

Title: Criteria for the encoding of script-specific dandas  
Date: March 2, 2008  
From: Unicode and US National Body  
To: SC2/WG2

In meeting 51, SC2/WG2 resolved to solicit input from national bodies on the formulation of a criteria for the encoding of script-specific dandas (resolution M51.7).

The Unicode Consortium and the US National Body recommends the following:

The recommendations for the encoding of danda characters are parallel but somewhat distinct in the case of already encoded scripts and for new scripts not yet encoded. For clarity, these recommendations are written out completely for each case.

A. For currently encoded scripts

1. If the orthographies using the existing script do not make use of dandas, do not encode any script-specific dandas for it. Example: Sinhala
2. If the orthographies using the existing script do make use of dandas, and there are already encoded script-specific dandas, use those dandas in the context of that script. List: Tibetan, Myanmar, Khmer, Balinese, Phags-pa, Lepcha, Ol-Chiki, Saurashtra, Kayah Li, Cham
3. If the orthographies using the existing script do make use of dandas, and there are no script-specific dandas already encoded, then a clear determination should be made between one of several possible alternatives. And to ensure the stable representation of text, that determination, once made, should not be reversed. The options are:
  - a. Specify the use of particular, already-encoded dandas from another block in the standard.
  - b. Encode new, script-specific dandas for use with the existing script.

c. If and only if it can be demonstrated that orthographies using the existing script have a plain text contrastive use between two types of dandas, a combination of option a) and option b), to represent the distinction.

To change existing practice, there should be demonstrable evidence that there is a need to change. Once the final determination is made, the recommendations should be clearly documented and not changed in the future.

Existing practice is as follows:

a. Use of already-encoded dandas from another block:

1735/1736: Tagalog, Buhid, Tagbanwa; and Hanunoo itself.

0964/0965: Bengali, Gurmukhi, Gujarati, Oriya, Syloti Nagri, Tamil, Telugu, Kannada, Malayalam; and Devanagari itself.

Existing practice notwithstanding, the existence of the use of dandas in orthographies for one of these scripts **might** be taken as supporting a determination 3b) for that script, i.e., encoding of script-specific dandas for the script, if evidence of formal difference is persuasive.

Note that use of dandas is not usual for South Indian scripts (Tamil, Telugu, Kannada, Malayalam), but is seen for Sanskrit texts rendered in those scripts (and Tamil Grantha).

Note that contrastive use of dandas is reported for the Bengali script, which should be taken into account for determining whether a 3c) decision is appropriate for that particular script.

B. For proposed new encoding of formerly unencoded scripts

1. If the orthographies using the proposed script do not make use of dandas, do not encode any script-specific dandas for it.

2. If the orthographies using the proposed script do make use of dandas, then a clear determination should be made between one of several possible alternatives. And to ensure the stable representation of text, that determination, once made, should not be reversed. The options are:

a. Specify the use of particular, already-encoded dandas from another block in the standard.

b. Encode new, script-specific dandas for use with the proposed script.

c. If and only if it can be demonstrated that orthographies using the proposed script have a plain text contrastive use between two types of dandas, a combination of option a) and option b), to represent the distinction.

The existence of the use of dandas in orthographies for a script proposed for encoding is generally taken as sufficient to justify a determination 3b), i.e., encoding of script-specific dandas for the script. However, there may be considerations that would favor determination 3a) instead. In any case, the determination **must** be made when the script is approved for encoding.