

Title: Proposal to add the provisional kTabooVariant property to the UniHan database
Source: Individual
Author: Roy Wang
Action Required: For consideration by UTC
Date: 2026-09-09

Taboo variants (避讳字) were used in ancient China, which followed a core principle that involved indirectly writing an important historical figure's name whereby the final stroke of a Han ideograph was generally omitted. We therefore propose adding to the UniHan database a new provisional property to explicitly identify such Han ideographs.

For China, the scope of this property should be restricted to taboo variants that were used to refer to important historical figures who were either emperors or were posthumously given the title of emperor. The scope may vary for other regions that use Han ideographs.

The table below shows a prototypical taboo variant case, which was used to avoid writing 玄, which is the name of the emperor of the Qing Dynasty (清朝): 玄燁 (玄烨 in Simplified Chinese).

Standard Form	Taboo Variant
U+7384 玄	U+248E5 𠂇

Taboo variants are sometimes created by other means, such as by changing to a homophone (同音字), by changing to a synonym (同义字), or by omitting a Han ideograph altogether. We recommend that the kTabooVariant property be used to identify such Han ideographs when it resulted in a completely new Han ideograph, with the most common means being to omit the final stroke.

In addition, taboo variants may also be used as components, include substituted components, or include additional components.

The table below shows another prototypical taboo variant case, which was used to avoid writing 麻 as a component of 歷.

Case	Standard Form	Taboo Variant
Normal	U+20AB1 𪛗	U+53AF 𪛗
Used as a component	U+53A4 𪛖	U+20A75 𪛖
Substituted component	U+66C6 曆	U+66A6 曆
Substituted component	U+6B77 𪛘	U+6B74 𪛘

Currently, the IRG unifies taboo variants with their standard forms, but there are still early disunified forms that have been encoded, along with special cases. Identifying these cases through the use of a new Unihan database property is useful.

Taboo variants must correspond to at least one Han ideograph in an important historical figure's name. The last page of this document provides the proposed property data in human-readable form, along with other key information.

The table below provides the fields for this property as it would appear in UAX #38:

Field	Text
Property	kTabooVariant
Status	Provisional
Category	Variants
Introduced	<i>TBD</i>
Delimiter	space
Syntax	U\[+23]?[0-9A-F]{4}
Description	One or more Unicode values for the taboo variants (避讳字) of this ideograph. This property is fully documented in Unicode Technical Note #63, "Unihan Database Property 'kTabooVariant'" [UTN63]

If accepted, and as implied by the Description in the table above, a new Unicode Technical Note would be published that documents this property, which would include additional information that is provided at the end of this document.

The format of the kTabooVariant property is shown below, which matches the format of other Variant properties.

Standard Form UCS	kTabooVariant	Taboo Form UCS
-------------------	---------------	----------------

Example: U+4E16 kTabooVariant U+2000D U+200D2

Note that some Han ideographs may be associated with more than one taboo variant.

Data File

The kTabooVariant.txt data file, which is a PDF attachment, provides everything that is necessary for adding this provisional property and its property data to the UniHan Database, and for adding a full description of the property to UAX #38, Unicode Han Database (UniHan). The file includes at the end human-readable mappings between the two forms.

Acknowledgements

We thank Dr. Ken Lunde, Eiso Chan, and Tao Yang for reviewing this document and giving suggestions.

End of Document

Proposed kTabooVariant property data in human-readable form

Name of Important Historical Figure: Separate the names with a semicolon when there are more than two.

Period and Region: Separate the ‘period and region’ with a semicolon when there are more than two, separate the contents (Period, Region) inside with commas.

Standard Form	Taboo Variant	Name of Important Historical Figure	Period and Region
U+4E16 世	U+2000D 卅	李世民	Tang Dynasty (唐朝), China
U+4E16 世	U+200D2 卅	李世民	Tang Dynasty (唐朝), China
U+53A4 麻	U+20A75 麻	愛新覺羅·弘曆	Qing Dynasty (清朝), China
U+5BE7 寧	U+21A6C 寧	愛新覺羅·旻寧	Qing Dynasty (清朝), China
U+5F18 弘	U+2239E 弘	愛新覺羅·弘曆	Qing Dynasty (清朝), China
U+6052 恒	U+22606 恒	趙恒;李恒	Song Dynasty (宋朝), China; Tang Dynasty (唐朝), China
U+65FB 旻	U+23150 旻	愛新覺羅·旻寧	Qing Dynasty (清朝), China
U+66C6 曆	U+66A6 曆	愛新覺羅·弘曆	Qing Dynasty (清朝), China
U+6B77 歷	U+6B74 歷	愛新覺羅·弘曆	Qing Dynasty (清朝), China
U+6DF3 淳	U+6DA5 淳	愛新覺羅·載淳	Qing Dynasty (清朝), China
U+71C1 燁	U+2435E 燁	愛新覺羅·玄燁	Qing Dynasty (清朝), China
U+7217 燁	U+2447B 燁	愛新覺羅·玄燁	Qing Dynasty (清朝), China
U+7384 玄	U+248E5 玄	愛新覺羅·玄燁	Qing Dynasty (清朝), China
U+7430 琰	U+2494E 琰	愛新覺羅·顥琰	Qing Dynasty (清朝), China
U+799B 禎	U+256C4 禎	愛新覺羅·胤禎	Qing Dynasty (清朝), China
U+80E4 胤	U+2664C 胤	愛新覺羅·胤禎	Qing Dynasty (清朝), China
U+80E4 胤	U+2664D 胤	愛新覺羅·胤禎	Qing Dynasty (清朝), China
U+8669 虢	U+2E4ED 虢	李虎	Tang Dynasty (唐朝), China

U+20AB1 𪛗	U+53AF 𪛗	愛新覺羅・弘曆	Qing Dynasty (清朝), China
-----------	----------	---------	--------------------------

The following entries are under consideration pending more investigation:

Standard Form	Taboo Variant	Name of Important Historical Figure	Period and Region
U+65E6 旦	U+2CEA3 𠂔	李旦	Tang Dynasty (唐朝), China
U+84C4 蓄	U+26BFE 茗	愛新覺羅・玄燁	Qing Dynasty (清朝), China