ISO/IEC JTC1/SC2/WG2 N3644
L2/09-186
2009-05-04

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 for encoding the Manichaean script in the SMP of the UCS**
Source: **UC Berkeley Script Encoding Initiative (Universal Scripts Project)**
Authors: **Michael Everson and Desmond Durkin-Meisterernst**
Status: **Individual Contribution**
Action: **For consideration by JTC1/SC2/WG2 and UTC**
Replaces: **N2544, N3378**
Date: **2009-05-04**

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 (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.

2. Structure. Manichaean is an alphabetic script written right-to-left, with spaces between words. The Manichaean script and Syriac Estrangelo are sister scripts, both having evolved from Aramaic. (This does not mean that Syriac behaviour is necessarily more relevant to Manichaean than to other cursive right-to-left scripts; it means only that some of the lettershapes are fairly closely related between the two scripts.) Because of its use by Manichaeans in Central Asia, the script has been called "Manichaean" by modern scholars. A number of consonants are distinguished from base consonants by the use of one or two dots; these eleven letters (seven with two dots and three with one dot) are encoded explicitly. These letters do not have decompositions. Five characters have final variant forms which are significant but unpredictable; a variation selector is specified to invoke this special shaping behaviour. There are three diacritical marks which indicate abbreviation, plurality, or the conjunction *ud*.

3. Names and ordering. The names used for the Manichaean characters are based on their Imperial Aramaic analogues. 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 𐫁 BETH *b*, 𐫐 KAPH *k*, 𐫉 ZAYIN *z*, 𐫋 JAYIN *j*, 𐫞 QOPH *q*, the marked forms are 𐫂 BHETH *β*, 𐫒 KHAPH *k̈*, 𐫊 ZHAYIN *ž*, 𐫌 JHAYIN, 𐫠 QHOPH *q̈*. For letters 𐫙 AYIN ʿ, and 𐫢 SHIN *š*, where -H- does not make sense, the initial letter has been doubled, 𐫚 AAYIN ʿ̈, 𐫣 SSHIN *ś*. The -H- is used in some other letters, such as 𐫄 GHIMEL *γ*, 𐫔 DHAMEDH *δ*, 𐫕 THAMEDH *θ* ~ *δδ* (from 𐫃 GIMEL *g*, 𐫓 LAMEDH *l*); in letters with a single dot a letter is simply changed, as in 𐫜 FE *f*, 𐫑 XAPH *k̇*, 𐫟 XOPH *q̇* (from 𐫛 PE *p*, 𐫐 KAPH *k*, 𐫞 QOPH *q*). The order of the letters in the code chart is the alphabetical order; dotted letters are considered separate letters and are not interfiled with the base characters.

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. The glyphs shown are X_n nominal, X_r right-joining, X_m dual-joining, and X_l left-joining.

Dual-joining Manichaean Characters

Character	X_n	X_r	X_m	X_l
BETH	𐫁	𐫁	𐫁	𐫁
BHETH	𐫂	𐫂	𐫂	𐫂
GIMEL	𐫃	𐫃	𐫃	𐫃
GHIMEL	𐫄	𐫄	𐫄	𐫄
LAMEDH	𐫓	𐫓	𐫓	𐫓
DHAMEDH	𐫔	𐫔	𐫔	𐫔
THAMEDH	𐫕	𐫕	𐫕	𐫕
MEM	𐫖/𐫖	𐫖/𐫖	𐫖	𐫖
SAMEKH	𐫘	𐫘	𐫘	𐫘
AYIN	𐫙	𐫙	𐫙	𐫙
AAYIN	𐫚	𐫚	𐫚	𐫚
PE	𐫛	𐫛	𐫛	𐫛
FE	𐫜	𐫜	𐫜	𐫜
QOPH	𐫞	𐫞	𐫞	𐫞
XOPH	𐫟	𐫟	𐫟	𐫟
QHOPH	𐫠	𐫠	𐫠	𐫠

Right-joining Manichaean Characters

Character	X_n	X_r
DALETH	𐫅/𐫅	𐫅
WAW	𐫇	𐫇
ZAYIN	𐫉	𐫉
ZHAYIN	𐫊	𐫊
TETH	𐫎	𐫎
YODH	𐫏	𐫏
KAPH	𐫐	𐫐
XAPH	𐫑	𐫑
KHAPH	𐫒	𐫒
SADHE	𐫝	𐫝
RESH	𐫡/𐫡	𐫡
TAW	𐫤	𐫤

Aleph-joining Manichaean Character

Character	A_n	A_{rb}	A_{mb}	A_l
ALEPH	𐫀	𐫀	𐫀	𐫀

Joining Class:		
	An:	ALEPH nominal
	Al:	ALEPH left-joining
	Arb:	ALEPH right-joining following BETH or BHETH
	Amb:	ALEPH dual-joining (medial) following BETH or BHETH

The joining behavior of ALEPH differs from that of other letters. ALEPH is dual joining, but it joins on the right *only* to BETH or BHETH. It does not join to the right to other dual-joining (or left-joining) letters, which means it also functions as a non-joining left context for letters of those joining classes. Compare 𐫀𐫁𐫂𐫀 *"m'm'* = 𐫀 *'* + 𐫂 *m* + 𐫀 *'* + 𐫂 *m* + 𐫀 *'* + 𐫀 *'* ← (where ALEPH does not join to preceding ALEPH or MEM or to following ALEPH but does join to following MEM) with 𐫀𐫁𐫀𐫁𐫀 *"b'b'* = 𐫀 *'* + 𐫁 *b* + 𐫀 *'* + 𐫁 *b* + 𐫀 *'* + 𐫀 *'* (where ALEPH does not join to preceding or following ALEPH but does join to preceding and following BETH).

Left-joining Manichaean Characters

Character	$\mathbf{X_n}$	$\mathbf{X_l}$
HETH	𐫅	𐫅
NUN	𐫃/𐫄	𐫄

Non-joining Manichaean Characters

Character	$\mathbf{X_n}$
HE	𐫆/𐫇
JAYIN	𐫉
JHAYIN	𐫊
SHIN	𐫋
SSHIN	𐫌

4.1 Five Manichaean characters have special forms which can be invoked by means of the character U+FE00 VARIATION SELECTOR-1. In instances other than nominal (and final for MEM) forms the VS-1 has no effect. The table below is displayed in logical order.

DALETH 𐫀	+	VS−1	VS 1	=	d_n 𐫀
HE 𐫇	+	VS−1	VS 1	=	h_n 𐫆
MEM 𐫂	+	VS−1	VS 1	=	m_n 𐫂
MEM 𐫈	+	VS−1	VS 1	=	m_r 𐫈
NUN 𐫄	+	VS−1	VS 1	=	n_n 𐫃
RESH 𐫍	+	VS−1	VS 1	=	r_n 𐫍

The shapes invoked by the VARIATION SELECTOR are not preditable, though they tend to occur at the end of lines. Because the shape is not predictable, a script-specific variation selector has not been proposed. Use of U+FE00 has data implications for the UCD: additions will be required for the files StandardizedVariants.txt and StandardizedVariants.html. The lines for StandardizedVariants.txt are as follows:

```
10AC5 FE00; alternate form; isolate # MANICHAEAN LETTER DALETH
10AC6 FE00; alternate form; isolate # MANICHAEAN LETTER HE
10AD6 FE00; alternate form; isolate # MANICHAEAN LETTER MEM
10AD6 FE00; alternate form; isolate final # MANICHAEAN LETTER MEM
10AD7 FE00; alternate form; isolate # MANICHAEAN LETTER NUN
10AE1 FE00; alternate form; isolate # MANICHAEAN LETTER RESH
```

Evidently these would also be the first non-BMP characters to make use of U+FE00. Whether this has implications for data is a matter for the UTC. We note that there is also a set of variation selectors in Plane 14; we do not know the implications for preferring either set over the other, and do not have a preference. We have briefly considerd, and quickly rejected, the idea of encoding alternate forms of MEM, DALETH, HE, NUN, and RESH; these standard variants are glyph variants only.

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

čy = YODH + SADHE ←

čn = NUN + SADHE ←

Note that if SADHE is in right-joining form the ligatures are also applied: *čy* , *čn* .

4.3. Manichaean makes use of a *kashida* to extend a word. The character U+ 0640 ARABIC TATWEEL is proposed to be used for this function. but it may be the case that the script property for that character needs to be changed to Common. Mandaic also has a similar requirement.

5. 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. Unfortunately very few Manichaean numbers are attested. The numbers 10, 20, and 100 take the form of Manichaean letters (HE, PE, MEM); their glyphs were re-analysed from the original Aramaic prototype. The following is an exhaustive list of numbers attested in Manichaean. The third column is displayed in visual order; the fourth column is the manuscript source.

1		1 ←	M283 II V 4
2		1 + 1 ←	
3		1 + 1 + 1 ←	M67 R ii 11
4		1 + 1 + 1 + 1 ←	M74 II R 18
7		1 + 1 + 5←	
8		1 + 1 + 1 + 5 ←	
12		1 + 1 + 10 ←	M14 R 1, 2, 4, 9, 10
15		5 + 10 ←	M5750 R ii 21
68		1 + 1 + 1 + 5 + 20 + 20 + 20 ←	M1 390
77		1 + 1 + 5 + 10 + 20 + 20 + 20 ←	M1 321
162		1 + 1 SPACE 20 + 20 + 20 SPACE 100 ←	M1 167
546		1 + 5 + 20 + 20 [linebreak] 100 + 5 ←	M1 160–161

Note that the height at which 1 and 5 are drawn is different when following 10 or 20: compare 17 and 27.

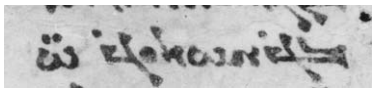
Dual-joining Manichaean Numbers

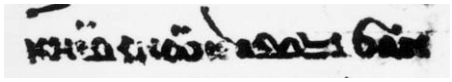
Character	X_n	X_r	X_m	X_l
ONE				
FIVE				
TEN				
TWENTY				

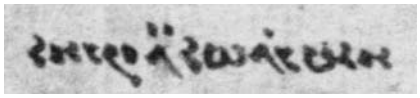
Right-joining Manichaean Number

Character	X_n	X_r
ONE HUNDRED		

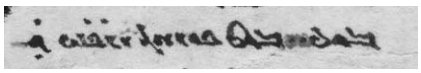
6. Diacritical marks. U+10AE5 ◌̈ MANICHAEAN COMBINING ABBREVIATION MARK is used with 𐫥 SHIN *š* and 𐫦 WAW *w*, in the combinations *š*, *š'n* and *w*. The dots indicate an abbreviation of the normal spellings *'wš*, *'wš'n* and *'wd*. The common factor here is the conjunction *ud* 'and' on its own or with the enclitic pronouns *-š* 'his, her, its' and *-šān* 'theirs' attached. As will be seen below, this character can also serve to indicate plurality, as a substitute for U+10AE6 ◌̤ MANICHAEAN COMBINING PLURAL MARK. The references before the transliterations (such as "M5/R/ii/18/") are to Manichaean scriptures and fragments.

š 𐫥̈  M5/R/ii/18/ *bgr'štygr š*

š'n 𐫡𐫦𐫥̈  M2/II/R/ii/34/ *'wṯ bst𐫵 š'n frh'*

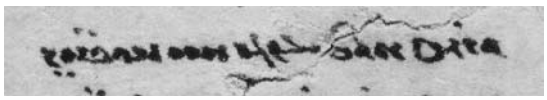
ẅ 𐫦̈  M7/II/V/i/27/ *'c rwšn ẅ yzd'n*

Like the previous character, U+10AC8 𐫫 MANICHAEAN LETTER UD is only used to indicate to the word *ud*, and it does not occur when enclitic pronouns other than *-š* and *-šān* are attached to it:

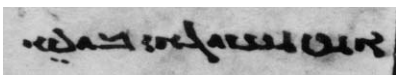
ụ̈ 𐫫̤  M39/R/i/7/ *bwj𐫵 bwṯ pyd'g drfš ụ̈*

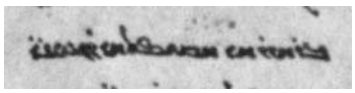
A combining diacritic has not been proposed because it would be used only with one character. This proposal is analogous to the Sindhi word abbreviation U+06FD.

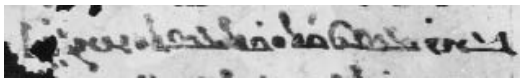
U+10AE6 ◌̤ MANICHAEAN COMBINING PLURAL MARK is also used to indicate that a spelling has been shortened; it is frequently used at the end of the manuscript line to indicate that the scribe has shortened a word to fit it in. The shortening frequently involves the plural ending in 𐫡𐫦- *-'n* which is reduced to *n* with dots placed below it. It is this usage from which the name for this character has been derived. Although the shortening very often involves leaving out an 𐫡 ALEF *'*, the dots cannot be taken to signify a missing ALEF because shortening occasionally involves leaving out other letters.

shortening *-y'* 𐫡𐫵 to *ÿ* 𐫵̤  M39/R/i/18/ *wrc 'wṯ 'zgd 'yy hwfrÿd* for *fry'd*

The position of the dots is important; compare this abbreviation with JHAYIN 𐫚̈ *j̈* above:

shortening of *-j'* to *j̤* 𐫚̤  M42/V/i/5/ *'c hndwg'n bẇjd* for *bwj'd*

shortening of *-'n* to *n̈* 𐫦̈  M1/368/ *br'dr'n 'mwst'n rwšn̈n* for *rwšn'n*

shortening other than of *-'-*  M34/V/7/ *b'r 'stftyft ° 'styh'g(f̈t)* for *'styh'gyft*

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. Punctuation. A variety of punctuation marks is used: 𐫰 𐫱 𐫲 𐫳 𐫴 𐫵 𐫶. Often part of the punctuation is written in red; this behaviour is outside the scope of character encoding. The punctuation system was elaborated quite clearly by the Manichaeans. The size and shape of dots was significant, and this has been taken over into Manichaean typography. The punctuation forms *a coherent set*. This set is unrelated to the punctuation which developed in the European typographic tradition. We can see no benefit to trying to unify some of these with existing characters (since others will certainly remain un-unified) and have a very strong preference for a single script-specific set to be encoded.

U+10AF0 𐫰 MANICHAEAN PUNCTUATION STAR is used to mark the beginning and end of headlines.

U+10AF1 𐫱 MANICHAEAN PUNCTUATION FLEURON (a black dot surrounded by petals often in red or blue) is used to mark the beginning and end of headlines and captions.

U+10AF2 𐫲 MANICHAEAN PUNCTUATION DOUBLE DOT WITHIN DOT (two black dots surrounded by red circles) is used to indicate larger units of text in a prose text or the end of a strophe in a verse text. This kind of division can *also* be indicated by using a sequence 𐫳𐫳 of U+10AF3 𐫳 MANICHAEAN PUNCTUATION DOT WITHIN DOT; we prefer to have the DOUBLE DOT WITHIN DOT ENCODED uniquely because without an explicit character one would have to resort to a ligation mechanism like ZWJ to form the joined pair—but this kind of ligation of punctuation would be unprecedented in the UCS. The user should be able to choose between 𐫲 and 𐫳𐫳.

U+10AF3 𐫳 MANICHAEAN PUNCTUATION DOT WITHIN DOT (one black dot surrounded by a red circle) is used to indicate smaller units of text in a prose text or the end of a half-verse in a verse text.

U+10AF4 𐫴 MANICHAEAN PUNCTUATION DOT is used to indicate sub-units of text, logical parts of a sentence or units in a list. It is not a word separator.

U+10AF5 𐫵 MANICHAEAN PUNCTUATION TWO DOTS is similar to U+10AF1 𐫱 MANICHAEAN PUNCTUATION FLEURON, just placed vertically, usually with red circles. It is used to mark the beginning and end of headlines and captions.

U+10AF6 𐫶 MANICHAEAN PUNCTUATION LINE FILLER is used as a sort of ellipsis to fill out a line. See Figures 6 and 7.

8. Unicode Character Properties

```
10AC0;MANICHAEAN LETTER ALEPH;Lo;0;R;;;;;N;;;;;
10AC1;MANICHAEAN LETTER BETH;Lo;0;R;;;;;N;;;;;
10AC2;MANICHAEAN LETTER BHETH;Lo;0;R;;;;;N;;;;;
10AC3;MANICHAEAN LETTER GIMEL;Lo;0;R;;;;;N;;;;;
10AC4;MANICHAEAN LETTER GHIMEL;Lo;0;R;;;;;N;;;;;
10AC5;MANICHAEAN LETTER DALETH;Lo;0;R;;;;;N;;;;;
10AC6;MANICHAEAN LETTER HE;Lo;0;R;;;;;N;;;;;
10AC7;MANICHAEAN LETTER WAW;Lo;0;R;;;;;N;;;;;
10AC8;MANICHAEAN LETTER UD;Lo;0;R;;;;;N;;;;;
10AC9;MANICHAEAN LETTER ZAYIN;Lo;0;R;;;;;N;;;;;
10ACA;MANICHAEAN LETTER ZHAYIN;Lo;0;R;;;;;N;;;;;
10ACB;MANICHAEAN LETTER JAYIN;Lo;0;R;;;;;N;;;;;
10ACC;MANICHAEAN LETTER JHAYIN;Lo;0;R;;;;;N;;;;;
```

```
10ACD;MANICHAEAN LETTER HETH;Lo;0;R;;;;;N;;;;;
10ACE;MANICHAEAN LETTER TETH;Lo;0;R;;;;;N;;;;;
10ACF;MANICHAEAN LETTER YODH;Lo;0;R;;;;;N;;;;;
10AD0;MANICHAEAN LETTER KAPH;Lo;0;R;;;;;N;;;;;
10AD1;MANICHAEAN LETTER XAPH;Lo;0;R;;;;;N;;;;;
10AD2;MANICHAEAN LETTER KHAPH;Lo;0;R;;;;;N;;;;;
10AD3;MANICHAEAN LETTER LAMEDH;Lo;0;R;;;;;N;;;;;
10AD4;MANICHAEAN LETTER DHAMEDH;Lo;0;R;;;;;N;;;;;
10AD5;MANICHAEAN LETTER THAMEDH;Lo;0;R;;;;;N;;;;;
10AD6;MANICHAEAN LETTER MEM;Lo;0;R;;;;;N;;;;;
10AD7;MANICHAEAN LETTER NUN;Lo;0;R;;;;;N;;;;;
10AD8;MANICHAEAN LETTER SAMEKH;Lo;0;R;;;;;N;;;;;
10AD9;MANICHAEAN LETTER AYIN;Lo;0;R;;;;;N;;;;;
10ADA;MANICHAEAN LETTER AAYIN;Lo;0;R;;;;;N;;;;;
10ADB;MANICHAEAN LETTER PE;Lo;0;R;;;;;N;;;;;
10ADC;MANICHAEAN LETTER FE;Lo;0;R;;;;;N;;;;;
10ADD;MANICHAEAN LETTER SADHE;Lo;0;R;;;;;N;;;;;
10ADE;MANICHAEAN LETTER QOPH;Lo;0;R;;;;;N;;;;;
10ADF;MANICHAEAN LETTER XOPH;Lo;0;R;;;;;N;;;;;
10AE0;MANICHAEAN LETTER QHOPH;Lo;0;R;;;;;N;;;;;
10AE1;MANICHAEAN LETTER RESH;Lo;0;R;;;;;N;;;;;
10AE2;MANICHAEAN LETTER SHIN;Lo;0;R;;;;;N;;;;;
10AE3;MANICHAEAN LETTER SSHIN;Lo;0;R;;;;;N;;;;;
10AE4;MANICHAEAN LETTER TAW;Lo;0;R;;;;;N;;;;;
10AE5;MANICHAEAN COMBINING ABBREVIATION MARK;Mn;230;NSM;;;;;N;;;;;
10AE6;MANICHAEAN COMBINING PLURAL MARK;Mn;220;NSM;;;;;N;;;;;
10AEB;MANICHAEAN NUMBER ONE;No;0;R;;;;1;N;;;;;
10AEC;MANICHAEAN NUMBER FIVE;No;0;R;;;;5;N;;;;;
10AED;MANICHAEAN NUMBER TEN;No;0;R;;;;10;N;;;;;
10AEE;MANICHAEAN NUMBER TWENTY;No;0;R;;;;20;N;;;;;
10AEF;MANICHAEAN NUMBER ONE HUNDRED;No;0;R;;;;100;N;;;;;
10AF0;MANICHAEAN PUNCTUATION STAR;Po;0;AL;;;;;N;;;;;
10AF1;MANICHAEAN PUNCTUATION FLEURON;Po;0;AL;;;;;N;;;;;
10AF2;MANICHAEAN PUNCTUATION DOUBLE DOT WITHIN DOT;Po;0;AL;;;;;N;;;;;
10AF3;MANICHAEAN PUNCTUATION DOT WITHIN DOT;Po;0;AL;;;;;N;;;;;
10AF4;MANICHAEAN PUNCTUATION DOT;Po;0;AL;;;;;N;;;;;
10AF5;MANICHAEAN PUNCTUATION TWO DOTS;Po;0;AL;;;;;N;;;;;
10AF6;MANICHAEAN PUNCTUATION LINE FILLER;Po;0;AL;;;;;N;;;;;
```

9. Bibliography

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.

Faulmann, Carl. 1990 (1880). *Das Buch der Schrift*. Frankfurt am Main: Eichborn. ISBN 3-8218-1720-8

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.

11. Acknowledgements. This project was made possible in part by a grant from the U.S. National Endowment for the Humanities, which funded the which funded the Universal Scripts Project (part of the Script Encoding Initiative at UC Berkeley) in respect of the Manichaean encoding. Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment of the Humanities.

Figures

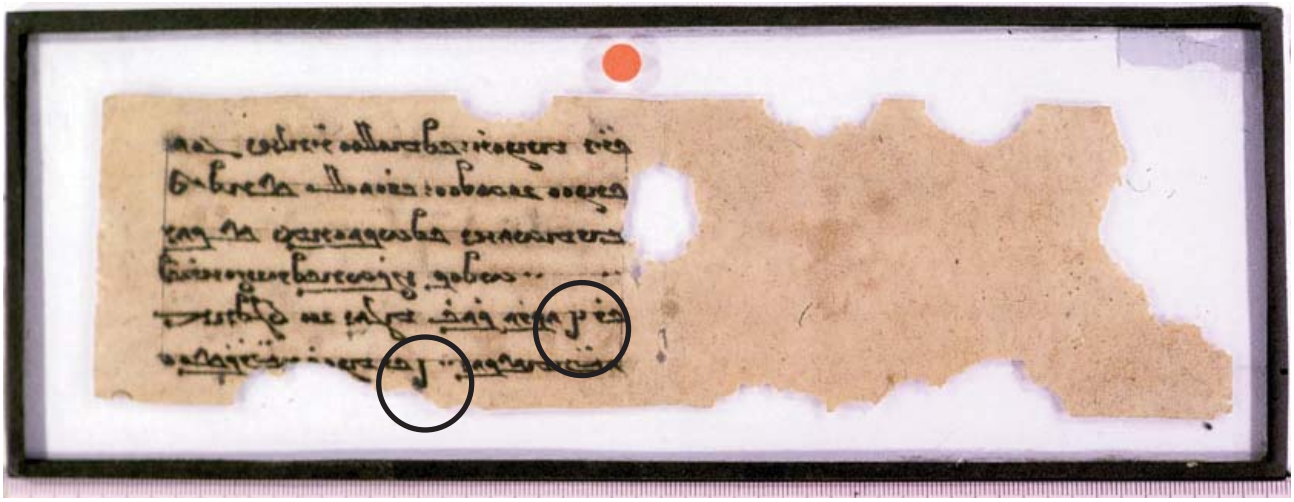


Figure 1. One side of the Manichaean manuscript page M113. The numbers 𐫁 1 and 𐫂 2 are circled.

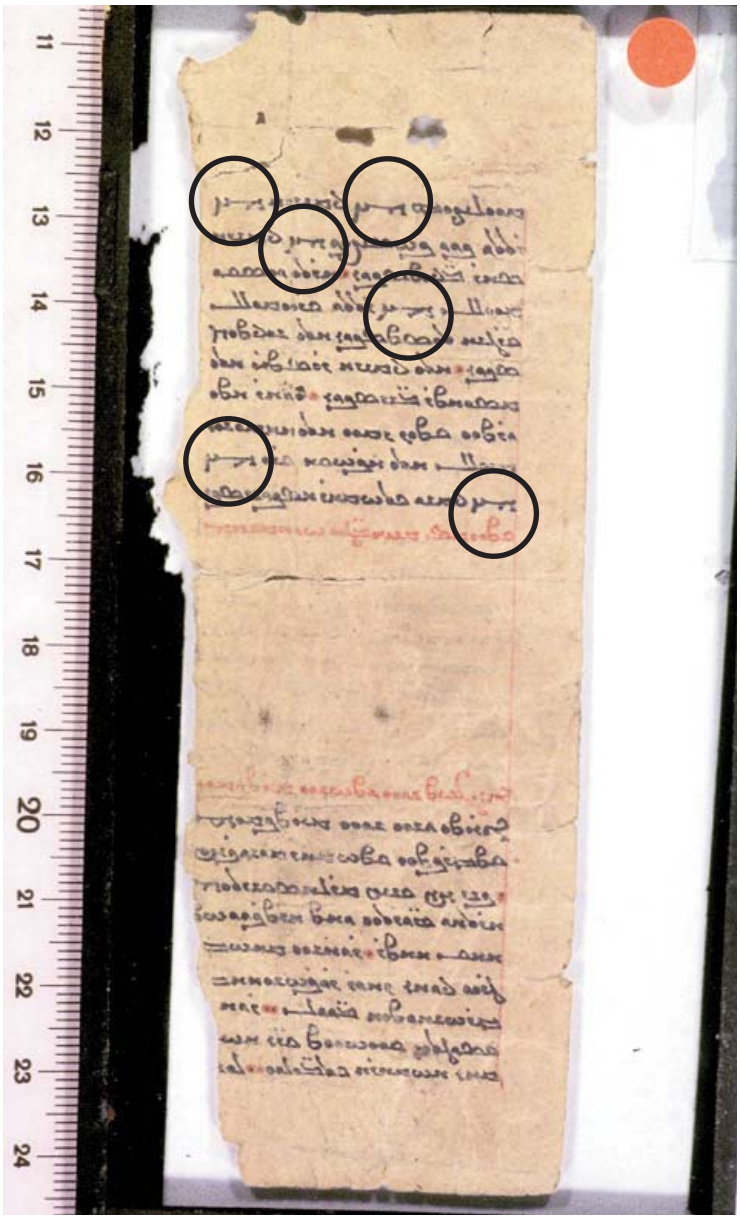


Figure 2. One side of the Manichaean manuscript page M14, showing the number 𐫄𐫁 12 in lines1, 2, 4, 9, and 10.

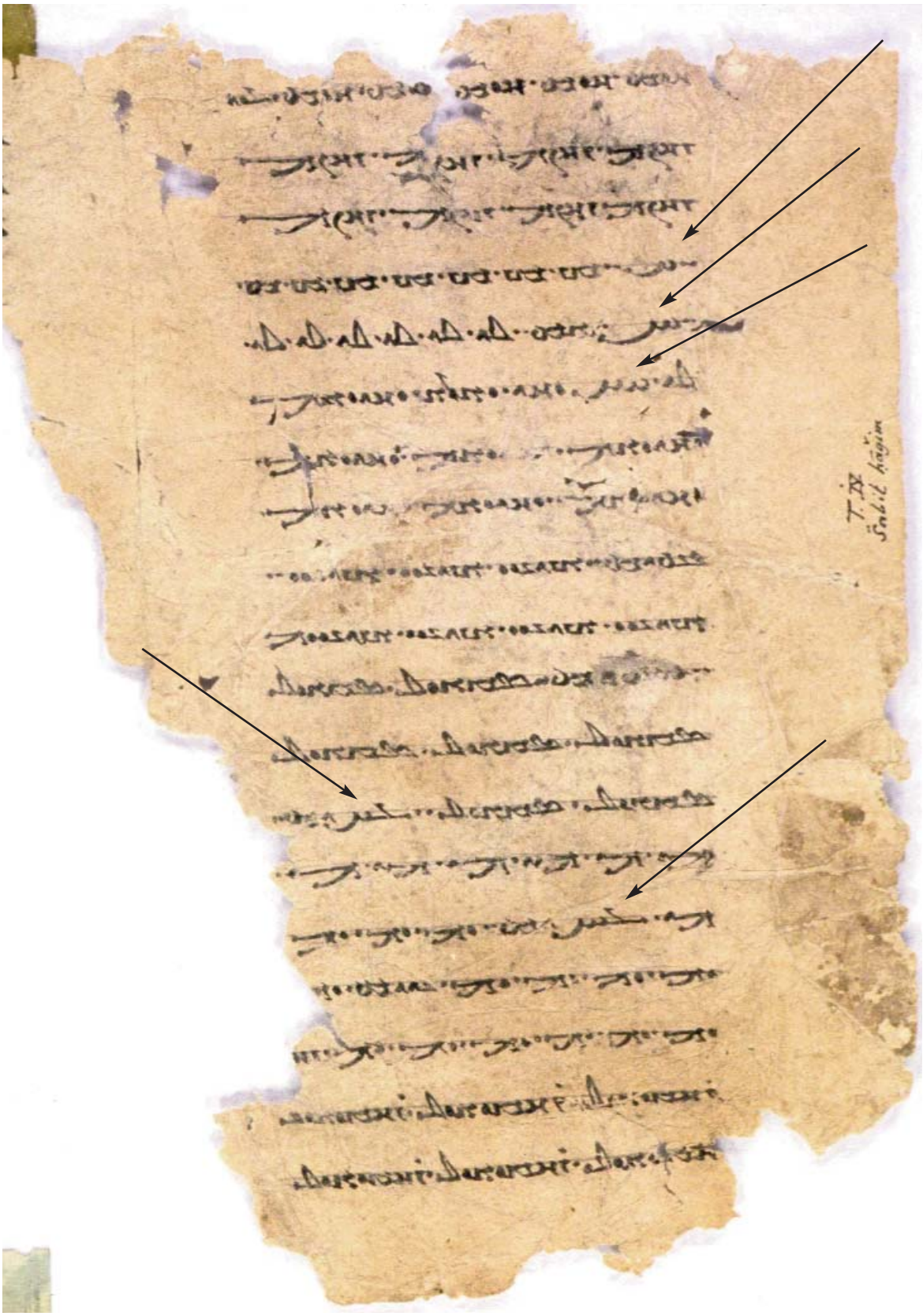


Figure 3. One side of the Manichaean manuscript page M8430, showing the numbers 𐫂 2, 𐫃 3, 𐫄 4, 𐫇 7, and 𐫈 8.

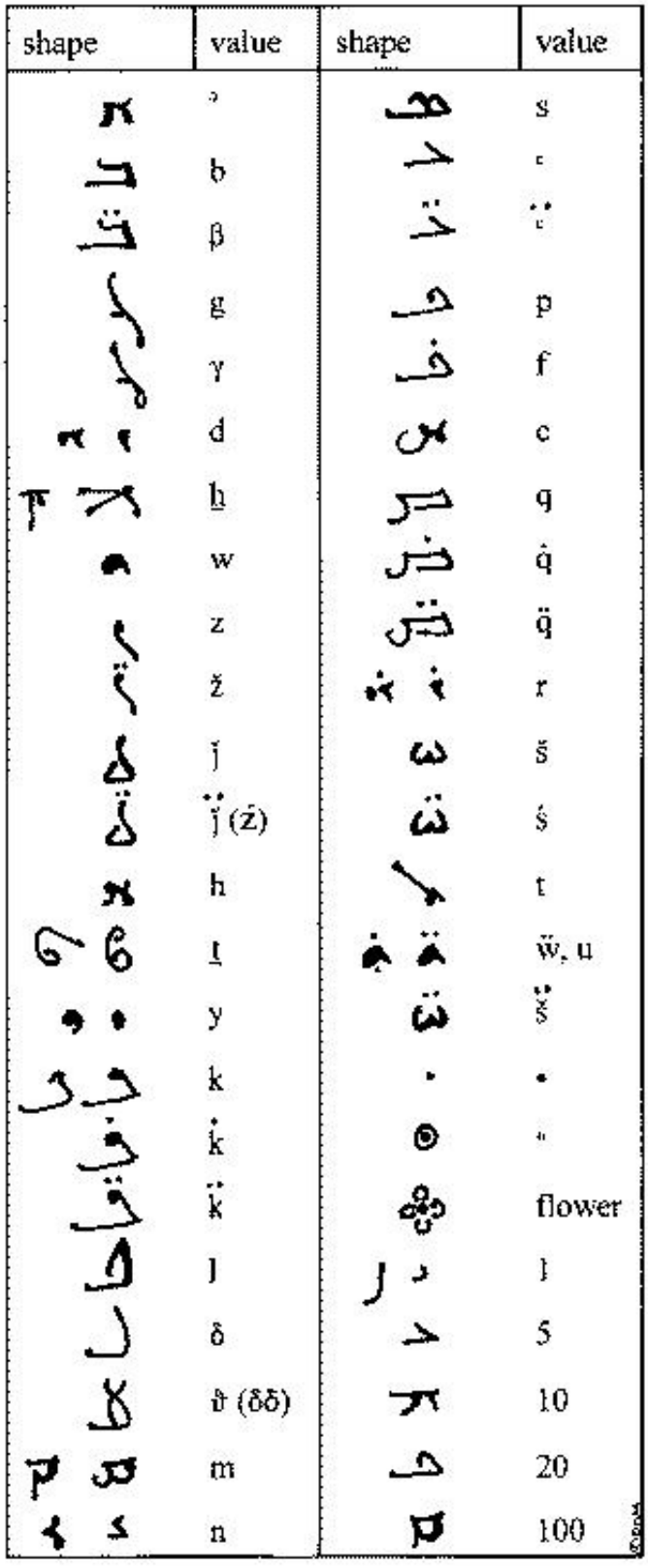
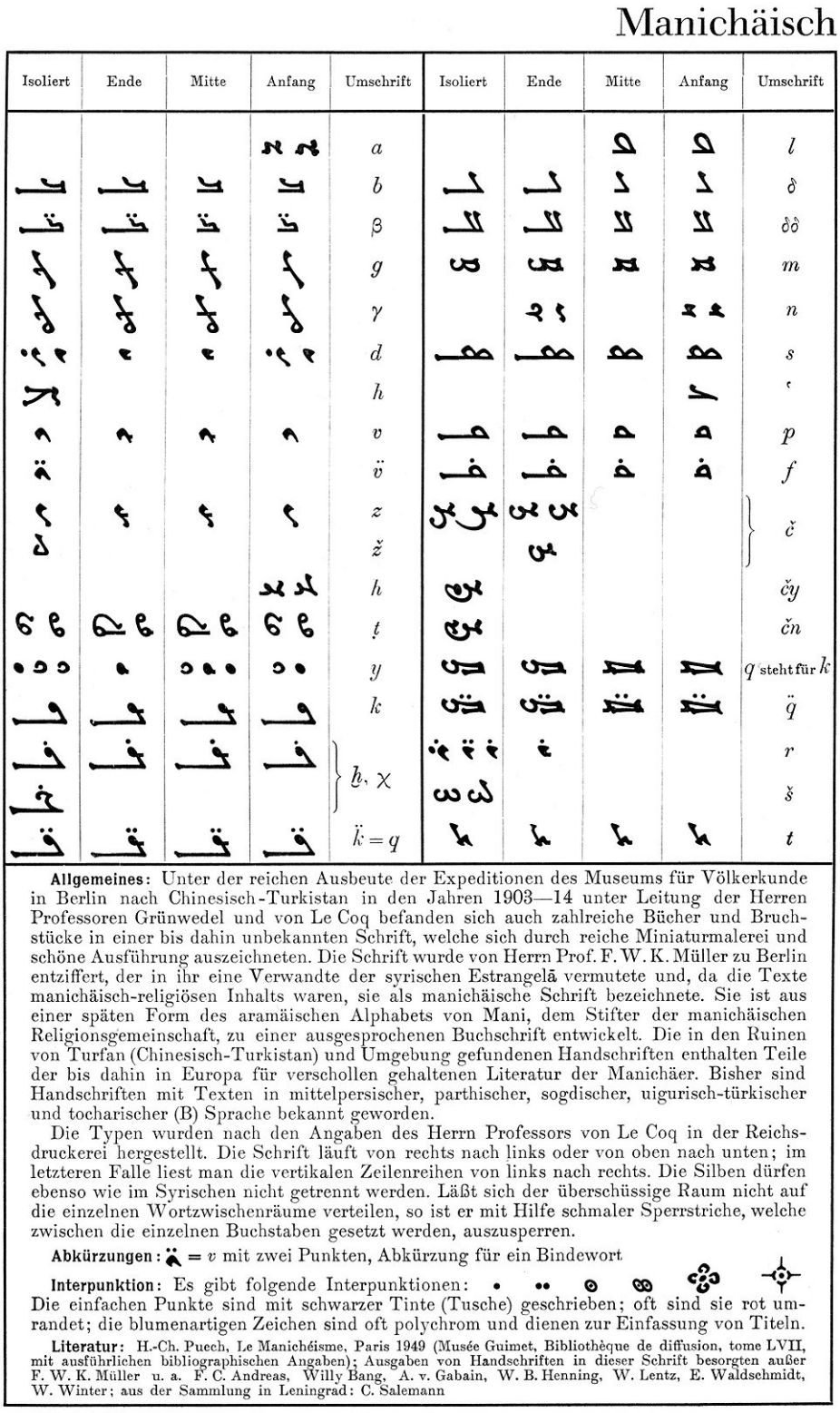


Figure 4. Table of the Manichaean script by Desmond Durkin-Meisterernst.



Manichäisch

Isoliert	Ende	Mitte	Anfang	Umschrift	Isoliert	Ende	Mitte	Anfang	Umschrift
				a					*l*
				b					*δ*
				β					*δδ*
				g					*m*
				γ					*n*
				d					*s*
				h					*ʿ*
				v					*p*
				v̈					*f*
				z					*č*
				ž					
				h					*čy*
				ṭ					*čn*
				y					*q* steht für *k*
				k					*q̈*
				h̲, *χ*					*r*
									š
				k̈ = *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 5. Description of Manichaean script from a German source. In the description of the punctuation a pair of thick dots is shown; in encoding this would be a sequence (••) of two U+10AF4 • MANICHAEAN PUNCTUATION SINGLE THICK DOT characters.

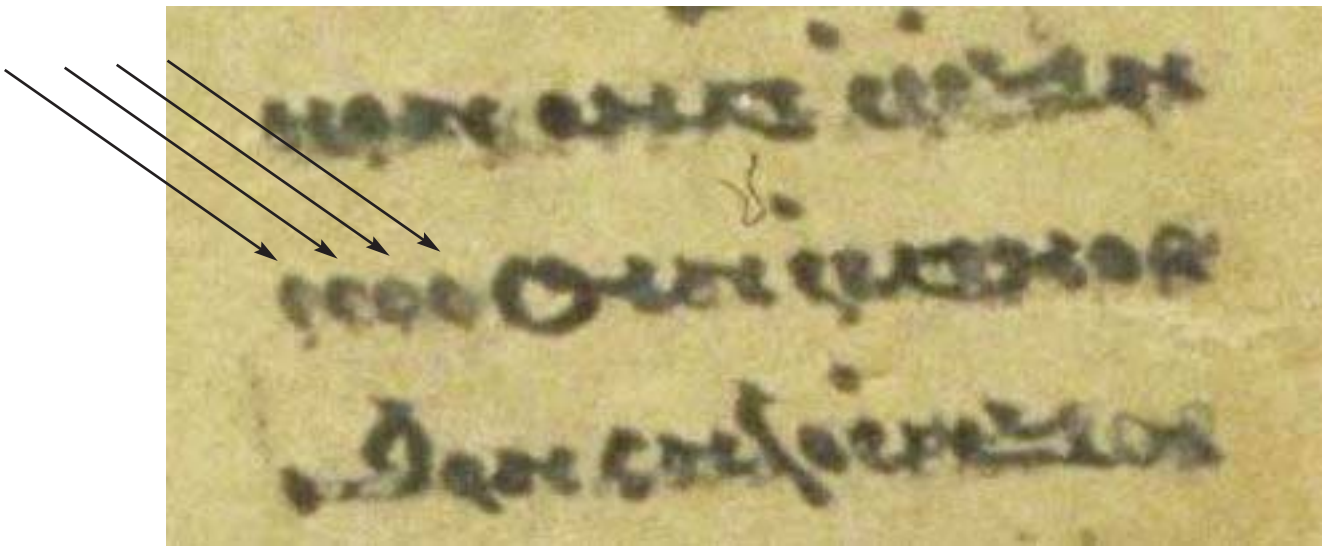


Figure 6. Example of the line filler in use in manuscript M7981/II/R/i/23/. The text reads

← / / ←

→ *’brdr mhy ’wd / wyhmdr ’c **** / ’b’ryg’n ’wl*

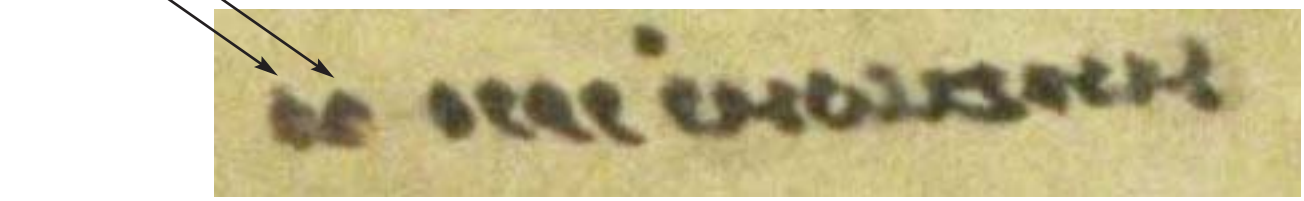
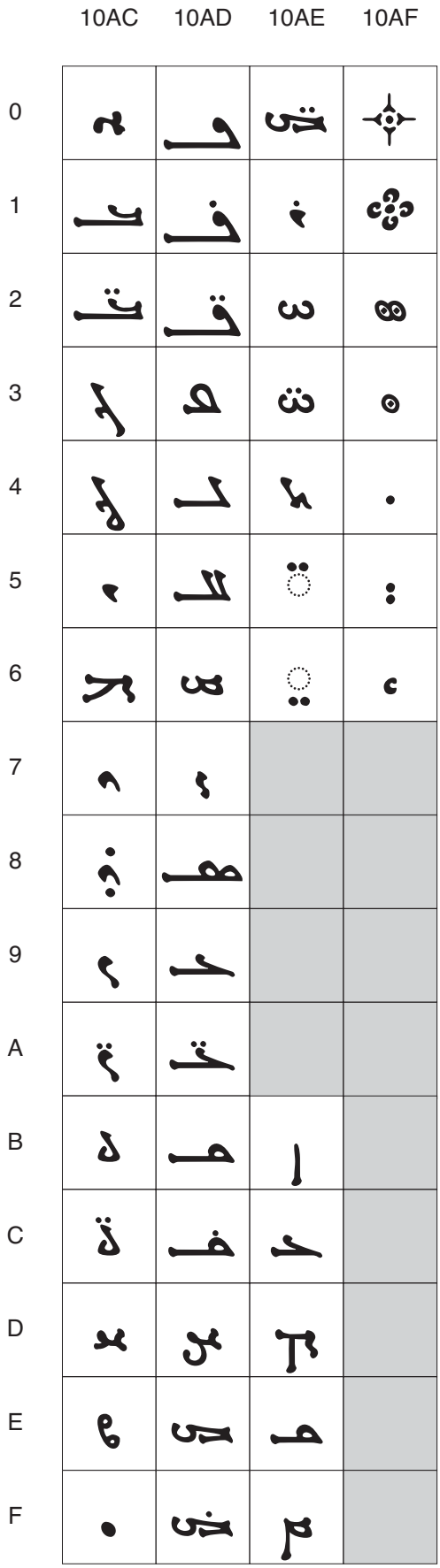


Figure 7. Examples of the line-filler in use in manuscript M7981/II/R/ii/29/. The text reads

←

→ *h’mšhr dwdy ***

Row 10A: MANICHAEAN DRAFT



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 UD
C9	MANICHAEAN LETTER ZAYIN
CA	MANICHAEAN LETTER ZHAYIN
CB	MANICHAEAN LETTER JAYIN
CC	MANICHAEAN LETTER JHAYIN
CD	MANICHAEAN LETTER HETH
CE	MANICHAEAN LETTER TETH
CF	MANICHAEAN LETTER YODH
D0	MANICHAEAN LETTER KAPH
D1	MANICHAEAN LETTER XAPH
D2	MANICHAEAN LETTER KHAPH
D3	MANICHAEAN LETTER LAMEDH
D4	MANICHAEAN LETTER DHAMEDH
D5	MANICHAEAN LETTER THAMEDH
D6	MANICHAEAN LETTER MEM
D7	MANICHAEAN LETTER NUN
D8	MANICHAEAN LETTER SAMEKH
D9	MANICHAEAN LETTER AYIN
DA	MANICHAEAN LETTER AAYIN
DB	MANICHAEAN LETTER PE
DC	MANICHAEAN LETTER FE
DD	MANICHAEAN LETTER SADHE
DE	MANICHAEAN LETTER QOPH
DF	MANICHAEAN LETTER XOPH
E0	MANICHAEAN LETTER QHOPH
E1	MANICHAEAN LETTER RESH
E2	MANICHAEAN LETTER SHIN
E3	MANICHAEAN LETTER SSHIN
E4	MANICHAEAN LETTER TAW
E5	MANICHAEAN COMBINING ABBREVIATION MARK
E6	MANICHAEAN COMBINING PLURAL MARK
E7	(This position shall not be used)
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 DOT
F5	MANICHAEAN PUNCTUATION TWO DOTS
F6	MANICHAEAN PUNCTUATION LINE FILLER
F7	(This position shall not be used)
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 for encoding the Manichaean script in the SMP of 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

2009-05-04

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

Manichaean.

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

53.

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

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?