

DATE: 2007 - 08 - 13

ISO/IEC JTC 1/SC 2/WG 2 Universal Multiple - Octet Coded Character Set (UCS) - ISO/IEC 10646 Secretariat: ANSI

Doc Type	Working Group Document
Title	Proposal for encoding Nüshu in the SMP of the UCS
Source	China
Action	For consideration by JTC1/SC2/WG2
Date	2007-08-13

I、Introduction

Nüshu is a kind of rural women-only Character Set, which comes down around Xiaoshui Valley in Jiangyong County, Hunan Province. It is widely used for narrative and private prosecution among local women. Customarily Nüshu characters appear to be italics ,with a rough shape as the Chinese character "多". In this sense, Nüshu is defined as a variant form of Chinese characters. A character of Nüshu represents an independent syllable that maybe records a group of homophone or a group of characters with similar pronunciation in their local dialect. Based on survey and statistics of the works of every Nüshu elderly woman, we found that everyone just used about four or five hundred characters to record a complete local dialect. That is, there are two separate writing systems in Jiangyong: one is Chinese characters used by men while the other is Nüshu which is written to communicate among local women.

Nüshu is a kind of one and only female characters in today's world, and is regarded as the miracle in the history of civilization with great linguistic、anthropological and other academic values. Now Nüshu has successfully been a member of the first national culture heritage list. The United States Library of Congress has already decided to collect related research results. We will introduce Nüshu in detail from the followings:

1. Nüshu is not a secret language, but free, open, natural and positive. The reason why its use is limited among the women is that local male society has paid no attention to the activity of women such as reading and singing the songs written in the paper and fans (that is typical Nüshu culture activity.)
2. Nüshu records the local dialect as a syllabic words system. A character signs a group of homophones and characters with similar pronunciation. So less than four hundred basic characters are enough to record and represent their local dialect of Xiaoshui in Jiangyong which is the area of Chenguan local dialect (according to Huang Xuezhen, 2003) . Every Nüshu elder woman just used no more than 400 basic characters (about from four hundred to five hundred with allograph included) to record more than 400 syllables (about 1200 syllables with tones included). Its most typical feature is to record many words with few characters. So it is a false view to create and add certain characters arbitrarily just because of limited existing characters.
3. Nüshu is a variant form in the development process of Chinese character. Specially, it

is a variant form of Kaishu (a regular Chinese script). After the statistics and research of the basic characters from traditional passed down works of anonymous writers, we have come to the conclusion that Nüshu originates from Chinese characters and there are three ways to change into Nüshu: basically direct borrowing from Chinese, modified variant and derivative new characters from common components.

4. Seven-character songbook is the main form besides few five-character poems. Autobiographies, letters, folk songs, monody, narration and translation constitute its most contents, which reflect the regional historical culture and civil life. The study of these songbooks has widely academic value such as history, nation, folk-custom, literature and linguistics.
5. Nüshu-writers usually write on cloth covers, scrip, fans, cloth-handkerchiefs and flowered belt.

Nüshu came to the world and developed with the development of history. But now the aboriginality of Nüshu has disappeared from the stage of the history. The local government hasn't cared about it until the end of the twentieth century. So most works were scattered and lost at home and abroad. Since the beginning of the twenty-first century, endangered Nüshu has suddenly become hot. Inauthentic characters, books, villages and some nonsensical sayings concealed the real Nüshu. So it is time for us to prevent the abnormal phenomena and keep it from distortion.

II、Process description

1. the collection and collation of Nüshu characters

< *Chinese Nüshu works collection* > which is published by *China Publishing House* at 2005 included more than 90% original ecological data. We identified every Nüshu character of the book by finding its corresponding Chinese character. Besides that, based on the *Chengguan accent* which is investigated and published as <*Jiangyong dialect research*> by *Huang Xuezhen, 1993*, we make out the pronunciation of each Nüshu character and establish a character table respectively in the order of strokes and phonetics.

2. the judgment and collation of Nüshu basic and variant characters

Nüshu variant characters differ from the concept of Chinese allograph characters that usually refers to the characters with the same phonetic and meaning but with different graphs. Nüshu originates from Chinese characters, but differs a lot. A Nüshu character represents a syllable which records a group of local homophone. That is to say, the meaning of a Nüshu character is uncertain and needs inferring by the context. The difference between basic characters and allograph is not rooted by different languages or dialects, but by personal writing habit which causes more than one graph for one character.

We divided basic characters and allograph on the basis of *character-cell theory* as follows:

- a. with the same origin
- b. with the same configuration methods
- c. with the same pronunciation
- d. with no distinction in the meaning.(substitutable)

Characters complied with the above features but with different graphs are regarded as allographs. The character with the highest frequency is called basic representative and others

as allograph. The final result of the study and the statistics can be referred to ‘*Nüshu basic characters and their origin verification*’ (published by *Nüshu character-comparison, 2006*) .

3. choosing the related allograph.

Nüshu allographs are classified four as following:

- a characters with different origin are regarded as different ones and given respective code.

Some Nüshu characters were borrowed from Chinese ones with different source and formed different-origin ones. But they have no distinction in their meaning and are substitutable. We still code these kinds of characters.

E.g. 我  / ie  ( (

mai  (

b. characters with tiny stroke distinction (such as the distinction between linked stroke and broken one) are seen as one graph, and the one that has the highest frequency is representative and coded.

E.g.  : nian ⁴²娘¹⁵⁸⁶⁻³³⁴⁻¹⁷,  : k'u ⁵刻¹³²⁻³, we coded the first graph of every group.

- c. characters with some distinction of minor part such as its position, quantity and shape but the same main structure and configuration contour are regarded different, and coded separately. Although they have the same meaning and are substitutable, they have their own codes.

E.g.  y ⁴²如⁵⁹⁰⁻⁸⁹⁻¹²;  sic ³⁵写¹²¹⁻⁶¹⁻⁷⁻²⁸

- d. sometimes characters with similar component and configuration are spelled confusedly. They are coded as allograph. E.g.  tou ³³度³⁸⁻³⁷, tu ³³独⁷⁶⁻⁷¹.

Supplement:

Some seem to come from changed Chinese character obviously and not in accordance with Nüshu spelling rules. They are later-created subjectively and not coded. E.g.  , .

4. order rules

The basic characters lie in the front and allograph following. The basic and allograph are in good order by the amount of strokes from small to multi separately. 449 characters are coded in all.

III、Related explanation

1. the font

There are no unified standards and rules for Nüshu and many graphs seem vague. So after we picked the chosen characters, we invited the local calligrapher to write the picked Nüshu characters one by one. Then we scanned the handwriting, processed it by Photoshop and use FONTCREATEOE professional tools to make a preliminary font.

the font of Nüshu characters is greatly improved by *Weifang Beida Jade Bird Huaguang Imagesetter Co., Ltd* which is famous and professional in China. And the codes of characters begin from 1B000.

2. the relationship between Nüshu and Chinese characters.

Customarily Nüshu characters appear to be italic with a rough shape as similar as the Chinese character "多". Its spelling order is from top-right to bottom-left. There are three kinds of situation for its origin from Chinese character.

- a) Borrow the whole character body, and change it from squared-shape to italic. There

are about 100. e.g. 𠂔 (甲) 𠂓 (如) 𠂔 (主) 𠂔 (良) 𠂔 (内). some of them

are reversed from Chinese character and italic. such as : 𠂔 (义) 𠂔 (己) 𠂔 (春)

𠂔 (吞) 𠂔 (天), and some are changed italic from simplified character, such as 𠂔

(后) 𠂔 (个) 𠂔 (来) 𠂔 (归) 𠂔 (听) 𠂔 (双).

- b) A little change occurs to the main body of Chinese character, but it is obvious to

keep its Chinese shape. e.g. 如 (借) 𠂔 (白) 𠂔 (苦) 𠂔 (母) 𠂔 (昔) 𠂔 (左)

Some changes look more obvious than the above, but still kept Chinese trail.

e.g. 𠂔 (羊) 𠂔 (光) 𠂔 (焦) 𠂔 (明) 𠂔 (去) 𠂔 (别) 𠂔 (心)

- c) Get some components as sound signs and derive some new characters which form a character-cluster. Some clusters contain more than 10 characters.

e.g. 𠂔 (声) -> 𠂔、𠂔、𠂔、𠂔

3. Character names

Nüshu has a close relationship with Chinese character. So we can use the Chinese naming rule to name Nüshu character. That is to say, the name of Nüshu character is its corresponding code sequence number: Nüshu-1BXXX.

Each of the symbols and marks in the proposed Nüshu character set ought to be of one-one correspondence with one code.

IV、Prospects

1. Nüshu is a kind of mature character symbol system. As a unique character type, Nüshu is a new category to various characters in the world, a new kind of datum to the research of linguistics, and some new subjects are put forward such as how to translate Chinese works into Nüshu.

2. There are large amount of literatures written in Nüshu, most of which record some important historic events of the time such as Taiping rebellion, some crime of Japan when invading into china, land reforming movement in China. Besides that, some folkways and anecdotes are included. All of them are valuable historical data.

3. Nüshu is a kind of one and only female characters in today's world, which is of great impetus to the research of female culture both in China and in the world.

4. The important function of Nüshu is to confide, hearken and communicate in the groups, which is made good use of to modern mind cure. It has a universal meaning in the world.

Should this proposed set be accepted as a part of the plane of ISO/IEC 10646, it will

facilitate the globalization progress of information processing with Nüshu. The protection and save of some important data and further related research are more available and convenient for Nüshu lovers.

V、Acknowledgement

This report is drawn up by the *Committee for Female's Script (Nüshu) of Chinese Folk Literature and Art Society* and the *School of Humanities and Social Sciences of Tsinghua University*. The government of Jiangyong in Hunan Province supports us greatly. At the same time, acknowledgement for China Electronic Standardization Institute, the Institute of Linguistics of the Chinese Academy of Social Sciences, Tsinghua University Computer Department and Computational Linguistics Laboratory.

VI、Reference

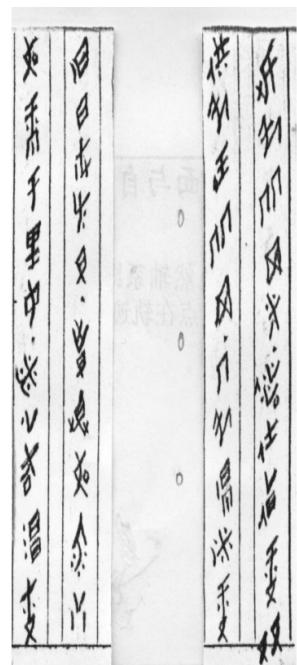
Chinese Nüshu works collection, China Publishing House, 2005

Nüshu Character comparison, Intellectual Property Press, 2006.

VII、Connections

If any questions, please send email to Zhao Liming(zhaolm@mail.tsinghua.edu.cn) and Zhu Cuifang(kalisiqi@gmail.com). At the same time, please send the feedback to Chen Zhuang(chenzh@cesi.ac.cn) who is the head of this program.

VIII、Nüshu sample



(on the right)

床头明月光 疑是地上霜
举头望明月 低头思故乡

Main meaning:

Moonlight shines the bed quietly so brightly that it is considered as the frost on the ground. I raised my head to watch the bright moon and my hometown came into my brain. So I was homesick when I looked down.

(on the left)

白日依山尽 黄河入海流
欲穷千里目 更上一层楼

Main meaning:

The sun disappears accompanied with the faraway mountain, and the yellow river comes into the sea. I go upstairs to see beautiful far sight as far as possible.



The cover of *Chinese Nüshu works Collection*



Nüshu works sample



Chinese translation:

眼泪双垂提笔坐 三傑回言奉□□

眼泪双垂提笔坐
拨开抑郁来完义
起眼看来万般好
身在位楼多为贵
当头独当是错了
嫂娘芳花多娇行
妹你听书尽想远
你会先离来配比
今日把文道出听
当今先离就没用
百起我气千时日
如今齐头回府住
当厅提言眼泪落
粗针传名出四对
刚好老成合心坐
拆边未陪怎不气
几时转归高楼坐
思远同陪日好过
设此有依好身曰
可寒可怜依空日
多时想言就是哭
三傑先离依曰在

三傑回言奉□□
请起妹娘步高厅
四傑结为义开个
根尾言上听父母
一明礼入当□□
样样胜如人四边
世海依身入礼行
已是不由就世间
只到妹娘请谅宽
到你完恩依过了
朝夜不离长弟身
相伴可立不宽行
坐义曰全要回宫
只气绣花色不全
就曰人来听我声
你东我西各自行
要乐逍遙伴听声
当夜父双开礼遥
一世同陪在驰边
一家恭贺身不高
只怨将身错度了
回家坐拢念不欢

TTABLE XX — Row 1B0—1B1

	1B00	1B01	1B02	1B03	1B04	1B05	1B06	1B07	1B08	1B09
0	丶	彳	𠂔	𠂇	𠂊	𠂓	𠂔	𠂎	𠂆	𠂔
1	𠂊	𠂔	𠂊	𠂇	𠂊	𠂔	𠂔	𠂎	𠂊	𠂔
2	𠂔	𠂇	𠂊	𠂇	𠂇	𠂔	𠂇	𠂇	𠂔	𠂇
3	亼	水	𠂔	𠂇	𠂇	𠂔	𠂇	𠂇	𠂔	𠂔
4	𠂊	𠂇	𠂔	𠂇	𠂇	𠂔	𠂇	𠂇	𠂊	𠂇
5	𠂔	𠂇	𠂊	𠂇	𠂔	𠂔	𠂔	𠂇	𠂇	𠂔
6	丶	𠂔	𠂔	𠂇	𠂇	𠂔	𠂔	𠂔	𠂇	𠂔
7	(𠂇	𠂇	𠂔	𠂔	𠂇	𠂇	𠂇	𠂇	𠂔
8	𠂔	𠂇	𠂇	𠂇	𠂇	𠂔	𠂇	𠂇	𠂇	𠂔
9	𠂊	𠂔	𠂔	𠂔	𠂇	𠂔	𠂔	𠂇	𠂇	𠂔
A	𠂇	𠂔	𠂇	𠂇	𠂇	𠂔	𠂔	𠂔	𠂇	𠂔
B	𠂇	𠂇	𠂇	𠂔	𠂇	𠂇	𠂇	𠂇	𠂇	𠂇
C	𠂇	𠂇	𠂇	𠂇	𠂔	𠂇	𠂇	𠂇	𠂇	𠂇
D	𠂔	𠂇	𠂇	𠂇	𠂔	𠂇	𠂔	𠂇	𠂇	𠂔
E	𠂔	𠂇	𠂇	𠂇	𠂇	𠂔	𠂇	𠂇	𠂇	𠂇
F	𠂔	𠂇	𠂔	𠂇	𠂔	𠂔	𠂔	𠂔	𠂇	𠂔

	1B14	1B15	1B16	1B17	1B18	1B19	1B1A	1B1B	1B1C	1B1D
0	𠂇	𠂉	𠂊	𠂔	𠂎	𠂏	𠂓	𠂔	𠂕	
1	𠂆	𠂈	𠂉	𠂊	𠂐	𠂑	𠂃	𠂄		
2	𠂅	𠂆	𠂈	𠂊	𠂐	𠂋	𠂌	𠂍		
3	𠂇	𠂈	𠂉	𠂊	𠂐	𠂒	𠂔	𠂕		
4	𠂆	𠂈	𠂉	𠂊	𠂐	𠂒	𠂃	𠂄		
5	𠂇	𠂈	𠂉	𠂊	𠂐	𠂒	𠂃	𠂄		
6	𠂆	𠂈	𠂉	𠂊	𠂐	𠂒	𠂃	𠂄		
7	𠂇	𠂈	𠂉	𠂊	𠂐	𠂒	𠂃	𠂄		
8	𠂆	𠂈	𠂉	𠂊	𠂐	𠂒	𠂃	𠂄		
9	𠂇	𠂈	𠂉	𠂊	𠂐	𠂒	𠂃	𠂄		
A	𠂆	𠂈	𠂉	𠂊	𠂐	𠂒	𠂃	𠂄		
B	𠂆	𠂈	𠂉	𠂊	𠂐	𠂒	𠂃	𠂄		
C	𠂇	𠂈	𠂉	𠂊	𠂐	𠂒	𠂃	𠂄		
D	𠂇	𠂈	𠂉	𠂊	𠂐	𠂒	𠂃	𠂄		
E	𠂇	𠂈	𠂉	𠂊	𠂐	𠂒	𠂃	𠂄		
F	𠂇	𠂈	𠂉	𠂊	𠂐	𠂒	𠂃	𠂄		

A. Administrative

1. Title:	Proposal for encoding Nüshu in the SMP of the UCS		
2. Requester's name:	China		
3. Requester type (Member body/Liaison/Individual contribution):	Member body		
4. Submission date:	2007-08-13		
5. Requester's reference (if applicable):			
6. Choose one of the following:			
This is a complete proposal:		Yes	
(or) More information will be provided later:			

B. Technical – General

1. Choose one of the following:			
a. This proposal is for a new script (set of characters):	Yes		
Proposed name of script:	Nüshu		
b. The proposal is for addition of character(s) to an existing block:			
Name of the existing block:			
2. Number of characters in proposal:	449		
3. Proposed category (select one from below - see section 2.2 of P&P document):	B		
A-Contemporary	B.1-Specialized collection)	(small	B.2-Specialized collection) (large)
C-Major extinct	D-Attested extinct	E-Minor extinct	
F-Archaic Hieroglyphic or deographic	G-Obscure or questionable usage symbols		
4. Is a repertoire including character names provided?	No		
a. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document?			
b. Are the character shapes attached in a legible form suitable for review?	Yes		
5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?	Weifang Beida Jade Bird Huaguang Imagesetter Co., Ltd.		
If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:	chenzh@cesi.ac.cn		
6. References:			
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	Yes		
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	Yes		
7. 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)?	No		

8. 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/UCD.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

If YES explain

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 *the Government of Yongzhou; The Committee for Female's Script whom? (Niushu) of Chinese Folk Literature and Art Society*

If YES, available relevant documents:

3. Information on the user community for the proposed characters (for example:

size, demographics, information technology use, or publishing use) is included? Yes

Reference: *Niushu was used by women in Jiangyong and now by local people.*

4. The context of use for the proposed characters (type of use; common or rare) common

Reference *It is widely used for communication and tourism development by local people*

5. Are the proposed characters in current use by the user community? Yes

If YES, where? Reference:

Jiangyong, Hunan, China

6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely

in the BMP?

No

If YES, is a rationale provided?

If YES, reference:

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

If YES, is a rationale for its inclusion provided?

If YES, reference:

9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? No

If YES, is a rationale for its inclusion provided?	
If YES, reference:	
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?	<i>No</i>
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
11. Does the proposal include use of combining characters and/or use of composite sequences?	<i>No</i>
If YES, is a rationale for such use provided?	
If YES, reference:	
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	<i>No</i>
If YES, describe in detail (include attachment if necessary)	
13. Does the proposal contain any Ideographic compatibility character(s)?	<i>No</i>
If YES, is the equivalent corresponding unified ideographic character(s) identified?	
If YES, reference:	