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

Doc Type: Working Group Document
Title: Proposal to add additional characters for Coptic and Latin to the UCS
Source: Michael Everson, Stephen Emmel (Universität Münster), Antti Marjanen (University of Helsinki), Ismo Dunderberg (University of Helsinki), John Baines (Oxford University), Susana Pedro (Universidade Lusófona de Humanidades e Tecnologia), António Emiliano (Universidade Nova de Lisboa)
Status: Individual Contribution
Action: For consideration by JTC1/SC2/WG2 and UTC
Date: 2007-05-12

0. Introduction. This proposal requests four additional cryptogrammic characters for Coptic, three Coptic-specific combining marks, and three generic combining diacritical marks for use with (at least) the Coptic and Latin scripts. If this proposal is accepted, the following characters will exist:

2CEB COPTIC CAPITAL LETTER CRYPTOGRAMMIC SHEI

2CEC COPTIC SMALL LETTER CRYPTOGRAMMIC SHEI

2CED COPTIC CAPITAL LETTER CRYPTOGRAMMIC GANGIA

2CEE COPTIC SMALL LETTER CRYPTOGRAMMIC GANGIA

2CEF COPTIC COMBINING NI ABOVE

2CF0 COPTIC COMBINING SPIRITUS ASPER

2CF1 COPTIC COMBINING SPIRITUS LENIS

FE24 COMBINING MACRON LEFT HALF

FE25 COMBINING MACRON RIGHT HALF

FE26 COMBINING CONJOINING MACRON

1. Cryptogrammic letters for Coptic. Document N2744 proposed a number of cryptogrammic letters used for secrecy in Coptic documents based on alphabet cyphers. Further research has turned up ꓳꓳ ꓳ COPTIC LETTER CRYPTOGRAMMIC SHEI (used for ⳝ ⳝ) and ꓳ ꓳ COPTIC LETTER CRYPTOGRAMMIC GANGIA (used for ⳤ ⳤ), which are proposed for encoding here at U+2CEB–U+2CEE.

2. Combining character additions for Coptic. The ꓳ COPTIC COMBINING NI ABOVE is typically used at the end of a line in Coptic to indicate a final NI after a vowel. It is unrelated to characters with similar use, such as the TILDE or MACRON used in Latin for a similar purpose, and its glyph typically hangs over the space to the right of the base character. The ꓳ COPTIC COMBINING SPIRITUS ASPER and ꓳ COPTIC COMBINING SPIRITUS LENIS are used relatively rarely, over the letter н, sometimes to indicate that it is the borrowed Greek conjunction η 'or'. The distinction between asper and lenis seems to be lost in Coptic, as both are used. We do not believe that for Coptic these can be unified with the generic U+0314 COMBINING

REVERSED COMMA ABOVE and U+0313 COMBINING COMMA ABOVE used for Greek, or with the U+0485 COMBINING CYRILLIC DASIA PNEUMATA and U+0486 COMBINING CYRILLIC PSILI PNEUMATA, neither of which has a shape appropriate for Coptic. If encoded, scholars in other disciplines who *do* require a non-comma-shaped spiritus, might find these characters useful.

3. Generic combining character additions. Coptic and medieval Latin both employ horizontal marks above letters to indicate abbreviations; Coptic also uses a horizontal mark to indicate that a consonant or group of consonants is syllabic. In N2639 and N2744, there was a discussion of the use of generic combining characters in Coptic. Of these, U+0304 COMBINING MACRON and U+0305 COMBINING OVERLINE have been specified for use in some contexts, with the MACRON having the syllabic use and the OVERLINE being used for abbreviation or to indicate numeric use of a letter. The problem is that Coptic has behaviour which goes beyond the specifications of those two marks, namely, that a run of characters may be marked with a superlinear line, and in Coptic the orthographic convention for the most frequent function of such lines is to indicate the beginning and end of such a run with a "short" macron.

3.1. Ordinary use of COMBINING MACRON and COMBINING OVERLINE. When used over a single character or over several characters in succession as a mark of individual syllabicity, the use of U+0304 is not problematic. In the word ⲛ̄ⲙ̄ⲕⲟⲟϩ *ṇmko'h* 'the pains', two individual letters each bear a COMBINING MACRON; there is no connection between them, and they are relatively short. When used over several characters in succession as a mark of abbreviation, the use of U+0305 is not problematic. In a *nomen sacrum* such as ⲡ̅ⲛ̅ⲁ̅ *pna* < *pneuma* 'spirit', the broad COMBINING OVERLINE covers the entire width of each letter and connects, as expected. Letters used as numbers are also written with U+0305.

3.2. Complex spellings where COMBINING MACRON and COMBINING OVERLINE do not suffice. Where there are problems is more complex spellings. In the preposition ⲙ̄ⲛ̄ *mṇ* 'with', the superlinear line does not behave in the same way as the COMBINING OVERLINE does in ⲡ̅ⲛ̅ⲁ̅; the line must be centred between the two characters and does not extend to their full width. One might propose the use of U+035E COMBINING DOUBLE MACRON, but this does not work either, because the behaviour expected is not confined to *pairs* of letters, but extends to triplets, quadruplets, quintuplets, and so on arbitrarily. Examples: ϥ̄ⲛ̄ⲧ̄ *fṇt* 'worm', ⲡ̅ϩ̅ⲣ̅ⲃ̅ *phṛb* 'the form'. The same convention is used to mark proper names of divinities and heroes: examples are ⲥ̅ⲁ̅ⲕ̅ⲗ̅ⲁ̅ *Sakla* (a name for the demonic ruler of this world) and ⲁ̅ⲛ̅ⲁ̅ⲥ̅ⲥ̅ⲏ̅ⲥ̅ ⲇ̅ⲩ̅ⲥ̅ⲏ̅ⲥ̅ *Anassês Dusês*. Because U+035E is a base character applying to the first of a pair of letters, its relation to the second of the pair is basically undefined. A font might contain contextual glyph variations for triplets including U+035E, but this doesn't address the requirements to handle strings of arbitrary length beyond two letters.

3.3. Proposed joining "half-macrons" and "conjoining macron" allow the representation of complex Coptic spellings. We propose the addition of two combining characters which can indicate correctly the beginning and end of a superlinear run while also permitting relatively simple rendering of the glyphs required. The characters are a hybrid between COMBINING MACRON and COMBINING OVERLINE: each will join to each other as in ⲙ̄ⲛ̄, and each will join to to a new COMBINING CONJOINING MACRON on one side only as an extender as in ϥ̄ⲛ̄ⲧ̄ or ⲡ̅ϩ̅ⲣ̅ⲃ̅. With the two new characters, ◌̄ COMBINING MACRON LEFT HALF and ̄◌ COMBINING MACRON RIGHT HALF, all of the expected behaviour in Coptic can be accounted for. For ⲙ̄ⲛ̄, the recommended spelling would be ⲙ MI + ◌̄ RIGHT-JOINING MACRON + ⲛ NI + ◌̄ LEFT-JOINING MACRON (not ⲙ MI + ◌͞◌ DOUBLE MACRON + ⲛ NI); for ϥ̄ⲛ̄ⲧ̄, the recommended spelling would be ϥ FEI + ◌̄ RIGHT-JOINING MACRON + ⲛ NI + ◌̄ CONJOINING MACRON + ⲧ TAU + ◌̄ LEFT-JOINING MACRON (no other sequence makes sense as traditional OVERLINE does not join to MACRON); for ⲡ̅ϩ̅ⲣ̅ⲃ̅, the recommended spelling would be ⲡ PI + ◌̄ RIGHT-JOINING MACRON + ϩ HORI + ◌̄ CONJOINING MACRON + ⲣ RO + ◌̄ CONJOINING MACRON + ⲃ VIDA + ◌̄ LEFT-JOINING MACRON. These characters allow for representation of

complex Coptic words and phrases, such as ⲛⲥⲁⲕⲗⲁ *ṇSakla* 'to Sakla' or ϩⲛⲙⲙⲛⲧⲙⲁⲓϩⲟⲙⲧ *hṇṃṃṇtmaïhoṃt* 'in the acts of greediness'.

3.4 Use of the joining macrons in other scripts. The use of the two joining macrons would not be limited to Coptic. In the Figures below, we give examples of the same behaviour in Latin texts from medieval Portugal. It cannot be said that the argument for the need for these characters in Latin script is as strong as it is for Coptic; Latin abbreviations may be structured somewhat differently, and the distinction of syllabic vs abbreviation does not apply. In some kinds of diplomatic editions of medieval texts, however, the same glyph behaviour does occur, particularly the macron-centred-between-two letters. We consider this potential use—as well as use with other scripts such as Cyrillic, Georgian, or Greek—to warrant the encoding of these two characters as generic characters.

4. Comparison of different spellings of Coptic superlinear strokes. In the examples below, orange indicates COMBINING MACRON, red indicates COMBINING MACRON LEFT HALF, blue indicates COMBINING MACRON RIGHT HALF, green indicates COMBINING CONJOINING MACRON, and magenta indicates COMBINING OVERLINE.

Note the following minimal pair; in the third example, COMBINING MACRON does not join with either end of the COMBINING OVERLINE:

5. Unicode Character Properties. Character properties are proposed here.

```
2CEB;COPTIC CAPITAL LETTER CRYPTOGRAMMIC SHEI;Lu;0;L;;;;;N;;;;2CEC;
2CEC;COPTIC SMALL LETTER CRYPTOGRAMMIC SHEI;Ll;0;L;;;;;N;;;2CEB;;2CEB
2CED;COPTIC CAPITAL LETTER CRYPTOGRAMMIC GANGIA;Lu;0;L;;;;;N;;;;2CEE
2CEE;COPTIC SMALL LETTER CRYPTOGRAMMIC GANGIA;Ll;0;L;;;;;N;;;2CED;;2CED
2CEF;COPTIC COMBINING NI ABOVE;Mn;232;NSM;;;;;N;;;;;
2CF0;COPTIC COMBINING SPIRITUS ASPER;Mn;232;NSM;;;;;N;;;;;
2CF1;COPTIC COMBINING SPIRITUS LENIS;Mn;232;NSM;;;;;N;;;;;

FE24;COMBINING MACRON LEFT HALF;Mn;230;NSM;;;;;N;;;;;
FE25;COMBINING MACRON RIGHT HALF;Mn;230;NSM;;;;;N;;;;;
FE26;COMBINING CONJOINING MACRON;Mn;230;NSM;;;;;N;;;;;
```

6. Reference glyphs for Coptic and urgency of this proposal. The International Association for Coptic Studies (IACS) has commissioned the font *Keft* (from Coptic Ⲕⲉϥⲧ for what is now *Qifṭ* in the Sahidic area of Egypt, called *Κόπτος* in Greek). This font will be made freely available to everyone. We propose that *Keft* replace the reference glyphs for Coptic in the UCS code charts. Shortcomings in the encoding for Coptic discovered during the development of this font have prompted the present proposal. In order to ensure timely completion of the font development and release, we request that the UTC and WG2 treat the addition *especially of the two macrons* as a matter of urgency, and we hope that WG2 will agree to add them to an existing ballot to assist the IACS in providing a working font to Copticists worldwide as soon as possible.

7. Bibliography.

Chassinat, Émile. 1911. *Le quatrième livre des entretiens et épîtres de Shenouti*. (Mémoires publiés par les membres de l'Institut Français d'Archéologie Orientale du Caire; 23) Le Caire: Imprimerie de l'Institut Français d'Archéologie Orientale.

Lantschoot, Arnold van. 1929. *Receuil des Colophons des manuscrits chrétiens d'Égypte. Tome I: Les colophons coptes des manuscrits sahidiques*. Louvain: J.-B. Ispas.

Layton, Bentley. 1992. "Two unpublished Shenute fragments *Against Kronos*: Layton, *Brit*. *Lib*., Nos. 90 and 91", in *Journal of Coptic Studies* 2. Louvain: Peeters.

Wisse, Frederik. 1979. "Language mysticism in the Nag Hammadi texts and in early Coptic monasticism I: cryptography", in *Enchoria: Zeitschrift für Demotistik und Koptologie*. Wiesbaden: Harrassowitz.

Young, Dwight W. 2001. "An unplaced fragment from Shenute's *Fourth Canon*", in *Journal of Coptic Studies* 3. Louvain: Peeters.

Acknowledgements

Figures

Figure 1. Sample from Lantschoot 1929 showing CRYPTOGRAMMIC SHEI; Lantschoot uses regular GANGIA here, either because he lacked a glyph or because he did not identify the CRYPTOGRAMMIC GANGIA; compare Figure 2.

Figure 2. Sample from Wisse 1979, showing CRYPTOGRAMMIC SHEI and CRYPTOGRAMMIC GANGIA in a re-edition of the same text given in Figure 1.

126   É. CHASSINAT.

[Inst. fr., Copte nᵒ 1, fol. 59, vᵒ.]

Figure 3. Sample from Chassinat 1911, showing COPTIC COMBINING NI ABOVE, COMBINING MACRON, COMBINING CONJOINING MACRON, COMBINING MACRON LEFT HALF, and COMBINING MACRON RIGHT HALF.

ENTRETIENS ET ÉPÎTRES DE SHENOUTI.   195

[Inst. fr., Copte n° 1, fol. 94, r°.]

Figure 4. Sample from Chassinat 1911, showing COMBINING MACRON, COMBINING CONJOINING MACRON, COMBINING MACRON LEFT HALF, and COMBINING MACRON RIGHT HALF.

ⲭⲟⲟⲩⲛ̄ϭⲓⲡ̄ⲡⲁⲛⲧⲟ
ⲕⲣⲁⲧⲱⲣ̄ⲙ̄ⲙⲉ̄ⲓ̄ⲥ,
ⲛ̄ⲥⲉⲧⲱⲣⲡ̄ⲙ̄ⲡⲟⲩ⳿ⲁ̂
15 ⲡⲟⲩⲁ̂ⲙ̄ⲙⲟⲛ⳿, ⲁⲩⲱ
ⲧⲟⲩⲉⲉⲓ̂ⲧⲟⲩⲉⲓ̂ⲙ̄ⲙⲱ
ⲧ̄ⲛ, ᵛ ⲛ̄ⲥⲉϥⲓⲧⲟⲩⲉ̄ⲛ
ⲟⲩϭⲱⲛ̄ⲧⲙ̄ⲛⲟⲩ
ⲟⲣⲅⲏ̂,ⲉⲣⲁⲧϥ̄ⲙ̄
20 ⲡⲉⲧⲙⲟⲥⲧⲉ ⲙ̄ⲧⲛ
ϭⲓⲛ ⲣ̂ϩⲱⲃ⳿ⲧⲏⲣ̄ⲥ,
ˢⁱᶜ ⲭⲓⲛⲛ̄ⲟⲩⲭⲱϩ̄ⲙ̄
ⲙ̄ⲛⲟⲩⲥⲱⲱϥ,
ϣⲁⲟⲩⲭⲓⲟⲩⲉ⳿ⲙ̄ⲛ
25 ⲟⲩϭⲟⲗ, ᵛ ⲁⲩⲱⲭⲓⲛ
ⲛ̄ⲟⲩⲕⲱϩ⳿ⲙ̄ⲛⲟⲩ
ⲙⲟⲥⲧⲉ, ᵛ ϣⲁⲟⲩⲧ̄
ⲧⲱⲛ⳿, ⲙ̄ⲛⲟⲩⲕⲁⲧⲁ
ⲗⲁⲗⲉⲓⲁ̂,ⲙ̄ⲛⲟⲩ
30 ⲕⲣ̄ⲙⲣⲙ̄, ᵛ ⲁⲩⲱ
ⲭⲓⲛⲟⲩⲭⲓ̂ⲙ̄ⲛⲟⲩ

ⲣⲟⲩ⳿ⲉⲧ̄ϩⲛ̄ⲁ̄ⲙⲛⲧⲉ,
45 Ⲟⲩⲛ̄ϩⲁϩ⳿ϩⲛ̄ⲛⲉ̄ⲛⲧⲁⲩ
ⲃⲱⲕ⳿ⲉⲡⲉⲥⲏⲧ⳿ⲉ
ⲣⲟϥ⳿ⲛ̄ϩⲏⲧ̄ⲛ̄ⲭⲓⲛ
ⲛ̄ϣⲟⲣⲡ̄, ᵛ ⲙ̄ⲛⲛⲉ
ⲧⲛⲁⲃⲱⲕ⳿ⲟⲛ⳿ⲛ̄ϩⲣⲁⲓ̄
50 ⲛ̄ϩⲏⲧ̄ⲛ̄ⲉϩⲣⲁⲓ̄ⲉⲣⲟϥ⳿
ⲭⲓⲛⲛ̄ⲧⲉⲛⲟⲩ,
ⲁⲩⲱϣⲁⲧⲥⲩⲛⲧⲉ
ⲗⲉⲓⲁ⳿ⲙ̄ⲡⲁⲓⲱⲛ⳿, ⲛⲁ
ⲡⲱϩ⳿ⲙ̄ⲡⲉⲩϩⲏⲧ⳿,
55 ⲁⲩⲱϩⲁⲙⲟⲓ̂ⲉⲛⲉϥ
ⲛⲁⲡⲱϩ, ⲁⲩⲱⲛ̄
ⲥⲉⲭⲱ̂ⲛ̄ϩⲉⲛⲛⲟϭ
ⲛ̄ϣⲁⲭⲉⲙ̄ⲙ̄ⲛⲧⲉ
ⲃⲓⲏⲛ⳿, ϩⲛ̄ⲧⲃⲁⲥⲁ
60 ⲛⲟⲥ⳿ⲙ̄ⲡⲕⲱϩ̄ⲧ
ⲙ̄ⲛ̄ⲡϥ̄ⲛⲧ,
ⲁⲩⲱϩ̄ⲙ̄ⲡⲣⲱ
ⲕ̄ϩ̄ⲙ̄ⲡⲉⲓⲃⲉ, ⲙ̄ⲛ̄

Figure 5. Sample from Young 2001, showing COMBINING MACRON, COMBINING OVERLINE, COMBINING MACRON LEFT HALF, and COMBINING MACRON RIGHT HALF.

ⲛⲁⲉⲩⲫⲣⲁⲛⲉ˙
ⲉⲩϣⲁⲛⲛⲁⲩⲉ
ⲡⲉⲕⲭⲓⲕⲃⲁ̂ⲉ
ⲧⲉⲣⲉⲡⲭⲟⲉⲓⲥ
ⲡⲡⲁⲛⲧⲟⲕⲣⲁ
ⲧⲱⲣ⳿ⲙ̄ⲁⲛⲧϥⲉ
ⲭⲱⲕ⳿ ᵛ.
ⲙ̄ⲡⲉϩⲟⲟⲩⲉ
ⲧ̄ⲙⲙⲁⲩⲛⲉⲧ
ⲛⲁⲩⲉⲣⲟⲕⲛⲁ
ⲣϣⲡⲏⲣⲉⲉϩⲣⲁⲓ̄
ⲉⲭⲱⲕ⳿,ⲛ̄ⲥⲉ
ⲭⲟⲟⲥⲭⲉⲡⲁⲓ̄
ⲡⲉⲡⲣⲱⲙⲉⲉⲧ
†ⲛⲟⲩϭⲥ̄ⲙ̄ⲡⲕⲁϩ⳿.
ⲁⲩⲱⲛⲉⲛⲧⲁⲕ
ⲧⲥⲟⲟⲩⲧⲏⲣⲟⲩ
ϣⲁⲛⲧⲟⲩ†ϩⲉ⳿
ϩⲛ̄ⲧⲙⲛ̄ⲧⲁⲧ
ⲛⲟⲩⲧⲉ, ᵛ ⲥⲉ
ⲛⲁⲭⲟⲟⲥⲛⲁⲕ⳿
ⲉⲩϩ̄ⲙⲡⲧⲁⲕⲟ̂

1

5

10

15

20

ⲉⲓ̄ⲉⲡⲉⲥⲏⲧ⳿ⲉ
ⲁⲙ̄ⲛⲧⲉ. ᵛ ⲡⲉⲕ
ⲟⲩⲛⲟϥⲉⲧⲛⲁ
ϣⲱϥ̂, ᵛ ⲥⲉⲛⲁ
ⲡ̄ⲣⲱϩⲁⲣⲟⲕⲛ̄
ⲟⲩⲭⲟⲟⲗⲉⲥ˝
ⲛ̄ⲥⲉϩⲟⲃⲥⲕⲛ̄
ⲟⲩϥ̄ⲛⲧ ᵛ =
ⲉⲧⲉⲛϥϭⲓⲣⲟ
ⲟⲩϣⲁⲛⲉⲧⲣⲉ
ⲓ̄ⲥⲟⲩⲱϩ⳿ⲛ̄ϩⲏⲧϥ
ⲙⲁⲣⲉϥϣⲱⲡⲉ
ⲉⲩⲧⲱⲣⲡ̄ⲙ̄ⲡⲥⲁ
ⲧⲁⲛⲁⲥ⳿, ᵛ ⲁⲩⲱ
ⲉⲟⲩϣⲱⲗ⳿ⲛ̄ⲛ̄
ⲇⲁⲓⲙⲟⲛⲓⲟⲛ⳿.
ⲕⲁⲛ⳿ⲙ̄ⲡ̄ⲛⲭⲟ
ⲟⲥⲛ̄ⲧⲉⲓ̈ϩⲉ,ⲑⲉ
ⲧⲉⲧⲁⲓ̈ⲉⲧⲟⲩϣⲟ
 ⲟⲡ̄ⲙ̄ⲙⲟⲥ ᵛ.
ⲧⲙⲛ̄ⲧⲁ†ⲙ̄ⲡϣⲁ̂
ⲙ̄ⲡⲣⲱⲙⲉⲧⲉ

Figure 6. Sample from Layton 1992, showing COMBINING MACRON, COMBINING CONJOINING MACRON, COMBINING MACRON LEFT HALF, and COMBINING MACRON RIGHT HALF.
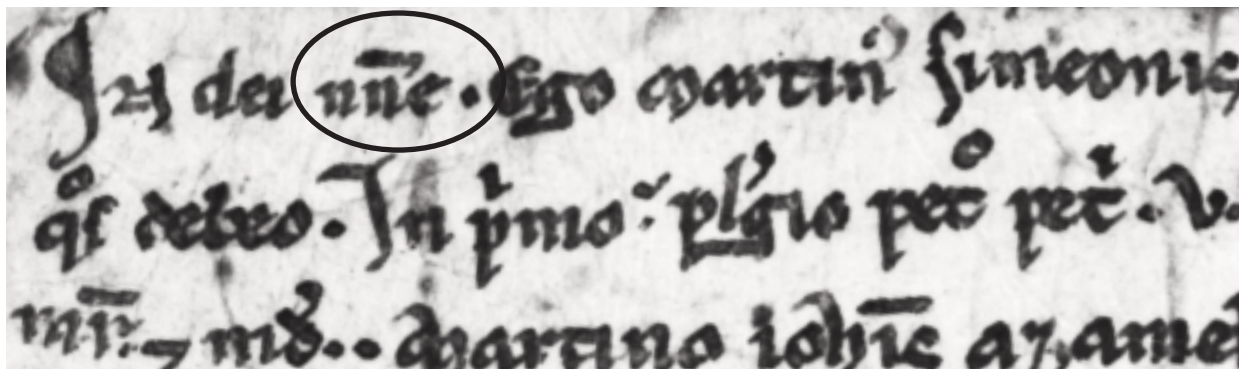
Figure 7. Sample from medieval Portuguese manuscript IAN/TT, Mitra de Braga, maço 1, nº. 6, showing an abbreviation for *nomine* using RIGHT-JOINING MACRON, COMBINING CONJOINING MACRON, and LEFT-JOINING MACRON.
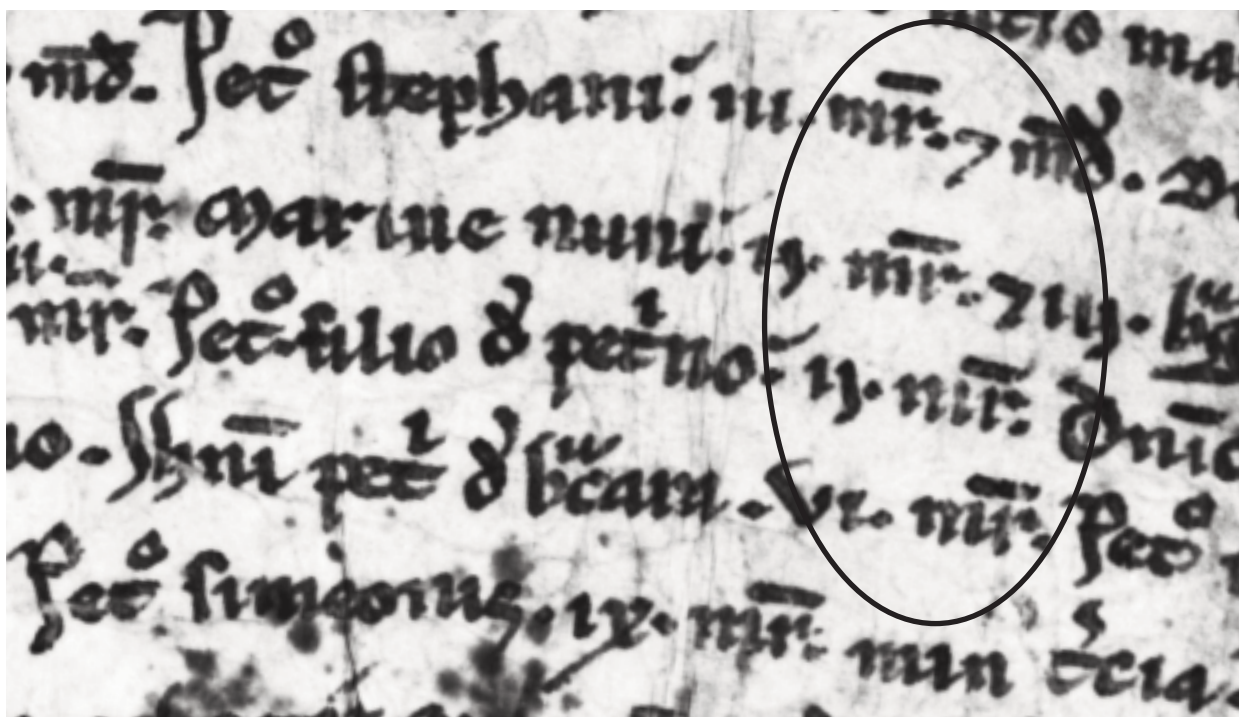
Figure 8. Sample from medieval Portuguese manuscript IAN/TT, Mitra de Braga, maço 1, nº. 6, showing an abbreviation for the currency *morabitino* using RIGHT-JOINING MACRON and LEFT-JOINING MACRON.
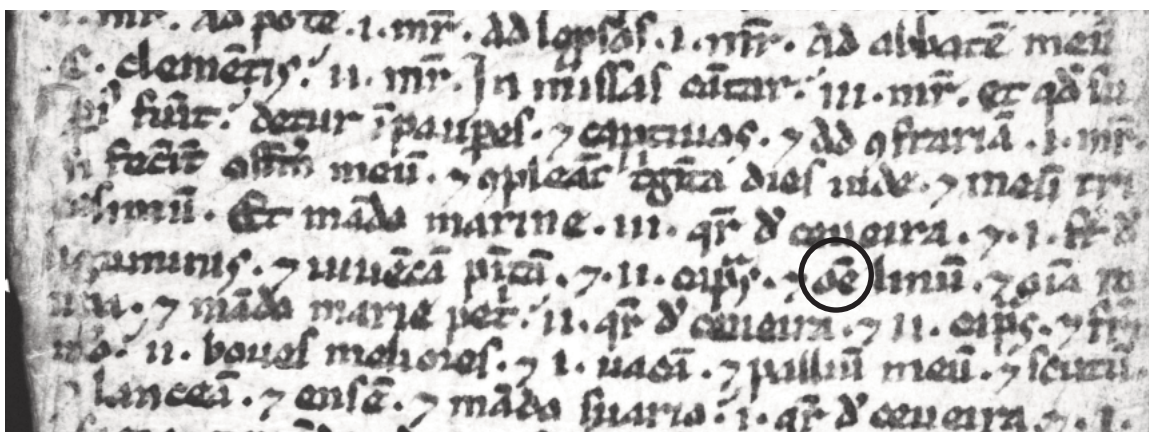
Figure 9. Sample from medieval Portuguese manuscript IAN/TT, Cabido da Sé de Coimbra, maço 6, nº. 45, showing an abbreviation for *omne* using RIGHT-JOINING MACRON and LEFT-JOINING MACRON.
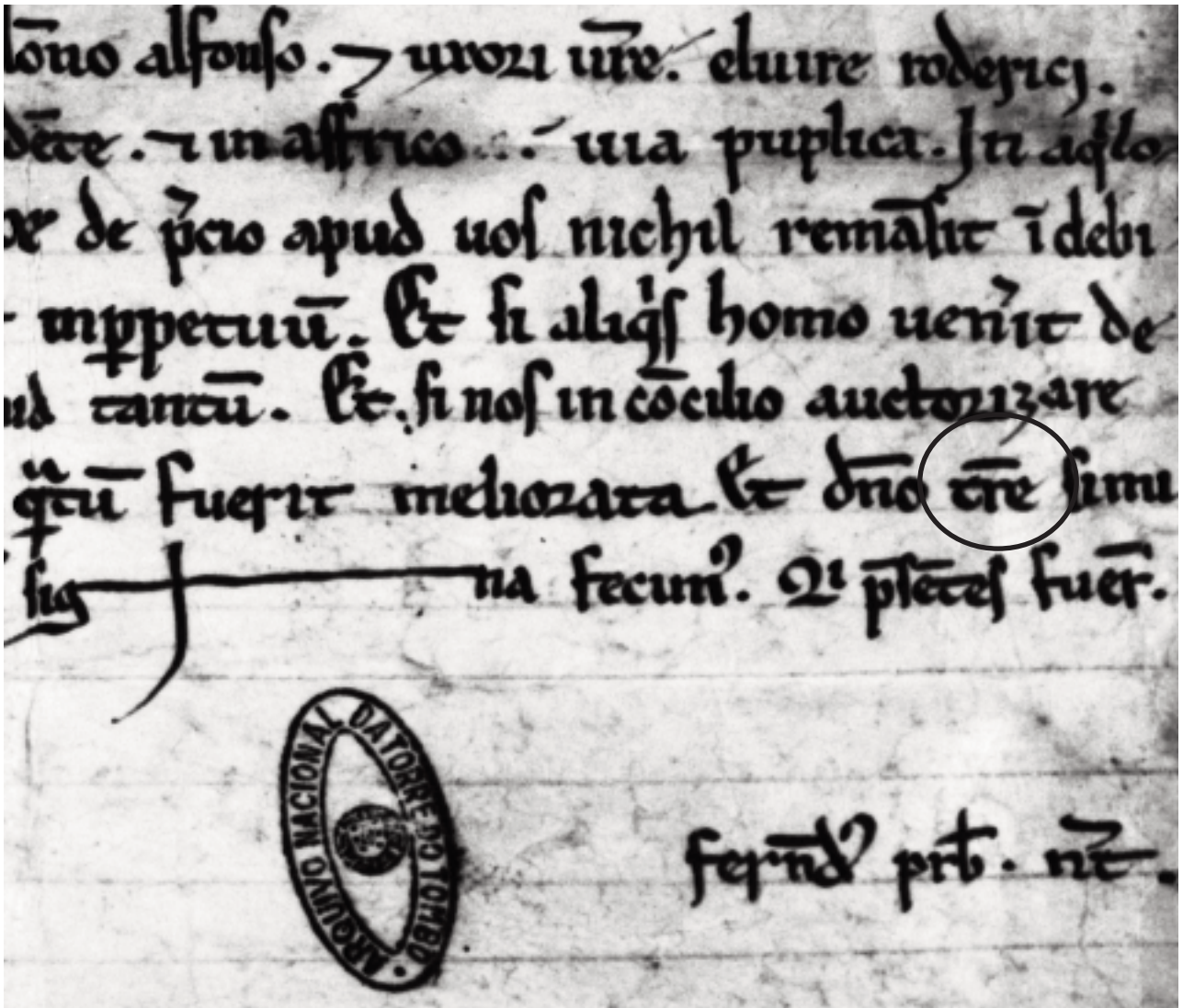
Figure 2. Sample from medieval Portuguese manuscript IAN/TT, Cabido da Sé de Coimbra, maço 8, nº.
23, showing an abbreviation for *terre* using RIGHT-JOINING MACRON, COMBINING CONJOINING MACRON, and
LEFT-JOINING MACRON.
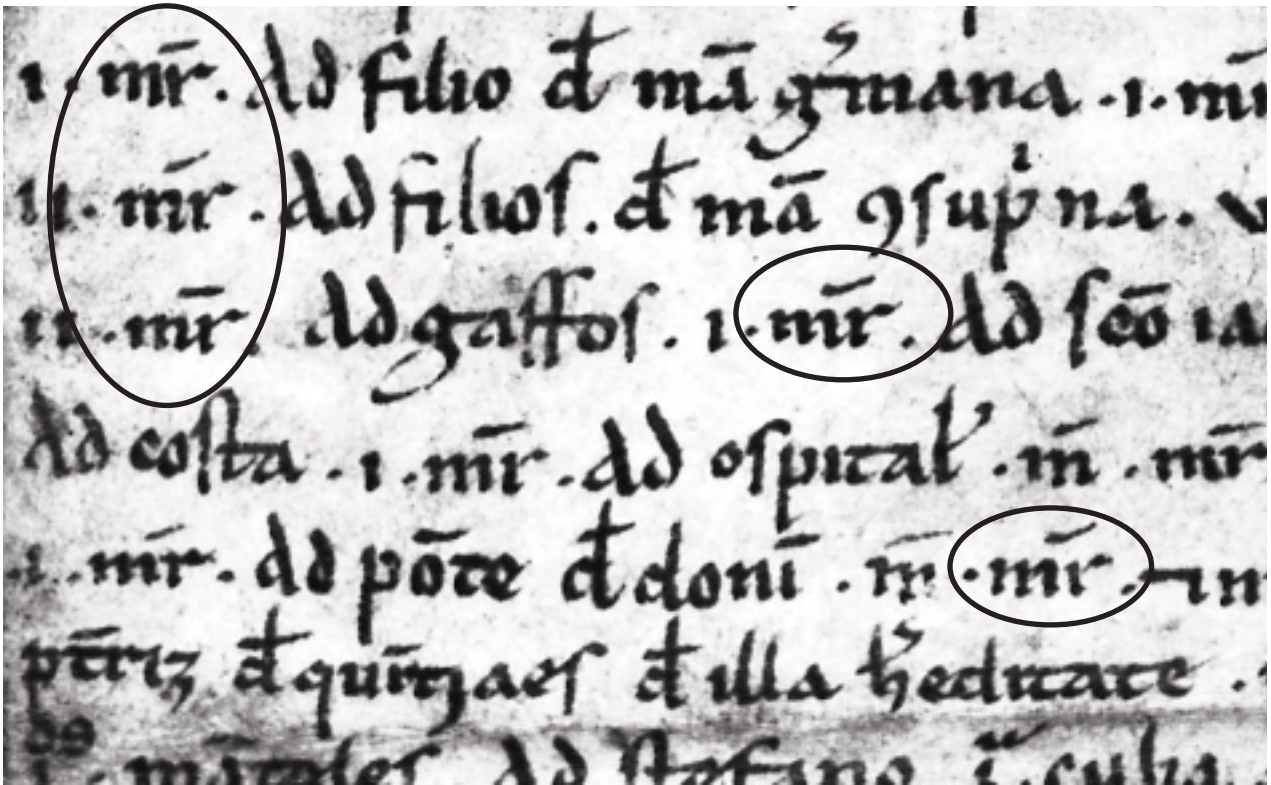
Figure 2. Sample from medieval Portuguese manuscript IAN/TT, Colegiada de Santa Maria da Oliveira de Guimarães, maço 4, nº. 44, showing an abbreviation for the currency *morabitino* using RIGHT-JOINING MACRON and LEFT-JOINING MACRON.
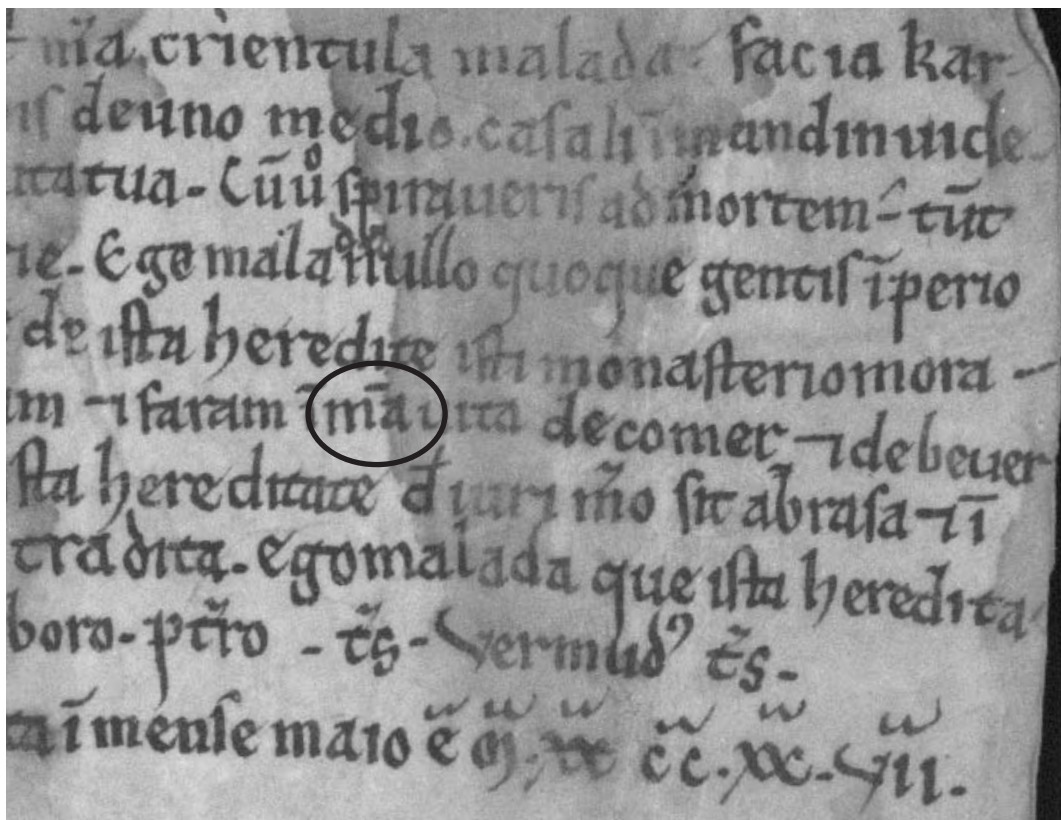
Figure 2. Sample from medieval Portuguese manuscript IAN/TT, Mosteiro de S. Salvador de Moreira, maço 8, nº. 28 showing an abbreviation for the word *mea* using RIGHT-JOINING MACRON and LEFT-JOINING MACRON.

COPTIC

	2C8	2C9	2CA	2CB	2CC	2CD	2CE	2CF		03E
0	ⲁ	ⲑ	ⲡ	ⲱ	ⲯ	ⳁ	ⳏ	◌̀		Ꝫ
1	ⲁ	ⲑ	ⲡ	ⲱ	ⲯ	ⳁ	ⳏ	◌̄		ꝫ
2	ⲃ	ⲓ	ⲣ	ⳋ	ⲵ	Ϭ	ⳓ			Ꝭ
3	ⲃ	ⲓ	ⲣ	ⳋ	ⲵ	ϭ	ⳓ			ꝭ
4	ⲅ	ⲕ	ⲥ	ⳛ	ⳝ	ⳟ	ⳤ			Ꝯ
5	ⲅ	ⲕ	ⲥ	ⳛ	ⳝ	ⳟ	ⳡ			ꝯ
6	ⲇ	ⲗ	ⲧ	Ⳛ	Ⳝ	Ⳟ	ⳡ			Ꝧ
7	ⲇ	ⲗ	ⲧ	Ⳛ	Ⳝ	Ⳟ	ⳣ			ꝧ
8	ⲉ	ⲙ	ⲩ	ⳉ	ⳉ	ⳗ	⳥			Ꝩ
9	ⲉ	ⲙ	ⲩ	ⳉ	ⳉ	ⳗ	⳥	⟍		ꝩ
A	ⲋ	ⲛ	ⲫ	⟋	ⳗ	ⳬ	⳪	⟋		Ꝺ
B	ⲋ	ⲛ	ⲫ	⟋	ⳗ	ⳬ	⳪	⟍		ꝺ
C	ⲍ	Ⳓ	ⲭ	ⳤ	ⳍ	ⳓ	⳪	⫽		Ꝼ
D	ⲍ	Ⳓ	ⲭ	ⳤ	ⳍ	ⳓ	Ⳮ	ꜱ		ꝼ
E	ⲏ	ⲟ	Ⲩ	⳥	ⲣ	ⳬ	Ⳮ	›		Ꝥ
F	ⲏ	ⲟ	ⲩ	⳥	ⲣ	ⳬ	◌̀	ꞌ		ꝥ

12

COPTIC

hex	Name		hex	Name
00	COPTIC CAPITAL LETTER ALFA		59	COPTIC SMALL LETTER OLD COPTIC DJA
01	COPTIC SMALL LETTER ALFA		5A	COPTIC CAPITAL LETTER OLD COPTIC SHIMA
02	COPTIC CAPITAL LETTER VIDA		5B	COPTIC SMALL LETTER OLD COPTIC SHIMA
03	COPTIC SMALL LETTER VIDA		5C	COPTIC CAPITAL LETTER OLD NUBIAN SHIMA
04	COPTIC CAPITAL LETTER GAMMA		5D	COPTIC SMALL LETTER OLD NUBIAN SHIMA
05	COPTIC SMALL LETTER GAMMA		5E	COPTIC CAPITAL LETTER OLD NUBIAN NGI
06	COPTIC CAPITAL LETTER DALDA		5F	COPTIC SMALL LETTER OLD NUBIAN NGI
07	COPTIC SMALL LETTER DALDA		60	COPTIC CAPITAL LETTER OLD NUBIAN NYI
08	COPTIC CAPITAL LETTER EIE		61	COPTIC SMALL LETTER OLD NUBIAN NYI
09	COPTIC SMALL LETTER EIE		62	COPTIC CAPITAL LETTER OLD NUBIAN WAU
0A	COPTIC CAPITAL LETTER SOU		63	COPTIC SMALL LETTER OLD NUBIAN WAU
0B	COPTIC SMALL LETTER SOU		64	COPTIC SYMBOL KAI
0C	COPTIC CAPITAL LETTER ZATA		65	COPTIC SYMBOL MI RO
0D	COPTIC SMALL LETTER ZATA		66	COPTIC SYMBOL PI RO
0E	COPTIC CAPITAL LETTER HATE		67	COPTIC SYMBOL STAUROS
0F	COPTIC SMALL LETTER HATE		68	COPTIC SYMBOL TAU RO
10	COPTIC CAPITAL LETTER THETHE		69	COPTIC SYMBOL KHI RO
11	COPTIC SMALL LETTER THETHE		6A	COPTIC SYMBOL SHIMA SIMA
12	COPTIC CAPITAL LETTER IAUDA		6B	COPTIC CAPITAL LETTER CRYPTOGRAMMIC SHEI
13	COPTIC SMALL LETTER IAUDA		6C	COPTIC SMALL LETTER CRYPTOGRAMMIC SHEI
14	COPTIC CAPITAL LETTER KAPA		6D	COPTIC CAPITAL LETTER CRYPTOGRAMMIC GANGIA
15	COPTIC SMALL LETTER KAPA		6E	COPTIC SMALL LETTER CRYPTOGRAMMIC GANGIA
16	COPTIC CAPITAL LETTER LAULA		6F	COPTIC COMBINING NI ABOVE
17	COPTIC SMALL LETTER LAULA		70	COPTIC COMBINING SPIRITUS ASPER
18	COPTIC CAPITAL LETTER MI		71	COPTIC COMBINING SPIRITUS LENIS
19	COPTIC SMALL LETTER MI		72	(This position shall not be used)
1A	COPTIC CAPITAL LETTER NI		73	(This position shall not be used)
1B	COPTIC SMALL LETTER NI		74	(This position shall not be used)
1C	COPTIC CAPITAL LETTER KSI		75	(This position shall not be used)
1D	COPTIC SMALL LETTER KSI		76	(This position shall not be used)
1E	COPTIC CAPITAL LETTER O		77	(This position shall not be used)
1F	COPTIC SMALL LETTER O		78	(This position shall not be used)
20	COPTIC CAPITAL LETTER PI		79	COPTIC OLD NUBIAN FULL STOP
21	COPTIC SMALL LETTER PI		7A	COPTIC OLD NUBIAN DIRECT QUESTION MARK
22	COPTIC CAPITAL LETTER RO		7B	COPTIC OLD NUBIAN INDIRECT QUESTION MARK
23	COPTIC SMALL LETTER RO		7C	COPTIC OLD NUBIAN VERSE DIVIDER
24	COPTIC CAPITAL LETTER SIMA		7D	COPTIC FRACTION ONE HALF
25	COPTIC SMALL LETTER SIMA		7E	COPTIC FULL STOP
26	COPTIC CAPITAL LETTER TAU		7F	COPTIC MORPHEME DIVIDER
27	COPTIC SMALL LETTER TAU			
28	COPTIC CAPITAL LETTER UA			
29	COPTIC SMALL LETTER UA			
2A	COPTIC CAPITAL LETTER FI			
2B	COPTIC SMALL LETTER FI			
2C	COPTIC CAPITAL LETTER KHI			
2D	COPTIC SMALL LETTER KHI			
2E	COPTIC CAPITAL LETTER PSI			
2F	COPTIC SMALL LETTER PSI			
30	COPTIC CAPITAL LETTER OOU			
31	COPTIC SMALL LETTER OOU			
32	COPTIC CAPITAL LETTER DIALECT-P ALEF			
33	COPTIC SMALL LETTER DIALECT-P ALEF			
34	COPTIC CAPITAL LETTER OLD COPTIC AIN			
35	COPTIC SMALL LETTER OLD COPTIC AIN			
36	COPTIC CAPITAL LETTER CRYPTOGRAMMIC EIE			
37	COPTIC SMALL LETTER CRYPTOGRAMMIC EIE			
38	COPTIC CAPITAL LETTER DIALECT-P KAPA			
39	COPTIC SMALL LETTER DIALECT-P KAPA			
3A	COPTIC CAPITAL LETTER DIALECT-P NI			
3B	COPTIC SMALL LETTER DIALECT-P NI			
3C	COPTIC CAPITAL LETTER CRYPTOGRAMMIC NI			
3D	COPTIC SMALL LETTER CRYPTOGRAMMIC NI			
3E	COPTIC CAPITAL LETTER OLD COPTIC OOU			
3F	COPTIC SMALL LETTER OLD COPTIC OOU			
40	COPTIC CAPITAL LETTER SAMPI			
41	COPTIC SMALL LETTER SAMPI			
42	COPTIC CAPITAL LETTER CROSSED SHEI			
43	COPTIC SMALL LETTER CROSSED SHEI			
44	COPTIC CAPITAL LETTER OLD COPTIC SHEI			
45	COPTIC SMALL LETTER OLD COPTIC SHEI			
46	COPTIC CAPITAL LETTER OLD COPTIC ESH			
47	COPTIC SMALL LETTER OLD COPTIC ESH			
48	COPTIC CAPITAL LETTER AKHMIMIC KHEI			
49	COPTIC SMALL LETTER AKHMIMIC KHEI			
4A	COPTIC CAPITAL LETTER DIALECT-P HORI			
4B	COPTIC SMALL LETTER DIALECT-P HORI			
4C	COPTIC CAPITAL LETTER OLD COPTIC HORI			
4D	COPTIC SMALL LETTER OLD COPTIC HORI			
4E	COPTIC CAPITAL LETTER OLD COPTIC HA			
4F	COPTIC SMALL LETTER OLD COPTIC HA			
50	COPTIC CAPITAL LETTER L-SHAPED HA			
51	COPTIC SMALL LETTER L-SHAPED HA			
52	COPTIC CAPITAL LETTER OLD COPTIC HEI			
53	COPTIC SMALL LETTER OLD COPTIC HEI			
54	COPTIC CAPITAL LETTER OLD COPTIC HAT			
55	COPTIC SMALL LETTER OLD COPTIC HAT			
56	COPTIC CAPITAL LETTER OLD COPTIC GANGIA			
57	COPTIC SMALL LETTER OLD COPTIC GANGIA			
58	COPTIC CAPITAL LETTER OLD COPTIC DJA			

A. Administrative
1. Title
Proposal to encode Medievalist and Iranianist punctuation characters to the UCS
2. Requester's name
Michael Everson, Stephen Emmel, Antti Marjanen, Ismo Dunderberg, John Baines, Susana Pedro, António Emiliano
3. Requester type (Member body/Liaison/Individual contribution)
Individual contribution.
4. Submission date
2007-03-15
5. Requester's reference (if applicable)
6. Choose one of the following:
6a. This is a complete proposal
Yes.
6b. More information will be provided later
No.

B. Technical – General
1. Choose one of the following:
1a. This proposal is for a new script (set of characters)
No.
1b. Proposed name of script
1c. The proposal is for addition of character(s) to an existing block
Yes
1d. Name of the existing block
Coptic and Combining Diacritical Marks Supplement
2. Number of characters in proposal
9 (7, 2).
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 B.1.
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.
No.
2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?
Yes.

2b. If YES, with whom?

The authors are members of the user community.

2c. If YES, available relevant documents

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

Copticists, Biblical scholars, Medievalists, Latinists, and other scholars.

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

Used historically and in modern editions.

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?

Yes.

6b. If YES, is a rationale provided?

Yes.

6c. If YES, reference

Accordance with the Roadmap. Keep with other Coptic and diacritical marks.

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?

Yes.

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

Yes.

10c. If YES, reference

See §3 above.

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, but examples of how the diacritics work are given in §4.

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?