

ACIP Sanskrit Input Code Standards

The following document presents the Asian Classics Input Project standards for the input of Sanskrit. Whenever inputting Sanskrit items to date, the Project has generally simply followed the system found in the particular document; that is, we have not attempted to modify every document input in order to bring it into an ACIP standard. This document is, therefore, meant first of all to describe the variations that one may expect in ACIP documents, and then finally to describe the input code standard which we have developed for general future use.

Please note that this document refers to ACIP input code standards, and not to the representation of pronunciation. ACIP treatment of Sanskrit pronunciation in written materials is described in a separate document entitled "ACIP Standards for Simplified Tibetan and Sanskrit Pronunciation."

Sanskrit with standard diacritics

In ACIP printed materials which we anticipate will be read mainly by specialists, and where we have typeset the material ourselves and can utilize the special fonts available to us, we use a standard set of diacritics which will immediately be clear to such specialists. One example would be the word *pramāṇa*.

Sanskrit with standard diacritics in next letter position

In applications intended primarily for specialists, but where the fonts available for the application do not allow all needed diacritics above or below the letter to which they refer, ACIP simply places the particular diacritic in the next letter position; for example, you will see *prama-n.a* for *pramāṇa*. This method of representation is used only in situations where there could be no ambiguity possible about the system used; for example, in the menu trees of the *AsiaView* user interface or in the "Sanskrit title" field of the ACIP Master Catalog, where no other hyphens or periods appear.

The following is a summary of how combinations are treated in this manner.

These letter combinations appear in the standard romanization of Sanskrit:

ā ī ū ṛ ṝ ḷ ḹ ṁ ḥ ṅ ñ ṭ ḍ ṇ ś ṣ

They are represented, respectively, as follows:

a- i- u- r. r.- l. l.- m. h. n* n~ t. d. n. sh s.

Sanskrit in Tibetan letters

The method traditionally used by Tibetans to represent Sanskrit in Tibetan letters is effective, but can initially be rather difficult to read when transliterated into roman letters. Wherever we are rendering such a case—for example, when in a Tibetan translation of a Sanskrit classic the original Sanskrit title has been given in-situ by the translator in Tibetan letters—we simply retain the Tibetan conventions of the normal ACIP Tibetan input code. (For a complete explanation of this code, please see the section entitled “ACIP Tibetan Input Code Standards”, given just above).

Most important, we make no attempt to correct any spellings or phrasing: the Tibetanized Sanskrit in the database is a massive tumbleweed that will have to be sorted out by scholars of later generations; we simply type what's there.

ACIP input code for the Tibetan rendering of our example word would thus be given as {PRA M'A nA}. In ACIP's experience, the only standard conventions here that scholars may find difficult are two:

1. The representation of the sub-joined 'a-*chung* (“small ah”) used for indicating long Sanskrit vowels is rendered as an apostrophe (') immediately **preceding** the roman letter for the Tibetan vowel, as in the case of {PRA M'A nA} above: since in this position it can never be confused with, 'a, the 23rd Tibetan letter, in suffix position (which is typed as an apostrophe **following** the roman letter for the vowel).
2. The representation of Tibetan transliterations of, for instance, the Sanskrit vocalic *r* through a combination of a sub-joined Tibetan *ra* and a reversed *giku* vowel: thus the word for Sanskrit itself is rendered: {SAm SKRi TA}.

The ACIP Sanskrit Input Code Standard

One of the most important goals of the entire ACIP effort is to allow scholars to search a large amount of material digitally, and find immediately the items they are looking for. Until a standard such as Unicode is universally built into computers by manufacturers around the world, there will be no universal method for displaying on a computer screen, or typing in conveniently from a computer keyboard, the combinations of letters and diacritical marks that are used by Sanskritists everywhere to represent certain Devanagari letters.

In the meantime, efforts like ACIP are forced to develop interim standards that will allow users to search Sanskrit documents easily from a normal keyboard. The following is a summary of the ACIP Sanskrit Input Code Standard. We distinguish between a “transcription standard” and an “input code standard” in that ACIP *input*

code standards, both for Sanskrit and Tibetan, are meant to be the most efficient, cost-effective, non-ambiguous methods of entering massive amounts of textual data, in a format which can easily be converted by users to their favorite *transcription standard*.

Any ACIP document can, within a few minutes, be converted to any transcription standard that a user desires. Because there are so many competing transcription standards, and because specialists are often passionately attached to their favorite one, we have seen our role as simply providing a base input code that can be converted to any other standard immediately.

We hope to convert all of our Sanskrit materials gradually to this ACIP input code standard, for easy searching, but will maintain a parallel version in our database where a different standard was used in the original input or publication. In some cases, the resulting ACIP standardized version will be the only easily searchable one; whereas the original version will include more detailed information, such as accents and so on, that would interfere with most kinds of digital searching.

The International Transliteration System for Sanskrit is as follows:

a ā i ī u ū ṛ ṝ ḷ ḹ e ai o au aṃ aḥ
ka kha ga gha ṅa ca cha ja jha ṅa
ṭa ṭha ḍa ḍha ṇa ta tha da dha na
pa pha ba bha ma ya ra la va
śa ṣa sa ha

The problematic letters are those which employ diacritical marks:

ā ī ū ṛ ṝ ḷ ḹ ṃ ḥ ṅ ñ ṭ ḍ ṇ ś ṣ

In the *standardized* electronic versions of Sanskrit reference works, ACIP employs the following character sequences for these signs, respectively:

a# i# u# r% r%# l% l%# m% h% n* n~ t% d% n% sh s%

When needed, the special lingual consonant l is represented by l~.