Appendix B

Sanskrit Library Phonetic Basic

The Sanskrit Library Phonetic Basic encoding scheme (SLP1) attempts to meet high standards of unambiguous encoding while restricting encoding to 76 codepoints in the ASCII character set. SLP1 utilizes 58 codepoints to encode segments: 53 to represent phonetic segments and five to represent punctuation ⟨′ . ? − ⟩. In addition SLP1 utilizes 18 codepoints to encode phonetic features: three to indicate stricture, six to indicate length, eight to indicate tone, and one to indicate nasalization. Although certain features are indicated by a sequence of codepoints, no codepoints double as both segments and features. While useful, SLP1 is not an ideal encoding. To its credit it is consistent in that it consistently encodes phonetic rather than graphic elements (with the exception of the punctuation signs). Yet it does not maintain a consistent basis of encoding because it mixes the encoding of phonetic segments and phonetic features. Nor does it satisfy the Fano condition because it utilizes a few codepoints as prefixes in code sequences. For example, the forward slash ⟨/⟩, back slash ⟨\⟩, and caret ⟨^⟩ indicate udātta, anudātta, and independent svarita accents by themselves but also serve as the prefixes in several sequences that indicate particular tones and tonal sequences realized in various Vedic traditions; and the digit ⟨1⟩, which by itself indicates short length, is used as a prefix in a sequence that serves to in-

151

dicate length of $1\frac{1}{2}$ morae. Nevertheless, single codepoints capture most phonetic segments commonly used in classical Sanskrit. The only commonly occurring phonetic segment that requires a sequence is nasalized *l,* i. e. ⟨**l~**⟩. Moreover, SLP1 does clearly define single codepoints or code sequences to capture a comprehensive set of phonetic distinctions in classical and Vedic Sanskrit.

B.1 Basic Segments

अ a **a**	आ ā **A**	इ i **i**	ई ī **I**	उ u **u**	ऊ ū **U**

	ऋ r̥ **f**	ॠ r̥̄ **F**	ऌ l̥ **x**	ॡ l̥̄ **X**	
	ए e **e**	ऐ ai **E**	ओ o **o**	औ au **O**	

क् k **k**	ख् kh **K**	ग् g **g**	घ् gh **G**	ङ् ṅ **N**
च् c **c**	छ् ch **C**	ज् j **j**	झ् jh **J**	ञ् ñ **Y**
ट् ṭ **w**	ठ् ṭh **W**	ड् ḍ **q**	ढ् ḍh **Q**	ण् ṇ **R**
		ळ् ḷ **L**	ळ्ह् ḷh **|**	
त् t **t**	थ् th **T**	द् d **d**	ध् dh **D**	न् n **n**
प् p **p**	फ् ph **P**	ब् b **b**	भ् bh **B**	म् m **m**

	य् y **y**	र् r **r**	ल् l **l**	व् v **v**	
	श् ś **S**	ष् ṣ **z**	स् s **s**	ह् h **h**	
	: ḥ **H**	⪥ h̲ **Z**	⪥ h̲̣ **V**	˙ṁ **M**	

B.2 Punctuation

Although punctuation does not properly belong to a phonetic encoding, a limited number of punctuation tokens are supported in this encoding, since they can be used to provide basic segmentation information. The question mark is used to indicate inaudible or illegible characters in transcription.

ऽ '	।.	‾?	- -	□ ⌴
'	.	?	—	⌴
avagraha	danda	question	hyphen	space

B.3 Modifiers

Modifiers are added after a character to indicate variations in segment stricture, length, accent, and nasalization, in the order stated. Prolonged length, accent, and nasalization occur in classical Sanskrit as well as Vedic. Modifiers are used in combination to indicate special features of stricture, length, accent, and nasalization in Vedic.

B.3.1 Stricture

- **_** heaviness [used for semivowels *y* or *v*]
- **=** lightness [used for semivowels *y* or *v*]
- **!** lack of release (*abhinidhāna*) [used for stops or semivowels *y, v,* or *l*]

B.3.2 Length

- ***** subsegmental epenthetic vowel (*svarabhakti*)
- **#** length of half a mora
- **1** length of one mora [used in Vedic after short agitated *kampa;* short *e, o;* and heavy *anusvāra*]
- **1#** slightly lengthened
- **2** length of two morae [used for *dvimātra anusvāra* in Vedic]

3 prolonged length of three morae [used for *pluta* vowels]

4 prolonged length of four or more morae [used in *raṅga*]

B.3.3 Accent

/ high pitch

**** low pitch

^ circumflex

6 extra low tone

7 low tone

8 high tone

9 extra high tone

+ sharpness

B.3.4 Nasalization

~ nasalization

B.4 Modifier combinations and usage notes

B.4.1 Stricture

y_ heavy *y*

v_ heavy *v*

y= light *y*

v= light *v*

k! unreleased (*abhinidhāna*) *k*

g! unreleased (*abhinidhāna*) *g*

... similarly for other unreleased stops

y! unreleased (*abhinidhāna*) *y*

v! unreleased (*abhinidhāna*) *v*

l! unreleased (*abhinidhāna*) *l*

B.4.2 Length

a⋆	epenthetic *a*
i⋆	epenthetic *i*
u⋆	epenthetic *u*
f⋆	epenthetic *r̥*
x⋆	epenthetic *l̥*
e⋆	epenthetic *e*
e1	short *e*
o1	short *o*
a1#	slightly lengthened short *a*
	. . . similarly for other slightly lengthened short vowels

B.4.3 Surface accent

Tonal contours in Vedic have numerous distinct varieties described in *Prātiśākhyas*. The indication of these requires the use of the accent signs for high pitch, low pitch, and circumflex (**/**, ****, and **^**) in conjunction with tonal modifiers **6**, **7**, **8**, **9** that indicate the features *extra low, low, high,* and *extra high* pitch respectively. The additional modifier **+** is used to indicate a distinction in sharpness or effort of uncertain phonetic significance described in the *Vājasaneyi* (1.125) and *Taittirīya* (20.9-12) *Prātiśākhyas*, in spite of the same length of vowel and same beginning and end pitches. The term 'aggravation' below translates *kampa:* 'aggravated' means with *kampa;* 'unaggravated' means without *kampa.* The following modifier sequences are used to indicate the tonal features described to their right:

/8	high tone (*udātta*)
\\7	low tone (*anudātta*)
\\6	extra low tone (*sannatara*)
^98	declining tone from extra high to high (dependent and unaggravated independent *svarita* according to the *R̥kprātiśākhya*)
^97	declining tone from extra high to low (aggravated independent *svarita* according to the *R̥kprātiśākhya*)

^87	declining tone from high to low (dependent *svarita* according to the *Vājasaneyi* (1.125) and *Taittirīya* (20.9–12) *Prātiśākhyas*)
^87+	sharp declining tone from high to low (independent *svarita* according to the *Vājasaneyi* (1.125) and *Taittirīya* (20.9–12) *Prātiśākhyas*)
^86	declining tone from high to extra low (aggravated independent svarita according to the *Vājasaneyiprātiśākhya*)

Vowel accent examples

a/8	high toned vowel *a*
a^97	the vowel *a* with short agitated circumflex as described in the *R̥kprātiśākhya*
a3^97	the vowel *a* with prolonged agitated circumflex as described in the *R̥kprātiśākhya*

B.4.4   Syllabified visarga and anusvāra accent

H/	high-pitched *visarga*
H	low-pitched *visarga*
H^	*svarita visarga*
M	low-pitched *anusvāra*

B.4.5   Nasals

Nasalization

Both SLP1 and SLP2 include means to encode 20 yamas (**k~**, **kh~**, …, **b~**, **bh~**) considered, on phonetic grounds, to be epenthetic nasalized segments that adopt features of both of the preceding stop and of the following nasal. Yet the preferred method of encoding yamas, in accordance with the phonological analysis of most ancient Indian phonetic treatises, is to employ characters for just four epenthetic nasals (**k~**, **kh~**, **g~**, **gh~**), or, on the minority view of the *R̥kprātiśākhya*, to employ yamas

(**k~**, **kh~**, ..., **b~**, **bh~**) in place of the non-nasal stop that precedes the nasal. (See p. 63 and p. 72 for discussion.)

l~	nasalized *l*
y~	nasalized *y*
v~	nasalized *v*
k~	nasalized offset (*yama*), after unvoiced unaspirated non-nasal stop when followed by a nasal stop
K~	nasalized offset (*yama*), after unvoiced aspirated non-nasal stop when followed by a nasal stop
g~	nasalized offset (*yama*), after voiced unaspirated non-nasal stop when followed by a nasal stop
G~	nasalized offset (*yama*), after voiced aspirated non-nasal stop when followed by a nasal stop
h~	nasalized offset (*nāsikya*), after *h* when followed by a nasal stop

Anusvāra

M#	short *anusvāra* (which follows a long vowel according to the *Ṛk* and *Vājasaneyi Prātiśākhyas*: *RPr.* 13.22, 13.29, 13.32–33; *VPr.* 4.148–149; the short *anusvāra* measures half a mora while the preceding vowel measures 1.5 morae)
M1#	long *anusvāra* (which follows a short vowel according to the *Ṛk* and *Vājasaneyi Prātiśākhyas*; the long *anusvāra* measures 1.5 morae while the preceding vowel measures 0.5 morae)
M1	heavy *anusvāra* (which is usually called *guru* and also by some *hrasva* and which occurs before a conjunct consonant according to Śikṣās)
M2	two-mora *anusvāra* (which is called *dvimātra* and occurs before a consonant followed by *ṛ* according to Śikṣās)

Raṅga

2~ two-mora *raṅga* (vowel two *mātras* in length
nasalized for the last half *mātra* with *kampa* in the
middle according to *Pāṇinīyaśikṣā* 26–30)

4~ *raṅga* (nasalized vowel four *mātras* in length fol-
lowed by a break according to *Mallaśarmakṛta-
śikṣā*; texts show a double danda to mark the break