

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

Doc Type: Working Group Document**Title: Revised proposal to encode the Manichaeian script in the UCS****Source: Michael Everson and Desmond Durkin-Meisterernst****Status: Individual Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Date: 2003-05-06****A. Administrative****1. Title**

Revised proposal to encode the Manichaeian script in the UCS.

2. Requester's name

Michael Everson and Desmond Durkin-Meisterernst

3. Requester type (Member body/Liaison/Individual contribution)

Individual contribution.

4. Submission date

2003-05-06

5. Requester's reference (if applicable)**6. Choose one of the following:****6a. This is a complete proposal**

Yes.

6b. More information will be provided later

No.

B. Technical -- General**1. Choose one of the following:****1a. This proposal is for a new script (set of characters)**

Yes.

Proposed name of script

Manichaeian.

1b. The proposal is for addition of character(s) to an existing block

No.

1b. Name of the existing block**2. Number of characters in proposal**

36

3. Proposed category (see section II, Character Categories)

Category B.1.

4a. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000)

Level 3.

4b. Is a rationale provided for the choice?

Uses combining characters.

4c. If YES, reference**5a. Is a repertoire including character names provided?**

Yes.

5b. If YES, are the names in accordance with the character naming guidelines in Annex L of ISO/IEC 10646-1: 2000?

Yes.

5c. Are the character shapes attached in a legible form suitable for review?

Yes.

6a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Jost Gippert and Michael Everson.

6b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson, Fontographer.

7a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

No.

7b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

No.

8. 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.

9. 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 Unicode Character Database <http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

U+1xx00..U+1xx05 (Manichaeian punctuation):

General Category "Po" (Punctuation, Other)

Bidi Category "AL" (Right-to-Left Arabic)

Combining Class "0" (Spacing)

U+1xx10..U+1xx28 (Manichaeian letters):

General Category "Lo" (Letter, Other)

Bidi Category "AL" (Right-to-Left Arabic)

Combining Class "0" (Spacing)

U+1xx29 (Manichaeian lengthener):

General Category "Lm" (Letter, Modifier)

Bidi Category "AL" (Right-to-Left Arabic)

Combining Class "0" (Spacing)

U+1xx2A..U+1xx2D (Manichaeian non-spacing marks)

General Category "Mn" (Mark, Non-Spacing)

Bidi Category "AL" (Right-to-Left Arabic)

Combining Class "230" (Above)

C. Technical -- Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

No.

2a. 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.

2b. If YES, with whom?

Desmond Durkin-Meisterernst and the TITUS project in Frankfurt.

2c. If YES, available relevant documents

<http://titus.fkidg1.uni-frankfurt.de/unicode/iranian/3tagung.htm>

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?

Iranianists and other scholars.

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

Uncommon; the script is important for students of the Manichaeian religion, as well as Middle and Early Modern Persian, Parthian, Sogdian, Bactrian, Uighur, and Tokharian.

4b. Reference

P. Oktor Skjærvø in Daniels & Bright 1996.

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

Yes.

5b. If YES, where?

Scholarly publications.

6a. After giving due considerations to the principles in Principles and Procedures document (a WG 2 standing document) must the proposed characters be entirely in the BMP?

No. Proposed allocation is U+10B00-U+10B2F

6b. If YES, is a rationale provided?

6c. If YES, reference

Accordance with the Roadmap.

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

Yes.

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

No.

10b. If YES, is a rationale for its inclusion provided?

10c. If YES, reference

11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?

No.

11b. If YES, is a rationale for such use provided?

11c. If YES, reference

12a. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

12b. If YES, reference

13a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

13b. If YES, describe in detail (include attachment if necessary)

14a. Does the proposal contain any Ideographic compatibility character(s)?

No.

14b. If YES, is the equivalent corresponding unified ideographic character(s) identified?

Manichaeism is a dualistic religion founded by Mani (216-274 or 277 CE) which flourished for a number of centuries and finally petered out in about the 14th century. Mani grew up in Babylonia and his religious system was designed to combine and bring to completion the various major religious systems (Judaeo-Christianity, Gnosticism, Zoroastrianism and even Buddhism) living side by side but opposed to each other in Mesopotamia and surrounding areas much of which was part of the vast Sasanian Empire. The main features of Manichaeism are dualism – the cosmic opposition of the good principle, light, and the evil principle, darkness – the gnostic awakening of the individual soul to its divine origins and the need to free the light trapped in matter in order to return it to its proper place in paradise. A particular feature of Manichaeism is Mani's decision to spread his teachings in any language available. This resulted in a body of Manichaean literature in many languages as Manichaeism spread eastwards and westwards. Since Manichaeism faced persecution in most places, much of its literature was destroyed, though significant Coptic and Greek Manichean sources have survived. Manichaeism became an official state religion in the Uighur kingdom in Central Asia (from 762 to 840 CE) and it is here, in the Turfan oasis on the Silk Road in Central Asia that the most significant Manichaean texts in the east were found. These are written in Manichaean script in the Iranian languages Middle and Early Modern Persian, Parthian, Sogdian, and Bactrian, as well as in the Turkic language Uighur and, to a lesser extent, the Indo-European language Tocharian.

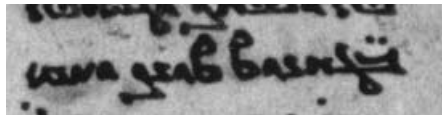
The Manichaean script evolved from Syriac Estrangelo, and is believed by some to have been devised by Mani himself. The script is written from right to left. A few of the characters have final forms in addition to the form which is used initially and medially. A set of punctuation marks exists, and there are four diacritical marks which, while they look similar, behave differently. The double dot used to differentiate consonants is the same size as the single dot used for a similar purpose; the other double dots are sometimes drawn in a larger size.

Rendering. U+1xx20 ◌̇ MANICHAEAN COMBINING SINGLE DOT is used to modify consonants to indicate additional sounds. This character is applied to the letters 𐎠 *d*, 𐎡 *k*, and 𐎢 *p*. The three different double dots in the Manichaean script behave differently from one another. U+1xx2A ◌̈ MANICHAEAN COMBINING DOUBLE DOT is used to modify consonants to indicate additional sounds. This character is applied to the letters 𐎣 *b*, 𐎤 *z*, 𐎥 *k*, 𐎦 *j*, 𐎧 *q*, and 𐎨 *x*. The dots are always placed horizontally above the letter.

𐎠 <i>d</i>	+	◌̇	→	𐎠̇ <i>r</i>
𐎡 <i>k</i>	+	◌̇	→	𐎡̇ <i>x</i>
𐎢 <i>p</i>	+	◌̇	→	𐎢̇ <i>f</i>
𐎣 <i>b</i>	+	◌̈	→	𐎣̈ <i>β</i>
𐎤 <i>z</i>	+	◌̈	→	𐎤̈ <i>]</i>
𐎥 <i>k</i>	+	◌̈	→	𐎥̈ <i>]</i>
𐎦 <i>j</i>	+	◌̈	→	𐎦̈ <i>[</i>
𐎧 <i>q</i>	+	◌̈	→	𐎧̈ <i>`</i>
𐎨 <i>x</i>	+	◌̈	→	𐎨̈ <i>^</i>

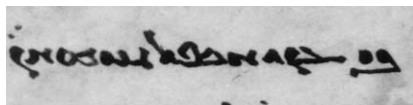
Examples

𐎣̈



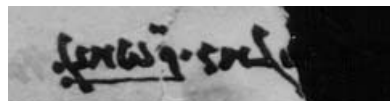
M1/117/ βγZnw\ \wnk whmn

𐎤̈



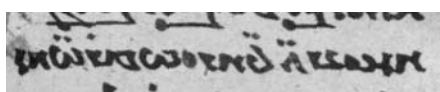
M42/V/ii/21/ ky YzwZst hym Z]

𐎣̈



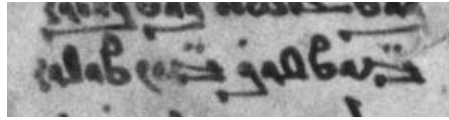
MIKIII4979/R/i/8/](.)gZn . _XZng

[𐎦̈ and 𐎦̈^



MIKIII8259/I/R/i/5/ zhynd a
 [ZdyXmr^Zn (apparently a mistake for
 [Zdy^mrXZn).

ف and x

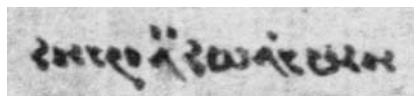


M1/136/ `w\lwx `yz \wlwn

U+1xx2B MANICHAEAN COMBINING ABBREVIATION UD is used with X and w, in the combinations X, XZn and w. The dots indicate an abbreviation of the normal spellings ZwX, ZwXZn and Zwd. The common factor here is the conjunction *ud* ‘and’ on its own or with the enclitic pronouns -X ‘his, her, its’ and -Xān ‘theirs’ attached. For the abbreviated form a some scholars have preferred to transliterate *u*. The use of this abbreviation restricted to the word *ud*, and it does not occur when enclitic pronouns other than -X and -Xān are attached to it. There is a further peculiarity in that a can be written with one dot above and one below the letter: d.

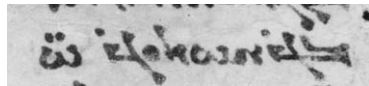
Examples

a



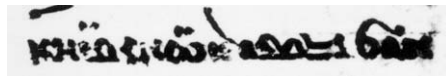
M7/II/V/i/27/ Zc rwXn a yzdZn

c



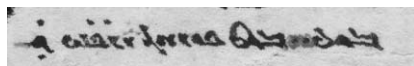
M5/R/ii/18/ bgrZXtygr c

cZn and f



M2/II/R/ii/34/ Zw\ bst °° cZn frhZ

d variant



M39/R/i/7/ bwj °° bw\ pydZg drfX d

U+1xx2B MANICHAEAN COMBINING PLURAL DOTS is also used to indicate that a spelling has been shortened, but the behaviour of this character is different from that of MANICHAEAN COMBINING ABBREVIATION UD, which is limited in application to X and w. Double dots are frequently used at the end of the manuscript line to indicate that the scribe has shortened a word to fit it in. The shortening frequently involves the plural ending in -Zn which is reduced to *n* with dots placed above, below or after it (it is this usage from which the name for this character has been derived). Although the shortening very often involves leaving out an Z, the dots cannot be taken to signify a missing Z because shortening occasionally involves leaving out other letters.

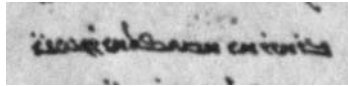
Examples

shortening -yZ to b



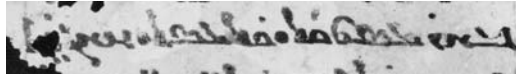
M39/R/i/18/ wrc Zw\ Yzgd Zyy hwfrbd for fryZd

shortening of -zn to f^{••}



M1/368/ brZdrZn ZmwstZn
rwXfn for rwXnZn

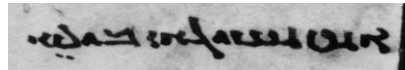
shortening other than of -z-



M34/V/7/ bZr Ys\ftyft °
YstyhZg(et) for YstyhZgyft

The position of the dots is important; compare this abbreviation with [above:

shortening of -jZ to g^{••}



M42/V/i/5/ Zc hndwgZn bwgd
for bwjZd

The illustrations here are taken from W. Sundermann, *Iranian Manichaean Turfan texts in early publications (1904-1934): Photo Edition*. London: School of Oriental and African Studies 1996 (CII Supplementary Series Vol. III).

TABLE XXX - Row 1xx: MANICHAEAN

	1xx0	1xx1	1xx2
0			
1			
2			
3			
4			
5			
6			
7			
8			
9			
A			
B			
C			
D			
E			
F			

G = 00
P = 01

