

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 Manichaean script in the UCS****Source: Michael Everson and Desmond Durkin-Meisterernst****Status: Individual Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Replaces: N2544****Date: 2008-01-28**

This is a preliminary proposal to encode the Manichaean script. Comment is invited.

1. Introduction. 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 (Judaean-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.

2. Processing. Manichaean is an alphabetic script written right-to-left, with spaces between words. The Manichaean script evolved from Syriac Estrangelo. Because of its use by Manichaeans in Central Asia, the script has been called “Manichaean” by modern scholars. 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.

3. Names and ordering. The names used for the Manichaean characters are based on their Imperial Aramaic analogues. The order of the characters in the code charts is their alphabetical order. Since Manichaean makes use a number of characters which are derived from Aramaic base-letters, new names based on the Aramaic letter-names have been devised in accordance with the usual UCS conventions, so that naming scheme is mnemonic and useful. For example, spirant letters using a double-dot diacritic are typically named using the letter -H-, so for 𐭪 BETH *b*, 𐭪 KAF *k*, 𐭪 ZAYIN *z*, 𐭪 JAYIN *j*, 𐭪 QOPH *q*, the

marked forms are  BHETH β ,  KHAFA $k̄$,  ZHAYIN $z̄$,  JHAYIN $j̄$,  QHOPH $q̄$. For letters  AYIN $ʿ$, and  SSHIN $ś$, where -H- does not make sense, the initial letter has been doubled,  AAYIN $ʿ$,  SSHIN $ś$. The H- is used in some other letters, such as  GHIMEL $γ$,  DHAMEDH $ð$,  THAMEDH $θ \sim ðð$ (from  GIMEL g ,  LAMEDH l); in letters with a single dot a letter is simply changed, as in  FE f ,  XAF k ,  QAPH q (from  PE p ,  KAF k ,  QOPH q).

4. Shaping. The Manichaean script as proposed for encoding has fully-developed joining behaviour. The table below shows the joining forms as well as noting which characters do not have joining behaviour.

Dual-joining Manichaean Characters

Character	X _n	X _r	X _m	X _l
BETH				
BHETH				
GIMEL				
GHIMEL				
LAMEDH				
DHAMEDH				
THAMEDH				
MEM				
SAMEKH				
AYIN				
AAYIN				
PE				
FE				
QOPH				
QAPH				
QHOPH				

Right-joining Manichaean Characters

Character	X _n	X _r
DALETH		
WAW		
ZAYIN		
ZHAYIN		
TETH		
YODH		
KAPH		
XAPH		
KHAPH		
SADHE		
RESH		
TAW		

Beth-joining Manichaean Character

Character	X _n	X _{rb}	X _{mb}	X _l
ALEPH				

The letter ALEPH is generally left-joining, but it has right joining behaviour when following the letters BETH and BHETH.

Left-joining Manichaean Characters

Character	X _n	X ₁
HETH	𐎡	𐎡
NUN	𐎢	𐎢

Non-joining Manichaean Characters

Character	X _n
HE	𐎣
JAYIN	𐎤
JHAYIN	𐎥
SHIN	𐎦
SSHIN	𐎧

It has been suggested that it is possible to make an analysis of Manichaean which has initial and final forms only. In this analysis, final forms do not join to the left and initial forms have long tails that sometimes touch and sometimes do not. This might in some sense be an analysis that the ductus of a pen stroke might support, but the problem with this model is twofold. First, it is not a model that is familiar to implementors of the UCS—the dual-joining model is well understood and easy to implement. In the second place, the standard analysis that was used for Arabic, Syriac, and Psalter Pahlavi does in fact work very well for Manichaean. Precise shaping can be handled with ZWJ and ZWNJ if necessary.

4.1. Manichaean makes use of two standard and obligatory ligatures. The table below is displayed in visual order:

$$\begin{aligned} \check{y} \text{ 𐎣𐎡 } &= \text{YODH } \bullet + \text{SADHE } \text{ 𐎣 } \leftarrow \\ \check{n} \text{ 𐎣𐎢 } &= \text{NUN } \text{ 𐎢 } + \text{SADHE } \text{ 𐎣 } \leftarrow \end{aligned}$$

Note that if SADHE is in right-joining form the ligatures are also applied: $\check{y} \text{ 𐎣𐎡𐎣 } , \check{n} \text{ 𐎣𐎢𐎣 } .$

5.2. Manichaean numbers. Manichaean has its own numbers, which have right-to-left directionality. Numbers are built up out of 1, 5, 10, 20, and 100; the number 1000 is not attested in Manichaean. The following is an exhaustive list of numbers attested in Manichaean. The third column is displayed in visual order.

1	𐎠	1 ←	M283 II V 4
2	𐎠𐎠	1 + 1 ←	
3	𐎠𐎠𐎠	1 + 1 + 1 ←	M67 R ii 11
4	𐎠𐎠𐎠𐎠	1 + 1 + 1 + 1 ←	M74 II R 18
7	𐎠𐎠𐎠𐎠𐎠	1 + 1 + 5 ←	
8	𐎠𐎠𐎠𐎠𐎠𐎠	1 + 1 + 1 + 5 ←	
12	𐎠𐎠𐎠𐎠𐎠𐎠𐎠	1 + 1 + 10 ←	
15	𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠	5 + 10 ←	M5750 R ii 21
68	𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠	1 + 1 + 1 + 5 + 20 + 20 + 20 ←	M1 390
77	𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠	1 + 1 + 5 + 10 + 20 + 20 + 20 ←	M1 321
162	𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠	1 + 1 SPACE 20 + 20 + 20 SPACE 100 ←	M1 167
546	𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠𐎠	1 + 5 + 20 + 20 [linebreak] 100 + 5 ←	M1 160

Unfortunately very few Manichaean numbers are attested. Several of them take the form of Manichaean letters, though it is probable that these forms were re-analysed from an original Aramaic prototype.

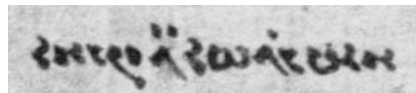
Dual-joining Psalter Pahlavi Numbers [needs editing]

Character	X _n	X _r	X _m	X _l
ONE	𐭠	𐭡	𐭢	𐭣
FIVE	𐭠𐭡	𐭡𐭢	𐭢𐭣	𐭣𐭤
TEN	𐭠𐭡𐭢	𐭡𐭢𐭣	𐭢𐭣𐭤	𐭣𐭤𐭥
TWENTY	𐭠𐭡𐭢𐭣	𐭡𐭢𐭣𐭤	𐭢𐭣𐭤𐭥	𐭣𐭤𐭥𐭦
ONE HUNDRED	𐭠𐭡𐭢𐭣𐭤𐭥𐭦	𐭡𐭢𐭣𐭤𐭥𐭦𐭧		[not dual joining?]

4.2 Other diacritical marks. U+1xx2B MANICHAEAN COMBINING ABBREVIATION UD is used with *š* and *w*, in the combinations *š*, *š'n* and *w*. The dots indicate an abbreviation of the normal spellings *'wš*, *'wš'n* and *'wd*. The common factor here is the conjunction *ud* 'and' on its own or with the enclitic pronouns *-š* 'his, her, its' and *-šān* 'theirs' attached. For the abbreviated form *w̄* some scholars have preferred to transliterate *u*. The use of this abbreviation is restricted to the word *ud*, and it does not occur when enclitic pronouns other than *-š* and *-šān* are attached to it. There is a further peculiarity in that *w̄* can be written with one dot above and one below the letter: *w̄*.

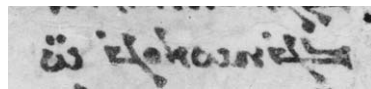
6. Punctuation.

w̄ 𐭣̄



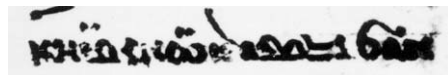
M7/II/V/i/27/ 'c *rwšn w̄ yzd'n*

š 𐭡̄



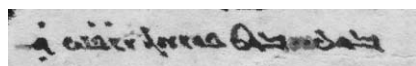
M5/R/ii/18/ *bgr'štygr š*

š'n 𐭡̄𐭣̄ and *f* 𐭠̄



M2/II/R/ii/34/ 'w̄ *bst š'n frh'*

w̄ variant 𐭣̄̇̈



M39/R/i/7/ *bwj bw̄t pyd'g drfš w̄*

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 *š* 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 *-n* 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 *'*, the dots cannot be taken to signify a missing *'* because shortening occasionally involves leaving out other letters.

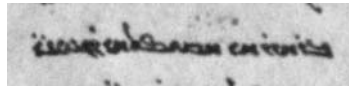
Examples

shortening *-y'* to *ȳ* 𐭣̄̇̈



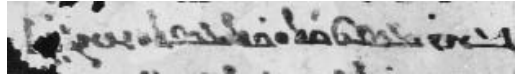
M39/R/i/18/ *wrc 'w̄t 'zgd 'yy hwfrj̄d for fry'd*

shortening of -'n to ṇ 𐭌



M1/368/ br'dr'n 'mwst'n rwšñn for rwšn'

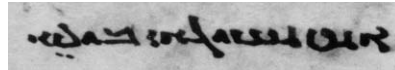
shortening other than of -'



M34/V/7/ b'r 'stfityft ° 'styh'g(fī)
for 'styh'gyft

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

shortening of -j' to j̄ 𐭌



M42/V/i/5/ 'c hndwg'n bwjd for bwj'd

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).

7. Bibliography

- Faulmann, Carl. 1990 (1880). *Das Buch der Schrift*. Frankfurt am Main: Eichborn. ISBN 3-8218-1720-8
- Driver, G. R. 1976. *Semitic writing from pictograph to alphabet*. Third edition edited by S. A. Hopkins. London: Oxford University Press for the British Academy.
- Ifrah, Georges. 2000. *The universal history of numbers. Volume 1: The world's first number-systems. Volume 2: The modern number-system*. Translated from the French by David Bellos, E. F. Harding. Sophie Wood, and Ian Monk. London: Harvill Press. ISBN 1-86046-790-3, ISBN 1-86046-791-1
- Naveh, Joseph. 1987. *Early history of the alphabet: an introduction to West Semitic epigraphy and palaeography*. Jerusalem: Magnes Press, the Hebrew University. ISBN 965-223-436-2
- Skjærvø, P. Oktor. 1996. "Aramaic scripts for Iranian languages" in *The World's Writing Systems*, ed. Peter T. Daniels & William Bright. New York; Oxford: Oxford University Press. ISBN 0-19-507993-0
- Taylor, Isaac. 1883. *The alphabet: an account of the origin and development of letters*. Vol. 1: Semitic alphabets; Vol. 2: Aryan alphabets. London: Kegan Paul.

8. Unicode Character Properties

10840;MANICHAEAN LETTER ALEPH;Lo;0;R;;;;N;;;;;
10841;MANICHAEAN LETTER BETH;Lo;0;R;;;;N;;;;;
10842;MANICHAEAN LETTER BHETH;Lo;0;R;;;;N;;;;;
10843;MANICHAEAN LETTER GIMEL;Lo;0;R;;;;N;;;;;
10844;MANICHAEAN LETTER GHIMEL;Lo;0;R;;;;N;;;;;
10845;MANICHAEAN LETTER DALETH;Lo;0;R;;;;N;;;;;
10846;MANICHAEAN LETTER HE;Lo;0;R;;;;N;;;;;
10847;MANICHAEAN LETTER WAW;Lo;0;R;;;;N;;;;;
10848;MANICHAEAN LETTER ZAYIN;Lo;0;R;;;;N;;;;;
10849;MANICHAEAN LETTER ZHAYIN;Lo;0;R;;;;N;;;;;
1084A;MANICHAEAN LETTER JAYIN;Lo;0;R;;;;N;;;;;
1084B;MANICHAEAN LETTER JHAYIN;Lo;0;R;;;;N;;;;;
1084C;MANICHAEAN LETTER HETH;Lo;0;R;;;;N;;;;;
1084D;MANICHAEAN LETTER TETH;Lo;0;R;;;;N;;;;;
1084E;MANICHAEAN LETTER YODH;Lo;0;R;;;;N;;;;;
1084F;MANICHAEAN LETTER KAPH;Lo;0;R;;;;N;;;;;
10850;MANICHAEAN LETTER XAPH;Lo;0;R;;;;N;;;;;
10851;MANICHAEAN LETTER KHAPH;Lo;0;R;;;;N;;;;;
10852;MANICHAEAN LETTER LAMEDH;Lo;0;R;;;;N;;;;;
10853;MANICHAEAN LETTER DHAMEDH;Lo;0;R;;;;N;;;;;
10854;MANICHAEAN LETTER THAMEDH;Lo;0;R;;;;N;;;;;
10855;MANICHAEAN LETTER MEM;Lo;0;R;;;;N;;;;;
10856;MANICHAEAN LETTER NUN;Lo;0;R;;;;N;;;;;
10857;MANICHAEAN LETTER SAMEKH;Lo;0;R;;;;N;;;;;
10858;MANICHAEAN LETTER AYIN;Lo;0;R;;;;N;;;;;
10859;MANICHAEAN LETTER AAYIN;Lo;0;R;;;;N;;;;;
1085A;MANICHAEAN LETTER PE;Lo;0;R;;;;N;;;;;
1085B;MANICHAEAN LETTER FE;Lo;0;R;;;;N;;;;;
1085C;MANICHAEAN LETTER SADHE;Lo;0;R;;;;N;;;;;

1085D;MANICHAEAN LETTER QOPH;Lo;0;R;;;N;;;;
 1085E;MANICHAEAN LETTER QAPH;Lo;0;R;;;N;;;;
 1085F;MANICHAEAN LETTER QHOPH;Lo;0;R;;;N;;;;
 10860;MANICHAEAN LETTER RESH;Lo;0;R;;;N;;;;
 10861;MANICHAEAN LETTER SHIN;Lo;0;R;;;N;;;;
 10862;MANICHAEAN LETTER SSHIN;Lo;0;R;;;N;;;;
 10863;MANICHAEAN LETTER TAW;Lo;0;R;;;N;;;;
 10B64;MANICHAEAN KASHIDA;Lm;0;AL;;;N;;;;
 10B64;MANICHAEAN COMBINING ABBREVIATION DOTS;Mn;230;NSM;;;N;;;;
 10B65;MANICHAEAN COMBINING ABBREVIATION UD;Mn;220;NSM;;;N;;;; — is this the right property???
 10B66;MANICHAEAN COMBINING PLURAL DOTS;Mn;220;NSM;;;N;;;;
 1086B;MANICHAEAN NUMBER ONE;No;0;R;;;1;N;;;;
 1086C;MANICHAEAN NUMBER FIVE;No;0;R;;;5;N;;;;
 1086D;MANICHAEAN NUMBER TEN;No;0;R;;;10;N;;;;
 1086E;MANICHAEAN NUMBER TWENTY;No;0;R;;;20;N;;;;
 1086F;MANICHAEAN NUMBER ONE HUNDRED;No;0;R;;;100;N;;;;
 10B6A;MANICHAEAN PUNCTUATION FLEURON;Po;0;AL;;;N;;;;
 10B6B;MANICHAEAN PUNCTUATION DOT WITHIN DOT;Po;0;AL;;;N;;;;
 10B6C;MANICHAEAN PUNCTUATION SINGLE THICK DOT;Po;0;AL;;;N;;;;
 10B6D;MANICHAEAN PUNCTUATION SINGLE THIN DOT;Po;0;AL;;;N;;;;
 10B6E;MANICHAEAN PUNCTUATION TWO DOTS;Po;0;AL;;;N;;;;
 10B6F;MANICHAEAN PUNCTUATION LINE FILLER;Po;0;AL;;;N;;;;

Figures

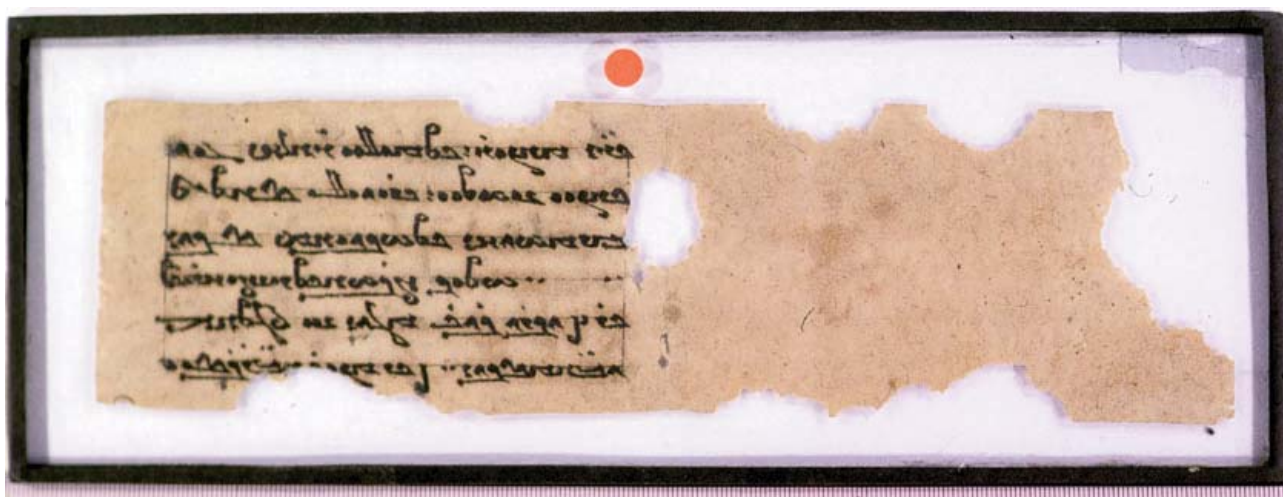


Figure 3. Description of Manichaean script from a German source.

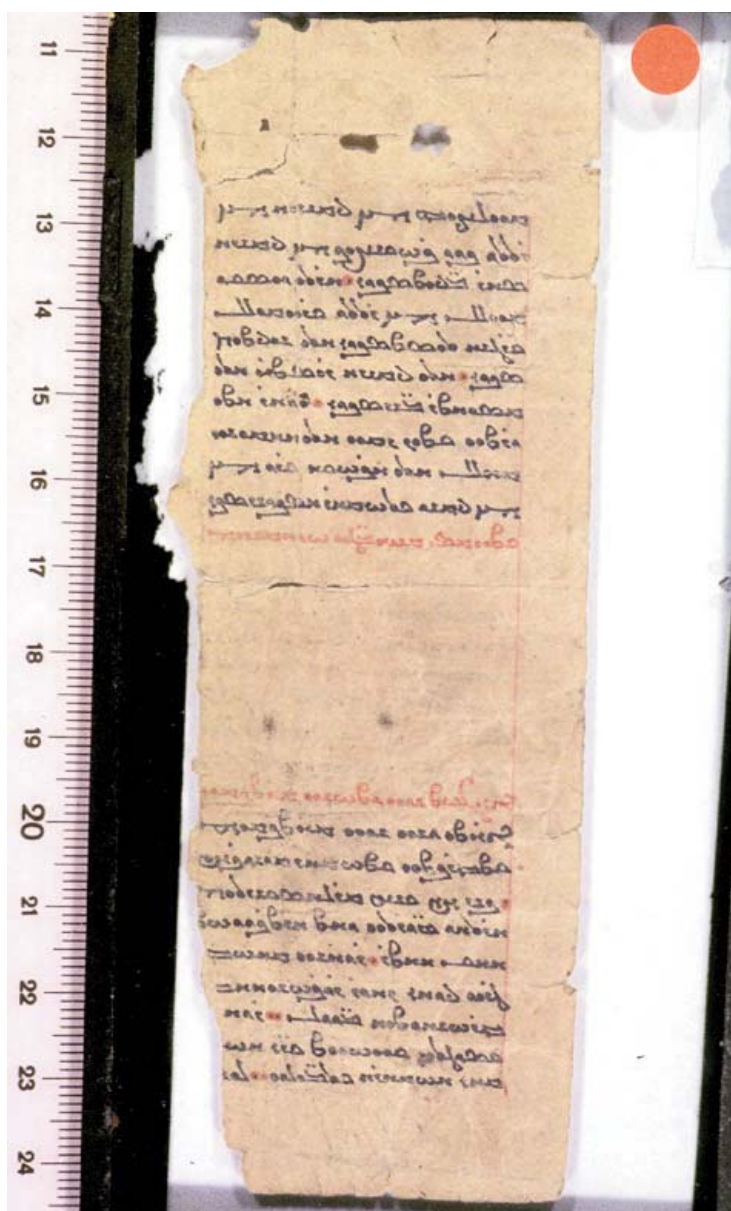


Figure 3. Description of Manichaean script from a German source.

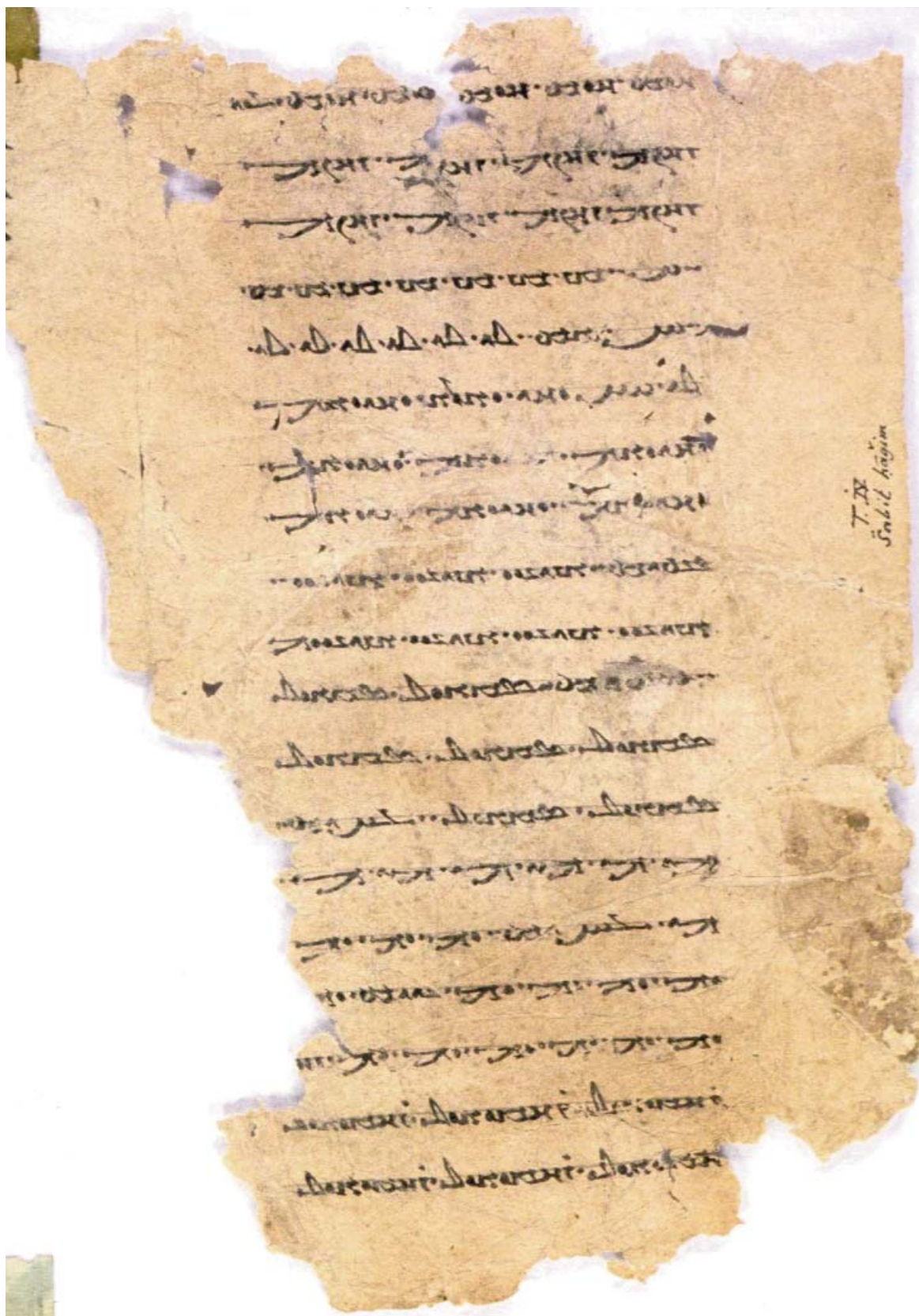


Figure 3. Description of Manichaean script from manuscript M8430 source.

shape	value	shape	value
ا	a	س	s
ب	b	ع	e
پ	p	ف	f
گ	g	ق	q
د	d	ک	k
ه	h	گ	g
و	w	ز	z
ز	z	ح	h
ژ	ž	ط	t
چ	č	ث	ṯ
ج (z)	j (z)	ج	g
ه	h	چ	č
ل	l	ح	h
ی	y	ط	t
ک	k	ث	ṯ
گ	g	ج	g
ک	k	چ	č
ک	k	ح	h
ل	l	ط	t
د	d	ث	ṯ
د (dd)	ḏ (dd)	ج	g
م	m	چ	č
ن	n	ح	h

Figure 2. Table of Iranian alphabets, from a table made by Desmond Durkin-Meisterernst.

Manichäisch

Isoliert	Ende	Mitte	Anfang	Umschrift	Isoliert	Ende	Mitte	Anfang	Umschrift
				a					l
				b					δ
				β					δδ
				g					m
				γ					n
				d					s
				h					c
				v					p
				v̄					f
				z					č
				ž					
				h					čy
				t					čn
				y	q steht für k				
				k					q̄
				b, x					r
									š
				k̄ = q					t

Allgemeines: Unter der reichen Ausbeute der Expeditionen des Museums für Völkerkunde in Berlin nach Chinesisch-Turkistan in den Jahren 1903—14 unter Leitung der Herren Professoren Grünwedel und von Le Coq befanden sich auch zahlreiche Bücher und Bruchstücke in einer bis dahin unbekannten Schrift, welche sich durch reiche Miniaturmalerei und schöne Ausführung auszeichneten. Die Schrift wurde von Herrn Prof. F. W. K. Müller zu Berlin entziffert, der in ihr eine Verwandte der syrischen Estrangelā vermutete und, da die Texte manichäisch-religiösen Inhalts waren, sie als manichäische Schrift bezeichnete. Sie ist aus einer späten Form des aramäischen Alphabets von Mani, dem Stifter der manichäischen Religionsgemeinschaft, zu einer ausgesprochenen Buchschrift entwickelt. Die in den Ruinen von Turfan (Chinesisch-Turkistan) und Umgebung gefundenen Handschriften enthalten Teile der bis dahin in Europa für verschollen gehaltenen Literatur der Manichäer. Bisher sind Handschriften mit Texten in mittelpersischer, parthischer, sogdischer, uigurisch-türkischer und tocharischer (B) Sprache bekannt geworden.

Die Typen wurden nach den Angaben des Herrn Professors von Le Coq in der Reichsdruckerei hergestellt. Die Schrift läuft von rechts nach links oder von oben nach unten; im letzteren Falle liest man die vertikalen Zeilenreihen von links nach rechts. Die Silben dürfen ebenso wie im Syrischen nicht getrennt werden. Läßt sich der überschüssige Raum nicht auf die einzelnen Wortzwischenräume verteilen, so ist er mit Hilfe schmaler Sperrstriche, welche zwischen die einzelnen Buchstaben gesetzt werden, auszusperren.

Abkürzungen: = v mit zwei Punkten, Abkürzung für ein Bindewort.

Interpunktion: Es gibt folgende Interpunktionen: Die einfachen Punkte sind mit schwarzer Tinte (Tusche) geschrieben; oft sind sie rot umrandet; die blumenartigen Zeichen sind oft polychrom und dienen zur Einfassung von Titeln.

Literatur: H.-Ch. Puech, *Le Manichéisme*, Paris 1949 (Musée Guimet, Bibliothèque de diffusion, tome LVII, mit ausführlichen bibliographischen Angaben); Ausgaben von Handschriften in dieser Schrift besorgten außer F. W. K. Müller u. a. F. C. Andreas, Willy Bang, A. v. Gabain, W. B. Henning, W. Lentz, E. Waldschmidt, W. Winter; aus der Sammlung in Leningrad: C. Salemann

Figure 3. Description of Manichaeian script from a German source.

Row 10A: MANICHAEAN DRAFT

	10AC	10AD	10AE	10AF
0	Ⲁ	ⲁ	Ⲃ	ⲃ
1	Ⲅ	ⲅ	Ⲇ	ⲇ
2	Ⲉ	ⲉ	Ⲋ	ⲋ
3	Ⲍ	ⲍ	Ⲏ	ⲏ
4	Ⲑ	ⲑ	Ⲓ	ⲓ
5	Ⲕ	ⲕ	Ⲗ	ⲗ
6	Ⲙ	ⲙ	Ⲏ	ⲏ
7	Ⲑ	ⲑ	Ⲓ	ⲓ
8	Ⲕ	ⲕ		
9	Ⲙ	ⲙ		
A	Ⲑ	ⲑ		
B	Ⲕ	ⲕ	Ⲓ	
C	Ⲙ	ⲙ	ⲕ	
D	Ⲑ	ⲑ	Ⲓ	
E	Ⲕ	ⲕ	ⲕ	
F	Ⲙ	ⲙ	Ⲑ	

hex	Name
C0	MANICHAEAN LETTER ALEPH
01	MANICHAEAN LETTER BETH
C2	MANICHAEAN LETTER BHETH
C3	MANICHAEAN LETTER GIMEL
C4	MANICHAEAN LETTER GHIMEL
C5	MANICHAEAN LETTER DALETH
C6	MANICHAEAN LETTER HE
C7	MANICHAEAN LETTER WAW
C8	MANICHAEAN LETTER ZAYIN
C9	MANICHAEAN LETTER ZHAYIN
CA	MANICHAEAN LETTER JAYIN
CB	MANICHAEAN LETTER JHAYIN
CC	MANICHAEAN LETTER HETH
CD	MANICHAEAN LETTER TETH
CE	MANICHAEAN LETTER YODH
CF	MANICHAEAN LETTER KAPH
D0	MANICHAEAN LETTER XAPH
D1	MANICHAEAN LETTER KHAPH
D2	MANICHAEAN LETTER LAMEDH
D3	MANICHAEAN LETTER DHAMEDH
D4	MANICHAEAN LETTER THAMEDH
D5	MANICHAEAN LETTER MEM
D6	MANICHAEAN LETTER NUN
D7	MANICHAEAN LETTER SAMEKH
D8	MANICHAEAN LETTER AYIN
D9	MANICHAEAN LETTER AAYIN
DA	MANICHAEAN LETTER PE
DB	MANICHAEAN LETTER FE
DC	MANICHAEAN LETTER SADHE
DD	MANICHAEAN LETTER QOPH
DE	MANICHAEAN LETTER XOPH
DF	MANICHAEAN LETTER QHOPH
E0	MANICHAEAN LETTER RESH
E1	MANICHAEAN LETTER SHIN
E2	MANICHAEAN LETTER SSHIN
E3	MANICHAEAN LETTER TAW
E4	MANICHAEAN KASHIDA
E5	MANICHAEAN COMBINING ABBREVIATION DOTS
E6	MANICHAEAN COMBINING ABBREVIATION UD
E7	MANICHAEAN COMBINING PLURAL DOTS
E8	(This position shall not be used)
E9	(This position shall not be used)
EA	(This position shall not be used)
EB	MANICHAEAN NUMBER ONE
EC	MANICHAEAN NUMBER FIVE
ED	MANICHAEAN NUMBER TEN
EE	MANICHAEAN NUMBER TWENTY
EF	MANICHAEAN NUMBER ONE HUNDRED
F0	MANICHAEAN PUNCTUATION STAR
F1	MANICHAEAN PUNCTUATION FLEURON
F2	MANICHAEAN PUNCTUATION DOUBLE DOT WITHIN DOT
F3	MANICHAEAN PUNCTUATION DOT WITHIN DOT
F4	MANICHAEAN PUNCTUATION SINGLE THICK DOT
F5	MANICHAEAN PUNCTUATION SINGLE THIN DOT
F6	MANICHAEAN PUNCTUATION TWO DOTS
F7	MANICHAEAN PUNCTUATION LINE FILLER
F8	(This position shall not be used)
F9	(This position shall not be used)
FA	(This position shall not be used)
FB	(This position shall not be used)
FC	(This position shall not be used)
FD	(This position shall not be used)
FE	(This position shall not be used)
FF	(This position shall not be used)

A. Administrative

1. Title

Revised proposal to encode the Manichaean 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

2008-01-28

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

No.

6b. More information will be provided later

Yes.

B. Technical – General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

Yes.

1b. Proposed name of script

Avestan.

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

No.

1d. Name of the existing block

2. Number of characters in proposal

64.

3. Proposed category (A-Contemporary; B.1-Specialized (small collection); B.2-Specialized (large collection); C-Major extinct; D-Attested extinct; E-Minor extinct; F-Archaic Hieroglyphic or Ideographic; G-Obscure or questionable usage symbols)

Category C.

4a. Is a repertoire including character names provided?

Yes.

4b. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document?

Yes.

4c. Are the character shapes attached in a legible form suitable for review?

Yes.

5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Michael Everson.

5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson, Fontographer.

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes.

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

7. 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)?

Yes.

8. 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.

See above.

C. Technical – Justification

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

Yes. See N2556, N1684.

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?

Jost Gippert, Desmond Durkin-Meisterernst, Günter Schweiger

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 Manichaean religion, as well as Middle and Early Modern Persian, Parthian, Sogdian, Bactrian, Uighur, and Tokharian.

4b. Reference

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 the P&P document must the proposed characters be entirely in the BMP?

No.

6b. If YES, is a rationale provided?

6c. If YES, reference

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

No.

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

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

11e. If YES, reference

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?

No.

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?