Meter identification of Sanskrit verse

KESHAV MELNAD, PAWAN GOYAL, *and* PETER SCHARF

Abstract: A significant portion of Sanskrit literature composed over more than three millennia beginning with the Vedic hymns is composed in metrical verse. Discussion of particular types of Sanskrit meter appears even in the oldest extant Vedic text, the *Ṛgveda*, and the science of meter documenting various poetic meters is mentioned in the oldest lists of disciplines. Based on an analysis of the standard classical works of the science of poetics, Velankar (1949) compiled an exhaustive list of more than six hundred different meters which was included by Apte, Gode, and Karve (1957–1959) as an appendix. The present paper presents Web-based software that analyzes Sanskrit metrical patterns and identifies meters. While using a precise phonetic encoding it yet allows numerous input methods and accepts either accented or unaccented text. The software has been successfully tested on a database of 1031 verses in the *Pañcākhyānaka*, including 291 verses in 23 different types of meters besides Anuṣṭubh. The software should be widely useful to Sanskrit students and scholars, especially those who focus on poetics.

Keywords: prosody, metrics, poetics, Sanskrit meter, meter identification, syllable parsing

1 Introduction

A major portion of Sanskrit literature is in the form of poetry. The proportion of Sanskrit literature composed in verse over the period of more than three millennia from which works are extant is

significant. The earliest Vedic works, *Ṛgveda*, *Sāmaveda*, *Atharvaveda*, and much of the *Yajurveda*, are composed in verse. The great epics *Mahābhārata* and *Rāmāyaṇa* are composed almost exclusively in verse. Works in every genre of classical Sanskrit literature from mathematics, linguistics, medicine, and philosophy to the dramatic and literary arts are composed in verse. There are two main purposes behind versification:

1. Composing a text in verse makes it easy to memorize.
2. Versified text is suitable for melodic and rhythmic chanting.

Much of Indic verse literature was intended for oral transmission, ritual, and public performance.

The large body of Sanskrit poetry is composed in specific metrical patterns. Several metrical patterns are identified and discussed even in the *Ṛgveda* itself. Among the six ancillary disciplines called *Vedāṅga*s associated with the study and understanding of the Vedas is the science of metrics (*chandas*). The *Pāṇinīyaśikṣā* calls this science the feet of the Veda, *chandaḥ pādau tu vedasya* (*PS.* 41).

Identification of meter presents a difficult task for students and scholars of Sanskrit. To offer assistance in this task, the present work develops a software tool to recognize them automatically. Recently Mishra (2007) developed software to analyze metrical patterns. His software, which he deemed a test version, recognizes 1,352 metrical patterns (`http://sanskrit.sai.uni-heidelberg.de/Chanda/HTML/`). Although the work is prodigious, a few deficiencies detract from it:

1. It recognizes only meters with a fixed number of syllables per verse quarter (*pāda*) and cannot recognize meters based upon the number of moræ (*mātrā*s).
2. It requires special treatment of contiguous vowels since the Kyoto-Harvard encoding which it requires as input fails to

distinguish sequences of the contiguous simple vowels *aï* and *aü* from the diphthongs *ai* and *au*.

3. It does not handle accented text, again due to limitations of the Kyoto-Harvard encoding.

This paper presents a Sanskrit metrical analyzer, which we call 'Meter Identifying Tool' (MIT), that improves upon Mishra's work in some respects. It analyzes a wide range of meters, uses a precise phonetic encoding that accommodates accented text, and allows numerous input methods. While at present our tool recognizes only 661 metrical patterns, these include several types of meters based upon the number of moræ. Processing is done in the Sanskrit Library Phonetic Basic encoding (SLP1) documented by Scharf and Hyman (2011: 151–58). SLP1 distinguishes diphthongs from sequential simple vowels and provides methods to indicate accented texts.

The paper is organized as follows. Section 2 describes the source texts on which this work is based. Section 3 describes the structure of Sanskrit meters. Section 4 describes MIT including our database of metrical forms (§4.1), the form of input and ouput (§4.2), and our algorithm for the identification of metrical patterns (§4.3). Section 5 evaluates the results of testing our tool. Conclusions and future work are discussed in section 6.

2 Sources describing Sanskrit meters

Ollett (2013) recently summarized the history of the Indian science of poetics. The standard classical works include Piṅgala's *Chandaḥśāstra* (c. 200 BCE) (Joseph 2011), Jayadeva's *Jayadevachandas* (c. 600 CE), Jayakīrti's *Chando'nuśāsana* (11th c.), Hemacandra's *Chando'nuśāsana* (12th c.) (Velankar 1949), and Kedārabhaṭṭa's *Vr̥ttaratnākara* (11–12th c.) (Kedārabhaṭṭa 1942) among others. Piṅgala's *Chandaḥśāstra* provides definitions of various

kinds of meters in a sūtra text consisting of eight chapters. After a general introduction to prosody in the first chapter and before discussing the origin of meters in the closing chapter, Kedārabhaṭṭa describes four types of meters in the central four chapters of his *Vr̥ttaratnākara*. Jayadeva introduced the brilliant mnemonic technique of composing the definition of a metrical pattern in the very metrical pattern to be defined. For example, the definition of the Indravajrā meter given in (2) below (p. 332) consists of a verse quarter in that meter. Such a definition is said to be endowed with the object to be defined as well as its definition and is thus termed in Sanskrit *lakṣyalakṣaṇasaṁyukta*. Later poetic works on Sanskrit meter employ this technique of definition as well.

Apte, Gode, and Karve (1957–1959) incorporated in Appendix A 'Sanskrit prosody' the list of Sanskrit metrical patterns compiled by Velankar (1949) from several of the classical Sanskrit poetic treatises. The second part of this appendix, called 'A classified list of Sanskrit meters,' contains 769 metrical definitions. MIT presently recognizes 661 of these, including all of the fixed-syllable (*varṇavr̥tta*) meters except those of the daṇḍaka variety. Although MIT includes several meters based upon the number of moræ, it has yet to include the bulk of these.

3 Sanskrit prosody

Sanskrit prosody is metrical. Numerous metrical patterns of several general types are based upon varying sequences of light and heavy syllables that constitute a verse quarter or a line constituting half a verse. After explaining the factors that determine syllable weight, we describe basic units for the two major types of meter, that based upon numbers of syllables (*varṇavr̥tta*) and that based upon number of moræ (*mātrāvr̥tta*).

3.1 Syllable weight

Phonetically, a syllable consists of a single sonorous peak surrounded by less sonorous elements. In Sanskrit, syllables consist of a single vowel or diphthong possibly preceded by up to five consonants and possibly followed by a coda consisting of one or two consonants. Indic scripts orthographically represent syllables by any initial consonants, the vowel, and possibly an anusvāra or visarga. For metrical purposes, syllable weight is determined by vowel length and the presence or absence of a subsequent consonant cluster. There are two weights for a syllable, viz. light (*laghu*) and heavy (*guru*). Pāṇini in his *Aṣṭādhyāyī* dedicated 3 aphorisms to explain the weight of vowels:

- *A.* 1.4.10 ह्रस्वं लघु। A short vowel is termed *laghu*.
- *A.* 1.4.11 संयोगे गुरु। (ह्रस्वम् 10) A short vowel immediately followed by a consonant cluster is termed *guru*.
- *A.* 1.4.12 दीर्घं च। (गुरु 11) A long vowel is also termed *guru*.

In classical Sanskrit poetry, the conditions for determining a light or heavy vowel, and a convention for marking them in writing, are described by Kedārabhaṭṭa in the following verse:

> सानुस्वारो विसर्गान्तो दीर्घो युक्तपरश्च यः।
> वा पादान्ते त्वसौ ग्वक्रो ज्ञेयोऽन्यो मात्रिको लृजुः॥
> (*VR.* 1.9 Kedārabhaṭṭa 1942: 6)
> 'A syllable is heavy if it has an anusvāra, a visarga, or a long vowel, or is followed by a consonant cluster, and optionally if it occurs at the end of a pāda. A heavy syllable is denoted by a curly symbol (ऽ) and a light by a straight line (।).'

Apte, Gode, and Karve (1957–1959: Appendix A, p. 1b) adds, "The consonant clusters प्र & ह्र, as also ब्र & क्र are said to be exceptions, before which the vowel may be short by a sort of poetical

license." In this paper, instead of the avagraha and daṇḍa, we use the modern notations macron for heavy and breve for light. We use *vā* in our database to indicate the optional heaviness of the final syllable of a pāda. For example, consider the Vasantatilakā verse quarter

(1) निन्दन्तु नीतिनिपुणा यदि वा स्तुवन्तु.

Here the last syllable *u* is laghu. But the definition of the Vasantatilakā meter requires the final syllable of each verse quarter to be heavy. In accordance with the rule concerning optional heaviness of the final syllable of a pāda (*vā pādānte*), however, it may be considered heavy and thus, the verse quarter matches the Vasantatilakā metrical pattern.

3.2 The basic unit of varṇavṛtta meter: gaṇa

The term *gaṇa* is a technical term in metrics for a sequence of three syllables. Every possible pattern of light and heavy syllables is designated by a compound beginning with a single character and ending in this term. Since there are two possible weights, laghu or guru, for three syllables, there are eight ($2^3 = 8$) types of gaṇas. A popular verse of unknown origin describes the eight possible trisyllabic weight patterns, designated by the same number of terms, by indicating where in the sequence of three syllables a light or heavy syllable appears.

> आदिमध्यावसानेषु यरता यान्ति लाघवम्।
> भजसा गौरवं यान्ति मनौ तु गुरुलाघवम्॥
> 'At the beginning, middle, and end respectively, *y*, *r*, and *t* go to lightness; *B*, *j*, and *s* go to heaviness, but *m* and *n* go to heaviness and lightness.'

Table 1 presents the eight patterns described in the verse.

Table 1
Gaṇa patterns

¯ stands for guru and ˘ for laghu

No.	गण	gaṇa	pattern
1	म्	m	¯¯¯
2	य्	y	˘¯¯
3	र्	r	¯˘¯
4	त्	t	¯¯˘
5	भ्	bh	¯˘˘
6	ज्	j	˘¯˘
7	स्	s	˘˘¯
8	न्	n	˘˘˘

3.3 The basic unit of mātrāvṛtta meter: caturmātrika

Groups of syllables measuring four moræ (*mātrā*) constitute the basic unit of meters based on the number of moræ. A laghu syllable is counted as one mātrā and a guru as two. Four mātrās constitute a *caturmātrika*. There are five possible patterns of a *caturmātrika* as shown in Table 2.

3.4 Types of meters

As mentioned above, there are two major types of meters: varṇavṛtta and mātrāvṛtta, the first of which is based upon patterns of gaṇas and the second of which is based upon patterns of caturmātrikas. The first type has three subtypes. Hence the four types of meters are as follows:

1. varṇavṛtta

Table 2
Mātrā patterns

No.	***Caturmātrā***	Pattern
1	sarvaguru (G)	¯¯
2	bhagaṇa (B)	¯˘˘
3	jagaṇa (j)	˘¯˘
4	sagaṇa (s)	˘˘¯
5	sarvalaghu (L)	˘˘˘˘

a. samavr̥tta
b. ardhasamavr̥tta
c. viṣamavr̥tta

2. mātrāvr̥tta

A samavr̥tta meter has the same gaṇa pattern in each of its four pādas. An ardhasamavr̥tta meter has two gaṇa patterns, one in its first and third pādas, and a different gaṇa pattern in its second and fourth pādas. A viṣamavr̥tta meter may have a different gaṇa pattern in each of its four pādas. A mātrāvr̥tta meter has a different pattern of caturmātrikas in its first and second lines.

Metrical patterns are defined by describing the sequences of gaṇas or caturmātrikas in a verse quarter or line. For example, the samavr̥tta meter Indravajrā is defined by stating that each verse quarter consists of the pattern of gaṇas *t t j g g*. Apte, Gode, and Karve (1957–1959: Appendix A, p. 4a) provides the following lakṣyalakṣaṇasaṁyukta definition of the Indravajrā meter:

(2) स्यादिन्द्रवज्रा यदि तौ जगौ गः
syādindravajrā yadi tau jagau gaḥ
¯ ¯ ˘|¯ ¯ ˘|˘ ¯ ˘| ¯ ¯
'That is called *indravajrā* if it consists of *t t j g g*.'

We provide additional examples of the patterns of gaṇas or caturmātrikas in verses while describing the output of MIT in §5 below.

4 Meter Identifying Tool (MIT)

MIT consists of a database of metrical definitions and a Java program. We describe the structure of the database, the input and output, and our algorithm for conducting metrical analysis in the following three subsections.

4.1 Meter database

Our meter database contains metrical definitions in an easily readable text file. The meter pattern conforms to a different prototype for each of the four types of meters described in §3.4. Because a samavr̥tta meter has the same pattern in all four pādas, it is sufficient for the database to contain just one pattern for a pāda to define all four pādas of a verse. Besides the pāda pattern, a Boolean variable indicates whether the metrical definition permits optional heaviness of the final syllable of a pāda (*vā pādānte*); the value is indicated as true when it does. Thus, a samavr̥tta metrical definition is stored in a prototype containing the meter name and a single pattern plus the Boolean value. Because an ardhasamavr̥tta meter has one pattern for its first and third pādas but a different pattern for its second and fourth pādas, the database must include two patterns in its metrical definition each of which is followed by the Boolean value. Because each of the four pādas of a viṣamavr̥tta meter may have a different pattern, the database must include in its metrical definition four patterns each with its Boolean value. For mātrāvr̥tta meters, the patterns of *caturmātrika*s for the first and second lines are different. Thus a mātrāvr̥tta meter is stored with two patterns each of which is followed by the Boolean value. Table 3 shows the prototypes of each of the four types of meters and

the number of metrical definitions of each type of meter contained in our database as of the date of publication.

Table 3

Prototypes of each of the four types of meters and the number of meters of each type in our database

Meter type	Prototype	Quantity
samavr̥tta	<meter name> <pattern1> <vA>	587
ardhasamavr̥tta	<meter name> <pattern1> <vA> <pattern2> <vA>	48
viṣamavr̥tta	<meter name> <pattern1> <vA> <pattern2> <vA> <pattern3> <vA> <pattern4> <vA>	14
mātrāvr̥tta	<meter name> <pattern1> <vA> <pattern2> <vA>	12
Total		661

4.2 Input and output

MIT is an interactive program that accepts a full verse, line, or verse quarter in Sanskrit and returns a five-part analysis. If the input consists of a line or verse, the first line must end in a single daṇḍa and the second in a double daṇḍa. While at present the input and display is in SLP1 (the encoding in which processing is performed), the tool will soon be provided with the Sanskrit Library transcoding preferences that include input from and out-

put to popular meta-encodings such as Kyoto-Harvard, ITrans, Titus, Velthuis, and Hyderabad-Tirupati (WX), as well as to Unicode Roman, and the Unicode representation of major Indic scripts Devanagari, Gurmukhi, Kannada, Bengali, Telugu, Malayalam, Gujarati and Oriya. The analysis returned consists of the following five parts:

1. name and type of meter linked to a definition of the meter
2. the string parsed into orthographic syllables
3. the scansion (*prastāra*) of the string showing the pattern of light and heavy syllables.
4. the pattern of gaṇas or caturmātrikas
5. the number of syllables or mātrās

4.3 Algorithm

After transcoding the input text to SLP1, the first task is to remove any non-phonetic characters, except the periods that represent daṇḍas at the end of lines. The second step is to divide the string into syllables. The definition of an orthographic syllable is $C^*VX?$, where C stands for a consonant, V stands for a vowel, X stands for the set {anusvāra, visarga, jihvāamūlīya, upadhmānīya}, and $*$ and ? are regular expression quantifiers representing zero or more and zero or one respectively. The regular expression used to match an orthographic syllable in SLP1 is therefore:

```
[yvrlYmNRnJBGQDjbgqdKPCWTcwtkpSzsh]*
[aAiIuUfFxXeEoO][HMZV]?
```

Consonants final in a line are grouped with the last orthographic syllable.

Once the syllables are found, we find the weights of the syllables, and then attempt to determine the pattern of gaṇas or caturmātrikas. More restrictive metrical patterns are checked before

less restrictive ones. The input verse is checked for the samavr̥tta, ardhasamavr̥tta and viṣamavr̥tta metrical patterns in that order. Only when the input verse does not match any of these patterns, do we search for a mātrāvr̥tta meter pattern.

If the input text lacks line-end markers, it is assumed to be a single pāda and to belong to the samavr̥tta type of meter. If line markers are present then an attempt is made to divide the lines into pādas since these are not demarcated in the input. The pāda boundary is required for the meters of the varṇavr̥tta types, not for the mātrāvr̥tta meters which are defined according to patterns per line.

We first check the number of syllables per line and proceed to search for the appropiate type of meter based upon the following conditions:

1. If the number of syllables in the first line is the same as in the second line and each line contains an even number of syllables, then we divide the lines in half, generate the gaṇa patterns for each pāda, and check whether the patterns of all pādas are the same. If so, we search for the pattern among the samavr̥tta meter definitions before proceeding to search for the pattern among the ardhasamavr̥tta and viṣamavr̥tta types.
2. If the number of syllables in the first line is the same as in the second line, but each line contains an odd number of syllables, then the verse cannot be of the samavr̥tta type. In this case, the search for a pattern proceeds directly to those among the ardhasamavr̥tta type and, if not found, to those among the viṣamavr̥tta type.
3. If the number of syllables in the first line is different from the number in the second line, then the verse cannot be of the samavr̥tta or ardhasamavr̥tta types. In this case, the search

for a pattern proceeds directly to those among the viṣamavr̥tta type.

Now, it is easy to find the pāda boundaries in samavr̥tta meters because the number of syllables is the same in each pāda and we can divide the set of syllables in each line into two equal parts. For the ardhasamavr̥tta meters containing different numbers of syllables in each pāda, or viṣamavr̥tta meters, on the other hand, we need a different method because there are numerous possible divisions of each line into pādas. If $\{x,y\}$ represents the number of syllables in pādas one and two respectively, then, for example, for an input line containing 17 syllables, it is possible to divide a line into pādas containing various numbers of syllables, $\{8,9\}$, $\{9,8\}$, $\{7,10\}$, $\{10,7\}$, etc. In general, for an input line containing x syllables, we will have $x-1$ possible pairs. This corresponds to the fact that the first pāda can contain any number of syllables from 1 to $x-1$, and the second pāda will contain the remaining syllables.

For each possible division of a meter into pādas as described above, we search for matching *gaṇa* patterns of the pādas in our database to determine whether there exists a meter definition that matches these patterns. If we find a match, we stop searching and output the result.

If the input verse does not match any of the samavr̥tta, ardhasamavr̥tta or viṣamavr̥tta type meters, we attempt to match it with the mātrāvr̥tta meters. The patterns for mātrāvr̥tta meters can be implemented using a regular expression search. For example, the pattern for the Āryā meter is: `J4J4JjJg` or `J4J4JLJg`, where `j` corresponds to *jagaṇa* as listed in Table 2, `g` corresponds to a guru syllable, `4` corresponds to any of the caturmātrika patterns, `J` corresponds to any caturmātrika pattern other than jagaṇa, and `L` corresponds to a sarvalaghu. Hence we first convert each of the meter definitions for mātrāvr̥tta meters into an equivalent regular expression automatically. Then, we try to match the input verse

with the regular expressions corresponding to the various definitions.

5 Results

In this section, we will discuss one example verse from each meter type and report the output produced by our algorithm when the verse was given as input to MIT. The original verse is shown in Devanāgarī, the input below it, and the output next in SLP1 encoding.

5.1 A samavṛtta verse in Vasantatilakā meter

निन्दन्तु नीतिनिपुणा यदि वा स्तुवन्तु
लक्ष्मीः समाविशतु गच्छतु वा यथेष्टम्।
अद्यैव वा मरणमस्तु युगान्तरे वा
न्याय्यात्पथः प्रविचलन्ति पदं न धीराः॥

```
nindantu nItinipuRA yadi vA stuvantu
   lakzmIH samAviSatu gacCatu vA yaTezwam.
adyEva vA maraRamastu yugAntare vA
   nyAyyAtpaTaH pravicalanti padaM na DIrAH..

The verse is in vasantatilakA meter which is a
samavftta.
The metrical patterns are...
......pAda 1:::::::::
Parsed String     : ni nda ntu nI ti ni pu RA ya
                    di vA stu va ntu
Syllable weights  : g g l g l l l g l l g l g g
Gana Pattern      : t B j j g g
No.  of Syllables : 14
......pAda 2:::::::::
Parsed String     : la kzmIH sa mA vi Sa tu ga cCa
                    tu vA ya Te zwam
Syllable weights  : g g l g l l l g l l g l g g
Gana Pattern      : t B j j g g
No.  of Syllables : 14
```

```
......pAda 3:::::::::
Parsed String     : a dyE va vA ma ra Ra ma stu yu
                    gA nta re vA
Syllable weights  : g g l g l l l g l l g l g g
Gana Pattern      : t B j j g g
No.  of Syllables : 14
......pAda 4:::::::::
Parsed String     : nyA yyA tpa TaH pra vi ca la
                    nti pa daM na DI rAH
Syllable weights  : g g l g l l l g l l g l g g
Gana Pattern      : t B j j g g
No.  of Syllables : 14
```

5.2 An ārdhasamavṛtta verse in Vegavatī meter

स्मर वेगवती व्रजरामा केशववंशरवैरतिमुग्धा।
रभसान्न गुरून्गणायन्ती केलिनिकुञ्जगृहाय जगाम॥

```
smara vegavatI vrajarAmA
   keSavavaMSaravEratimugDA.
raBasAnna gurUngaRayantI
   kelinikuYjagfhAya jagAma..

The verse is in vegavatI meter which is a arDasama-
vftta.
The metrical patterns are...
......pAda 1:::::::::
Parsed String     : sma ra ve ga va tI vra ja rA
                    mA
Syllable weights  : l l g l l g l l g g
Gana Pattern      : s s s g
No.  of Syllables : 10
......pAda 2:::::::::
Parsed String     : ke Sa va vaM Sa ra vE ra ti mu
                    gDA
Syllable weights  : g l l g l l g l l g g
Gana Pattern      : B B B g g
No.  of Syllables : 11
......pAda 3:::::::::
```

```
Parsed String     : ra Ba sA nna gu rU nga Ra ya
                    ntI
Syllable weights  : l l g l l g l l g g
Gana Pattern      : s s s g
No.  of Syllables : 10
......pAda 4:::::::::
Parsed String     : ke li ni ku Yja gf hA ya ja gA
                    ma
Syllable weights  : g l l g l l g l l g g
Gana Pattern      : B B B g g
No.  of Syllables : 11
```

5.3 A viṣamavṛtta verse in Lalita meter

नयुगं सकारयुगलं च भवति चरणं तृतीयकम्।
तदुदितमुरुमतिभिर्ललितं यदि शेषमस्य खलु पूर्वतुल्यकम्॥

```
nayugaM sakArayugalaM ca
   Bavati caraRaM tftIyakam .
taduditamurumatiBirlalitaM
   yadi Sezamasya Kalu pUrvatulyakam .

The verse is in lalitam meter which is a vizama-
vftta.
The metrical patterns are...
......pAda 1:::::::::
Parsed String     : na yu gaM sa kA ra yu ga laM
                    ca
Syllable weights  : l l g l g l l l g g
Gana Pattern      : s j s g
No.  of Syllables : 10
......pAda 2:::::::::
Parsed String     : Ba va ti ca ra RaM tf tI ya
                    kam
Syllable weights  : l l l l l g l g l g
Gana Pattern      : n s j g
No.  of Syllables : 10
......pAda 3:::::::::
```

```
Parsed String     : ta du di ta mu ru ma ti Bi rla
                    li taM
Syllable weights  : l l l l l l l l g l l g
Gana Pattern      : n n s s
No.  of Syllables : 12
......pAda 4:::::::::
Parsed String     : ya di Se za ma sya Ka lu pU
                    rva tu lya kam
Syllable weights  : l l g l g l l l g l g l g
Gana Pattern      : s j s j g
No.  of Syllables : 13
```

5.4 A mātrāvṛtta verse in Āryā meter

कृष्णः शिशुः सुतो मे वल्लवकुलटाभिराहृतो न गृहे।
ह्नणमपि वसत्यसाविति जगाद गोष्ठां यशोदार्या॥

```
kfzRaH SiSuH suto me
   vallavakulawABirAhfto na gfhe .
kzaRamapi vasatyasAviti
   jagAda gozWyAM yaSodAryA ..
```

```
The verse is in AryA meter which is a mAtrAvftta
......pUrvArdha:::::::::
Parsed string   : kf zRaH Si SuH su to me va lla
                  va ku la wA Bi rA hf to na gf he
Syllable weight : g g - l g l - g g - g l l - l l
                  g - l g l - g l l - g
Pattern         : J 4 J 4 J j J g
No.  of morae   : 30
......uttarArdha:::::::::
Parsed String   : kza Ra ma pi va sa tya sA vi ti
                  ja gA da go zWyAM ya So dA ryA
Syllable weight : l l l l - l g l - g l l - l g l
                  - g g - l - g g - g
Pattern         : J 4 J 4 J l J g
No.  of morae   : 27
```

Recall that the pattern shown for mātrāvr̥tta meters is different from the varṇavr̥tta meters; the pattern is explained in Table 2 (p. 332) and at the end of §4.3 (p. 332).

5.5 Performance

To determine how useful the tool would be on a real dataset, we evaluated the program's performance on all of the verses in Pūrṇabhadra's *Pañcākhyānaka* (Hertel 1908) extracted from the Sanskrit Library's digital edition of the work (Scharf 2011). The database of these verses consists of 1031 verses, 291 of which contain 23 types of metrical patterns other than Anuṣṭubh. This database, which also includes the name of the meter of each verse, served as the gold standard for our evaluation. We used our tool to identify the metrical pattern corresponding to each verse. MIT correctly recognized 1018 out of the 1031 verses (98.7%) and 287 out of the 291 non-Anuṣṭubh verses (98.6%). There were no cases in which a meter was recognized incorrectly. On the contrary, our tool discovered several data-entry errors in our digital text and in the annotations of meter types that we then corrected by reference to the original printed edition.

We analyzed the 13 verses in our database whose meters were not identified to discover the reasons behind the failure. Our gold-standard identified 9 of the 13 verses not recognized as Anuṣṭubh, 3 as Upajāti, and 1 as Śārdūlavikrīḍita. For the last and for one of the Anuṣṭubh verses, our meter definition did not include the exceptional parameter described in §3.1 that permits syllables to be considered short even if they occur before one of the four conjunct consonants *kr*, *pr*, *br*, or *hr* (that is, *r* preceded by *k*, *p*, *b*, or *h*). Our tool included only the most restrictive definition of Upajāti meter Apte, Gode, and Karve (1957–1959: Appendix A, p. 4a) list according to which the meter consists of any combination of pādas in the two meters Indravajrā and Upendravajrā. However, other

definitions they mention permit combinations of Indravaṁśā and Vaṁśasthā, Smr̥ti and Śruti, and indeed combinations of any samavr̥tta meters. Our tool did identify combinations of Indravaṁśā and Vaṁśasthā in two of the unidentified meters, and a combination of three samavr̥tta meters in one of the unidentified meters. Of the remaining 8 Anuṣṭubh meters not recognized by our tool, 3 have an extra syllable in one pāda, 3 contain a prohibited sagaṇa pattern in the 5th through 7th syllables of the first or third pāda, 1 contains a prohibited nagaṇa pattern in the 2nd through 4th syllables in the first pāda, and the last unidentified verse lacks a required jagaṇa pattern in its second pāda. In every case in which the meter of a verse was not identified, our tool correctly identified the metrical pattern of each pāda individually.

6 Conclusions and future work

In this paper, we described a tool for the analysis of Sanskrit prosody and identification of meter type. The tool covers a wide range of meters and has been successfully tested on a considerable number of verses. The evaluation of this tool over a database of 1031 verses from the *Pañcākhyānaka* gave impressive results with greater than 98.6% accuracy.

At present the tool is limited by the ability to recognize just the 661 meters in our database. Future work involves extending our database to cover all kinds of meters and to deal with certain peculiarities. As an example of one such peculiarity, various meter definitions specify restrictions on the position of a caesura (*yati*) and require that the preceding Sanskrit word end before it. For example, the Śārdūlavikrīḍita meter has 19 syllables in each pāda. An additional condition is that there is a caesura after 12 syllables and that the preceding Sanskrit word end prior to it. To recognize whether these conditions are satisfied requires a parser able

to determine word boundaries. In future work we would like to integrate the output of a text segmentation tool such as the Sanskrit Heritage Reader (Huet 2005; Huet and Goyal 2013) with our system in order to check this condition.

Among meters not adequately handled in our tool at present are Vedic meters. Generally Vedic meters conform to classical definitions for meters bearing the same names regarding the number of syllables per pāda but not regarding the patterns of light and heavy syllables. In other cases, the definition of a meter in Vedic differs from its classical definition. For example, the Gāyatrī meter in Vedic has three pādas each of which consists of eight syllables whereas in the classical definition it has four pādas of six syllables each.

As mentioned in the previous section, our tool is currently over restrictive in the realm of Upajāti definitions. Likewise, only a few of the several Āryā meters have been implemented. There are several subdivisions of Āryā meter such as *Gīti*, *Vaitālīya*, *Vaktra* and *Mātrāsamaka* with different metrical patterns. We plan to include Vedic meters, additional Upajāti patterns, Āryā subdivisions and other metrical patterns in future work.

Although the Web version of the program currently handles just one verse at a time, a command-line version of the program run locally is able to analyze any number of verses in a file. We intend to provide this facility on the Web as well.

At present, if a string of text without line markers is submitted for analysis, it is assumed to be a pāda of a samavṛtta meter and the program quits if a matching samavṛtta metrical pattern is not found. We intend to add procedures to subdivide the string into lines and pādas to check for possible matches to other types of meters as well. In this way one could submit any string to discover whether it is metrical or not. After implementing this feature, we plan to test MIT against a known database of non-metrical text to see whether it falsely identifies metrical patterns. We also have

plans to extend our tool to highlight probable spelling mistakes in the input verse.

The tool has many potential applications other than identifying the metrical pattern of an input verse. The feature of discovering whether an input text contains any metrical pattern could have applications to syntactic parsing. As Scharf, Goyal, Ajotikar, and Savardekar describe in the preceding paper in this volume, poetry and prose differ significantly in their syntax. If a tool could determine whether the input text is prose or poetry, it would improve the performance of dependency parsers if they provide a different set of penalties for poetry than they do for prose.

References

Apte, Vaman Shivaram, Parshuram Krishna Gode, and Cintamana Ganesa Karve. 1957–1959. *Revised and enlarged edition of Prin. V. S. Apte's The practical Sanskrit-English dictionary*. 3 vols. Pune: Prasad Prakashan.

Hertel, Johannes, ed. 1908. *The Panchatantra: a collection of ancient Hindu tales in the recension, called* Panchakhyanaka*, and dated 1199 A.D., of the Jaina monk, Purnabhadra*. Harvard Oriental Series 11. Cambridge, Mass.: Harvard University.

Huet, Gérard. 2005. "A functional toolkit for morphological and phonological processing: application to a Sanskrit tagger." *Journal of Functional Programming* 15.4: 573–614. URL: `yquem.inria.fr/~huet/PUBLIC/tagger.pdf`.

Huet, Gérard and Pawan Goyal. 2013. "Design of a lean interface for Sanskrit corpus annotation." *Proceedings of ICON 2013, the 10th International Conference on NLP*, pp. 177–86.

Joseph, George Gheverghese. 2011. *The crest of the peacock: non-European roots of mathematics*. Princeton: Princeton University Press.

Kedārabhaṭṭa. 1942. *vr̥ttaratnākara: with the gloss Saubhāgyavatī and commentary Ratnaprabhā*, ed. and comm. by Nr̥siṁhadeva Śāstrin. Lahore: Meharacandra Lakṣmaṇadāsa.

Mishra, Anand. 2007. *Sanskrit metre recognizer*. URL: `http://sanskrit.sai.uni-heidelberg.de/Chanda/HTML/`.

Ollett, Andrew. 2013. "The gaṇacchandas in the Indian metrical tradition." *Contributions to current research in Indology: proceedings of the first International Indology Graduate Research Symposium, September 2009, Oxford*; vol. 1, *Puṣpikā: tracing ancient India through texts and traditions*, ed. by Nina Mirnig, Péter-Dániel Szántó, and Michael Williams, ch. 15. Oxford: Oxbow Books.

Scharf, Peter M., ed. 2011. *The* Pañcākhyānaka *of Pūrṇabhadra: first XML edition*. URL: `http://sanskritlibrary.org`.

Scharf, Peter M. and Malcolm D. Hyman. 2011. *Linguistic issues in encoding Sanskrit*. Delhi: Motilal Banarsidass.

Velankar, H. D., ed. 1949. *Jayadāman: a collection of ancient texts on Sanskrit prosody and a classical list of Sanskrit meters with an alphabetical index*. Haritoṣamālā 1. Includes the following texts: *Jayadevachandas* by Jayadeva, *Chando'nuśāsana* by Jayakīrti, *Vr̥ttaratnākara* by Kedārabhaṭṭa, and *Chando'nuśāsana* by Hemacandra. Bombay: Haritosha Samiti.