

JTC1/SC2/WG2 N3449
2008-09-15

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






Doc Type: Working Group Document
Title: Review of Nushu Proposal
Source: United Kingdom
Status: National Body Contribution
Action: For consideration by JTC1/SC2/WG2 and UTC
Date: 2008-09-15

1. Introduction

The UK welcomes the revised Nushu proposal by China (N3463), and the multi-column charts (N3462). However there are still some issues that still need to be addressed. In particular, there are worrying inconsistencies between N3463 and N3462. For example, N3462 Appendix B and N3463 Appendix E "Nushu basic characters and their origin verification" are both stated to be the 4th draft of February 2008 (see N3462 page 44 and N3463 page 90), but there are a significant number of differences between them, not least that the table in N3462 gives a total of 395 basic characters, whereas the table in N3463 gives a total of 398 basic characters. These discrepancies suggest that the character set is still in development and the proposal is not yet mature.



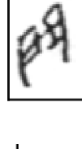

2. Some Problematic Characters

A major issue that needs to be addressed is how to deal with Nushu characters that are mentioned in the proposal, but that are not proposed for encoding.









A good example of a character whose encoding status is unclear is the character *ying* 英 that occurs in the personal name Ying Tai 英台. The images below, from pages 58 and 60 of N3463, show the phrase *tingchang Ying Tai nü* 听唱英台女 "Listen to the girl Ying Tai sing" written in three different hands. The characters for *ting* 听, *chang* 唱, *tai* 台 and *nü* 女 are respectively proposed for encoding as U+1B06C  PHENG, U+1B176  QIANG, U+1B00B  TEOA  and U+1B011  NYUA, but the highlighted character *ying* 英 is not proposed for encoding. In short the characters proposed in N3473 are insufficient to represent this sample text.






Although the Nüshu character for *ying* 英 shown above is not proposed for encoding, it is given in Appendix E of N3463 (page 86). Other Nüshu characters given in Appendix E that are not shown in the chart of proposed characters include the characters corresponding to *jiu* 酒 (page 86), *men* 門 (page 87) and *mo* 莫 (page 87).

 (英)	yn ⁴⁴	英 ⁹¹
 (酒)	tsiou ³³	酒 ⁹
 (門)	mai ⁴²	门 ²³ 闻 ⁷⁸
 (莫)	mu ³³	莫 ¹¹

When we look at the book *Nüshu Yongzi Bijiao* 女书用字比较 "A Comparison of characters used for writing Women's Script" we find that the entry for characters pronounced *nioŋ*⁴⁴ (Mandarin *ying*), corresponding to the characters 英 and 鸚櫻嬰纓, includes four distinct glyphs for *ying* 英 (circled in red, and labelled "1" to "4" in the figure below).

<i>nioŋ</i> ⁴⁴	鸚櫻英 嬰纓	 1 英 ₁	 4 英 ₄₄	 4 英 ₅	 2 英 ₁₃ 嬰 ₁	 4 英 ₄₇
			 4 嬰 ₂		 3 英 ₁	 4 嬰 ₄

Glyph forms 1  and 2  are unified (see N3463 page 46) and proposed for encoding as U+1B156, and glyph form 3 is possibly unifiable with U+1B0E1 , but glyph form 4 is neither proposed for encoding separately nor explicitly unified with U+1B156. As this "italic" form of *ying* 英 is apparently quite common, but has a completely different glyph form to the proposed U+1B156, it is not evident why it is not encoded separately. This example demonstrates the difficulty of properly reviewing the proposed character repertoire without knowing what the criteria for unification and encoding of Nüshu characters is.