

The Phonemic approach for Sanskrit text

Prof. R. K. Joshi
Visiting Design Specialist
Centre for Development of Ad-
vanced Computing, (C-DAC)
Juhu, Mumbai, India.
Email:rkjoshi@cdacmumbai.in

Dr. T. N. Dharmadhikari
Ex-Director, Vaidika
Sanshodhan
Mandal, Pune, India.

**Dr. Vijay Vasudev
Bedekar**
President, Institute for
Oriental Study, Thane,
India.

1 Varna (phoneme) and Akshara (syllable)

Sanskrit Grammar has distinguished the terms varna (phoneme) and akshara (syllable). Both these terms are used in the context of spoken languages and can be extended to written languages.

Since the oral tradition in India was of a higher order, the stress on right pronunciation was laid at most on the spoken language. To represent such speech nuances in written language, various chinhas (signs) were evolved as to strike the equivalence in spoken and written expressions. This extra-ordinary activity is part of the Indian tradition.

Diagram 1

Proposed Varnamala for computerization of Sanskrit									
Vowels:									
अ a 0001	आ ā 0101	इ i 0003	ई ī 0103	उ u 0075	ऊ ū 0105	ऋ ṛ 1036	ॠ ṝ 1037	ऌ ḷ 1037	ॡ ḹ 1038
ऐ e 0005	ए ē 0113	ऎ æ 0008	ॠ ḥ 0103	ऋ ṛ code needed	ओ o 0047	ॠ ō 0140	ॡ œ code needed	ॢ ṃ code needed	ॣ ṃ̄ code needed
Vowel signs:									
Anusvara ṁ 1041			Candrabindu ṃ 0002 0170			Visarga ḥ 1028			
Consonants:									
क k 0008	ख kh 0008 0208	ग g 0201	घ gh 0201 0204	ङ ṅ 0148	च c 0003	छ ch 0003 0204	ज j 0004	झ jh 0004 0204	ञ ñ 0272
ट ṭ 0208	ठ ṭh 0208	ड ḍ 0208	ढ ḍh 0208	ण ṇ 1047	त t 0074	थ th 0074 0204	द d 0004	ध dh 0004 0204	न n 0006
प p 0070	फ ph 0070 0204	ब b 0002	भ bh 0002 0204	म m 0002	य y 0074	र r 0072	ल l 0002	ळ ḷ 0002	व v 0074
श ś 0002	ष ṣ 0002	स s 0072	ह h 0008						

Therefore, the realization of such phonemic system in the context of new technology seems to be imperative where writing is talked in the context of speech and speech in the context of writing. The attempt is made to identify varnamala comprising of basic speech sound units as vowel phonemes (swara varna) and consonant phonemes (vyanjan varna) (Diagram 1).

These phonemes (varnas) when combined as C...C + V or only V, form complete phonetic cluster. The correspondence in spoken and written syllables must be preserved through the Phonemic scheme firstly by giving each phoneme a distinct identity and secondly by giving each chinha -denoting nuances of speech -a distinct identity in the form of standardisation.

2 Sanskrit Phonology and Orthography

Presently, Devanagari script is used for writing classical Sanskrit as well as Vedic Sanskrit. This includes the multi-tier usage of diacritic marks of complex compositions, above, below and at the sides of the base glyphs. Therefore, as compared to modern historical derivatives from Sanskrit such as Hindi, Marathi, Nepali etc., the Sanskrit text demands adequate range of characters as well as exhaustive rendering rules to achieve the advanced typographic quality in Classical as well as Vedic Sanskrit text.

3 Standardisation Principles

The effective smallest unit of the Sanskrit writing system can be the phoneme (varna). The range of phonemes (Varnamala) consists of 'SwaraVarna' (Vowel Phoneme) and 'Vyanjan Varna' (Consonant phoneme). While 'Swara Varna' is self-powered and it is not dependent on any other element, the 'Vyanjan Varna' however, needs an addition of 'Swara Varna' to compose its syllabic entity. While 'Swara Varna' (V) can be written down as syllables ('akshara'), other syllables are the outcome of the combination of 'Vyanjan Varna' and 'Vowel Varna'.

4 Phoneme (varna) to Syllables (aksharas)

As mentioned earlier phonemes are divided into two types: vowel phonemes (swara varna) and consonant phonemes (vyanjan varna). They together broadly constitute the Varnamala which has been referred as a varna-samamnaya. The orthographic representation of these varnas is done in a systematic way. The combination of consonant phoneme and a vowel phoneme produces a syllable (akshara). A cluster of glyphs emerges as an outcome of this process.

For example,

/k/ + /a/ = /ka/ syllable which is written as ... क्+अ=क

/p/ + /aa/ = /paa/ syllabic akshara is /paa/ प्+आ=पा

Please note that corresponding to each swara phoneme there is an akshara which is its syllabic form.

Vowel phoneme अ आ इ ई etc.

Vowel syllable अ आ इ ई etc.

This similarity has unfortunately caused the non differentiation in between Varnamala and Aksharamala in the present times. Recently some grammar books in Indian languages are attempting to explain the difference between Varna and Akshara.

In the written text when the combination of CV occurs, only then the Swara Varna is rendered into a Swara Matra sign. Otherwise the Vowel Varna is written in the text as it is (as independent vowel). Therefore vowel Matra is just the rendering form of a sound of a Swara Varna in CV. However in the context of V, VV, VC, VCCC.... The Swara Varna remains as is ओ, आई, अच्

5 Rendering of aksharas (Syllables)

k-phoneme + /a/ = k-akshar क्+अ=क

The syllables formed by adding vowel phonemes /a/, /aa/, /i/, etc. to the consonant phoneme are written by creating aksharas. All swara phonemes are added to one consonant phoneme one by one. This concept is called a baaraakhadi.

Thus the concept of extended range of 'Barakhadi' (18 syllables) is achieved in the following way.

K(d) + vv1 = K + A = KA क्+अ=क

K(d) + vv2 = K + AA = KAA क्+आ=का

K(d) + vv3 = K + I = KI क्+इ=कि

K(d) + vv4 = K + II = KII क्+ई=की

K(d) + vv5 = K + U = KU क्+उ=कु

K(d) + vv6 = K + UU = KUU क्+ऊ=कू

K(d) + vv7 = K + Vocalic R = K(Vocalic)R क्+ऋ=कृ

K(d) + vv8 = K + Vocalic RR= K(Vocalic)RR क्+ऌ=कृ

K(d) + vv9 = K + Vocalic L = K(Vocalic)L क्+ऍ=कृ

K(d) + vv10 = K + Vocalic LL = K(Vocalic)LL क्+ऎ=कृ

K(d) + vv11 = K + E = KE (Short) क्+ए=के

K(d) + vv12 = K + EE = KE क्+ए=के

K(d) + vv13 = K + E = K(Candra)E क्+ऐ=कै

K(d) + vv14 = K + AE = KAI क्+ऐ=कै

K(d) + vv15 = K + O = KO (Short) क्+ओ=को

K(d) + vv16 = K + O = KO क्+ओ=को

K(d) + vv17 = K + O = K(Candra)O क्+औ=कौ

K(d) + vv18 = K + AU = KAU क्+औ=कौ

The combination of two forms (C&V) into a syllable, at times creates a new integrated shape or retains partial identity of both the forms.

Syllables can also be formed by adding vowel phonemes to a sequence of more than one consonant phonemes. These syllables are called jodaksharas or sanyuktaksharas. For example:

k-phoneme + y-phoneme + aa-phoneme = kyaa

क्+य्+आ=क्या


s-phoneme + t-phoneme + u-phoneme = stu

स्+त्+उ=स्तु

Sample text output (Diagram 2)

IndiX

Output of Sanskrit and Vedic Sanskrit text
through inputs of Sanskrit Varnamala



Centre for Development
of Advanced Computing
Gulmehar Cross Road No. 9
Jyoti, Mumbai 40
Tel: +91-22-262 0160
Fax: +91-22-262 3120

Level A : Ordinary Reading

त्यम् पु वाजिनं देवजुतं सहोवानं तरुतारं रथानाम् ।
अरिष्टनेमिं पृतनाजमाशुं स्वस्तये तादृर्मिहा हुवेम ॥
RV. 10. 178.
- तादृर्मिसाम

Level B : Reading with High Stress and Low Stress

त्यम् पु वाजिनं देवजुतं सहोवानं तरुतारं रथानाम् ।
अरिष्टनेमिं पृतनाजमाशुं स्वस्तये तादृर्मिहा हुवेम ॥
RV. 10. 178.
- तादृर्मिसाम

Level C : Speaking (Chanting) with full tonal variations

ॐ ऽ ६ म् ॥ त्यम्पु ॥ वाजिं ॥ ना ऽ २३४५ म् ॥ देवजुता ऽ २३४ म् ॥
५२ ३ २ १ २ ३ ३ ४ ५
सहोवानं ता ॥ स्ता ऽ ३ ॥ रै रथानाम् ॥
२ ३ ३ ४ ५ ६ ७ ८ ९ १० ११ १२ १३ १४ १५ १६ १७ १८ १९ २० २१ २२ २३ २४ २५ २६ २७ २८ २९ ३०
आरिष्ट ऽ ना ऽ २३४ इमीम् ॥ पृतना ऽ ३४३ जमाशुम् ॥ स्वस्त ॥ याइ ॥
१२ ३ ३ ४ ५ ६ ७ ८ ९ १० ११ १२ १३ १४ १५ १६ १७ १८ १९ २० २१ २२ २३ २४ २५ २६ २७ २८ २९ ३०
तादृर्मिहा ऽ ३४३ ॥ ह् ऽ ३ वा ऽ ५ इ मा ऽ ६ ६ ॥ २ ॥

A part from "Naradya Sāsa"
by Usha R. Bhatia,
Bharatpur Oriental Research Institute,
Poonia 411 004, INDIA

It is important to note that the invariant element in this process is the set of phonemes. The variation occurs in the shape of glyphs written in various Indian scripts. For example, the phoneme /k/ and /0/ will result in the glyph shape where graphic element is added in front in Devanagari where as in Bengali script, graphic shape will be added in front and prior to the base glyph. Therefore this model can be extended to most of the Indian languages which have phonetic base. To sum up the proposed scheme calls for code points for consonant phoneme k as compared to the existing Devanagari code which provides code points for the Akshara - glyph ka. The proposed scheme is of additive nature (k + a) as compared to subtractive model. This scheme would allow unambiguous representation of the entire repertoire of characters required for creating the exhaustive Devanagari script syllabic range along with its phonetic values.

6 Considerations: Varnamala

- 6.1 Through the Varnamala approach the IPA equivalence for Sanskrit text (as well as other Indian language text) can be established as one to one correspondence. And hence can be mapped very easily.
- 6.2 Through the Varnamala-Phonemic approach lexical order and sorting operation in the areas of dictionary etc. can be done in the logical and more efficient way.
- 6.3 The Phonemic scheme will help grammatically in context to Samaas and Sandhi features in Sanskrit.
- 6.4 Under the phonemic scheme the keyboard input procedure will be simplified by reducing keys for vowel matras.
- 6.5 The range of Vedic Sanskrit accent/tonal marks when identified, can be positioned with vowel varnas.
- 6.6 The essential Swaraadi Anusvar and Swaraadi Visarga can be added in the standardization.
- 6.7 If necessary, Swara Varna and Vyanjana Varna Bhedak Chihnas can be added in extended Devanagari Varnamala.
- 6.8 Essential punctuation marks as used in the modern times can also be added if necessary.

7 Conclusion

- 7.1 The new scheme of Phonemes (vowels varnas and consonants varnas) as basic characters, is nearer to the linguistic model of Sanskrit and serves all the linguistic needs.
- 7.2 The text-processing operations like indexing and sorting which are very important for in

formation storage and retrieval on computers can be performed efficiently.

- 7.3 Speech synthesis can be facilitated as the nuances of speech are preserved through phoneme standardization.
- 7.4 An absolute requirement for any script standardization is that it should facilitate a computer system to take any valid sequence of underlying character codes and algorithmically render the appropriate visual form from a given repertoire of surface glyphs. In the case of Phonemic scheme the required character shaping rules are well-formulated and therefore essential rendering engine can be built based on this concept.
- 7.5 (Alekha 1984 NCST, Vividha 1985 NCST, Vidura 1987 NCST). Text processing applications based on phonemic approach have been successfully implemented in Turnkey jobs and are in use with Sanskrit Institutions like Bharati Samskrit Vidya Niketanam, Lonavla and is considered favorably by other Sanskrit Institutions and scholars. Vedic Sanskrit text has been enabled on IndiX at C-DAC Mumbai using Phonemic approach (*Diagram 2*).

Acknowledgements

Technology Development for Indian Languages (TDIL), Govt. of India.

LC Group, C-DAC Mumbai (www.cdacmumbai.in)

Vinod Kumar, Sr. Software Specialist, C-DAC Mumbai
Jui Mhatre, TypeFont Designer, Language Computing Group, C-DAC Mumbai.

Supriya Kharkar, Assistant TypeFont Designer, Language Computing Group, C-DAC Mumbai.

References

- Dr. Kelkar A.R. 1989. *Transliteration of South Asian languages, a brief review and a proposal for a standard*. Centre of advanced study in linguistics, Deccan College, Pune.
- Handbook of the International Phonetic Association*. 1999. Cambridge University Press.
- Naravane V.D. 1961. *Bharatiya Vyavahar Kosh*. Triveni Sangam.
- Peter Ladefoged, 2001, *Vowels and Consonants*. Blackwell publishers, UK.
- Chandrakant More, Arvind Kulkarni, *Chetna Marathi Vyakaran Va Lekhan*, Chetna Publications, Mumbai, India.
- Prabodh Primer*, Department of official languages, Ministry of home affairs, Govt. of India.

R. K. Joshi, Prague 2003, A unified phonemic code based scheme for effective processing of Indian languages. 23rd Internationalization and Unicode.

R. K. Joshi. October 2002, Vedic Code, a draft, Vishwabharat No. 7.

R. Shama Sastri, K. Rangacharya, 1985 reprint, *The Taittiriya Pratishakhya*, Motilal Banarasidas, Delhi.

Shrinath Shanbaug, Durgesh Rau, R. K. Joshi 2002, An intelligent multi-layered input scheme for phonetic scripts. ACM International conference proceeding series, Hawthorne, New York.

W.S. Allen, 1961, *Phonetics in Ancient India*, London, Oxford University Press.

Supporting References – Note 1

अ इ उ ण् । ऋ ऌ क् । ए ओ ङ् । ऐ औ च् । ह य व र ट् ।
ल ण् । ञ म ङ ण न म् । झ भ ञ् । घ ढ ध ष् । ज ब ग ड ड श् ।
ख फ छ ठ थ च ट त व् । क प य् । श ष स र् । ह ल् ।

इति सुत्राण्यणादिसंज्ञार्थानि । हकारादिष्वकार उच्चारणार्थः ।
१ हलन्त्यम् ॥ १.३.३ ॥
उपदेशेऽन्त्यं हल् इत् स्यात् । उपदेश आद्योच्चारणम् ।

These aphorisms are meant for the formation of 'an' and such other technical designations. The vowel 'a' in the letters 'ha' etc. (in the aphorisms) is (appended) for (case of) pronunciation.

There are 42 phonemes enumerated in these 14 Shivasutras, wherein there are 9 Vowels, 1 Mahaprana, 4 Semi-vowels, 5 Nasals, 20 Consonants and 3 Fricatives. The addition of phoneme 'A' (Akaara) in phonemes other than vowels such as 'Ha' etc. is only for the convenience of utterance and as such the seemingly resemblance of these phonemes (such as 'Ha' etc.) to syllables can be explained.

Supporting References – Note 2

...तेषां विभागाः पञ्चधा स्मृतः
स्वरतः कालतः स्थानात् प्रयत्नानुप्रदानतः
इति वर्णविदः प्राहुर् निपुण तन्निबोधत

by time Length Place of articulation Process of articulation

सव्यञ्जनः सानुस्वरः शुद्धोवापि
with consonants with nasalization of pure vowel alone vowel

स्वरोऽक्षरम्
is syllable

A vowel with a consonant, or even by itself, forms a Syllable
शुक्लः प्राणिशास्त्रस्य
with the commentary of Uvata- उपट्ट

स्वयं राजते इति स्वरः ।
अन्वग् भवति व्यञ्जनम् ॥ -पतञ्जलि

पाणिनीय शिक्षा
8.10
Edited & translated by A. Weber

The concept of varna and its features such as Svara, Kaala, Sthaana and Prayatna etc. have been already identified in Paniniya Shiksha. The definition of Akshara (with consonant, with nasalization or a pure vowel can be called Akshara) is observed from Uvata (Rik Pratishakhya) and Patanjali's bhashya.

Supporting References – Note 3

(त्) विवृतकण्ठीस्थित विवाराघोषारूपप्राणाख्य
बाह्यप्रयत्नविशिष्ट श्वासद्वनिजनित उत्तरदन्तमूलाघो-
भागस्थान जिह्वाग्रकरण स्पृष्टप्रयत्नार्धमात्रिक परांग-
भूत वायु देवताक ब्राह्मण जातिक तकार ॥

(ई) संवृत कण्ठीस्थित संवाराख्यबाह्यप्रयत्न
सहित नादद्वनिजनित तालुस्थानात्सुपसंहृत कल्पोष्ठ-
सहित जिह्वा मध्यकरण विवृतप्रयत्न द्विमात्रिक ॥

(प्रचयस्य) सूर्यदेवताक शूद्रजाति तमोगुणसहित
सद्यमांगुलि मध्यरेखान्यासयोग्य उच्चकल्पतद्विलक्षण
क्रौञ्चवर्णणतुल्य मध्यमस्वरहेतुभूत सर्वांगस्थानोत्पन्न
प्रचयस्वरगुणक अग्निदेवताक ब्राह्मण जातिक ईकार
द्विमात्रिक विरामाः ॥

जिह्वाग्रेण तवर्गे दन्तमूलेषु ॥ ३८ ॥

'तवर्गे' कार्ये 'जिह्वाग्रेण' वर्णान् 'दन्तमूलेषु' स्पर्शयेत् ॥३८
जिह्वाग्रेण यथावस्थितेन उत्तरदन्तमूलेष्वधोभागे स्पर्शयति ।
शिक्षायां तु—तवर्गे दन्त्यं नकारं तु दन्तमूलीयं चोक्त्वा न-
कारस्यापि कश्चिदन्यत्वमुच्यते । तदस्माभिर्नाश्रयितव्यम् । प्रा-
तिशाख्यविप्रतिषिद्धत्वात् । तस्मान्नकारोपि सर्वत्र दन्तमूलीयः ॥

तज्जिह्वामम्

In Vedalakshanam various types of Varnakramas are documented for Ghanapathas. In the example above, the detail phonetic profile of consonant phoneme T and vowel phoneme Ii is self explanatory and vivid. The terms such as T Kaar, Ii Kaar are used to indicate the concept of Varna. In the Uccaranakalpa – section of Taittiriya Pratishakhya the formation of articulate sounds and their production is described in detailed way.