

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

Doc Type: Working Group Document
Title: Preliminary feedback on Khitan Large Script encoding (WG2 N5319)
Source: Eiso Chan (陈永聪, Culture and Art Publishing House)
Jerry You (游程宇, CCAMC)
Status: Individual Contribution
Action: For consideration by WG2 and UTC
Date: 2025-06-16

China submitted WG2 N5319 to encode Khitan Large Script (KLS) in UCS/Unicode. Andrew West and Viacheslav Zaytsev submitted their reviewing document as WG2 N5323. They raised many meaningful issues. This document could be treated as the supplement.

The following documents are related to KLS.

WG2 Doc №	UTC Doc №	Title	Source
N4559	L2/14-233	Preliminary Review of Proposal on Encoding Khitan Large Script in UCS (N4631)	Andrew West, Viacheslav Zaytsev
N4631	L2/14-234	Proposal on Encoding Khitan Large Script in UCS, <i>Appendix: Part 1, 2, 3, 4, 5</i>	China
N4642	L2/14-246	Ad Hoc Reports for Tangut and Khitan Large Script	Deborah Anderson
N5319		Preliminary Proposal on the Khitan Large Script	Sun Bojun, et al. (China)
N5323	L2/25-166	Review of Preliminary Proposal on the Khitan Large Script (WG2 N5319)	Andrew West, Viacheslav Zaytsev

When we discussed the encoding model of Khitan Small Script (KSS), Suzuki-San (Toshiya) once raised if we need to unify some KSS characters to CJKUI for the same glyphs in his document [WG2 N3934](#)(= [L2/10-453](#)). However, KSS is one kind of phonetic not ideographic script, so it is not better to do the unification. On the other hand, in WG2 N4642, the ad-hoc group got the agreements to unify some KLS characters to CJKUI.

After the preliminary review on WG2 N5319, we found the following types on the possible unification with CJKUI¹.

Type	Form	Reading	Definition	Example
A	✓	✓	✓	#16 vs U+5929 天
B	✓		✓	#1 vs U+4E00 一
C	✓			#23 vs U+7396 玖 #61 vs VN-F0165 𐰚𐰢 ² #66 vs WS2024-00379 𐰚一𐰢 ³
D		✓		#1599 vs U+342C 𐰚

For Type A, the character form, pronunciation and meaning of the example are totally the same as or matches the corresponding Hanzi. We can also treat them as the real Hanzi, that the situation is similar to Kanji (on'yomi) used in the modern Japanese writing system.

¹In the following table, “**Form**” means the character form is the same as one Hanzi, “**Reading**” means the pronunciation matches one Hanzi, “**Definition**” means the meaning is the same as or close to one Hanzi. The examples are just mentioned based on the current proposal.

²This character could be unified to 州 (U+5DDE), but there is not UCV now.

³IRG decided to not unify this character to 𐰚 (U+342C).

For Type B, the character form and meaning of the example are the same as the corresponding Hanzi, but the pronunciation is not relative properly, that the situation is similar to Kanji (kunyomi) used in the modern Japanese writing system.

For Type C, there is only the character form of the example which is the same as the corresponding Hanzi, that means the pronunciation and meaning are not relative.

For Type D, there is only the pronunciation of the example which matches the corresponding Hanzi. This situation is similar to the case of 馭 which reads “ngự” and means “horse” (not “to drive, to control”) in Vietnamese usages, and Prof. Hồng marks this situation as A2 in his dictionary *Tự Điển Chữ Nôm Dẫn Giải*. For #1599, it is used for the Chinese loan word of 留 “to stay, to keep” based on the proposal.

If WG2 decides to unify some KLS characters to CJKUI, we suggest Type A, B, D could be included for the unification works, and the experts need to consider how to handle Type C. After unification works, maybe one or more properties could be added to [Unihan DB](#) to mark which Hanzi are also needed to be used in Khitan written system later.

(End of Document)