

Title: Proposal to Add IDEOGRAPHIC TABOO VARIATION INDICATOR to ISO/IEC 10646
 Source: UTC
 Status: Liaison Member Contribution
 Action: For consideration by SC2/WG2
 Date: 2002-05-17

**ISO/IEC JTC 1/SC 2/WG 2
 PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
 FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646**

Please fill all the sections A, B and C below.

(Please read Principles and Procedures Document for guidelines and details before filling this form.)

See <http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html> for latest Form.

See <http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for latest Principles and Procedures document.

See <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest roadmaps.

A. Administrative


1. **Title:** *Proposal to Add IDEOGRAPHIC TABOO VARIATION INDICATOR to ISO/IEC 10646*
2. Requester's name: *Unicode Technical Committee*
3. Requester type (Member body/Liaison/Individual contribution): *Liaison*
4. Submission date: *17 May 2002*
5. Requester's reference (if applicable): *N/A*
6. (Choose one of the following:) *This is a complete proposal*

B. Technical - General

1. (Choose one of the following:)
 - b. The proposal is for addition of character(s) to an existing block:

Name of the existing block: *KANGXI RADICALS at U+2FDF*
 2. Number of characters in proposal: *1*
 3. Proposed category (see section II, Character Categories): *B.1 Specialized (Small Collections of Characters)*
 4. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000): *1*
 - Is a rationale provided for the choice? *Yes*
 - If Yes, reference: *There are no combining marks used*
 5. Is a repertoire including character names provided? *Yes*

IDEOGRAPHIC TABOO VARIATION INDICATOR

 - a. If YES, are the names in accordance with the character naming guidelines in Annex L of ISO/IEC 10646-1: 2000? *Yes*
 - b. Are the character shapes attached in a legible form suitable for review? *Yes*
- 
6. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard? *John H. Jenkins, Apple Computer, Inc.*

If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used: *N/A*
 7. References:
 - a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *Yes, below*
 - b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? *Yes, below*
 8. Special encoding issues:

Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

¹ Form number: N2352-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09)

No

9. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see <http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

C. Technical - Justification

- | | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| 1. Has this proposal for addition of character(s) been submitted before? | No |
| 2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? | Yes |
| If YES, with whom? | UTC, IRG |
| If YES, available relevant documents: | IRG N890 |
| 3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? | |
| <i>Taboo variants are rarely used in modern Chinese; the character would be of use mostly to scholars publishing transcriptions of earlier texts</i> | |
| 4. The context of use for the proposed characters (type of use; common or rare) | |
| Reference: | See below |
| 5. Are the proposed characters in current use by the user community? | No |
| 6. After giving due considerations to the principles in <i>Principles and Procedures document</i> (a WG 2 standing document) must the proposed characters be entirely in the BMP? | Yes |
| If YES, is a rationale provided? | |
| <i>It's only one character and other characters with similar function are in the BMP; it can be added with minimal disruption</i> | |
| 7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? | Yes |
| 8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? | No |
| 9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? | No |
| 10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? | No |
| 11. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)? | No |
| 12. Does the proposal contain characters with any special properties such as control function or similar semantics? | No |
| 13. Does the proposal contain any Ideographic compatibility character(s)? | No |

There is no distinction within Chinese between names and other words written with ideographs, it is entirely possible that the ideograph used for any individual's name will occur in a normal text. This can be awkward when the individual is someone of high rank or otherwise deserving of special respect; using the ideographs from their name in ordinary writing was traditionally seen as disrespectful and unacceptable. To get around this, the practice developed of using "taboo" forms (避諱) for these ideographs in normal texts. The ideographs would be deliberately distorted (usually, but not always by omitting the final stroke) to avoid writing the personal name of the individual in question. The precise form of the taboo form is not specified and not necessarily predictable.

The purpose of the IDEOGRAPHIC TABOO VARIANT INDICATOR is to mark places in a text where this has been done. An IDEOGRAPHIC TABOO VARIANT INDICATOR followed by an ideograph indicates that the ideograph was written with a taboo form in the original text. The precise nature of the taboo form is not specified (and is, indeed, irrelevant for purposes of transcribing the text).

If the font provides for this, the combination of the IDEOGRAPHIC TABOO VARIANT INDICATOR plus ideograph may be treated as a ligature and directly rendered by a taboo form glyph for the given ideograph. Otherwise, it should have a visual appearance such as specified in this proposal.

The following examples regarding the use of the taboo forms are provided by Richard Cook of UC Berkeley:

Here're 3 examples of [U+907f][U+8af1] (bihui) from p. 202 of <Song Ben Guang Yun> (SBGY) the Song Dynasty rhyming dictionary (YU Nae-wing, Chinese U. of Hong Kong, 1993):

<http://linguistics.berkeley.edu/~rscook/images/SBGY/SBGY-YN-202.jpg>

The 3 green arrows indicate taboo-deformed head entries, specifically, deformations of [U+6046], [U+63ef] and [U+7dea]. You can see the 1st 2 of these same 3 hanzi deformed also in the definitions.

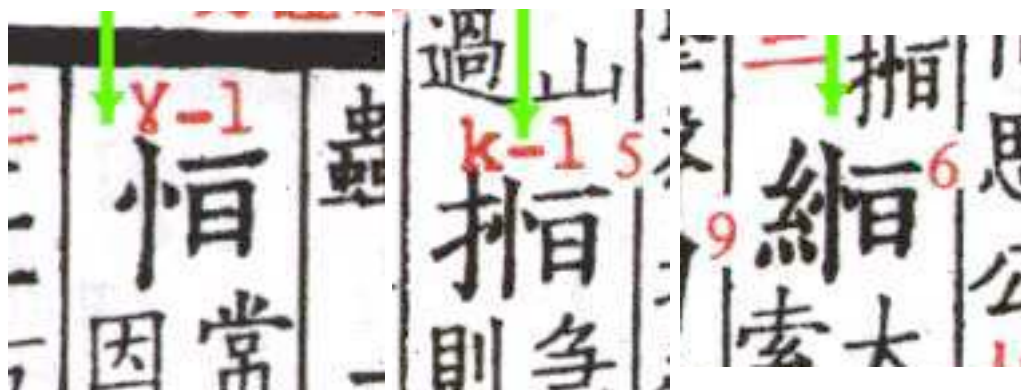
The tabbo deformation in all 3 is omission of the last heng stroke.

The "correct" (undeformed) writings of [U+63ef] and [U+7dea] can be seen in the (modern) footnotes 5 and 6 at the bottom of the page.

Note that [U+63ef] and [U+7dea] both have [U+6046] as phonetic, and so Taboo Deformation of [U+6046] spreads to all hanzi in which [U+6046] is a component. This is a general truth of TD, which we might call the "Taboo Deformation Spreading Principle" (TDSP).

Note that the TD forms of [U+6046], [U+63ef] and [U+7dea] are all non-hanzi, which is to say that omission of the last stroke does not result in confusion with preexisting non-TD hanzi. I have yet to see a case in which regular omission of the final stroke would result in confusion, and I have also not seen a case in which a TD (or TDSP) hanzi omits a stroke other than the final stroke. But I may just not have looked hard enough yet.

Here are the three taboo forms from Richard's scan:



These are the two non-taboo forms in the footnotes:

