

Transliterating Devanagari

Rupert Snell

The whole business of *trānslītereśan* and diacritical marks may seem like the most tedious subject in the world, but it has an important purpose: it allows the reader to envisage the correct sound and Hindi (Devanagari) spelling of a word. Popular rough-and-ready systems of transliteration, often using doubled characters for long vowels and an “n” to show vowel nasality (as in दीवारें *deewaaren*) etc., aren’t good enough for scholarly use because they leave too many ambiguities. So we have to have some rules!

Italics. As a rule of thumb, use of *italics* indicates a literal transcription, requiring the full set of Indic transliteration conventions.

Consonants and vowels follow the standard Indic charts below. (The inherent *a* vowel is shown in the consonant table, but not thereafter.)

अ	आ	इ	ई	उ	ऊ	ऋ	ए	ऐ	ओ	औ
<i>a</i>	<i>ā</i>	<i>i</i>	<i>ī</i>	<i>u</i>	<i>ū</i>	<i>ṛ</i>	<i>e</i>	<i>ai</i>	<i>o</i>	<i>au</i>

क <i>ka</i>	ख <i>kha</i>	ग <i>ga</i>	घ <i>gha</i>	ङ <i>ṅa</i>
च <i>ca</i>	छ <i>cha</i>	ज <i>ja</i>	झ <i>jha</i>	ञ <i>ña</i>
ट <i>ṭa</i>	ठ <i>ṭha</i>	ड <i>ḍa</i>	ढ <i>ḍha</i>	ण <i>ṇa</i>
त <i>ta</i>	थ <i>tha</i>	द <i>da</i>	ध <i>dha</i>	न <i>na</i>
प <i>pa</i>	फ <i>pha</i>	ब <i>ba</i>	भ <i>bha</i>	म <i>ma</i>
य <i>ya</i> र <i>ra</i> ल <i>la</i> व <i>va</i>				
श <i>śa</i> ष <i>ṣa</i> स <i>sa</i> ह <i>ha</i>				

Variations

The little subscript circle used in transliterating ऋ as *ṛ* is often replaced with a simple dot (an easier option on most word-processors), and although this technically conflicts with its use in *ṙ* as a transliteration for ढ, no actual ambiguity results in practice.

Some people prefer to write च and छ as *ch* and *chh* respectively — though this erodes the nice parallel between च *c* and ज *j*.

Avoid transliterating व as *w*, as this leads to inevitable inconsistency.

श and ष are sometimes written as *sh*, losing the distinction between them.

An apostrophe may be used to differentiate the transliteration of गृह from that of घ, as in the month name अग्रहन *ag'han* (the tadbhava form of अग्रहायण *agrahāyaṇ*, November-December).

Dotted consonants

Undotted and dotted pairs of Devanagari consonants are distinguished thus:

क, क़	<i>k, q</i>
ख, ख़	<i>kh, <u>kh</u></i>
ग, ग़	<i>g, <u>g</u></i>
ज, ज़	<i>j, z</i>
ड, ड़	<i>ḍ, ṛ</i>
ढ, ढ़	<i>ḍh, ṛh</i>

The use of underlining in transcribing ख़ and ग़ is admittedly not very clear or satisfactory, but it's hard to find a better alternative. (Sometimes we see *x* for ख़, as in ख़याल *xyāl*; but although it's a neat use of the only roman character that's unused elsewhere in transliterating Devanagari, it doesn't exactly recommend itself.) Popular spellings such as "Ghalib" for ग़ालिब may be horribly confusing (*gh* normally transliterates घ !), but are grandfathered and therefore sacrosanct.

Vowel nasality and nasal consonants

It's important to distinguish vowel nasality (as in हॉँ, नहॉँ) from nasal consonants (as in अंडा, हिन्दी). We'll look at these two separately.

The most consistent way of showing **vowel nasality** is with the tilde, ~ . By sitting directly on the vowel, this sign unambiguously marks a quality of the vowel itself; and a further advantage is that it takes no extra space in the line. But some people prefer to use a dotted *m* (as in the second options below) particularly if it's difficult to type tilde and macron together as in *ã, î, ù* !

नहीं	<i>nahī</i>	<i>nahīm</i>
हँसना	<i>hāsnā</i>	<i>hamsnā</i>
खींचना	<i>khīcnā</i>	<i>khīmcnā</i>
हाँ, माँ वही हैं, गाँव में।	<i>hā, mā vahī hai, gāv mē.</i>	<i>hām, mām vahīm haiṁ, gāṁv meri.</i>
नहीं, मैं कुँ में नहीं हूँ!	<i>nahī, maī kuē mē nahī hū!</i>	<i>nahīm, maiṁ kueri meri nahīm hūṁ!</i>

As we have seen above, **Nasal consonants** are romanized thus:

ङ	<i>ṅ</i>
ण	<i>ṇ</i>
ञ	<i>ñ</i>
न	<i>n</i>
म	<i>m</i>

Neither ङ nor ञ appears independently in Hindi, and in conjuncts they are nearly always substituted by *anusvār* (the nasal dot); so we don't have to worry about them. Words written with *anusvār* can be transliterated either by the roman character in the list above, or with a dotted "m", as below. Using a subscript dot distinguishes this from the vowel nasality discussed above, shown with a superscript dot as in हँ *hām*. A less formal third option is to use an undotted "n" or "m", as in the last column below; but this is definitely less fully "correct"!

अंग	<i>aṅg</i>	<i>aṁg</i>	<i>ang</i>
अंडा	<i>aṅḍā</i>	<i>aṁḍā</i>	<i>andā</i>
अंजन	<i>aṅjan</i>	<i>aṁjan</i>	<i>anjan</i>
अंदर	<i>andar</i>	<i>aṁdar</i>	<i>andar</i>
अंबा	<i>ambā</i>	<i>aṁbā</i>	<i>ambā</i>

Inherent "a" vowel. This is a real teaser, particularly if you reference both Hindi and Sanskrit sources. Is it *Rāmāyaṇa* or *Rāmāyaṇ*? Or maybe *Ramayana*, or *Ramayan*? You will have to work out your own compromises, keeping a careful list of words in each category so that you can maintain consistency. (In the case of this example, my own preference is to use roman "Rāmāyaṇa" when referring to the cultural phenomenon of the epic in general terms, but to

italicize it when referring to a specific text such as the Vālmīki version, dropping the final *a* for a vernacular version: *Rāmāyaṇa*, *Rāmāyaṇ.*)

It is often helpful to write the inherent *a* after a conjunct that is difficult to pronounce without it: योग्य *yogyā*, मित्र *mitra*, साहित्य *sāhitya*, पंचमी *pañcamī*. The *Oxford Hindi English Dictionary*, OHED, marks such occurrences with the “short vowel” sign ˘ (*sāhityă*, *pañcāmī*), but this convention is mostly restricted to the specialized worlds of lexicographers and technical linguists and need not be followed by us lesser mortals. When to include the inherent *a* in the medial position in a word such as नागरी is a matter of choice and depends on how one hears the word: my personal preference would be for *nāgarī* rather than *nāgrī* — three syllables rather than two. (OHED has (*nāgārī*). Transliteration will not show the difference between the the spellings परदा / पर्दा or कुरसी / कुर्सी, since these are orthographic differences only, i.e. share the two spellings share the same pronunciation.

In pre-modern Hindi, different rules apply. Pre-modern literature in languages such as Braj Bhāṣā and Avadhī is mostly in metrical verse, where the inherent vowel counts metrically and should therefore be written when transliterating into roman. This does not mean that associated words such as the names of poets need follow suit: Tulsidas doubtless pronounced his name with just three syllables, not as *Tulasīdāsa* with five! Similarly, many scholars writing in English would prefer the simplicity and elegance of a spelling such as *Rāmcaritmānas* (an accurate indication of how the word is pronounced in everyday use) over the “full” transliteration *Rāmacaritamānasa* that has to appear in a transliteration of a metrical line such as the following:

रामचरितमानस एहि नामा । सुनत श्रवन पाइअ विश्रामा ॥
rāmacaritamānasa ehi nāmā; sunata śravana pāia biśrāmā.

Pre-modern poets exploited the difference between vowel nasality and a nasal continent to their own metrical and musical ends. The word आनंद in its standard spelling produces a sequence of two long syllables and one short; but written as आनँद it has one long and two shorts.

Hindi / Sanskrit words. Many Hindi words borrowed from Sanskrit are well established in the context of English-medium scholarly prose, usually spelled with the inherent *a*, and can be used freely: रस is usually written as *rasa* rather than *ras*, योग is *yoga* rather than *yog*, वेद is *Veda* rather than *Ved*. But in transcribing full sentences of Hindi, write *ras*, *ved*, *yog* etc. — as in आरम्भ में केवल तीन ही वेद थे *ārambh mē keval tīn hī ved the*.

Modern authors’ names: use the spellings that the authors themselves are likely to use in English (these are subjective choices, and finding out the details may call for some research): that is, use the right-hand column below. Don’t italicize personal names when referring to individuals in English prose.

बच्चन	baccan	Bachchan
वर्मा	varmā	Varma or Verma, according to the individual
चौधरी	caudharī	Chaudhari, Chowdhuri, Chaudhary, etc. etc. etc.
सिंह	siṃh	Singh
पांडे	pāṇḍe	Pandey or Pande

Characters' names in fiction etc. may follow the above pattern in the context of English-medium academic prose (e.g. "Premchand's antagonist Hori in *Godān* is a model for many later characters in Hindi fiction".) However, होरी would be written *horī* in a direct transcription from the narrative.

Place names. Indian place-names almost always have well-established English spellings, usually traceable through Google, and you should follow these. They often differ quite substantially from literal transcriptions: ग्वालियर is "Gwalior" rather than Gvāliyar, इलाहाबाद is "Allahabad" rather than Ilāhābād, महु is (bizarrely) "Mhow" rather than Mahū ! Some of the old British spellings have been rationalized: कानपुर > Cawnpore > Kanpur. Different contexts call for different usages, and you will need to decide on the convention that best suits your context, and stick to it like glue.

Language names. Most of these have well-established roman spellings — Hindi, Urdu, Gujarati, Bengali or Bangla, Panjabi or Punjabi, all written without diacritics. Less well-known names of dialects etc. can be used with either type of spelling: ब्रज भाषा is Braj Bhāṣā or Braj Bhasha, अवधी is Avadhī or Awadhi (usually not Avdhī etc.) Just be consistent and everyone will be happy.

Upper and lower case. Upper-case characters are of course fairly meaningless in transliterating Devanagari, which doesn't have capitals (or rather, has *only* capitals and no lower case!), so in transliterating items such as text titles it's best to restrict upper case to the initial of the first word only — as in the titles *Basere se dūr*, *Dīvār mẽ ek khiṛkī rahtī thī*, *Antim araṇya*, *Lāhaur se lakhnaū tak*.

Many personal names have a separate life as abstract nouns: आशा "hope", शान्ति "peace", प्रकाश "light" etc. (and of course there are many English parallels such as Joy, Hope, Constance, Felicity...); for this reason there are some contexts where personal names might usefully be spelt with a capital. This is what I do in my Hindi course-books, to help learners see what is a name and what isn't. For example, the sentence ओ अरुण, ज़रा प्रताप के कमरे में जाकर एक कुरसी ले आना would be transliterated thus: *o Aruṇ, zarā Pratāp ke kamre mẽ jākar ek kursī le ānā*.

There are times when Devanagari's single-case arrangement can catch the reader unawares quite wonderfully. In the opening lines of Premchand's novel *Godān*, the word गोबर gobar occurs twice: firstly as the name of a character (an abbreviation of गोवर्धन / गोवर्धन), and secondly as a noun meaning "cow-dung" (from Sanskrit गोर्वर):

होरीराम ने दोनों बैलों को सानी-पानी देकर अपनी स्त्री धनिया से कहा – गोबर को ऊख गोड़ने भेज देना । मैं न जाने कब लौटूँ । ज़रा मेरी लाठी दे दे ।
धनिया के दोनों हाथ गोबर से भरे थे । उपले पाथकर आयी थी । बोली – अरे, कुछ रस-पानी तो कर लो । ऐसी जल्दी क्या है ।

A *candra* without a *bindu*? To the best of my knowledge, there is no consistently-established sign to transliterate the *candra* that represents an "English" vowel sound in loanwords such as कॉफ़ी *kāfī* "Coffee", ऑल इण्डिया रेडियो *āl inḍiyā reḍiyo* "All India Radio".

(This is effectively the transcription of the Indian pronunciation of these English sounds, i.e. the of the sounds as uttered by a Hindi-speaker; thus Hindi makes no distinction between "lawyer" and "liar" in the word लॉयर *lāyar*. Very appropriate!)

We could just invent a transliteration for the undotted *candra* right here and now! For example, we could underline the vowel, as underlining is easy to do on any computer. This new convention would of course have to be explained to readers...and, sadly, it's unlikely to catch on! But it would at least help us distinguish between the following:

वह कॉफ़ी पीती है ।	<i>vah kāfī pītī hai.</i>	She drinks coffee.
वह काफ़ी पीती है ।	<i>vah kāfī pītī hai.</i>	She drinks quite a bit.

Hypercorrectness? You'll have to make your own decision about when the pursuit of literal correctness tips over into pedantry! Be guided by practicality here. Publishers' names often include an English element, such as पब्लिशर्ज़, पब्लिशिंग, ऐंड संस (or संज़). English spellings seem best here, as *pablišarz*, *pablišing*, *aīḍ sans* (*sanz*) look horribly bizarre.

Establishing consistency. If you suspect that you may have more than one spelling of a word in your document, do a global check. But take care if you use automatic global changes: I once had a contract from a publisher who had globally changed "author" to "editor", unintentionally inventing the word "editorize".

The English spellings of Indian words that appear in standard English dictionaries, such as "nirvana", "pandit", and so on, can of course be used freely in English prose; but be aware that some of them (including these two examples) may have different implications in the two languages, as a result of which that you may prefer to maintain their Indic senses by literal transliterations in italics — *nirvāṇ*, *paṇḍit*.

Quoting other authors. Remember that when you quote a published passage you must always preserve that author's own chosen conventions — you have a duty to preserve the detail of what was written, and you may not change it to comply with your preferred conventions.

An example of transliterated Hindi prose:

आज तक मैं दूसरों की ज़िन्दगी पर आधारित कहानियाँ ही 'रचती' आई थी, पर इस बार मैंने अपनी कहानी लिखने की जुर्रत की है। है तो यह जुर्रत ही क्योंकि हर कथाकार अपनी रचनाओं में भी दूसरों के बहाने से कहीं न कहीं अपनी ज़िन्दगी के, अपने अनुभव के टुकड़े ही तो बिखेरता रहता है। कहीं उसके विचार और विश्वास गुँथे हुए हैं तो कहीं उसके उल्लास और अवसाद के क्षण...कहीं उसके सपने और उसकी आकांक्षाएँ अंकित हैं तो कहीं धिक्कार और प्रताड़ना के उद्गार।

āj tak maĩ dūsrõ kī zindagī par ādhārit kahāniyā̃ hī 'ractī' āī thī, par is bār maĩne apnī kahānī likhne kī jurrat kī hai. hai to yah jurrat hī kyōki har kathākār apnī racnāō mē bhī dūsrõ ke bahāne se kahī na kahī apnī zindagī ke, apne anubhav ke ṭukṛe hī to bikhertā rahtā hai. kahī uske vicār aur viśvās gūthe hue hāī to kahī uske ullās aur avsād ke kṣaṇ ... kahī uske sapne aur uskī ākāṅkṣāē aṅkit hāī to kahī dhikkār aur pratāṛnā ke udgār.

āj tak maiṁ dūsroṁ kī zindagī par ādhārit kahāniyāṁ hī 'ractī' āī thī, par is bār maiṁne apnī kahānī likhne kī jurrat kī hai. hai to yah jurrat hī kyōmki har kathākār apnī racnāoṁ mē bhī dūsroṁ ke bahāne se kahīṁ na kahīṁ apnī zindagī ke, apne anubhav ke ṭukṛe hī to bikhertā rahtā hai. kahīṁ uske vicār aur viśvās gūthe hue haiṁ to kahīṁ uske ullās aur avsād ke kṣaṇ ... kahīṁ uske sapne aur uskī ākāṅkṣāēṁ aṅkit haiṁ to kahīṁ dhikkār aur pratāṛnā ke udgār.

मन्तू भंडारी, एक कहानी यह भी (दिल्ली, राधाकृष्ण, सन् २००९, पृ० ७)

And a pre-modern couplet – a *dohā* from the *Satsaī* of Bihārīlāl:

भूषन भारु सँभारिहै कयौँ इह तन सुकुमार।
सूधे पाइ न धर परें सोभा ही कें भार ॥

bhūṣana bhāru sābhārihai kyaū iha tana sukumāra;
sūdhe pāi na dhara paraī sobhā hī kaī bhāra.

bhūṣana bhāru saṁbhārihai kyaum̐ iha tana sukumāra;
sūdhe pāi na dhara paraim̐ sobhā hī kaim̐ bhāra.

Rupert Snell, *Bihari: the Seven Hundred Poems* (forthcoming)