

To: Unicode Technical Committee

From: Stephen Emmel, emmstel@nwz.uni-muenster.de

Re: **Changes to ISO/IEC JTC1/SC2/WG2 N2636 and N2676 recommended by the International Association for Coptic Studies** (represented by Prof. Dr. Stephen Emmel [Professor für Koptologie, Westfälische Wilhelms-Universität Münster, Institut für Ägyptologie und Koptologie, Schlaunstrasse 2, D-48143 Münster, Germany], with input from a number of colleagues over the years, especially Bentley Layton, Leo Depuydt, Tito Orlandi, Rodolphe Kasser, and Gregor Wurst)

Status: Expert contribution

Date: January 26, 2004

Background: At its business meeting on September 2, 2000, the International Association for Coptic Studies (IACS) passed a resolution “request[ing] the Unicode Technical Committee and the Working Group 2 in ISO/IEC JTC1/SC2 to disunify Coptic from Greek in the Universal Character Set” and offering “to provide expert advice and guidance in the matter of defining a Coptic character set, for the mutual benefit of Unicode and Coptic studies.” In 2002, the IACS gave its endorsement to N2444, “Coptic supplementation in the BMP,” an expert contribution prepared by Michael Everson and Kamal Mansour, on the basis of which it was decided to disunify Coptic from Greek. Subsequently, N2636, “Revised proposal to add the Coptic alphabet to the BMP of the UCS” (dated 2003-10-01), was prepared without any further input or contribution from the IACS. While the IACS is pleased that N2636 was approved by WG2 in 2003 (see N2676, block 2C80–2CCF), it finds the proposal N2636 to be incomplete and inaccurate in certain respects. Most importantly, there are characters that must be added (as well as a few that should be removed or renamed), as explained in detail in this memo.

Outline of this document:

- I. Coptic and its numerous dialects
- II. Specific comments on N2636
 1. Comments on N2636, pp. 3-6 (§ D, “Proposal”)
 2. Recommended changes to N2636, pp. 16–17 (“Table XX - Row 2C: Coptic”; see also N2676, block 2C80–2CCF): Characters to be removed, renamed, or added
- III. Summary of recommendations
 1. Characters to be removed from N2636/N2676
 2. Character to be renamed in N2636/N2676
 3. Characters to be added to N2636/N2676

N.B.: Characters are referred to here according to the codepoint assignments as given in N2676 (but without the notation U+), which are mostly identical with the assignments initially proposed in N2636; characters already in the UCS are marked U+.

I. Coptic and its numerous dialects

1. Coptic manuscripts as a whole, covering a period of about 2,000 years and the entirety of Egypt from Nubia in the south to the Nile Delta in the north, display a significant number of different dialects, two of which stand out as supra-regional standards: “Sahidic” (the ancient standard) and “Bohairic” (the medieval and modern standard). Bohairic, by virtue of having survived the Arab Conquest of 641 C.E., eventually developed a graphic tradition that sets it apart visually (as well as linguistically) from Sahidic and the other dialects. This Bohairic tradition became the basis for the European typographical tradition (the earliest movable Coptic font was created in Rome in 1629). Later, as Coptic studies in Europe progressed and manuscript discoveries in Egypt expanded scholars’ knowledge of the history of Coptic, designers of Coptic fonts began rather to imitate the oldest known manuscripts, which from the point of view of graphics are rather of the Sahidic variety. (At the beginning of the history of written Coptic, around the beginning of the Common Era, everything was done in relatively close imitation of the Greek uncial scripts that were in use at the time. Even Bohairic manuscripts from this classical period look “Sahidic” in this purely graphic sense.)

2. From the point of view of graphics within the manuscript tradition, dialects other than Bohairic are often distinguishable from one another only by the occurrence of certain dialect-unique characters (such as “Akhmimic kheî” [2CB9, peculiar to dialect *A*, which is also known as the Akhmimic dialect], “crossed shai” [peculiar to dialect *I*], “zetaoid kappa” [peculiar to dialect *P*], etc.), but sometimes also by general features of design that lend an entire dialect a distinctive appearance. Other than Bohairic, the best example of such a dialect-unique script is “Manichaean Lycopolitan,” the Manichaeans in Egypt having developed their own variant of one Coptic dialect (dialect *L*, also known as Lycopolitan) and also their own distinctive script for that dialect, just as they did for some languages in other parts of the world.

3. Font designers since the eighteenth century have used various manuscript models as the basis for their designs (whether movable or computerized), but it is only rarely that we find a font that is meant specifically to represent a dialect other than “general Bohairic” or “general Sahidic.” (The glyphs used for illustrative purposes in N2636/N2676 are immediately recognizable as “general Bohairic”; those used in this memo are mostly “general Sahidic.”) In fact, however, the manuscript tradition offers a large number of potential models for standardized fonts, some of which would be associated with specific dialects, while others might represent different historical periods, or different regions, or different types of literature, and so on. Coptologists regard it as necessary to have at least two such different fonts (Bohairic and Sahidic), and desirable to have still more. Some existing Coptic fonts are strange hybrids that never existed in the manuscript tradition, and no presently existing computerized Coptic font is fully adequate for all Coptological needs.

4. It follows from what has been said so far that the designer of a Coptic font must first decide which dialect she wants to represent. The IACS believes that the UCS should include a complete set of Coptic characters, so that the designer can represent any given Coptic dialect thoroughly and efficiently. We do not expect the UCS to encode any character more than once, even if that character occurs in the alphabets of more than one dialect, and even if the glyphs of that character vary considerably from dialect to dialect. On the other hand, the Bohairic dialect developed a distinct set of upper case (capital) letters, so that the letters of the Bohairic alphabet need to be represented in the UCS in casing pairs. All other characters are effectively “caseless.” (Manuscripts in many of the Coptic dialects display enlarged characters as initials, but only in Bohairic were such initials transformed systematically into distinctive glyphs, thereby becoming upper-case characters.)

5. The illustrations in N2636/N2676 that occur in cased pairs are mostly typical Bohairic glyphs, but all—especially the lower case and caseless characters—could easily be replaced by many other glyphs representing either different dialects (Sahidic, Lycopolitan, Faiyumic, Akhmimic, Mesokemic, etc.) or different styles (on the order of differences among, say, Times, Garamond, Arial, etc.). To a certain extent, also combining characters and punctuation marks differ from dialect to dialect, as specified in § II.2 below).

II. Specific comments on N2636

1. Comments on N2636, pp. 3-6 (§ D, “Proposal”)

1a. In the **introductory paragraph** to § D, it is “recognized that there will be further additions required by scholars for special purposes, but the set here . . . is the core character set for Coptic.” This statement is not accurate. The set of characters proposed in N2636 (in addition to U+03E2 – U+03EF, “Greek and Coptic: Coptic-unique letters”) is the core character set only for the best known dialects of Coptic. It does not fully represent all the Coptic dialects. The recommended changes specified below (§ II.2) fill this deficiency by adding eighteen caseless characters not given in N2636/N2676. This recommendation, in its entirety, may be taken to be the foreseen “further additions required by scholars” (including certain script-unique symbols), with the clarification that the “special purposes” here in view are merely the normal needs of professional Coptologists who work in a technical manner with Coptic in all its varieties and not just with some more or less restricted part or parts of it.

The proposal’s parenthetical reference to “alphabetical characters representing the sounds of Old Nubian” is not entirely out of place here, since the Old Nubian alphabet consists of the borrowed Coptic alphabet with several additional letters. However, the IACS takes the position that there is no more reason to unify Coptic and Old Nubian in a single character set than there was to unify Coptic and Greek. Hence, while the IACS recommends that Old Nubian be included in the UCS, Old Nubian is left entirely out of further consideration in the present set of recommendations, which concern only Coptic.

1b. Combining characters (N2636, pp. 3–4).

- The “diaeresis” (Coptologists generally prefer the term “trema” in view of the various functions of this character in Coptic in comparison with Greek; furthermore, the description “in syllable-initial position” is incomplete and inadequate) can occur over more letters than just the two listed in N2636. In dialect *F7*, for example, the trema is regularly used instead of the “jinkim” (on which see below) for marking syllabic /m/ and /n/. Examples: ṭ ṡ ṇ ṁ
- The “apostrophe” is not always a combining character, and therefore it belongs also under the heading punctuation.
- The “circumflex” occurs already in the earliest Coptic manuscripts, that is, pre-seventh century, and not just over syllables formed by a single vowel. Examples: ֿ ֿ ֿ ֿ ֿ.
- The “acute accent” is not only a modern didactic usage, but occurs sporadically in Coptic manuscripts of all periods to indicate stress especially in Greek loan words.

- The name “jinkim” refers specifically to the superliteral dot (or some other mark) used in the Bohairic dialect, where this mark was given a name by medieval Coptic grammarians (examples: ⲛ ⲛ̇). The equivalent mark in Sahidic is called the “superlinear stroke” by Coptologists (examples: ⲙ ⲙ̇ ⲙ̇ⲙ̇), who prefer not to describe the function of these marks as involving “a reduced vowel.” All the evidence suggests that consonants so marked were recognized by the creators of the Coptic writing system as what we would now call syllabic consonants (“the syllabic pronunciation of a character,” as N2636 puts it), whence these combining characters might be called generally “marks of syllabicity” or “syllabicity markers”. These marks were an essential part of the Coptic writing system from the very beginning, and the behavior especially of the superlinear stroke poses problems for font design (see § II.2g below).
- On additional combining characters not mentioned here, see § II.2h below.

1c. Punctuation (N2636, p. 4). The list that is given here of punctuation characters used in Coptic is not complete. Further clarification of the “special oblique double hyphen” in Coptic and other languages (now approved in N2676 as 2E17 ≠ DOUBLE OBLIQUE HYPHEN) has been provided by D. Anderson in her memo L2/03-347, “Background on DOUBLE OBLIQUE HYPHEN” (individual contribution, dated October 19, 2003).

1d. Numerals and abbreviations (N2636, p. 4). The character RO WITH STROKE (2CBA ρ) belongs here (rather than under “additional symbols,” q.v. in the next paragraph, § II.1e) because it is in fact the Coptic numeral 900 (and not the same as the symbol for *tau-rho*, on which see the next paragraph).

1e. Additional symbols (N2636, p. 4). As stated in the preceding paragraph (§ II.1d), 2CBA ρ RO WITH STROKE (the numeral 900) must be distinguished from the “standard abbreviation” (ⲑ̇) for the character sequence TAU-ALFA-UA-RO, which character we recommend adding to the UCS and naming (1) COPTIC SYMBOL TAU RO (see § II.2f below). This latter symbol belongs with (2) COPTIC SYMBOL MI RO (2CBB ⲙ̇) and (3) COPTIC SYMBOL KHI RO (2CBC ⲕ̇) as another of six such symbols that are essential for Coptic text processing. The fourth, fifth, and sixth symbols, also needing to be added to N2636/N2676, are: (4) a combination of PI and RO (ⲑ̇ⲑ̇) standing for the Greek loan word *pros* (the preposition, as in the title of St. Paul’s epistles “to” the Romans, “to” the Galatians, etc.); (5) a combination of O and UA (ⲟⲩ) standing for the character sequence O-UA and possibly functioning as an alphabetic character in some Coptic dialects; and (6) a compendium (ⲕ̇) standing for the letter sequence KAPA-ALFA-IAUDA or KAPA-EIE and possibly functioning as an alphabetic character in some Coptic dialects. Essential for Bohairic is a ligatured combination of SHIMA and SIMA (ⲟⲩ) standing for the word *cōis* “lord” (compare the symbol 2CBD ⲩ̇, for *stauros* “cross”). See § II.2f below. There were (and still are) additional symbols, compendia, and monograms in use among Copts (for example: ⲙ̇ⲙ̇ⲙ̇ ⲙ̇ⲙ̇ⲙ̇ⲙ̇ ⲙ̇ⲙ̇ⲙ̇ⲙ̇ⲙ̇), but we do not regard it as scientifically necessary to request that all such characters be encoded in the UCS.

1f. Character names (N2636, pp. 4–5). The IACS regards the (more than five) additions recommended below as *core* additions, not as “*non-core* additions.” In discussing these additions to the core Coptic character set, we use the names already proposed (N2636) and accepted (N2676) as a model for the additional names, except that these latter are underlined and no codepoints are given for them. We assume that these provisional names will be changed, at least in some cases, in accordance with Unicode’s rules for naming characters.

1g. Bibliography (N2636, pp. 5–6). Additional references:

- Emmel, Stephen. 1992. “Coptic Language.” In: *The Anchor Bible Dictionary*, vol. 4, pp. 180–188. New York: Doubleday.
- Kasser, Rodolphe. 1980. “Prolégomènes à un essai de classification systématique des dialectes et subdialectes coptes selon les critères de la phonétique II. Alphabets et systèmes phonétiques.” *Muséon* 93:237–297.
- Kasser, Rodolphe. 1988. “Nommer les principaux graphemes vieux-coptes?” *Bulletin de la Société d’Égyptologie*, Genève 12:53–57.
- Kasser, Rodolphe. 1990. “A Standard System of Sigla for Referring to the Dialects of Coptic.” *Journal of Coptic Studies* 1:141–151.
- Kasser, Rodolphe. 1991. “Alphabets, Coptic” and “Alphabets, Old Coptic.” In: *The Coptic Encyclopedia*, vol. 8, pp. 32–45. New York: MacMillan.
- Layton, Bentley. 1973. “The Text and Orthography of the Coptic *Hypostasis of the Archons* (CG II, 4 Kr.).” *Zeitschrift für Papyrologie und Epigraphik* 11:173–200, with one plate.

2. Recommended changes to N2636, pp. 16-17 (“Table XX - Row 2C: Coptic”; see also N2676, block 2C80–2CCF): Characters to be removed, renamed, or added

2a. The characters 2C99 COPTIC CAPITAL LETTER AKHMIMIC KHEI and 2C9A COPTIC CAPITAL LETTER RO WITH STROKE **should be removed**. Because 2C99 is not a Bohairic character (see § I.4 above), and because 2C9A is a non-alphabetic character, neither one occurs in upper-case (capital) form.

2b. Character 2CBD Ⲭ COPTIC SYMBOL KHRISTOS **should be renamed**. The illustration for this character is not a symbol for *khristos* “Christ,” as the character name would seem to indicate, but rather for *stauros* “cross.” Hence 2CBD should be renamed COPTIC SYMBOL STAUROS.

2c. The minimal Coptic character set should include:

- 81 alphabetic characters: 32 characters in casing pairs (U+03E2 – U+03EF, plus 2C80–2C98, plus 2CA0–2CB8) and 17 caseless (2CB9 plus the additions specified in § II.2d below)
- 2 non-alphabetic numerals (2CBA and 2CBF)
- 2 cryptograms (which function as alphabetic characters in encrypted texts; see § II.2e below)
- 8 symbols (which can replace certain strings of alphabetic characters; 2CBB–2CBD plus the additions specified in § II.2f below)
- at least 3 combining “syllabicity markers” (see § II.2g below)
- certain additional combining characters (see § II.2h below)
- certain punctuation marks (see § II.2i below)
- certain editorial symbols (all already in the UCS)

In order for this minimal set of characters to be available to the designer of a Coptic font, certain characters need to be added to the UCS, as described in the following paragraphs.

2d. Additional alphabetic characters. In order to complete the repertoire of 81 alphabetic characters given in N2636/N2676 (and including U+03E2 – U+03EF), 16 more characters must be added (these characters have no known Coptic names, for which reason the following list gives the names that Coptologists have assigned to them, underlined, as explained in § II.1f above; the term “grapheme” in some of the following names is simply a synonym for “character” and could be omitted from the Unicode names). These characters are relatively little known to non-specialists because they occur in very early Coptic manuscripts (the so-called “Old Coptic texts,” designated in the following table by the abbreviations “Schm.,” “Hor.,” etc.) or only in the sparsely attested dialects *P* and *I*. The following table (from Kasser 1991, 42) shows the distribution of the characters proposed for approval below across the most important Old Coptic (listed horizontally across the top of the table) texts and gives (vertically, in col. 1) their approximate phonemic values in the alphabets in which they were used. Note that these Old Coptic characters occur alongside the better known Coptic characters in the alphabets in which they were used (most of the better known characters are simply not shown here; see rather Kasser 1991, 38–39).

TABLE 1. *Synoptic Table of Old Coptic Alphabets, with Commentary*

		Schm.	Hor.	Mich.	Mun.	Ox.	Mum.	Mim.	DMP.	Par.1	Par.2	<i>P</i>	<i>I</i>
1	/k/	ⲕ	Ⲛ	ⲕ?	?	ⲕ	?	ⲕ	Ⲛ	ⲕ	ⲕ	Ⲛ	ⲕ
2	/ŋ/	ⲛ	–	ⲛ	?	ⲛ	ⲛ	ⲛ	ⲛ	ⲛ	ⲛ	–	ⲛ
3	/š/	/	ⲟ	ⲟ	?	ⲟ	ⲟ	Ⲓ	ⲟ	ⲟ?	Ⲓ	ⲟ	ⲟ
4	/f/	ⲑ	ⲑ	ⲑ	?	ⲑ	?	ⲑ	ⲑ	ⲑ	ⲑ	ⲑ	ⲑ
5	/h/	ⲡ	ⲡ	ⲡ	?	ⲡ	?	ⲟ	?	ⲟ	Ⲓ	ⲡ	ⲡ
6	/č/	ⲛ	ⲛ	ⲛ	?	ⲛ	?	Ⲓ	ⲛ	ⲛ?	?	ⲛ	ⲛ
7	/c/	ⲕ	ⲕ	?	?	Ⲓ	?	?	ⲕ	?	?	ⲕ	Ⲓ
8	/ti/	=	=	=	?	=	?	=	=	=	=	†	†
9	/ʾ/		(Ⲓ)?			?	?	?	?		gem.	Ⲓ	gem.
10	/ç/		Ⲓ	(ⲛ)?	Ⲓ	Ⲓ	?		Ⲓ	Ⲓ		ⲛ	ⲛ
11	/x/	?	ⲛ	ⲛ	?	ⲛ	?		??	?		ⲛ	ⲛ
12	/čh/								ⲛ/ⲡ/ⲛ		(ⲛ)		
13	/h/	ⲟ			ⲛ	ⲛ	?		ⲛ?				

Notes: The character COPTIC Y-SHAPED GRAPHEME in line 13 under “DMP.” occurs alongside COPTIC P-SHAPED GRAPHEME with a question mark only because its phonemic value (either /ú/ or /h/) is uncertain, and because in this text it varies with ρ (perhaps under phonetically conditioned circumstances). The character COPTIC STRETCHED-CAPITAL-SIGMA-SHAPED GRAPHEME in line 10 under “Mich.” occurs in parentheses and with a question mark because there is but one example, in a context that renders its phonemic interpretation obscure; the same character seems to occur once also in “DMP.,” there too perhaps with the value /ç/. The shading in the table serves various purposes in the context in which it was published, but they are not relevant here (cf. Kasser 1980, 255–270). Most important is that “in this presentation of the graphemes typical of Old Coptic, each is given in a standardized form and no account is taken of its numerous particular graphic variants (sometimes very perceptibly remote from one another). . . . Furthermore, the signs are limited to those that appear regularly in these texts (or nearly so); it has not been judged indispensable to include also certain rare forms of Old Coptic graphemes whose use is occasional (or even, in most cases, exceptional) and does not seem to have any particular phonological significance” (Kasser 1991, 43). That is to say, the following list (which is based on the preceding table and related studies) is minimal and represents not merely variant glyphs of the better known characters, but entities that Coptologists believe to be alphabetic characters in their own right. These characters occur alongside the better known Coptic alphabetic characters in the earliest manuscripts, but most of them simply fell out of use by the end of the fourth century, as a standardized representation of the Sahidic alphabet came to be used all over Egypt. In many cases, the forms of these extinct characters can be traced back to their graphic antecedents in the pre-Coptic writing system known as Demotic. When describing the Egyptian language either synchronically (during the Hellenistic period and after) or diachronically, Coptologists and Egyptologists must use the entire Coptic alphabetic character set, including the characters proposed in the list below.

The following illustrations are printed in 48 point type, rather than in 10 point (for example: ⲁⲓⲣ ⲛⲓⲁ ⲁⲉⲥ –ⲥⲱ
ⲛⲏⲑ ⲉⲓ ⲓⲕ), for the sake of clarity in this document:

Ⲛ

COPTIC REVERSED-TAU-SHAPED ALEPH (see table, line 9)

ⲛ

COPTIC ZETAOID KAPPA (see table, line 1)

ⲛ

COPTIC HYPHEN-SHAPED NU (see table, line 2)

ⲱ

COPTIC CROSSED SHAI (see table, line 10)

ⲉ

COPTIC 9-SPIRALED GRAPHEME (see table, lines 5, 10, and 11)

ⲛ

COPTIC STRETCHED-CAPITAL-SIGMA-SHAPED GRAPHEME (see table, line 10)

ⲉ

COPTIC 3-SHAPED GRAPHEME (see table, lines 3, 5, and 13)

6

COPTIC 6-SPIRALED GRAPHEME (see table, lines 3 and 10)COPTIC FRACTION-STROKE-SHAPED GRAPHEME (see table, line 3)

P

COPTIC P-SHAPED GRAPHEME (see table, lines 5 and 13)

L

COPTIC L-SHAPED GRAPHEME (see table, line 5)COPTIC HOOK-SHAPED GRAPHEME (see table, line 13)COPTIC Y-SHAPED GRAPHEME (see table, line 13)COPTIC MINUSCULE-ALPHA-SHAPED GRAPHEME (see table, lines 6 and 12)

11

COPTIC REVERSED-PI-SHAPED GRAPHEME (see table, line 6)COPTIC DIVIDED-TRIANGLE-SHAPED GRAPHEME (see table, line 7)

2e. Cryptograms “appear to have been particularly popular among Coptic monks, who scribbled them in their religious books and on cliffs, walls and doorposts. They date from the early days of monasticism until the end of the Coptic era” (Frederik Wisse, “Language Mysticism in the Nag Hammadi Texts and in Early Coptic Monasticism I: Cryptography,” *Enchoria* 9 [1979] 101–120, at p. 103). The several known systems of encryption are all basically substitution codes, where one letter of the alphabet is replaced by another. But in two particular cases (*epsilon* [EIE] and *nu* [NI], two letters that occur with very high frequency in Coptic), special characters were often used to further encrypt characters that according to the basic system would otherwise be encrypted by themselves, hence not really encrypted at all. These two characters, which must be added to N2636/N2676, are:

COPTIC VERTICAL CRYPTOGRAMCOPTIC HORIZONTAL CRYPTOGRAM

2f. In order to complete the essential repertoire of **Coptic-unique symbols** given in N2636/N2676 (that is, in addition to 2CBB and 2CBC; see § II.1e above), the following five characters must be added:



COPTIC SYMBOL TAU RO



COPTIC SYMBOL PI RO



COPTIC SYMBOL KAI



COPTIC SYMBOL O UA



COPTIC SYMBOL COIS

Notes: COPTIC SYMBOL TAU RO is distinct from the character 2CBA ρ RO WITH STROKE (see §§ II.1d and 1e above). COPTIC SYMBOL PI RO stands for the Greek loan word *pros* (the preposition, as in the title of St. Paul's epistles "to" the Romans, "to" the Galatians, etc.). COPTIC SYMBOL KAI stands for the letter sequence KAPA-ALFA-IAUDA or KAPA-EIE and possibly functions as an alphabetic character in some Coptic dialects. COPTIC SYMBOL O UA stands for the character sequence O-UA and possibly functions as an alphabetic character in some Coptic dialects. COPTIC SYMBOL COIS is a ligatured combination of SHIMA and SIMA that stands for the (Bohairic) word *cōis* "lord."



2g. As stated above (§ II.1b, fifth bullet "jinkim"), there are certain **Coptic-unique syllabicity markers** that can be treated as **combining characters**, whose appearance and behavior will now be described.




(i) Coptic employs several combining characters, for several different purposes. Peculiar to Coptic is a set of superlinear (or superliteral) marks for indicating that a given consonant or group of consonants is or are syllabic. The same marks may also occur with single vowels that form syllables in and of themselves. Coptologists distinguish three basic forms of this "syllabicity marker" (or "mark of syllabicity"): a superlinear (often simply superliteral) stroke (ⲛ ⲛⲁ ⲛⲁⲛⲁ); a superliteral dot (properly called a "jinkim" in Bohairic: ⲙ); and something like a grave accent mark (ⲕ). To a great extent, these forms are associated with particular dialects and groups of dialects: the jinkim and grave accent mark are typically Bohairic, while the superlinear stroke is typically Sahidic. However, just as the variety of forms occurring in the manuscripts is greater than three, so their distribution is not cleanly definable in terms of dialects: there are Sahidic manuscripts and other dialects with jinkims, and there are Bohairic manuscripts and other dialects with superlinear strokes. Furthermore, various forms often occur in one and the same manuscript.

(ii) The behaviors of the superlinear stroke, the jinkim, and the grave accent mark are partly more complex than might first be thought. First, and most obviously, the width of a superlinear stroke, or the placement of a jinkim or accent, depends on the shape (width and height) of the character over which it occurs. The necessary adjustments are easy to make by hand in a manuscript, but they are not so easy to make with a printed font. The most successful Coptic fonts have employed multiple sets of the alphabetic characters, with customized superlinear strokes etc. designed already in combination with each character. In fonts where these marks are treated as combining characters, it has usually been found necessary to provide several different marks to choose from (different widths, heights, etc.), leaving it to the user to find the best choice for combining with a given character or string. The result is sometimes far from satisfactory. Second, further complexity results from the fact that the shape and placement of these marks can vary even over one and the same character in a single manuscript. Coptologists are very often forced to standardize (or regularize) the usage of a given manuscript when reproducing it in print. Third, and most complex, the superlinear stroke can occur either over a single character, or over two or more (relatively seldom over three or more) characters. In such cases, the manuscripts show a great deal of fluctuation in terms of the exact placement and width of the stroke.

(iii) Even if it is unrealistic for a font to provide sufficient flexibility to imitate every manuscript phenomenon of this type exactly, any good Coptic font must include the possibility of certain basic placements, taking into account the different widths of different characters. Here are examples of the most basic needs (using 36 point Coptic type for the sake of clarity):

“Coptic combining superlinear stroke”:

 (wide single-letter stroke, with 2CAC MI)
 (narrow single-letter stroke, with 2CA9 IAUDA)
 → COMBINING MACRON (U+0304) and/or COMBINING OVERLINE (U+0305)

 or  or  etc. (two-letter “Bindestrich”) → COMBINING DOUBLE MACRON


 or  or  etc. (three-letter “Bindestrich”)

 or  etc. (four-letter “Bindestrich”)


“Coptic combining jinkim”:

 (centered jinkim) → COMBINING DOT ABOVE (U+0307)

 (right-shifted jinkim) → COMBINING DOT ABOVE RIGHT

 (left-shifted jinkim; the dot should be farther to the left) → COMBINING DOT ABOVE LEFT

“Coptic combining grave accent”:

 (centered grave accent mark) → COMBINING GRAVE ACCENT (U+0300)

(iv) Let us look at characters already existing in the UCS with which the Coptic superlinear stroke, jinkim, and grave accent mark might be unified (compare N2636, p. 4 top).

As characters, COMBINING DOT ABOVE (U+0307) and COMBINING GRAVE ACCENT (U+0300) may be taken as equivalents of the Coptic centered jinkim (ⲓ) and the Coptic centered grave accent mark (ⲙ), respectively. In the case of the jinkim in the shape of a dot, however, the placement (centered, left-shifted, right-shifted) varies even on a single page, and Coptologists believe that this variation might be significant. Hence they need to have also be a COMBINING DOT ABOVE LEFT (as in: ⲓ̀) and a COMBINING DOT ABOVE RIGHT (as in: ⲓ́), which we therefore recommend, if they do not exist already in the UCS. (Compare U+0315 COMBINING COMMA ABOVE RIGHT alongside U+0313 COMBINING COMMA ABOVE.) Note that the COMBINING DOT ABOVE RIGHT that is being requested can be unified with a Coptic combining punctuation mark described in § II.2h below, namely the “Coptic combining dot Trenner.”

For the more complex characteristics of the Coptic superlinear stroke, it is possible that the existing COMBINING MACRON (U+0304), COMBINING OVERLINE (U+0305), and COMBINING DOUBLE MACRON (U+035E) will be sufficient for the single- and 2-letter combinations. But for combinations involving three or

more characters (including long strings of overlined characters as in: ⲓⲁⲗⲗⲁⲛⲉⲛⲱⲙⲉ), we assume that the designer of a Coptic font will have to find a creative solution not depending just on this or that specific combining character. **Indeed, for the 2-, 3-, 4-, and multi-character “Bindestrich” superlinear stroke, the IACS would be glad to have technical advice about how an optimal Coptic font might best solve the font-design problem posed by this feature of the Coptic writing system.**

2h. Other combining characters that are essential to a Coptic font for the most part already exist in the UCS. However, for the **trema (diaeresis, dialytika)** U+0308 ˘ COMBINING DIAERESIS, we recommend that the two most frequent combinations, namely COPTIC SMALL LETTER IAUDA WITH TREMA and COPTIC SMALL LETTER UA WITH TREMA, be encoded as separate characters (as has been done for Greek at U+03CA ι GREEK SMALL LETTER UPSILON WITH DIALYTIKA and U+03CB υ GREEK SMALL LETTER UPSILON WITH DIALYTIKA), as follows (note the difference in the two tremas, determined by the differing widths of the two characters):

ï COPTIC SMALL LETTER IAUDA WITH TREMA

ÿ COPTIC SMALL LETTER UA WITH TREMA

Three further combining characters are different shapes of what Coptologists call the “**Trenner**” (German for “divider”) or “division mark” or “apostrophe,” which can occur in three basic forms in a single text, namely: “straight apostrophe,” “hook,” and “point” (Layton 1973, 190–200). These forms are here called (in closer conformity with Unicode’s terminology): “grave accent,” “apostrophe,” and “dot.” The Coptic Trenner is a punctuation mark for dividing words (what counts as a “word” is variable, sometimes even within a single manuscript), and it belongs here with other combining characters only because sometimes (especially in early manuscripts) it takes the form of a combined “attachment” to certain characters (ⲓ GAMMA, ⲕ KAPA, Ⲛ KSI, ⲛ PI, and ⲣ TAU) rather than stands just after it (as a spacing character: cf. U+0060 GRAVE ACCENT, U+0027 APOSTROPHE, and U+02D9 DOT ABOVE). For example:

 COPTIC SMALL LETTER TAU with COMBINING GRAVE ACCENT ABOVE RIGHT

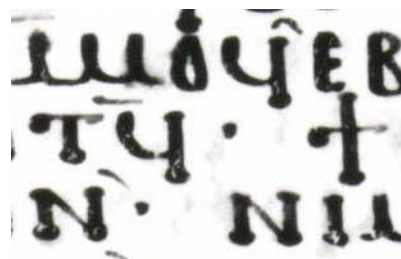
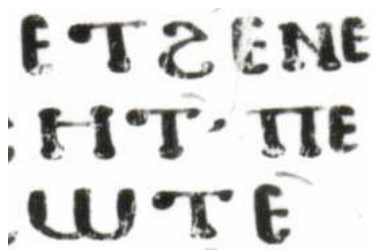
 COPTIC SMALL LETTER TAU with COMBINING COMMA ABOVE RIGHT (U+0315)

 COPTIC SMALL LETTER TAU with COMBINING DOT ABOVE RIGHT

For the “Coptic combining grave accent Trenner,” compare U+0300 COMBINING GRAVE ACCENT, except that the Trenner must be right-shifted; hence we need a COMBINING GRAVE ACCENT ABOVE RIGHT. For the “Coptic combining apostrophe Trenner,” we suppose that this character can be unified with U+0315 COMBINING COMMA ABOVE RIGHT. For the “Coptic combining dot Trenner,” obviously a COMBINING DOT ABOVE RIGHT can be the same character as one of the three forms of the jinkim (see § II.2g[iv] above), even though the functions of the characters are quite different.

2i. Punctuation. Turning now to non-combining marks of punctuation, Coptic requires a number of marks that already exist in the UCS, but also one that apparently does not, namely a MIDDLE COMMA, that is, a comma

raised well above the line of writing (standing in the same relation to a normal comma as a middle dot stands in relation to a normal dot or period), for example:



vs. the MIDDLE DOT, as in:

III. Summary of recommendations

1. Characters to be removed from N2636/N2676:

- 2C99 COPTIC CAPITAL LETTER AKHMIMIC KHEI
- 2C9A COPTIC CAPITAL LETTER RO WITH STROKE

2. Character to be renamed in N2636/N2676:

- 2CBD COPTIC SYMBOL KHRISTOS → COPTIC SYMBOL STAUROS

3. Characters to be added to N2636/N2676

16 caseless alphabetic characters

- Ⲁ COPTIC REVERSED-TAU-SHAPED ALEPH
- ⲁ COPTIC ZETAOID KAPPA
- Ⲃ COPTIC HYPHEN-SHAPED NU
- ⲃ COPTIC CROSSED SHAI
- Ⲅ COPTIC 9-SPIRALED GRAPHEME
- ⲅ COPTIC STRETCHED-CAPITAL-SIGMA-SHAPED GRAPHEME
- Ⲇ COPTIC 3-SHAPED GRAPHEME
- ⲇ COPTIC 6-SPIRALED GRAPHEME
- Ⲉ COPTIC FRACTION-STROKE-SHAPED GRAPHEME
- ⲉ COPTIC P-SHAPED GRAPHEME
- Ⲋ COPTIC L-SHAPED GRAPHEME
- ⲋ COPTIC HOOK-SHAPED GRAPHEME
- Ⲍ COPTIC Y-SHAPED GRAPHEME
- ⲍ COPTIC MINUSCULE-ALPHA-SHAPED GRAPHEME
- Ⲏ COPTIC REVERSED-PI-SHAPED GRAPHEME
- ⲏ COPTIC DIVIDED-TRIANGLE-SHAPED GRAPHEME

2 cryptograms

- Ⲑ COPTIC VERTICAL CRYPTOGRAM
- ⲑ COPTIC HORIZONTAL CRYPTOGRAM

5 symbols

- Ⲓ COPTIC SYMBOL TAU RO
- ⲓ COPTIC SYMBOL PI RO
- Ⲕ COPTIC SYMBOL KAI
- ⲕ COPTIC SYMBOL O UA
- Ⲍ COPTIC SYMBOL COIS

3 combining characters

- ⲍ COMBINING DOT ABOVE LEFT *
- Ⲏ COMBINING DOT ABOVE RIGHT *
- ⲏ COMBINING GRAVE ACCENT ABOVE RIGHT *

2 characters pre-combined with combining characters

- Ⲑ COPTIC SMALL LETTER IAUDA WITH TREMA
- ⲑ COPTIC SMALL LETTER UA WITH TREMA

1 punctuation mark

- Ⲓ MIDDLE COMMA *

* marks four characters that need not be encoded as Coptic-unique characters, but which could be added either as general punctuation or in connection with, for example, “Ancient Greek Textual Symbols” as now approved in N2676, block 2E00–2E7F.