedings of the Computational Sanskrit & Digital Humanities: Selected papers presented at the 18th World Sanskrit Conference*.

Rudradev Tripathi. 1972. *Sanskṛta Sāhitya me śabdālaṅkāra*. Shri Lal Bahadur Shastri Kendriya Sanskrit Vidyapeeth, Delhi.

Vrajaratnadas. *Kāvyādarśa*. Vrajaratnadas, Shrikamalamani Granthamaka Karyalaya, Kashi.