Doc Type: **Working Group Document**
Title: **Second 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, Desmond Durkin-Meisterernst, Roozbeh Pournader,**
 and Shervin Afshar
Status: **Liaison Contribution**
Action: **For consideration by JTC1/SC2/WG2 and UTC**
Replaces: **N2544, N3378, N3644R**
Date: **2011-05-10**

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 sometime after 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 from 762 until the beginning of the 11th century 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. Like Syriac, with which Manichaean shares some glyph shapes, the Manichaean script evolved from Aramaic. 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 ten letters (seven with two dots and three with one dot) are encoded explicitly. These letters do not have decompositions. Five characters have variant forms which are significant but unpredictable; a variation selector is specified to invoke this special shaping behaviour. There are two diacritical marks which indicate abbreviation, plurality, or the conjunction *ud*.

3. Names and ordering. Letter names for Manichaean not attested. For this encoding, 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 ﺐ BETH *b*, ﺑ KAPH *k*, � ZAYIN *z*, ﺑ JAYIN *j*, ﺑ QOPH *q*, the marked forms are ﺘ

BHETH *β*, ‏ﻘ‎ KHAPH *k̈*, ‏ﺫ‎ ZHAYIN *ž*, ‏ﺩ‎ JHAYIN *j̇*, ‏ﺨﻗ‎ QHOPH *q̈*. For the letters ‏ﺣ‎ AYIN ʻ, and ‏ﺳ‎ SHIN *š*, where -H- does not make sense, the initial letter has been doubled, ‏ﺣ‎ AAYIN *ʺ*, ‏ﺳ‎ SSHIN *ś*. The -H- is used in some other letters, such as ‏ﻍ‎ GHIMEL *γ*, ‏ﺪ‎ DHAMEDH *δ*, ‏ﺍ‎ THAMEDH *θ ~ δδ* (from ‏ﻍ‎ GIMEL *g*, ‏ﻝ‎ LAMEDH *l*); in letters with a single dot a letter is simply changed, as in ‏ﻒ‎ FE *f*, ‏ﻨ‎ XAPH *k*, ‏ﺨﻗ‎ XOPH *q̇* (from ‏ﻊ‎ PE *p*, ‏ﻙ‎ KAPH *k*, ‏ﻗ‎ 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. Note that Manichaean has two characters of Joining_Type=Left_Joining—the first characters to be encoded with this property value. Although this property value is foreseen in *The Unicode Standard*, section 8.2, some implementations of the Arabic/Syriac/N'Ko/Mandaic joining may have been using some optimizations without considering that such property values may exist. Such implementations may need to change in order to be able to handle Manichaean properly. (This is somewhat comparable to old implementations not supporting non-BMP characters until graphical characters were encoded outside the BMP.)

Dual-joining Manichaean Characters

Character	X_n	X_r	X_m	X_l
ALEPH				
BETH				
BHETH				
GIMEL				
GHIMEL				
LAMEDH				
DHAMEDH				
THAMEDH				
MEM				
SAMEKH				
AYIN				
AAYIN				
PE				
FE				
QOPH				
XOPH				
QHOPH				

Right-joining Manichaean Characters

Character	X_n	X_r
DALETH		
WAW		
ZAYIN		
ZHAYIN		
TETH		
YODH		
KAPH		
XAPH		
KHAPH		
SADHE		
RESH		
TAW		

Left-joining Manichaean Characters

Character	X$_n$	X$_l$
HETH	ᵚ	ᵚ
NUN	ʒ /ʓ	ʒ

Although ALEPH is dual joining, it usually joins on the right *only* to BETH or BHETH (though sometimes also GIMEL). It does not usually join to the right to other dual-joining or left-joining letters; in order to break this joining, U+200C ZERO WIDTH NON-JOINER should be used to break the connection. It is likely that ZERO WIDTH NON-JOINER will be a fairly common character in Manichaean texts, similar to the frequent use of the same character in various languages written in the Arabic script, like Persian, Urdu, Kurdish, etc. This will simplify the implementation of Manichaean script in computers. In this way, Manichean will follow the exact same joining and shaping algorithm as specified in the standard for scripts such as Arabic, Syriac, N'Ko, and Mandaic. It might be appropriate for intelligent user-friendly keyboards for entering Manichaean data to automate the insertion of some ZERO WIDTH NON-JOINER characters, for example when ALEPH follows a right join-causing character other than BETH, BHETH, or TATWEEL.

Similarly, where occasional "touching" between a joining character and a non-joining characters is desired, U+200C ZERO WIDTH JOINER should be used to change the shape of the joining character. Although this will not change the shape of the non-joining character or make the two characters ligate, it could simulate the desired behavior.

Non-joining Manichaean Characters

Character	X$_n$
HE	ᴛᴦ/ᴧ
JAYIN	ᕍ
JHAYIN	ᕍ̈
SHIN	ɯ
SSHIN	ɯ̈
UD	ᷡ

4.1 Five Manichaean characters have special alternate forms that occur in text.

Character	Normal form	Alternate form
DALETH isolate	᷍	᷍
HE isolate	ᴧ	ᴛᴦ
MEM isolate	ɯ	ᴘ
MEM final	ɯ	ᴘ
NUN isolate	ᷤ	ʒ
RESH isolate	᷍	᷍

These alternate forms tend to occur at the end of lines, though their occurrence is not predictable—occurrence is not conditioned by line ending, word position or other character contexts. Since they can co-occur in text with the corresponding "normal" forms, they cannot be considered font variants; both the normal and alternate forms need to be supported in a single font.

These variants are considered significant to Manichaean researchers involved in digitizing historic texts. For this reason, distinct encoded representations in plain text is required. Font rendering mechanisms (for example, discretionary OpenType features or other layout mechanisms) are not considered adequate for their needs. While this could be achieved by encoding these as distinct characters, that would break

recognized character unity. A variation-selector mechanism is preferred. For this reason, the following variation sequences using U+FE00 VARIATION SELECTOR-1 are proposed:

DALETH   + VS-1 = *alternate-form daleth*

HE   + VS-1 = *alternate-form he*

MEM   + VS-1 = *alternate-form mem*

NUN   + VS-1 = *alternate-form nun*

RESH   + VS-1 = *alternate-form resh*

The various shaping behaviours of these five characters was described above (DALETH is right-joining, HE is non-joining, etc.). The variation sequences interact with shaping behaviours in that alternate variant forms occur only in certain word-position contexts. Details of the shaping behaviour for these five characters is repeated here, only now clarifying the interaction with VS-1:

Character	X_n	X_r	X_m	X_l
DALETH	ٮ	ٮ		
DALETH + VS-1	٠ٮ	ٮ		
HE	ﮥ			
HE + VS-1	ﮥ			
MEM	ﻤ	ﻤ	ﻤ	ﻤ
MEM + VS-1	ﻤ	ﻤ	ﻤ	ﻤ
NUN	ٮ			ٮ
NUN + VS-1	ٮ			ٮ
RESH	ٮ	ٮ		
RESH + VS-1	٠ٮ	ٮ		

Note that the variation sequences do not change the basic shaping behaviours (joining type and joining group) of the characters; only the specific glyphs for particular contexts is changed.

The 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 final # MANICHAEAN LETTER MEM
10AD7 FE00; alternate form; isolate # MANICHAEAN LETTER NUN
10AE1 FE00; alternate form; isolate # MANICHAEAN LETTER RESH
```

The text for StandardizedVariants.html is as follows:

Rep Glyph	Character Sequence	Context	Alt Glyph	Description of variant appearance
ٮ	10AC5 FE00	isolate	٠ٮ	MANICHAEAN LETTER DALETH alternate form
ﮥ	10AC6 FE00	isolate	ﮥ	MANICHAEAN LETTER HE alternate form
ﻤ	10AD6 FE00	isolate	ﻤ	MANICHAEAN LETTER MEM alternate form
		final	ﻤ	
ٮ	10AD7 FE00	isolate	ٮ	MANICHAEAN LETTER NUN alternate form
ٮ	10AE1 FE00	isolate	٠ٮ	MANICHAEAN LETTER RESH alternate form

We have briefly considered, 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 obligatory ligatures; this means that the combinations SADHE + YODH and SADHE + NUN always results in a ligature.

SADHE$_n$ ꞃ	+	YODH ●	=	$čy_n$ ꞃ
SADHE$_r$ ꞃ	+	YODH ●	=	$čy_r$ ꞃ
SADHE$_n$ ꞃ	+	NUN ،	=	$čn_n$ ꞃ
SADHE$_r$ ꞃ	+	NUN ،	=	$čn_r$ ꞃ

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. Mandaic also has a similar requirement. The data file ScriptExtensions.txt would need to be changed to say:

```
0640          ; Arab Mand Mani Syrc # Lm       ARABIC TATWEEL
```

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. Unfortunately very few Manichaean numbers are attested. The numbers ꞃ 10, ꞃ 20, and ꞃ 100 are similar in shape to Manichaean letters (ꞃ HE, ꞃ PE, ꞃ MEM) but are different in behaviour; their glyphs were re-analysed from the original Aramaic prototypes. 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	ꞃ	1 ←	M283 II V 4
2	ꞃ	1 + 1 ←	
3	ꞃ	1 + 1 + 1 ←	M67 R ii 11
4	ꞃ	1 + 1 + 1 + 1 ←	M74 II R 18
7	ꞃ	1 + 1 + 5←	
8	ꞃ	1 + 1 + 1 + 5 ←	
12	ꞃ	1 + 1 + 10 ←	M14 R 1, 2, 4, 9, 10
15	ꞃ	5 + 10 ←	M5750 R ii 21
68	ꞃ	1 + 1 + 1 + 5 + 20 + 20 + 20 ←	M1 390
77	ꞃ	1 + 1 + 5 + 10 + 20 + 20 + 20 ←	M1 321
162	ꞃ	1 + 1 SPACE 20 + 20 + 20 SPACE 100 ←	M1 167
546	ꞃ	1 + 5 + 20 + 20 [linebreak] 100 + 5 ←	M1 160–161

Note that the height at which ꞃ 1 and ꞃ 5 are drawn is different when following ꞃ 10 from the way they are drawn when following ꞃ 20: compare ꞃ 17 (which shows the 7 drawn high) and ꞃ 27 (which shows the 7 drawn on the baseline with ꞃ 20); *ꞃ is incorrect for 17. (A font could use higher glyph shapes or technologies like OpenType GPOS tables for these numbers: ꞃ 11, ꞃ 12, ꞃ 13, ꞃ 14, ꞃ 15, ꞃ 16, ꞃ 17, ꞃ 18, and ꞃ 19.)

Dual-joining Manichaean Numbers

Character	X$_n$	X$_r$	X$_m$	X$_l$
ONE	ꞃ	ꞃ	ꞃ	ꞃ
FIVE	ꞃ	ꞃ	ꞃ	ꞃ
TEN	ꞃ	ꞃ	ꞃ	ꞃ
TWENTY	ꞃ	ꞃ	ꞃ	ꞃ

Character	X_n	X_r
ONE HUNDRED	ℙ	ℙ

Note that some of the glyphs in the table above have not been attested in the historical corpus of Manichaean. These are

Character	X_n	X_r	X_m	X_l
FIVE	(⌣)			
TEN	(𝗧𝗕)	(𝗧ᵉ)		
TWENTY	(⌣)			

The reconstructed forms here of TEN_n and TEN_r are based on reasonable expectations given the similarity of the base number and the letter HE. Note also that, while 100 is currently right-joining based on the limited evidence of the historical corpus, there is some chance that it could be dual joining, in which case its medial and left-joining forms could look something like the analogous forms for MEM.

6. Diacritical marks. U+10AE5 ⃛ MANICHAEAN ABBREVIATION MARK ABOVE is used with ᴡ SHIN *š* and ᴧ 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 ABBREVIATION MARK BELOW. 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 ⃛   M2/II/R/ii/34/ *'wt bst꙰ š'n frh'*

ẅ ⃛   M7/II/V/i/27/ *'c rwšn ẅ yzd'n*

U+10AC8 ⃛ MANICHAEAN SIGN UD is a particular spelling for the word *ud* 'and'. This character is not used when enclitic pronouns other than -*š* and -*šān* are attached to it (in which case U+10AE5 ⃛ MANICHAEAN ABBREVIATION MARK ABOVE is used as described above). An imperfect analogy might be English *'n*, *an'*, and *&*, all of which mean *and*).

ẇ ⃛   M39/R/i/7/ *bwj꙰ bwt pyd'g drfš ẇ*

A combining diacritic has not been proposed because it would be used only with one character.

U+10AE6 ⃛ MANICHAEAN ABBREVIATION MARK BELOW 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 ꙰ -*'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 ꙰ ALEF *'*, the dots cannot be taken to signify a missing ALEF because shortening occasionally involves leaving out other letters.

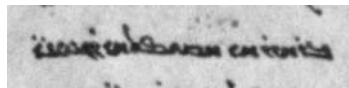|		
shortening -y' ✦ to ÿ ⠂		M39/R/i/18/ *wrc 'wt̠ 'zgd 'yy hwfrÿd* for *hwfry'd*
shortening of -j' to j̈ ◗		M42/V/i/5/ *'c hndwg'n bwj̈d* for *bwj'd*
shortening of -'n to n̈ ✦		M1/368/ *br'dr'n 'mwst'n rwšn̈n* for *rwšn'n*
shortening other than of -'-		M34/V/7/ *b'r 'st̠ftyft ° 'styh'g(f̈t)* for *'styh'gyft*
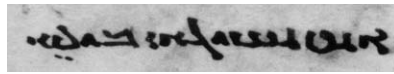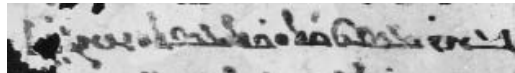
The illustrations here are taken from W. Sundermann, *Iranian Manichaean Turfan texts in early publications (1904-1934): Photo Edition*. London: School of Oriental and African Studies 1996 (CII Supplementary Series Vol. III).

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. It can be used in pairs •• as well as singly.

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. Line-Breaking. The letters and digits behave like letters, and will be the line breaking class AL (Alphabetic). The abbreviations marks will have the line breaking class CM (Combining Mark). The punctuations STAR, FLEURON, and TWO DOTS should have the line break class QU (Quotation), while DOUBLE DOT WITHIN DOT, DOT WITHIN DOT, and DOT should have the line break class EX (Exclamation/Interrogation). The LINE FILLER is a leader character, so it should have the property IN (Inseparable).

9. 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 SIGN UD;So;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 ABBREVIATION MARK ABOVE;Mn;230;NSM;;;;;N;;;;;
10AE6;MANICHAEAN ABBREVIATION MARK BELOW;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;R;;;;;N;;;;;
10AF1;MANICHAEAN PUNCTUATION FLEURON;Po;0;R;;;;;N;;;;;
10AF2;MANICHAEAN PUNCTUATION DOUBLE DOT WITHIN DOT;Po;0;R;;;;;N;;;;;
10AF3;MANICHAEAN PUNCTUATION DOT WITHIN DOT;Po;0;R;;;;;N;;;;;
10AF4;MANICHAEAN PUNCTUATION DOT;Po;0;R;;;;;N;;;;;
10AF5;MANICHAEAN PUNCTUATION TWO DOTS;Po;0;R;;;;;N;;;;;
10AF6;MANICHAEAN PUNCTUATION LINE FILLER;Po;0;R;;;;;N;;;;;
```

10. Unicode Joining Types and Groups. Note that although the basic structure of RESH looks very much like DALETH, they are not considered part of the same joining group because they look very different in their alternate form (when folllowed by VS-1).

```
10AC0; MANICHAEAN ALEPH; D; MANICHAEAN ALEPH
10AC1; MANICHAEAN BETH; D; MANICHAEAN BETH
10AC2; MANICHAEAN BETH WITH 2 DOTS ABOVE; D; MANICHAEAN BETH
```

```
10AC3; MANICHAEAN GIMEL; D; MANICHAEAN GIMEL
10AC4; MANICHAEAN GIMEL WITH ATTACHED RING BELOW; D; MANICHAEAN GIMEL
10AC5; MANICHAEAN DALETH; R; MANICHAEAN DALETH
10AC6; MANICHAEAN HE; U; No_Joining_Group
10AC7; MANICHAEAN WAW; R; MANICHAEAN WAW
10AC8; MANICHAEAN UD; U; No_Joining_Group
10AC9; MANICHAEAN ZAYIN; R; MANICHAEAN ZAYIN
10ACA; MANICHAEAN ZAYIN WITH 2 DOTS ABOVE; R; MANICHAEAN ZAYIN
10ACB; MANICHAEAN JAYIN; U; No_Joining_Group
10ACC; MANICHAEAN JAYIN WITH 2 DOTS ABOVE; U; No_Joining_Group
10ACD; MANICHAEAN HETH; L; MANICHAEAN HETH
10ACE; MANICHAEAN TETH; R; MANICHAEAN TETH
10ACF; MANICHAEAN YODH; R; MANICHAEAN YODH
10AD0; MANICHAEAN KAPH; R; MANICHAEAN KAPH
10AD1; MANICHAEAN KAPH WITH DOT ABOVE; R; MANICHAEAN KAPH
10AD2; MANICHAEAN KAPH WITH 2 DOTS ABOVE; R; MANICHAEAN KAPH
10AD3; MANICHAEAN LAMEDH; D; MANICHAEAN LAMEDH
10AD4; MANICHAEAN DHAMEDH; D; MANICHAEAN DHAMEDH
10AD5; MANICHAEAN THAMEDH; D; MANICHAEAN THAMEDH
10AD6; MANICHAEAN MEM; D; MANICHAEAN MEM
10AD7; MANICHAEAN NUN; L; MANICHAEAN NUN
10AD8; MANICHAEAN SAMEKH; D; MANICHAEAN SAMEKH
10AD9; MANICHAEAN AYIN; D; MANICHAEAN AYIN
10ADA; MANICHAEAN AYIN WITH 2 DOTS ABOVE; D; MANICHAEAN AYIN
10ADB; MANICHAEAN PE; D; MANICHAEAN PE
10ADC; MANICHAEAN PE WITH DOT ABOVE; D; MANICHAEAN PE
10ADD; MANICHAEAN SADHE; R; MANICHAEAN SADHE
10ADE; MANICHAEAN QOPH; D; MANICHAEAN QOPH
10ADF; MANICHAEAN QOPH WITH DOT ABOVE; D; MANICHAEAN QOPH
10AE0; MANICHAEAN QOPH WITH 2 DOTS; D ABOVE; MANICHAEAN QOPH
10AE1; MANICHAEAN RESH; R; MANICHAEAN RESH
10AE2; MANICHAEAN SHIN; U; No_Joining_Group
10AE3; MANICHAEAN SHIN WITH 2 DOTS ABOVE; U; No_Joining_Group
10AE4; MANICHAEAN TAW; R; MANICHAEAN TAW
10AEB; MANICHAEAN ONE; D; MANICHAEAN ONE
10AEC; MANICHAEAN FIVE; D; MANICHAEAN FIVE
10AED; MANICHAEAN TEN; D; MANICHAEAN TEN
10AEE; MANICHAEAN TWENTY; D; MANICHAEAN TWENTY
10AEF; MANICHAEAN HUNDRED; R; MANICHAEAN HUNDRED
```

11. "Confusability". Roozbeh Pournader wrote this section and is responsible for its content.

In-script:

```
10AC2 ≈ 10AC1 10AE5
10ACA ≈ 10AC9 10AE5
10AC5 FE00 ≈ 10AC9 10AF4
10ACC ≈ 10ACB 10AE5
10AD2 ≈ 10AD0 10AE5
10AD5 ≈ 10AD4 10AD4
10AD8 ≈ 10ADB 10ADB
10ADA ≈ 10AD9 10AE5
10ADB ≈ 10AD3
10AE0 ≈ 10ADE 10AE5
10AE1 FE00 ≈ 10AE1 10AF4
10AE3 ≈ 10AE2 10AE5
10AED ≈ 10AC6 FE00
10AEE ≈ 10ADB
10AEF ≈ 10AD6 FE00
10AF2 ≈ 10AF3 10AF3
10AF4 ≈ 10ACF
10AF6 ≈ 10AC5
```

Select similarities with other scripts:

```
10AD0 ≈ 06A1
10AD0 ≈ 0726
10AD1 ≈ 10AD0 0307
10AD1 ≈ 0641
10AD2 ≈ 0642
10ADB ≈ 06A1
10ADC ≈ 0641
10ADC ≈ 10ADB 0307
10ADF ≈ 10ADE 0307
10AE1 ≈ 10AC5 0307
10AE2 ≈ 03C9
10AE5 ≈ 0308
10AE6 ≈ 0324
10AF4 ≈ 002E
10AF5 ≈ 003A
```

12. 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.

	10AC	10AD	10AE	10AF
0	10AC0	10AD0	10AE0	10AF0
1	10AC1	10AD1	10AE1	10AF1
2	10AC2	10AD2	10AE2	10AF2
3	10AC3	10AD3	10AE3	10AF3
4	10AC4	10AD4	10AE4	10AF4
5	10AC5	10AD5	10AE5	10AF5
6	10AC6	10AD6	10AE6	10AF6
7	10AC7	10AD7		
8	10AC8	10AD8		
9	10AC9	10AD9		
A	10ACA	10ADA		
B	10ACB	10ADB	10AEB	
C	10ACC	10ADC	10AEC	
D	10ACD	10ADD	10AED	
E	10ACE	10ADE	10AEE	
F	10ACF	10ADF	10AEF	

Letters

10AC0   MANICHAEAN LETTER ALEPH
10AC1   MANICHAEAN LETTER BETH
10AC2   MANICHAEAN LETTER BHETH
10AC3   MANICHAEAN LETTER GIMEL
10AC4   MANICHAEAN LETTER GHIMEL
10AC5   MANICHAEAN LETTER DALETH
10AC6   MANICHAEAN LETTER HE
10AC7   MANICHAEAN LETTER WAW
10AC8   MANICHAEAN SIGN UD
10AC9   MANICHAEAN LETTER ZAYIN
10ACA   MANICHAEAN LETTER ZHAYIN
10ACB   MANICHAEAN LETTER JAYIN
10ACC   MANICHAEAN LETTER JHAYIN
10ACD   MANICHAEAN LETTER HETH
10ACE   MANICHAEAN LETTER TETH
10ACF   MANICHAEAN LETTER YODH
10AD0   MANICHAEAN LETTER KAPH
10AD1   MANICHAEAN LETTER XAPH
10AD2   MANICHAEAN LETTER KHAPH
10AD3   MANICHAEAN LETTER LAMEDH
10AD4   MANICHAEAN LETTER DHAMEDH
10AD5   MANICHAEAN LETTER THAMEDH
10AD6   MANICHAEAN LETTER MEM
10AD7   MANICHAEAN LETTER NUN
10AD8   MANICHAEAN LETTER SAMEKH
10AD9   MANICHAEAN LETTER AYIN
10ADA   MANICHAEAN LETTER AAYIN
10ADB   MANICHAEAN LETTER PE
10ADC   MANICHAEAN LETTER FE
10ADD   MANICHAEAN LETTER SADHE
10ADE   MANICHAEAN LETTER QOPH
10ADF   MANICHAEAN LETTER XOPH
10AE0   MANICHAEAN LETTER QHOPH
10AE1   MANICHAEAN LETTER RESH
10AE2   MANICHAEAN LETTER SHIN
10AE3   MANICHAEAN LETTER SSHIN
10AE4   MANICHAEAN LETTER TAW

Combining marks

10AE5   MANICHAEAN ABBREVIATION MARK ABOVE
10AE6   MANICHAEAN ABBREVIATION MARK BELOW

Numbers

10AEB   MANICHAEAN NUMBER ONE
10AEC   MANICHAEAN NUMBER FIVE
10AED   MANICHAEAN NUMBER TEN
10AEE   MANICHAEAN NUMBER TWENTY
10AEF   MANICHAEAN NUMBER ONE HUNDRED

Punctuation

10AF0   MANICHAEAN PUNCTUATION STAR
10AF1   MANICHAEAN PUNCTUATION FLEURON
10AF2   MANICHAEAN PUNCTUATION DOUBLE DOT WITHIN DOT
10AF3   MANICHAEAN PUNCTUATION DOT WITHIN DOT
10AF4   MANICHAEAN PUNCTUATION DOT
10AF5   MANICHAEAN PUNCTUATION TWO DOTS
10AF6   MANICHAEAN PUNCTUATION LINE FILLER

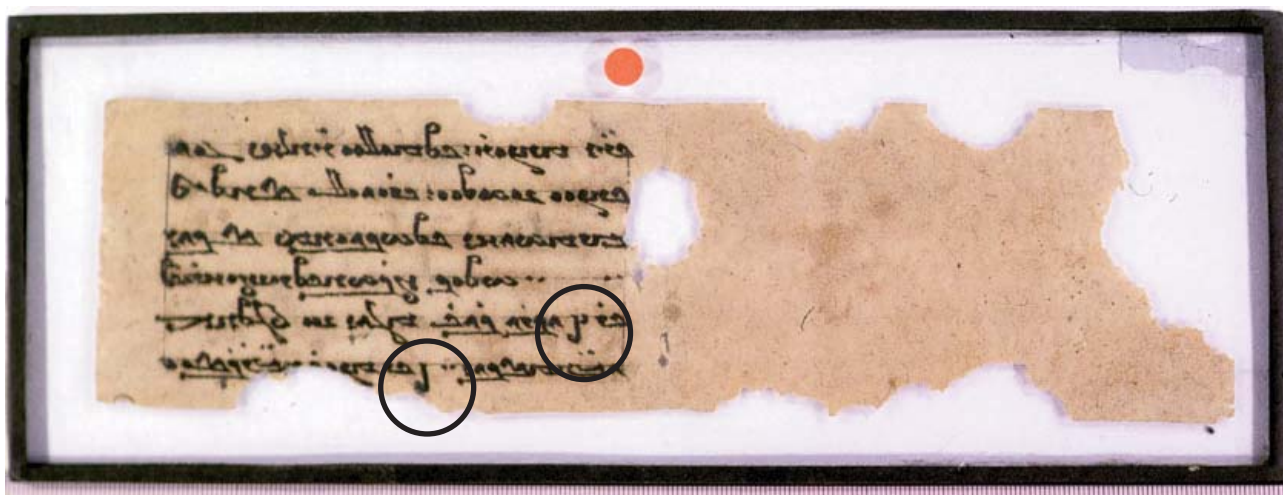**Figure 1.** One side of the Manichaean manuscript page M113. The numbers ﮊ 1 and ﭖ 2 are circled.
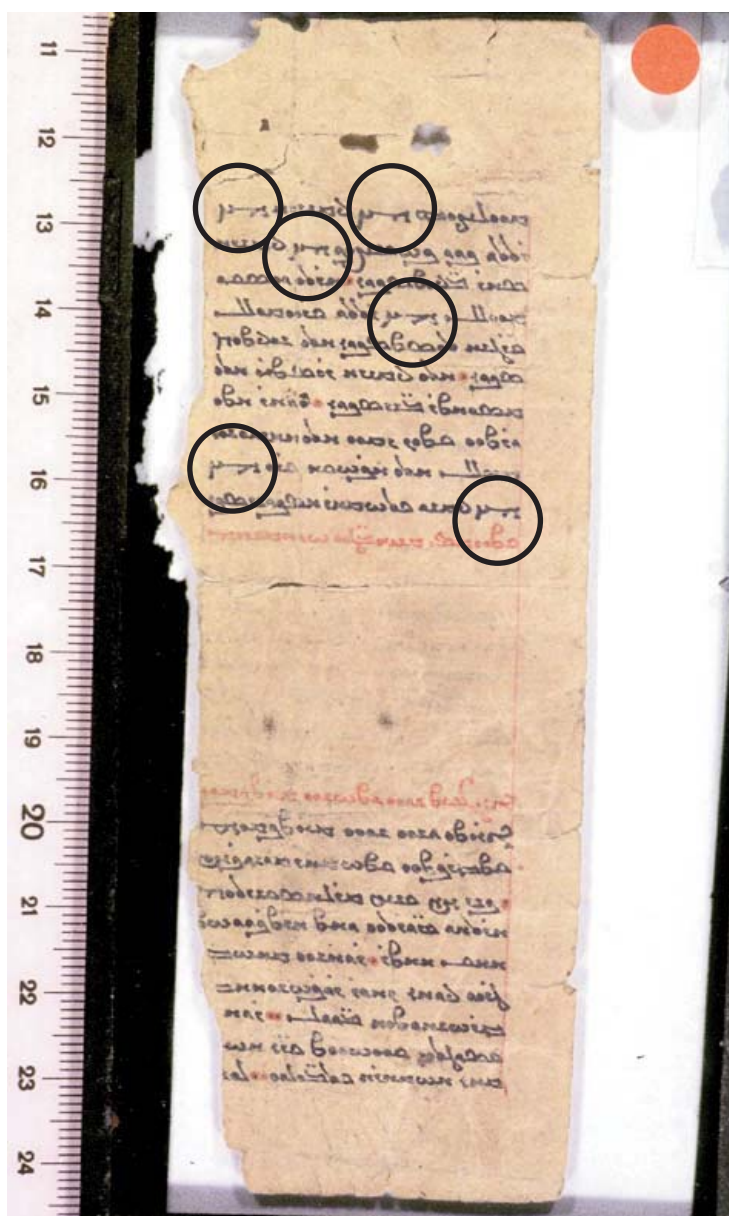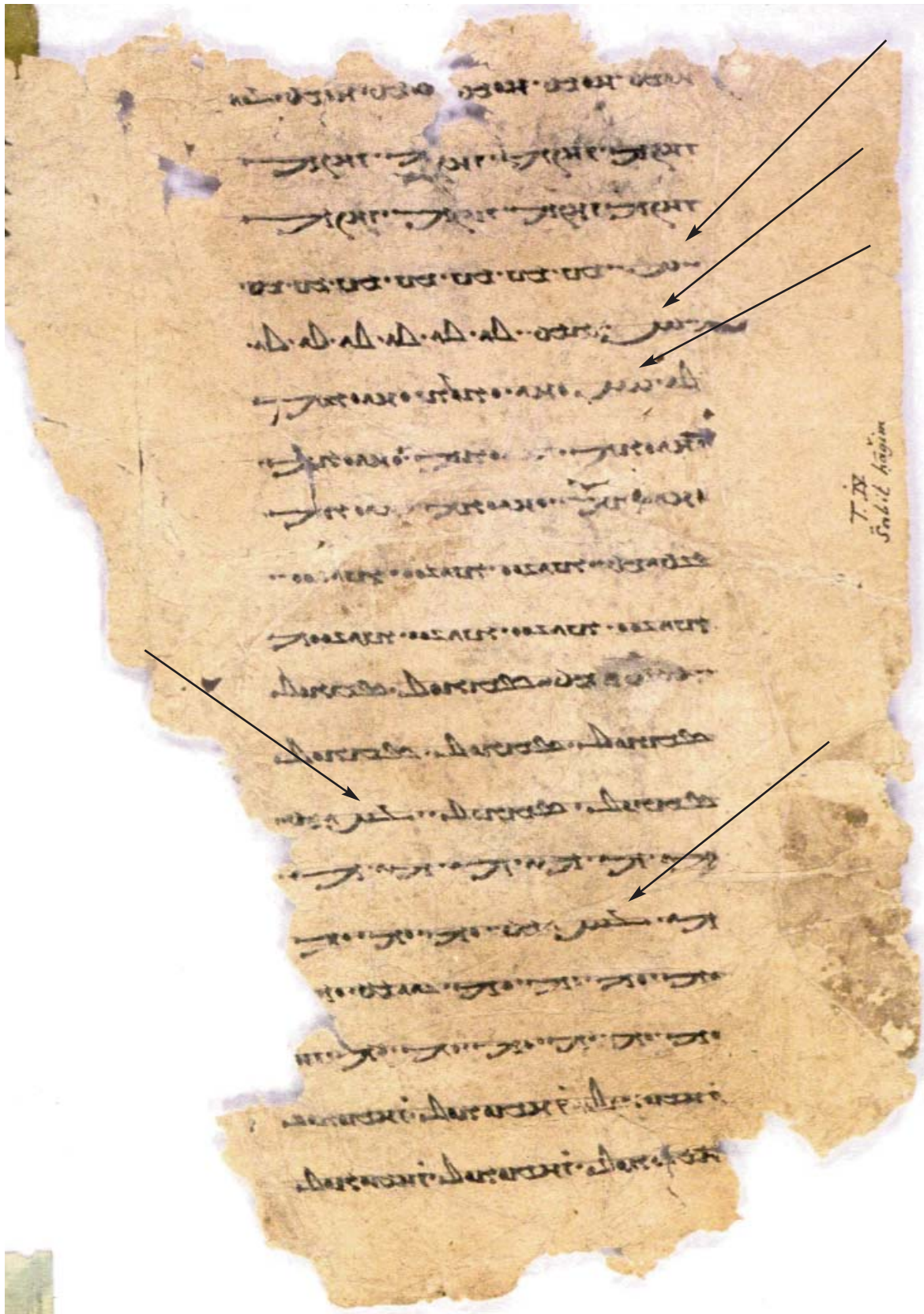

Figure 2. One side of the Manichaean manuscript page M14,

showing the number ﯓﭧ 12 in lines1, 2, 4, 9, and 10.

Figure 3. One side of the Manichaean manuscript page M8430, showing the numbers ܚܡܫ 2, ܣܪ 3, ܣܣܪ 4, ܚܡܪ 7, and ܡ 8.

shape	value	shape	value
𐫀	ʾ		s
	b		ʿ
	β		c̈
	g		p
	γ		f
	d		c
	ẖ		q
	w		q̂
	z		q̈
	ž		r
	ǰ		š
	j̈ (ẑ)		ś
	h		t
	ṯ		ẅ, u
	y		š̈
	k		.
	k̇		ᵕ
	k̈		flower
	l		1
	δ		5
	ϑ (δδ)		10
	m		20
	n		100

Figure 4. Table of the Manichaean script by Desmond Durkin-Meisterernst.
Note the transcriptions of ᴌ LAMEDH *l*, ᴌ DHAMEDH *δ*, and ᴌ THAMEDH *θ* (*δδ*); THAMEDH is in origin a combination of two DHAMEDHs, but it is not a typographic or decomposable ligature.

14

Manichäisch

Isoliert	Ende	Mitte	Anfang	Umschrift	Isoliert	Ende	Mitte	Anfang	Umschrift
				a					*l*
				b					*δ*
				β					*δδ*
				g					*m*
				γ					*n*
				d					*s*
				h					*ꜥ*
				v					*p*
				v̈					*f*
				z					*č*
				ž					
				h					*čy*
				t					*čn*
				y					*q* steht für *k*
				k					*q̈*
				ḫ, χ					*r*
									š
				k̈ = q					*t*

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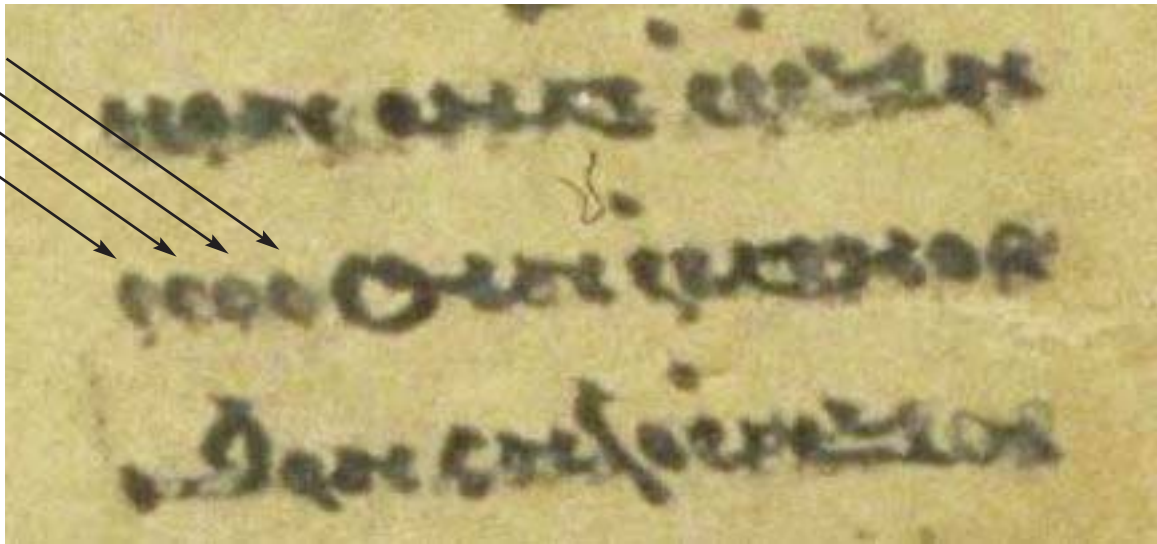
Figure 6. Example of the line filler in use in manuscript M7981/II/R/i/23/. The text reads

← ܪܫܝܪ̈ܐ ܐܝܪܒ ܐܝܪ / ܪܡܗܝܘ ܝܢܒ cccc / ܪܡܢ̇ܝܓܪ̈ܐ ܝܗܡ ܐܡܕ

→ *'brdr mhy 'wd / wyhmdr 'c **** / 'b'ryg'n 'wl*
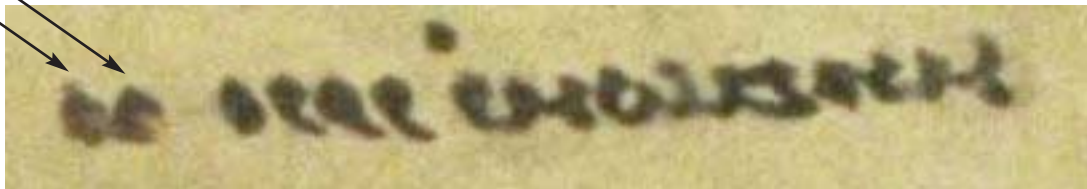
Figure 7. Examples of the line-filler in use in manuscript M7981/II/R/ii/29/. The text reads

← ܐܝܕܘܕ ܪܗܫܡ̇ܐܗ ܘܪ̈ܐ cc

→ *h'mšhr dwdy ***

A. Administrative

1. Title
Revised proposal for encoding the Manichaean script in the SMP of the UCS.
2. Requester's name
UC Berkeley Script Encoding Initiative (Universal Scripts Project)
(Authors: Michael Everson, Desmond Durkin-Meisterernst, Roozbeh Pournader, Shervin Afshar)
3. Requester type (Member body/Liaison/Individual contribution)
Liaison contribution.
4. Submission date
2011-05-10
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
51.
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 N3644R, 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)?
Yes.

11b. If YES, is a rationale for such use provided?
No.

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?