YAMAKA IDENTIFIER AND CLASSIFIER: A COMPUTATIONAL TOOL FOR THE ANALYSIS OF SANSKRIT FIGURE OF SOUND

By

Amruta Barbadikar and Amba Kulkarni

The science dealing with various aspects of the beauty of a literary text is termed as Alaṅkāraśāstra in Sanskrit. The literal meaning of 'Alaṅkāra' is an ornament or a decoration. In the context of literature, these ornaments are either related to the sounds of the phonemes in the words or to the meanings associated with the words or to both. The beautiful rhythmic patterns of sound are termed as Śabdālaṅkāras (ornaments of sound) and the patterns that give rise to pleasure due to the meanings of the words are called Arthālaṅkāras. There are constructions with rhythmic sound patterns that are also meaningful. These fall under the category of Ubhayālaṅkāras. Śabdālaṅkāras being governed by the rhythmic, repetitive sound patterns, are very easy to capture and can be identified with the pattern matching algorithms.

Repeated use of the same sequence of letters is one of the techniques that poets use for enhancing the beauty and charm of poetry. Anuprāsa (alliteration) and Yamaka are two popular śabdālaṅkāras. Both use repetitive patterns. Among these two, in alliteration typically a single letter or a syllable is repeated. For example, in the following phrase, the syllable 'ra' is repeated and so, it is an example of Anuprāsa.

'rāmo rājamaṇiḥ sadā vijayate
rāmaṁ rameśaṁ bhaje |'

In a variation of Anuprāsa called Lāṭānuprāsa, one or more nominal stems are repeated. The nominal stem may occur with different case markers.

Yamaka also shows the repetition of sound patterns. But these patterns are larger sequences of syllables, typically consisting of at least two syllables. For example, "samarasamaraso 'yam" Here, the sequence 'samara' is repeated. The first occurrence is the left part of a compound and the second occurrence is a part of the word 'samarasaḥ.'

The word 'Yamaka' is derived from the nominal stem 'yama' which means twin or coupled. Therefore, the literal sense of Yamaka is repetition. In the construction of a Yamaka, a group of phonemes is used more than once in a stanza to bring in a special rhythm. The meaning of these groups of phonemes at different occurrences may be the same or may be different. The effect it creates due to the sound patterns followed by the meaning, leads to joy in the minds of the reader. The construction of Yamaka and understanding of the poems with Yamaka leads to a delightful experience both for the composer as well as the listener.

Notable theoretical work regarding poetics, especially Yamaka, can be traced in various studies done so far. Söhnen [1] examined various definitions and classifications of Yamaka in five different treatises viz. Nāṭyaśāstra, Bhaṭṭikāvya, Kāvyālaṅkāra, Kāvyādarśa and Agnipurāṇa. The aim of this study was to find the influence of the definitions and classifications by the earlier scholars on the latter. The influence is presented in the form of a tree where the work at the child node inherits the properties from the work at the parent node. A Similar attempt has also been made by Hattori [2] where he compared the three contemporary texts Bhaṭṭikāvya, Kāvyālaṅkāra and Kāvyādarśa. He mainly looked at the various patterns of repetition accepted by the three texts on rhetoric.

Automatic processing of rhetorical language is assumed to be one of the toughest tasks in the area of Natural Language Processing as it is heavily dependent on the word meaning. Although, several works are found on the same. Shutova [3] presented a computational approach to process metaphor using statistical methods. Englard [4] used the rhetorical analysis of text to predict the author. For Hindi, Audichya and Saini [5], worked out the alaṅkāras in Hindi to present the hierarchical structure with a taxonomical listing of alaṅkāras. However, the computational implementation was not exercised. Naaz and Singh [6] was 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 in detail, from various applications perspective, by different scholars. Melnad et al [7], Rajagopalan [8], Neil [9], Terdalkar and Bhattacharya [10] are some of the notable contributors to the available state-of-the-art Sanskrit Meter Identification systems.

Despite rich rhetorical tradition, we do not find any computational

tool which analyzes Sanskrit from its aesthetic, semantic and figurative point of view. Except for Meter analysis, all the other aspects are untouched. This research presents a first-of-its-kind computational tool based on the Indian rhetorical tradition.

2. Yamaka

The use of various alaṅkāras can be traced in the Sanskrit literature as early as Vedas and Purāṇas. Agnipurāṇa and *Viṣṇudharmottarapurāṇa*[1] contains description of a śabdālaṅkāra called Yamaka in brief. The Sanskrit rock edicts of Aihole by the king Pulakeśin II (7 th c. CE) have many instances of Yamaka. Poets like Kālidāsa, Bhāravi, Māgha and Daṇḍin deployed Yamaka in their poetry.

The concept of Yamaka has evolved over a period of time ranging from Bharata (1st c. CE) to Mammaṭa (11th c. CE). We present below a sketch of definitions and classification schemes by various scholars.

Before proceeding a short note on the terminology is in order. A verse or stanza consists of four lines called foot or pāda (We will use these terms alternately). In what follows we consider only the repetitions within a verse. The Yamaka due to consecutive similar verses being the same is not considered here.

2.1 Bharata

For the first time, Yamaka as an alaṅkāra was presented by Bharata (1st c. CE), the pioneer of Sanskrit poetics, in Nāṭyaśāstra. He has described four alaṅkāras,[2] viz. Yamaka, Rūpaka, Dīpaka and Upamā out of which only Yamaka is a śabdālaṅkāra. According to Bharata, repetition of a word or a syllable is a characteristic of the figure of sound Yamaka. Bharata classified Yamaka into ten sub classes determined by the number of repetitions and the place of repetition in a line of a stanza. The names and patterns of these ten types are given below.

1. *Pādānta* - Once at the end of each foot (pāda). For example,

*dinakṣayāt saṃhṛtaraśmi**maṇḍalaṃ***
*divīva lagnaṃ tapanīya**maṇḍalam*|
*vibhāti tāmraṃ divi sūrya**maṇḍalaṃ***
*yathā taruṇyāḥ stanabhāra**maṇḍalam*||

2. *Kāñcī* - At the end and the beginning of the same foot. For

---

[1] *ādau madhye tathaivānte pādasya tu tadiṣyate*|
sandaṣṭakasamudgākhyau tathaiva yamakau matau ‖ 14.28*, Viṣṇudharmottarapurāṇa.*

[2] *upamā rūpakaṃ caiva dīpakaṃ yamakaṃ tathā* |
alaṅkārāstu vijñeyā catvāro nāṭakāśrayāḥ ‖ *16.41, Nāṭyaśāstra.*

example,

*yāmaṃ yāmaṃcan**dravatīnāṃ dravatīnāṃ***
*vyaktāvyaktā sā**rajanīnāṃ rajanīnām*|
*phulle phulle sam**bhramare vābhramare vā***
*rāmā rāmā vi**smayate ca smayate ca*‖

3. *Samudgaka* - The first half is repeated in the second half (the half may refer to a stanza or a foot). For example,

ketakīkusumapāṇḍuradantaḥ
śobhate pravarakānanahastī|
ketakīkusumapāṇḍuradantaḥ
śobhate pravarakānanahastī‖

4. *Vikrānta* - Alternate feet are the same. For example,

sa pūrvaṃ vāraṇo bhūtvā
dviśruṅga iva parvataḥ|
abhavaddantavaikalyā-
dviśruṅga iva parvataḥ‖

5. *Cakravāla* - The pattern at the end of the preceding foot is repeated at the beginning of the next foot. Alternatively, if two adjacent feet are the same then also it is called Cakravāla. For example,

*śailā yathā śatrubhirāhatā **hatā***
hatā**śca bhūyastvanupuṅkhapuṅ**khagaiḥ|
khagai**śca sarvairyudhi sañcitā**ścitā-
***ścitā**dhirūḍhā nihitāstalaistalaiḥ*‖

6. *Sandaṣṭaka* - The repetition is at the beginning of foot. For example,

***paśya paśya** ramaṇasya me guṇān*
***yena yena** vaśagāṃ karoti mām*|
***yena yena** hi sameti darśanaṃ*
***tena tena** vaśagāṃ karoti mām*‖

7. *Pādādi* - The pattern is repeated once at the beginning of any two feet. For example,

***viṣṇuḥ** sṛjati bhūtāni*
***viṣṇuḥ** saṃharate prajāḥ*|
***viṣṇuḥ** prasūte trailokyaṃ*
***viṣṇu**rlokādhidaivatam*‖

8. *Pādāntāmreḍita* - The repetition is at the end of any foot. For example,

*vijṛmbhitaṃ niḥśvasitaṃ **muhurmuhuḥ***

kathaṃ vidheyaṃ smaraṇaṃ ***pade pade*** |
yathā ca te dhyānamidaṃ ***punaḥ puna-***
rdhruvaṃgatā te rajanī ***vinā vinā*** ||

9. *Caturvyavasita* - All four feet of a stanza are the same. For example,

vāraṇānāmayameva kālo
vāraṇānāmayameva kālaḥ |
vāraṇānāmayameva kālo
vāraṇānāmayameva kālaḥ ||

10. *Mālā* - Repetition of syllables like a garland. (This type is further elaborated by the later rhetoricians as Anuprāsa alaṅkāra.) For example,

asau hi ***rā****mā* ***ra****tivig****ra****hap****ri****yā*
ra*haḥp****ra****galbhā* ***ra****maṇaṃ manogatam* |
ra*tena* ***rā****t****ri****ṃ* ***ra****mayet pa****re****ṇa vā*
*na cedudeṣyatta****ru****ṇaḥ pa****ro*** ***ri****puḥ* ||

2.2 Bhāmaha

Bhāmaha brought a valuable reform in the definition of Yamaka. While Bharata looked at only the sound patterns, Bhāmaha insisted on the patterns that are similar in sound but different in meaning.[3] He divided the foot of a stanza into three parts - ādi (beginning), madhya (middle) and anta (end) and classified Yamaka into five types as follows.

1. Ādi - at the beginning of a foot
2. Madhyānta - at the middle-end of a foot
3. Pādābhyāsa - Repetition of a foot in a stanza
4. Āvalī - a series of repetitions without considering the position in the foot
5. Samasta-pāda - repetition of a sequence in all feet

Bhāmaha believed that the types viz. sandaṣṭaka (repetition of alternate feet), samudgaka (repetition of adjoining feet), etc. of Bharata are nothing but the consequences of this fivefold classification. He directed that the repetition can occur either in the adjoining feet or the alternate feet or in all four feet.

2.3 Daṇḍin

Daṇḍin furnished Yamaka with new insights. While defining the Yamaka, he referred only to the sound patterns. He looked at the repetitions from three different angles.

---

[3] *tulyaśrutīnāṃ bhinnānāmabhidheyaiḥ parasparam* |
varṇānāṃ yaḥ punarvādo yamakaṃ tannigadyate || 2.17, *Bhāmaha-Kāvyālaṅkara*.

Positions of the repeated patterns - within a pāda and across the pādas. The patterns may repeat either in the same pāda or across the pādas, as described below. The number in the parenthesis indicates the possible patterns of that category in a verse.

Repetition in the same pāda. (4)

Pattern repeated in two different pādas. (6)

Pattern repeated in three different pādas. (4)

Pattern repeated in all four pādas. (1)

This results in 15(4+6+4+1) patterns.

The second dimension that Daṇḍin brought in, refers to the intervention of other syllables in the repetition. This is more concerned with the repetition in the same pāda, rather than across the pādas. If there is an intervention of other syllables between the two similar patterns, the pattern is labeled as 'Vyapeta', and otherwise it is called 'Avyapeta'. An example of Vyapeta Yamaka,

kamaleḥ samakeśānte
  kamalerṣyākaraṃ mukham |
kamalekhyaṃ karoṣi tvaṃ
  kamalevonmadiṣṇuṣu ||

And that of Avypeta Yamaka is,

ramaṇī ramaṇīyā me
  pāṭalā pāṭalāṃśukā |
vāruṇīvāruṇībhūta-
  saurabhā saurabhāspadam ||

Of course, one can have both intervened as well as unintervened patterns in the same pāda and such an instance is called 'avyapeta-vyapeta'. For example,

sālaṃ sālambakalikā
  sālaṃ sālaṃ na vīkṣitum |
nālī nālīnavakulā
  nālī nālīkinīrapi ||

Thus, there are three possible variations due to the presence and/or the absence of intervention.

A pāda is considered to be made up of three parts viz. beginning (ādi), middle (madhya) and end (antya). Since there are three parts, patterns may repeat at least twice and a maximum of thrice in the same pāda. Accordingly, there are seven patterns viz. -

Beginning (Ādi) - both the patterns are in the beginning.

Middle (Madhya) - the repetition is in the middle.

End (Antya) - both the patterns are at the end.

Beginning-middle (Ādimadhya) - once at the beginning and once in the middle.

Middle-end (Madhyānta) - once in the middle and once at the end.

Beginning-end (Ādyanta) - once at the beginning and once at the end.

Beginning-middle-end (Ādimadhyānta) - The pattern is found thrice in a pāda viz. beginning, middle and the end part.

This results into 15×3×7=315 possible patterns of Yamaka.

In addition to the Yamaka within a verse, Daṇḍin also provided examples of Yamaka across the verses. When two verses have the same sound patterns but they having different meanings then the Yamaka is termed 'Ślokābhyāsa Yamaka', and if only a half of the verse is the same, then it is termed 'Śloka-ardha-abhyāsa Yamaka'. Also, when there are two pādas with the same sound pattern but have different meanings, this pattern is called 'Pādābhyāsa Yamaka'. Another variety of Yamaka discussed by Daṇḍin is called Pratiloma Yamaka. In this variety, the repetition is in reverse order. The reversal may be within a pāda or between two pādas or it may cover a complete stanza. Here is an example of Pratiloma Yamaka covering a pāda.

yāmatāśa kṛtāyāsā sā yātā kṛśatā mayā |
ramaṇārakatā testu stutetākaraṇāmara ||

We observe that the part
yā | ma | tā | śa | kṛ | tā | yā | sā
is repeated in the reverse order in the first line and the part
ra | ma | ṇā | ra | ka | tā | te | stu
is repeated in the reverse order in the second line.

2.4 Vāmana

Vāmana underlined the 'different meaning' factor in the definition of Yamaka which was not explicitly mentioned in Kāvyādarśa of Daṇḍin. Also in the definition, he used the term 'Sthānaniyama' which emphasizes that the words need to be repeated in some well-defined places only.[4] Thus, he insisted, that there cannot be randomness in the repetition of patterns. Rather the patterns are to be in specific positions in different pādas.

2.5 Rudraṭa

Rudraṭa has done significant work on Yamaka on the same grounds as Daṇḍin. In Kāvyālaṅkāra, Rudraṭa elaborately explained Yamaka and

---

[4] *padamanekārthamakṣaraṃ cāvṛttaṃ sthānaniyame yamakam | pādāḥ pādasyaikasyānekasya ca ādimadhyāntabhāgāḥ sthānāni | 4.1-2, Kāvyālaṅkārasūtravṛtti*

its types. The Pratiloma Yamaka of Daṇḍin is not accepted by Rudraṭa, since for him the order of syllables is also important.[5]

Rudraṭa's main contribution is towards providing a multilevel classification with a systematic naming scheme. At the top level, he classified Yamaka into three categories depending on whether a complete verse is repeated or the repetition is only within a half-verse or the repetition occurs within one quarter of a verse. In the case of complete verse, there is no further sub-division. The other two patterns have further sub classifications depending on the position of the repeated patterns. The repetitions can be either across the pādas or within a pāda.

Samasta-pādāvṛtti Yamaka (Repetition across pādas)

In the case of repetitions across the pādas, either the pattern is repeated in two or all the four pādas or two different patterns are repeated in pairs of pādas resulting in 10 different possible combinations which are shown in [table 1] table. Here, the 'pi' denotes ith pāda.

Type	Pattern	Example
Paṅkti	$p_1 = p_2 = p_3 = p_4$	***sabhājanenopari pūritāsau sabhājane noparipūritāsau*** | ***sabhā janeno'paripūritāsau sabhājane no'paripūritāsau*** ‖
Samudgaka	$p_1 = p_3$, $p_2 = p_4$	***nanāma loko vidamānavena*** *mahī na cāritramudāradhīram* | ***na nāmalo' kovidamānavena*** *mahīnacāritramudāradhīram* ‖
Parivṛtti	$p_1 = p_4$, $p_2 = p_3$	***mudā ratāsau ramaṇī yatā yāṃ*** *smarasyado'laṃ kurutena vodhā* | *smarasyado'lamkurute'navodhā*-***mudāratāsau ramaṇīyatāyām*** ‖
Yugmaka	$p_1 = p_2$, $p_3 = p_4$	*vināyameno nayatā'sukhādinā vinā yamenonayatā sukhādinā* | ***mahājano'dīyata mānasādaraṃ mahājanodī yatamānasādaraṃ*** ‖
Mukha	$p_1 = p_2$	***cakraṃ dahatāraṃ cakranda hatāram*** | *khaḍgena tavājau rājannarinārī* ‖
Puccha	$p_3 = p_4$	*uttuṅgamātaṅgakulākule yo vyajeṣṭa* *śatrūnsamare sadaiva* |

[5] *kramagrahaṇāt pratilomānulomasarvatobhadrānuprāsādināṃ yamakatvanirāsaḥ* | *nahi teṣu tulyaśrutisadbhāve'pi tulyakramo vidyate* | Commentary of Namisādhu on 3.1, *Kāvyālaṅkāra.*

		sa sāramānīya mahāri cakraṃ sasāra mānī yamahāricakram ‖
Garbha	$p_2 = p_3$	*yo rājyamāsāya bhavatyacintaḥ* ***samudratārambharataḥ sadaiva*** | ***samudratārambharataḥ sa daiva****pramāṇamārabhya payasyudāste* ‖
sandaṃśa	$p_1 = p_3$	***sannārībharaṇomāya-****mārādhya vidhuśekharam* | ***sannārībharaṇo'māya-****statastvaṃ pṛthivīṃ jaya* ‖
Sandaṣṭaka	$p_2 = p_4$	*idaṃ ca yena svayamātmabhogyatāṃ* ***samastakāñcīkamanīyatākulam*** | *nitambabimbaṃ kathamastu no nṛṇāṃ* ***sa mastakāñcī kamanīyatākulam*** ‖
Avṛtti	$p_1 = p_4$	***mudāratāḍī samarājirājitaḥ*** *pravṛddhatejāḥ prathamo dhanuṣmatām* | *bhavānbibhartīha nagadha medinī-* ***mudāratāḍīsamarājirājitaḥ*** ‖

Table 1: *Samasta-pada-vṛtti Yamaka*

2. Ekadeśaja[6] Yamaka (Repetition within a pāda)

In the case of repetition within a pāda, he divides a quarter into two and three parts each and then considers the repetitions within and across each half or 1/3rd. And finally, he divides the half part and 1/3rd part into two halves resulting into 1/4th and 1/6th of a pāda and names the repeating patterns.

1.a Pādārdhāvṛtti Yamaka (Repetition in the half pāda)
(The number in parenthesis indicates the number of possible Yamakas with that pattern.)
- a.i Pādādi-paṅkti-ādi (10)
- a.ii Pādānta-paṅkti-ādi (10)
- a.iii Ādyantika (6)
- a.iv Antādika (6)
- a.v Pādasamudgaka (15)
- a.vi Ardhaparivṛtti (1)
- a.vii Anya-bheda - Pādacaturthārdhāvṛtti Yamaka (Repetition in the quarter of a pāda)
  - vii.1 Vaktra
  - vii.2 Śikhā
  - vii.3 Mālā

[6] This is also known as '*pādabhāgāvṛtti*'.

vii.4 Madhya
vii.5 Ādyanta
vii.6 Kāñcī

1.b Pāda-tṛtīyārdhāvṛtti Yamaka (Repetition 1/3rdof a pāda)
b.i Pādādi-paṅkti-ādi (10)
b.ii Pādamadhya-paṅkti-ādi (10)
b.iii Pādānta-paṅkti-ādi (10)
b.iv Ādyantika (6)
b.v Antādika (6)
b.vi Pādasamudgaka (15)
b.vii Ardhaparivṛtti (1)
b.viii Anya-bheda - Pādatṛtīyārdhāvṛtti (Repetition in 1/6th of a pāda)
viii.1 Ādimadhya
viii.2 Madhyānta
viii.3 Ādyanta

Rudraṭa admitted the possibility of innumerable permutations of these repetitions.[7] Thus, the combinations of the repetitive patterns, which do not follow any specific pattern regarding the position are labeled as 'Aniyata-pāda-bhāgāvṛtti Yamaka'.

2.6 Mammaṭa

Mammaṭa brought forward a perfectly well-defining definition of Yamaka. He defined Yamaka as a repetition of similar sounding syllables, having different meanings when the sequence of letters is meaningful.[8] If both or any of the utterance of the Yamaka is meaningless that also is considered as having different meanings. He accepted the same scheme as Rudraṭa's for the repetition across the pādas. But for the repetition within a pāda, he mentioned only the first two from 2a and the first three from 2b of Rudraṭa. post-Mammaṭa, no remarkable up-gradation or proposal of any innovative classification scheme of Yamaka is observed.

3 Yamaka Identifier and Classifier Tool

We have seen various definitions of Yamaka and various classification schemes provided by the Indian rhetorical tradition. The main difference in the variations of the definition is concerning the

---

[7] *yamakānāṃ gatireṣā deśāvayavāvapekṣamāṇānām* |
aniyatadeśāvayavaṃ tvaparamasaṃkhyaṃ sadevāsti ‖ 3.56, *Kāvyālaṅkāra.*
[8] *arthe satyarthabhinnānāṃ varṇānāṃ sā punaḥśrutiḥ yamakam* | 9.83.1, *Kāvyaprakāśa.*

meaning of the repeated patterns. While Mammaṭa insisted on the repeated patterns that have different meanings, others are silent on it when they define Yamaka. From the computational implementation point of view, we decide to ignore the meaning, as it is heavily dependent on the complete context. Regarding the classification, we decide to consider three classifications viz. of Bharata, Daṇḍin and Rudraṭa. We have chosen these three among the many for the following reasons. The three rhetoricians represent the development of the concept of Yamaka and the difference in the factors on which the classification is defined. Bharata is the foremost person providing a crude classification of Yamaka. Daṇḍin has provided a very elaborate classification scheme where he brings in three dimensions viz. position of the repeated pattern across or within a pāda, number of repetitions and the continuity. Rudraṭa, though followed Daṇḍin closely, provides a completely different classification following only the repetition across and within a pāda, considering the two-fold, three-fold, four-fold and six-fold symmetries within a pāda. Of course, the four-fold and six-fold symmetries make sense only if the pāda is sufficiently long.

Finally, a note about a small deviation due to the defects in pronunciation. Sounds such as 'r' and 'd' or 'l' are pronounced interchangeably in different parts of the country. Similarly, the sounds 'v' and 'b' or 'n' and 'ṇ'. This has been observed centuries ago by the Alaṅkārins,[9] and hence words with such differences are treated alike. Thus, for example, even though 'abalaṃ vitatāravam' and 'avalambita-tāravam' are two different words, for the purpose of Yamaka, these are treated alike. And hence the following two verses with these words fall under the category of Yamaka alaṅkāra.

sa tvārambharato'vaśyam-
  abalaṃ vitatāravam |
sarvadā raṇamānaiṣīd-
  avānalasamasthitaḥ ||
satvārambharatovaśyam-
  avalambitatāravam |
sarvadāraṇamānaiṣī
  davānalasamasthitaḥ ||

In the current implementation, the tool does not handle this kind of example.

---

[9] *ralayorḍalayostadvadvalayorbavayorapi* |
namayornaṇayoścānte savisargāvisargayoḥ || 7.6, *Ekāvalī.*

4 Normalization

Yamaka is a repetition of the same sequence of letters. Hence extracting various sequences of n-letters called 'n-grams' and choosing the high-frequency longest matches should help in identifying the Yamaka and its classification. However, there are certain linguistic and prosodic structure related concerns we need to address. These concerns and the solutions we adopted for resolving them are stated below.

The repeated patterns in Yamaka, are typically the words. Though not every definition of Yamaka insists on the 'meaningfulness' of the repeated pattern, these patterns typically are meaningful and hence they are the words. But these words can be a part of a compound as well, and thus these can be sometimes bound morphemes and not isolated words. Sanskrit is very rich in compound formation and often in Yamaka we see repeated patterns to be part of compounds. Therefore, instead of taking the word as a unit, we consider the syllable as a unit for generating the n-grams. A syllable is zero or more consonants followed by a vowel. We ignore the vowel modifiers viz. anusvāra and visarga.

Since the Yamaka is identified based on sound patterns, the variations in spelling need to be taken care of. Sanskrit allows some spelling variations with regard to nasalization. The nasal stops viz. ṅ, ñ, ṇ, n and m can be written as an anusvāra (ṃ) when it is followed by a homorganic stop. For example, 'ambuja' versus 'aṁbuja', 'aṅka' versus 'aṁka'. 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. Since the Yamaka pattern may cross the word boundaries, we decided to normalize all the nasal stops to anusvāra.

A special character that needs special attention is the avagraha (represented as a single QUOTE (') in IAST (International Alphabet for Sanskrit Texts). The avagraha is a writing convention to indicate the elided 'a' during the sandhi operation. Since for the purpose of Yamaka identification, we look at the sandhied text only, we ignore the avagraha if it is present in the input text.

Similarly, the pāda boundaries and the spaces between the words also do matter during pattern matching. In the oral tradition, the spaces between the words do not carry any significance. Yamaka deals with the sound patterns, and as such, we ignore the spaces between the words. However, we do retain the boundary between the two halves of the verse, since the pattern does not get carried over from one half of the verse into the other.

From the very definition of Yamaka, it follows that the repeated

words are ambiguous having more than one meaning and the meaning of such words gets decided only based on the context, the relation of the word with other words in the sentence and so on. In the absence of any computational tool that helps us in the disambiguation of such words, we decided to rely only on pattern matching, ignoring the meaning component. It is assumed that the input provided to the system is a real literary work by a human author and not a random piece of text generated by an automatic machine such as chat-GPT and as such, we can blindly assume that the repeated patterns do really have different meanings in different instances they are found.

5 Algorithm for Yamaka identification and classification

Taking into consideration all the concerns and the solutions mentioned, we now present below the algorithm for the identification and classification of Yamaka following the three schemes- viz. that of Bharata, Daṇḍin and Rudraṭa.

Since for computational purposes we ignore the word meaning and depend solely on the sound patterns, this tool can be used for identifying Yamaka from modern Indian languages as well. Also, as a fact, almost all major Indian scripts are used to write Sanskrit. Hence, we decide to allow the input in all major Indian scripts including IAST. Further, with the advent of computer technology without any proper keyboard and font support for Devanagari script or Sanskrit language in the early days, several transliteration schemes for keying in Sanskrit were evolved, such as Itrans, Velthuis, Harward Kyoto (HK), etc. From the processing point of view, transliteration schemes such as WX[10] and Sanskrit Library Phonetic basic encoding scheme (SLP) were developed. We decided to provide support for all such transliterations schemes as well as major Indian scripts. Thus, the tool takes the input in several encodings like Devanagari, IAST, SLP, WX, HK along with other Indian scripts such as Gujarati, Bengali, Odia, Malayalam, Tamil, Telugu and Kannada. The input is then converted into WX notation for processing. The WX notation provides a one-to-one mapping between the Sanskrit phonemes and the roman letters. The Unicode Devanagari encoding is not suitable for any phonemic-level operations since it does not have a one-to-one mapping between the sound and its Unicode representation. Hence, we decided to use the WX notation for internal processing.

After converting the input into WX notation, we generate the n-grams for each half of the verse separately and sort them on their length,

[10] *{https://en.wikipedia.org/wiki/WX_notation}*

retaining only the n-grams that have occurred at least two times. In the case of one-grams, even if their frequency is more than or equal to 2, they are ignored, since repetition of any one-gram can just be accidental. Remove the n-grams which are sub patterns of the larger n-grams from the list.

Now for classification, we check the position of each repeated verse with various parameters such as beginning, middle, end, the pāda numbers, whether the repetition is with intervention or without, and then name them following the three different schemes viz. of Bharata, Daṇḍin and Rudraṭa.

A web interface has been designed that allows one to input the text in the script of choice or the transliteration of choice if the input is in Roman. The output shows the verse by highlighting the repeated pattern, followed by the classification following the three schemes above.

The interface is available at https://sanskrit.uohyd.ac.in/yamakav3/

We provide verse as an input without any special symbols except for the full stops or pipe symbols (Daṇḍa) to mark the two different lines of the verse. Sample input and output are shown in Figure 1 and Figure 2 respectively. In the given input, the first and the third feet are identical. Hence, the output for the verse having this condition of Yamaka is called 'samasta-pādāvṛtti-sandaṃśa-yamaka' in Rudraṭas classification.

According to Daṇḍin it a case of '(1,3)-dvipādagata-yamaka'. Here, the numbers in the bracket are the pāda number of identical feet. In Bharata's scheme, the same condition is named as 'vikrānta-yamaka.'

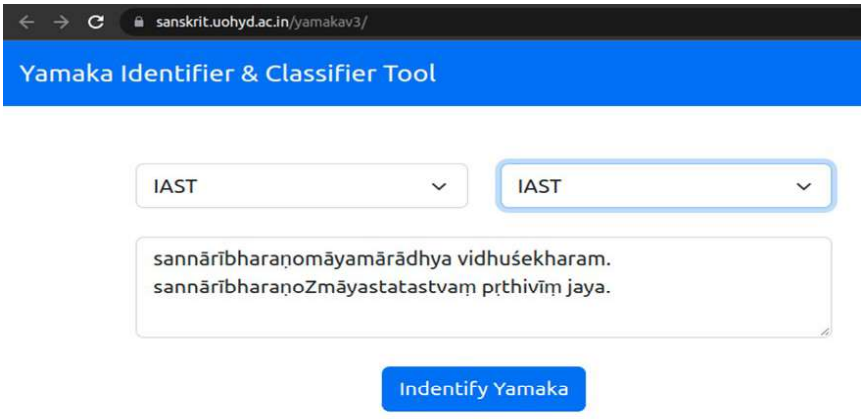

Figure 1: Interface of Yamaka Identifier and Classifier Tool: Input

Figure 2: Interface of Yamaka Identifier and Classifier Tool: Output

6 Performance evaluation

We have collected 185 different verses from Nāṭyaśāstra, Bhāmaha's Kāvyālaṅkāra, Kāvyādarśa and Kāvyālaṅkāra of Rudraṭa along with the verses from Raghuvaṃśa, Ghaṭakarpara, Śrimahābhāratatātparya of Ānandaīrtha, etc. Out of which 163 were analyzed correctly. In 22 cases the tool exhibited some over-generation, by producing multiple solutions, only one of which was correct and the other solutions incorrect. The extra solutions which were wrong were produced due to the following reasons. To classify Yamaka from the verses constructed in mātrāvṛttas is currently beyond the scope of the classifier program. Although, the Yamaka identification task is carried out flawlessly with such constructions. The verses with the patterns having interchangeable sounds mentioned before and the patterns ending with a consonant like 'd' or 't' (Ex. kamalaṃ kamalaṅkurvadalimad-dali matpriye) gets changed in their second repetition due to sandhi. In such cases, the tool identifies the repetition of dalima but fails to identify the repetition of alimat.

7 Conclusion

We have presented a computational tool that identifies the figure of sound 'Yamaka'. Since this figure of speech primarily depends on the

sound and not meaning, the tool can be used effectively to identify the Yamaka alaṅkāra from any other Indian language such as Telugu, Marathi, etc.

This tool can be used as a teaching aid to highlight the differences between various classification schemes, and also demonstrate various examples under different categories

This tool can also be deployed for other important tasks related to Sanskrit manuscripts. The errors produced by the OCR (Optical Character Recognition) due to the limitations of OCR or the damaged manuscripts may be fixed by seeking suggestions from this tool as well as the tools for metrical analysis.

References

Renate Söhnen. 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, 1995.

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

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

2 Benjamin Englard. A rhetorical analysis approach to natural language processing. In ArXiv, 2013.

3 Milind Kumar Audichya and Jatinderkumar R. Saini. Towards natural language processing with figures of speech in hindi poetry. In International Journal of Advanced Computer Science and Applications, 2021.

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

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

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

7 Tyler Neil. 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, 2023.

8 Hrishikesh Terdalkar and Arnab Bhattacharya. 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, 2023.

9 Acarya Vishveshvara. Kāvyaprakāśa. Janamandala Limited, Varanasi, 2017.

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

11 C. Shankara Rama Sastri. Kāvyālaṅkāra of bhāmaha. The Sri Balamanorama Press, Mylapore, Madras, 1956.

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

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

14 Durgaprasad and Kashinath Parab. Kāvyālaṅkāra. Nirnayasagar Press, Mumbai, 1886.

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

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

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

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

19 Rudradev Tripathi. Sanskrit sahitya me shabdalankara. Shri Lal Bahadur Shastri Kendriya Sanskrit Vidyapeeth, Delhi, 1972.

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