

# Comments on PRI 59: Script Specific Danda and Double Danda

Eric Muller,  
January 29, 2005

1. [Introduction](#)
2. [Arguments in favor of encoding script specific characters](#)
3. [Arguments against](#)
4. [Proposal 1: Encoding of script specific dandas](#)
5. [Proposal 2: Clarification of the existing characters.](#)
6. [Related Document History](#)

## 1. Introduction

The subject of script specific dandas was discussed on the [indic@unicode.org](mailto:indic@unicode.org) mailing list. Because there was no clear consensus, the next two sections attempt to summarize the arguments in favor and against disunification; there is no attempt to judge the validity of arguments or rebuttals, only to capture what has been said in the email discussion as accurately as possible. The two sections after that are alternate proposals.

The following individuals have participated in the discussion on the [indic@unicode.org](mailto:indic@unicode.org) list: Omi Azad, Vinod Balakrishnan, Stefan Baums, Peri Bhaskararao, Varghese Chacko, Gihan Dias, Ketaki Kushari Dyson, Micheal Everson, Soleiman Karim, Nishad Kaypally, Jonathan Kew, Antoine Leca, Rick McGowan, Mike Meir, Eric Muller, Paul Nelson, Mahesh T. Pai, Hariram Pansari, Dr. U. B. Pavanaja, Rajkumar S., Deepayan Sarkar, Rajeev J. Sebastian, Gautam Sengupta, Sukhjinder Sidhu, Steve Smith, Kevin Sooryan, Sinnathurai Srivas, K. G. Sulochana, Owen Taylor, Anirban Udr, Uma Umamaheswaran, Ken Whistler.

## 2. Arguments in favor of encoding script specific characters

F1: this pattern has already been followed for other scripts, such as Myanmar and Hanunoo. which have their own danda. It has also been followed in the "core" Indic scripts for characters such as U+0950 ॐ DEVANAGARI OM / U+0AD0 ૐ GUJARATI OM.

Rebuttal: Actually, the dandas encoded in the Hanunoo block are not specific to that script, but are shared by the Philippine scripts: they appear under the header "Generic Punctuation for Philippine scripts", and have generic names: U+1735 PHILIPPINE SINGLE PUNCTUATION and U+1736 PHILIPPINE DOUBLE PUNCTUATION; the names are not HANUNOO ... So the pattern already in place is one of sharing among closely related scripts. In fact, if separate characters are encoded for the Indic scripts, then it would be awkward to not also encode separate characters for the Philipinne scripts.

F2: the typographic behaviour of the danda and the double danda changes across the scripts. Not only are the shapes somewhat different, but the positioning is also different.

Rebuttal: in a font that supports a single script, this is not a problem at all: whether the glyphs are mapped from the Devanagari block or the Bengali block has no impact. Font technologies which support multiple scripts also support some mechanism by which the same code point can result in different glyphs, possibly with different layout behavior.

F3: in multi-script text, handling an occurrence of a danda between characters of different scripts (not an uncommon occurrence) requires some care, so that it is rendered using the same font as the text it belongs to. This is more difficult than for most other punctuation, because the dandas are not just at the end of a

text, they are also used to bracket a text (think about the consequences of unified quotation marks).

Rebuttal: this is not that difficult: the danda goes with the text that is before it. And anyway, the problem is not specific to the danda, it also applies to other punctuation marks (comma, period, etc), so a solution, which is applicable to the danda, has to exist anyway.

F4: users are often confused when they don't see the danda and double danda in the code chart for the Bengali (or Gurmukhi, ...) block.

Rebuttal: for most users, the code charts are not the first tool they do or should use. The layout of their keyboard is much more relevant to them, and it will show the danda and the double danda. We can assume that users who do need to dig in the code charts are also willing to read, e.g., chapter 9.

F5: users can be misled when they look at the Devanagari code chart, and see "DEVEVANAGARI DANDA"; they can easily interpret the name as restricting the use of the character to Devanagari.

Rebuttal: this can easily be addressed by an annotation of the character, right next to the name, such as "not script specific despite its name, also use with the other Indic scripts.", as well as text in chapter 9.

F6: Unlike the Western punctuation marks (coma, period, even space) which are clearly borrowed across writing systems, there is no reason to believe that the Devanagari dandas have been borrowed by the other Indic scripts, though they undoubtedly originate in a common ancestor.

### 3. Arguments against

A1: this pattern (shared characters) is already being followed for other punctuation characters which are used across the scripts: e.g. period, comma, question mark,

A2: the shared nature of U+0964 and U+0965 has already been recognized, in data and in implementations. Thus, encoding script specific characters would amount to a disunification, will the problems this usually causes. Even if there is today a small amount of data that would suffer from disunification, because of fairly long time it takes for the standard process, the amount of data with problematic representation would be significant by the time the new characters can be used.

Rebuttal: Unicode will be used for centuries. Ten years of Devanagari-encoded dandas in Oriya text is of little consequence. The disunification costs for this fix are insignificant compared to the benefits.

### 4. Proposal 1: Encoding of script specific dandas

The proposal is to encode the following characters ("P+" is used instead of "U+" to emphasize that those characters are proposed).

- P+09E4 BENGALI SIGN PURNACCHED
- P+09E5 BENGALI SIGN DEERGH VIRAM
- P+0A64 GURMUKHI SIGN PURNA VIRAM
- P+0A65 GURMUKHI SIGN DEERGH VIRAM
- P+0AE4 GUJARATI SIGN PURNA VIRAM
- P+0AE5 GUJARATI SIGN DEERGH VIRAM
- P+0B64 ORIYA SIGN PURNVIRAM
- P+0B65 ORIYA SIGN DEERGH VIRAM
- P+0C64 TELUGU SIGN PURN VIRAMA
- P+0C65 TELUGU SIGN DEERGH VIRAMA

The scripts covered are those for which TDIL proposed the encoding the dandas, and does not include Tamil, Kannada, Malayalam nor Sinhala.

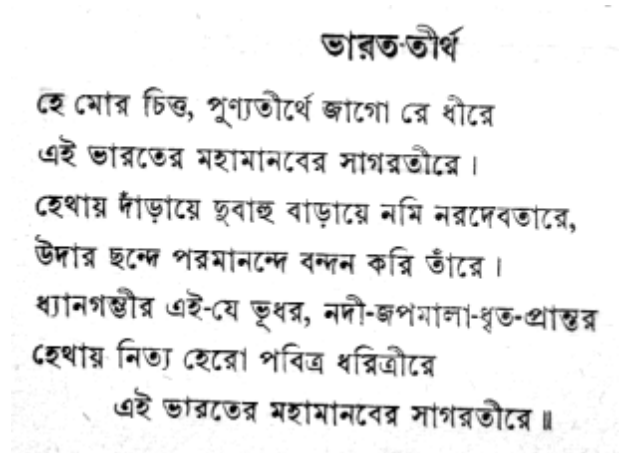
The names are those proposed by the TDIL. Alternatively, the model currently in place for Devanagari could be followed: "<script> DANDA" and "<script> DOUBLE DANDA", with annotations using the names above.

The properties of those characters would similar to those of the Devanagari equivalents:

- general category: Po
- combining class: 0
- no decomposition
- no numeric value
- bidi category: L, not mirrored
- no case conversion
- joining type: U
- east asian width: N
- linebreaking : AL
- script: Beng, Guru, Gujr, Orya or Telu, as appropriate

In addition, the script of U+0964 and U+0965 should be changed from Zyyy to Deva.

Here are examples of both dandas in Bengali, from *Learn Bengali yourself*, Bidhu Bhusan Das Gupta, Calcutta: Das Gupta Prakashan, 3rd edition, page 147. The danda occurs twice (vertical line at the end of the second and fourth line) and the double danda occurs at the end of the last line.



Here are examples of the Gurmukhi danda, from a newspaper, reprinted in *Writing Systems of the World*, Akira Nakahishi, Rutland: Charles E. Tuttle Company, 1980. The danda is the vertical line which occurs, among other positions, at the very end of the text.

ਭਾਈ ਵੀਰ ਸਾਹਿਬ ਦੁਪਹਿਰ ਮੁਖ ਮੰਤਰੀ  
ਸ: ਬਾਬਲ ਅਠਾ ਕਰਨਗੇ । ਸ: ਸੇਖੰ  
1971 ਵਿਚ ਕਸ਼ਮੀਰ ਸੀਮਾ ਤੇ ਜੰਗ ਚ  
ਸ਼ਹੀਦ ਹੋ ਗਏ ਸਨ । ਰਾਸਟਰਪਤੀ  
ਕੁਰਨਾ ਨੂੰ ਪਰਮਵੀਰ ਚਕਰ ਦਿਤਾ ਸੀ।  
ਇਸ ਦਿਨ ਦੇ ਹੋ ਰਹੇ ਸਮਾਗਮ ਵਿਚ  
ਸ਼ਹੀਦ ਦੀ ਮਾਤਾ ਤੇ ਕੁਸ ਦੀ ਧਰਮ  
ਪਤਨੀ ਦਾ ਸਨਮਾਨ ਕੀਤਾ ਜਾਏਗਾ ।  
ਸ਼ਹੀਦ ਦੇ ਬੁਤ ਦੁਆਲੇ ਰੰਗੀ ਦਾ  
ਮੰਚ ਬਣਾ ਦਿਤਾ ਗਿਆ ਹੈ । ਇਸ ਤੋਂ

ପୁସ୍ତକ ଉପରେ ଲିଖିତ ଶବ୍ଦ ଓ ଚିହ୍ନ  
ଫିଲାଇଟା ଗଠାମ ଗଠାଫିଲାଇଟା ହିଁ ସଫାଫି  
ନା ଚାହିଁ ।

Here are examples of the Oriya danda, from a newspaper, also reprinted in *Writing Systems of the World*. Again, the danda appears at the end of the text.

କିମ୍ପାସନ, ତା ୮ । ୧-ଅନୁକ୍ରମ ଦେଖିଲେ ଧନୀ ଦେଖିଲେ  
ଠାରୁ ଏବେ ଯେତେ ସାହାଯ୍ୟ ପାଇଛନ୍ତି, ତା'ଠାରୁ ଅଧିକ ସାହାଯ୍ୟ  
ପାଇବା ଲାଗି ଅନୁକ୍ରମ ଦେଖିଲେ ଦାବା କରିଛନ୍ତି । କିନ୍ତୁ କଣେ  
ମୁଖପାତ୍ରଙ୍କଠାରୁ ସୂଚନା ମିଳେ ଯେ ଏହି ଦାବା ହାସଲ କରିବା ଲାଗି  
ଅନୁକ୍ରମ ଦେଖିଲେ ସାମର୍ଥ୍ୟ ନାହିଁ ।

## 5. Proposal 2: Clarification of the existing characters.

Put the two characters U+0964 DEVANAGARI DANDA and U+0965 DEVANAGARI DOUBLE DANDA, under a new header “Generic Punctuation for Indic scripts” (on the model of the Philippine equivalents). Furthermore, add an annotation to each character, to compensate for the “DEVANAGARI” in their names.

Add some text in chapter 9, in the introduction of the chapter and/or in sections 9.1 Devanagari, 9.2 Bengali, 9.3 Gurmukhi, 9.4 Gujarati, 9.5 Oriya and 9.7 Telugu, indicating that U+0964 and U+0965 are the appropriate characters to use for the dandas of those scripts.

## 6. Related

Note that the danda characters U+104A MYANMAR SIGN LITTLE SECTION and U+104B MYANMAR SIGN SECTION have the script Zyyy. If they are specific to Myanmar, they should probably have the script Mymr.

---

## Document History

Author: Eric Muller

Revision	Date	Comments
1	January 29, 2005	First version