

JTC1/SC2/WG2 N5265

**Universal Multiple-Octet Coded Character Set  
International Organization for Standardization  
Organisation Internationale de Normalisation  
Международная организация по стандартизации**

**Doc Type:** Working Group Document  
**Title:** Feedback on the Proposal to Revise and Add Manchu Character U+1879 (L2/24-025)  
**Source:** China  
**Status:** Member Body Contribution  
**Action:** For consideration by JTC1/SC2/WG2  
**Date:** 2024-06-04

The currently effective encoding standards for Manchu are, at the moment, insufficiently capable of fully meeting the demands and requirements arising during the process of digitizing historical Manchu texts and documents. The Manchu script is a phonetic alphabet originally created by Jurchen rulers and scholars during the late sixteenth century. The script underwent several distinct steps during its development, beginning first with Old Manchu (without dots and circles), then entering a temporary Transitional Manchu stage (with some dots and circles), before ending at New Manchu (with dots and circles). Optimization and improvements to the Manchu script were for the most part achieved by the addition of dots or circles to vowels or syllables, which then clarified how these letters were to be read. Especially throughout the course of the Transitional phase, we can point to the occasional existence of various nonstandard habits in the personal handwriting of certain individuals, meaning that the Transitional Manchu period saw the creation of a number of nonstandard variant vowel and syllable letters.

### 1. Analysis of the Causes of the Issue

There could be any number of specific reasons explaining how the ᡩ glyph in question here came into use. Among the most probable of these possible explanations would involve situations where some arbitrary, personal-level decision was made when writing out Old Manchu letters (which did not have dots or circles), to start marking the dots and circles that were being developed during the reforms towards New Manchu (which did indicate dots and circles). Such habits could have been formed among individuals who were first trained with familiarity in Old Manchu, but then were later charged to master the rules of New Manchu. When writing out a word by hand, after having written a particular letter according to the conventions of Old Manchu- whether for aesthetic purpose or as a habitual action picked up from learning New Manchu- a New Manchu-style dot could have been marked to the right of the letter. For readers accustomed to both Old Manchu and New Manchu, an extra dot placed on the right-hand side of a letter would not create a “mistake” situation that would render

comprehension confusing or difficult. Instead, such a dot would simply provide additional emphasis on exactly which vowel was intended.

Another possibility is that, when the Manchu script was being developed from its foundation in the Mongolian script, all of the Mongolian script's vowel letters (a total of 7) were carried over for continued use. Looking at the writing system itself, this vowel letter may indeed exist, but over the course of actual practical usage, whether due to the particular characteristics of the Manchu vowel inventory, or due to whatever changes in pronunciation that may have occurred, it was not necessary to make use of all 7 of these vowel letters. As a result, during the transition to New Manchu, vowel letters that were no longer needed were then dropped. It may then be the case that this glyph in question was a vowel letter that had been discarded in this fashion.

## **2. On the Proposed Solution**

Regardless of the factors that may have caused this issue, there are essentially two solutions that could be proposed for dealing with this new letter: either add a new FVS, or add a new character (provisionally designated as U+1879). It is imperative that we be exceedingly cautious when considering a proposal to add a new vowel character. Adding a new vowel letter to accommodate a nonstandard, habitual handwriting practice (in most cases where this glyph occurs, it basically appears as a handwriting mistake) would not be reasonably justifiable from the perspective of academic linguistics. It would also not be in conformance with standardized Manchu orthography. Effectively, doing this would be equivalent to engaging in script reform. Adding a new FVS, on the other hand, would be a more moderate and reasonable approach. This approach would be preferable because when submitting a proposal, we should be thinking about the possibility that future studies of historical texts and documents may uncover even greater numbers of erroneous handwriting habits and practices. As such, adoption of a proposal to add a new FVS may make it easier to deal with any unpredictable situations that may arise in the future.

## **3. Suggestions and Recommendations**

The phenomenon described in the proposal does indeed exist. However, giving consideration to the present overall situation, at this stage it is not suggested that we rush to adopt this or any other proposal to solve the issue in question. Instead, the suggestion is that we first take further steps towards conducting fuller investigations and designing more complete demonstrations of proof; and then determine which specific revision proposals should be adopted. For example, we should be performing more extensive searches of the historical texts we have access to for any other similar issues that may also need to be addressed. We should also be looking into other relatively feasible revision plans, and then only after further analysis and demonstrative proof work is completed should we make a final decision on what solution proposals to adopt. The reasoning for our position is as follows:

(1) Adding a new vowel letter to the Manchu script in order to accommodate an erroneous handwriting habit would violate the integrity of Manchu's orthographic system. Compared to such an extreme measure, adding a new FVS would likely be a

safer and more moderate solution.

(2) The First Historical Archives of China (FHAC) currently houses a collection of over two million Manchu texts and documents. Beginning in November 2018, and in accordance with the currently effective Unicode standards, the FHAC has developed and launched a suite of related information technology and software systems, including OCR software for use with Manchu document image files, Manchu text input software, an electronic Manchu lexical corpus, software for converting between Romanized Manchu and the Standard Manchu script, as well as a Manchu archive data management program. These software programs have been used to create and build up a large electronic database of the historical Manchu documents held in the archives. As of October 2023, approximately 140,000 Manchu documents have been digitized and stored in the database system in their entirety, comprising some 1.88 million individual document image files, and a current total of over 110 million Manchu words. In order to uncover any other phenomena similar to the vowel letter question at hand, we have sent a research team to the FHAC to comb through the relevant materials and data, with the aim of identifying any other unusual glyphs or letters that the currently effective standards for encoding Manchu are unable to resolve. Although it will not be possible to discover every single one of these unusual or idiosyncratic letters at once, we will be putting every effort into finding as many problematic letters as we can, and then, once this work is completed, submitting a comprehensive solution proposal.

(3) Were the proposal to revise the standards and add the new vowel character (U+1879) to be adopted, we would then also have to consider the problem of transliterating Manchu into romanization. At present, the FHAC, along with several other institutions, have all adopted the standards stipulated in 《满文档案著录名词与术语汉译规则》 "Guidelines for Translating Manchu Archive Catalog Terms and Special Terminology into Chinese" (DA/T30-2019) when constructing their Manchu text databases. Since the romanization system for Manchu is based on phonetic letters and not on individual glyphs, the proposal to add the new vowel character (U+1879) then isn't simply unjustifiable with reference to Manchu orthography, it would also create unnecessary and otherwise avoidable problems when transliterating Manchu into romanization.

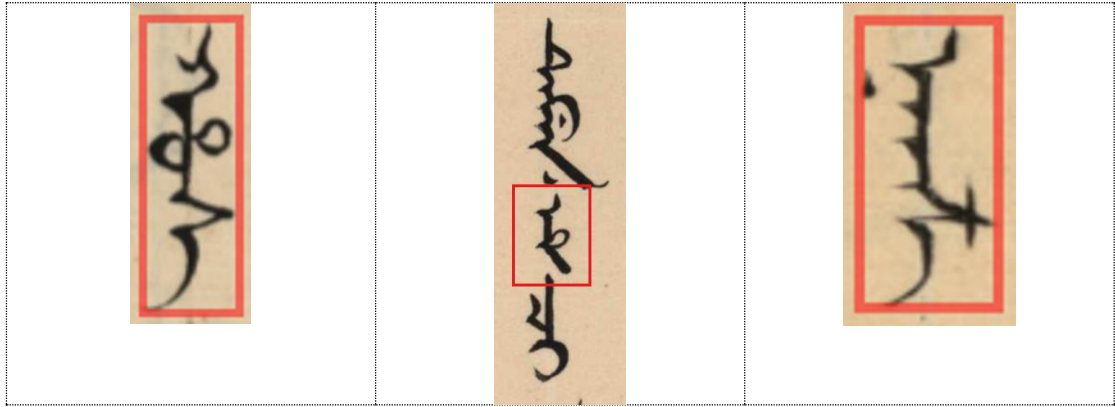
(4) The currently effective national standards for the encoding of Manchu, 《GB/T36645-2018 信息技术 满文名义字符、变形显现字符和控制字符使用规则》 "Information Technology—Manchu Nominal Characters, Presentation Characters and Use Rules of Controlling Characters" have been in place for many years now. Several electronic Manchu database systems, including the FHAC's full-text Manchu archives database, are all making use of these current standards. In order to minimize the impact that a revision of the Unicode standards may have on the work being done by these organizations, the suggestion is that comprehensive surveys and investigations first be done on the corporations and other institutions applying the national standards towards product development before any revisions to the standards are enacted. We have already dispatched investigative survey teams to these corporations and institutions, and

preliminary investigative work prior to potential revisions to the encoding standards for the Manchu script have already underway. As such, we do not suggest that any solution proposals be adopted for the time being. Instead, our suggestion is to wait until all of our research and investigations have been completed, after which point we can begin discussions on all the other potential encoding issues that may arise, and then discuss which solution plans may be most appropriate.

#### 4. Other Similar Issues

After initiating our preliminary investigative work prior to revisions to the encoding standards for Manchu, at the time of writing this feedback response, we have already found a few other instances of handwritten phenomena the status of which calls for further discussion with respect to the rules of Manchu orthography and the digitization of the historical Manchu archives. These phenomena are not of the exact same type as the glyph currently under discussion (ᡩᡠ, provisionally designated as U+1879), but should be considered together with it when revising and perfecting the encoding standards for Manchu. Further, these investigations and searches of the archives are still ongoing.





Given all of the above reasons, our suggestion is to not rush into a revision of Manchu's encoding standards, and in particular to not be hasty with adopting the proposal to add this new vowel character. Once our research investigations have been completed and we are able to present a list of all the possible other encoding problems that should be addressed, we hope at that point to examine all these issues together and then discuss what the best plan of action should be.

End.