ISO/IEC JTC1/SC2/WG2 N3886
L2/10-333
2010-09-08

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

Doc Type: **Working Group Document**
Title: **Towards the encoding of a complete set of Coptic numbers in the UCS**
Source: **Michael Everson, Stephen Emmel (Universität Münster, American University in Cairo, International Association of Coptic Studies)**
Status: **Individual Contribution**
Action: **For consideration by JTC1/SC2/WG2 and UTC**
Date: **2010-09-08**

The “Coptic Numerals” requested in the document N3843 “Final Proposal to Encode Coptic Numbers in ISO/IEC 10646” by Anshuman Pandey, are what the Copts call the “epact” numerals, “epact” being the Greek word ἐπακτός ‘imported’. In general, these numerals are found only in documentary contexts, that is, in texts written for sundry ephemeral purposes, as opposed to literary manuscripts (though we give below the evidence of the one literary manuscript which we happen to know where the epact numerals occur, namely among the page numbers). The relevant documents (and there seem to be very many of them) are mainly from the centuries after the Arab Conquest, that is, from the eighth century CE and later (see Abd El-Shaheed Abd El-Nour 1993).

These are things with which papyrologists (rather than philologists) work. Greek documents from Egypt of the same period look very similar, and this common script has affinities with the Greek miniscule script that was developed in Byzantium for use in literary manuscripts (which became the basis for the standard tradition of modern printed Greek). It appears also that many of the relevant documents are written in Arabic, with the numbers written in Coptic, using the epact numerals.

What we are dealing with is a medieval Coptic borrowing of a then new cursive Greek script, at a time (post-Arab Conquest) when the use of Coptic for documentary purposes began really to flourish widely for the first time. As a part of the new script, numerals also got a new shape, the “epact” numerals. (We do not know how far back this designation for these numerals is attested.) Eventually, it seems, when Coptic died out as a living language, surviving only in a literary and liturgical tradition within the Coptic Church, these “epact” numerals seem to have gained their status as a distinct set of characters alongside the older uncial characters in their use as numerals. See Figure 1 below, an illustration from Abd El-Shaheed Abd El-Nour 1993, 19 (the epact numerals are in col. 4 of that table).

Given that the epact numerals are cursive characters, it is not surprising that they occur in a number of glyph variants in the corpus of documentary texts. We are not certain which variants to choose for a standard, nor do we know to what extent (if any) a standard is even discernible somewhere. It is interesting in this regard that someone already in the mid-nineteenth century was able to print such a clean-looking set of epact numerals as in Pihan 1860 (according to Pandey’s Figure 2, reproduced here also as Figure 2). We wonder what Pihan used for a model.

In Figure 3 below we show an example of the repertoire of a digital font based on a hot-metal font that was created for the Institut Français d’Archéologie Orientale.

In Figure 4 below we show the actual glyphs of a variety of epact numbers used as page numbers in the White Monastery codex GI.

With regard to another aspect of the current proposal, it must be immediately obvious that the proposed COPTIC NUMBER MARK brings with it the same technical problem that we have had with the Coptic superlinear stroke, namely that a single character in the UCS would be useful only over a single numeral, not over two or more numerals. Furthermore, this COPTIC NUMBER MARK is really just a glyph variant of the COMBINING MACRON and/or the relevant COMBINING MACRON LEFT HALF, COMBINING MACRON RIGHT HALF, and COMBINING CONJOINING MACRON. It is the view of the International Association for Coptic Studies (IACS) that only these existing characters should be used for representing "superlinear strokes" over the epact numbers. If there *is* a requirement to encode this "supertending mark" at U+0605 as a single character ARABIC COPTIC NUMBER MARK, then we do not believe that a standard Coptic font either will be able or should be required to implement the technology to get such a mark to behave as it would have to behave in order to be rendered correctly (compare the necessity that was already recognized for the encoding of three characters COMBINING MACRON LEFT HALF, COMBINING CONJOINING MACRON, and COMBINING MACRON RIGHT HALF, for use over two or more Coptic characters, in addition to COMBINING MACRON for use over a single Coptic character). Perhaps fonts in use for rendering the Arabic script routinely implement such a feature, but we are not aware of that.

Figure 6 below is the key reason for this discussion paper, however. It shows a table from Stern 1880, which shows a variety of shapes for the Coptic numbers. Moreover, it shows a number of fraction signs, and it is these fraction signs which we believe require further study so that a complete set of characters can be encoded.

An analysis of the fractions in Figure 6 shows some patterns:

- The fractions for ⅓ and ¼ and ⅔ appear to be new characters.
- The fractions for ⅕ and ⅐ and ⅑ and ⅒ appear to be the numbers 5, 7, 9, and 10 with a diacritic.
- The fractions for ⅙ and ⅛ do not quite look like the numbers 6 and 8 though the first appears to have the same diacritic that is used with 5, 7, 9, and 10, as noted above.
- The fraction for ⅖ appears to be the number 5 with two of the diacritics.
- The fraction for ⅘ appears to be (4 × ⅕) with the multiplication represented by an overline.
- The sum (⅕ + ⅐) is just ⅕ and ⅐ side-by-side.
- The product (⅕ × ⅐) is ⅕ and ⅐ side-by-side with the multiplication represented by an overline.

Clearly there is a pattern here, but the rules and details are not yet clear to us. Anshuman Pandey has done excellent preliminary work—particularly in examining some of the Arabic-script materials—but we believe that further research is needed, especially with regard to the fractions, with input from Coptic papyrologists who are deeply engaged with the texts in which numbers of all sorts occur most frequently. Therefore we would like to request that WG2 and the UTC not yet encode the Coptic numbers. We think it would be best for the numbers, including the fractions, all to be encoded in a single block, for which purpose we need first to be as certain as possible that the entire repertoire of distinct characters has been collected, and we need to be as certain as possible that the system of usage of those characters has been fully analyzed and understood. For example, since it is not certain what the mechanism for generating (for instance) other ⅖-type fractions, and since it is uncertain whether the diacritic in the ⅕-type fractions is productive or not, further research is indicated.

Since the International Association for Coptic Studies will be an implementor of this block of numbers and fractions, we would very much prefer to have a comprehensive set ballotted all at once.

Bibliography.

Abd El-Shaheed Abd El-Nour, Samiha. 1993. "Epact Numerals," in: *Acts of the Fifth International Congress of Coptic Studies, Washington, 12–15 August 1992*, vol. 2, *Papers from the Sections*, edited by David W. Johnson, (Rome: Centro Italiano Microfiches, 1993), part 1, 13–21.

Pihan, Antoine Paulin. 1860. *Exposé des signes de numération usités chez les peuples orientaux anciens et modernes*. Paris: L'imprimerie impériale.

Stern, Ludwig. 1880. *Koptische Grammatik*. Leipzig.

Figures

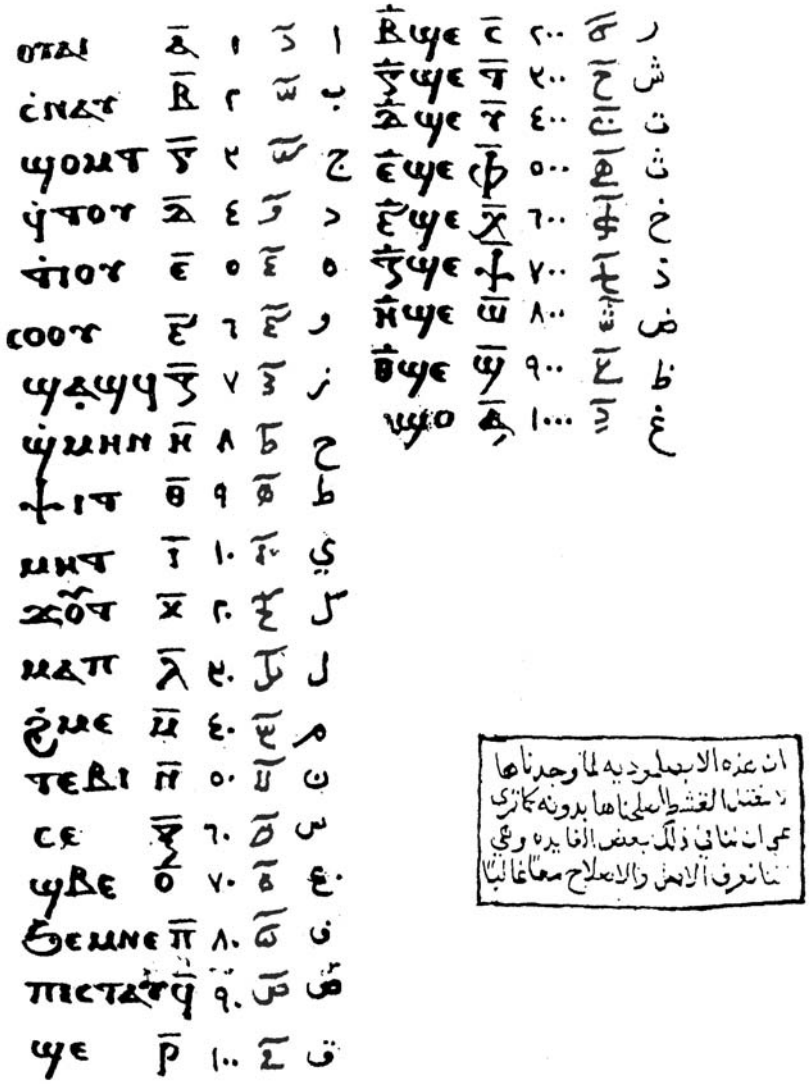

PLATE I: Table of numerals, giving the names of the Coptic numerals in Bohairic, followed by the Coptic, Arabic, epact, and *al-ğummal* numerals. (Coptic Museum MS 13 Liturgy (Serial No. 284))

Figure 1. Illustration from Abd El-Shaheed Abd El-Nour 1993:19 (the epact numerals are in col. 4 of the table).

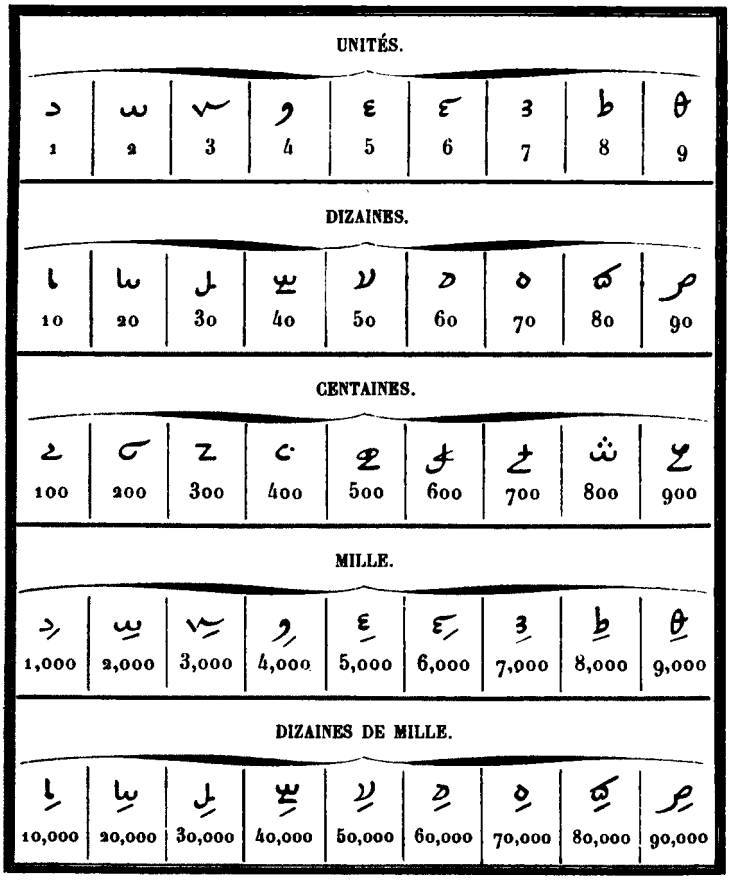

Figure 2. Table showing the forms of Coptic numbers (from Pihan 1860:213)

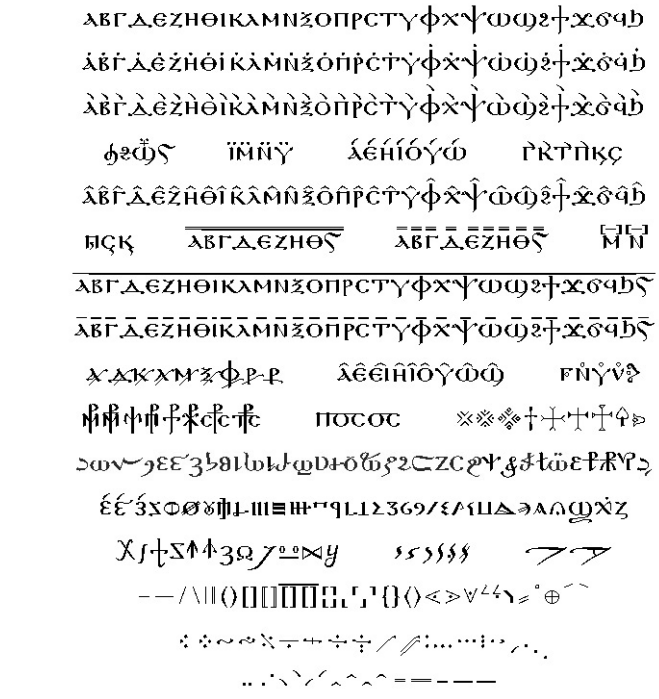

Figure 3. The font specimen for the recently-created electronic version of the early twentieth-century Coptic font that was designed for use at the Institut Français d'Archéologie Orientale in Cairo. The epact numerals are in line 11.

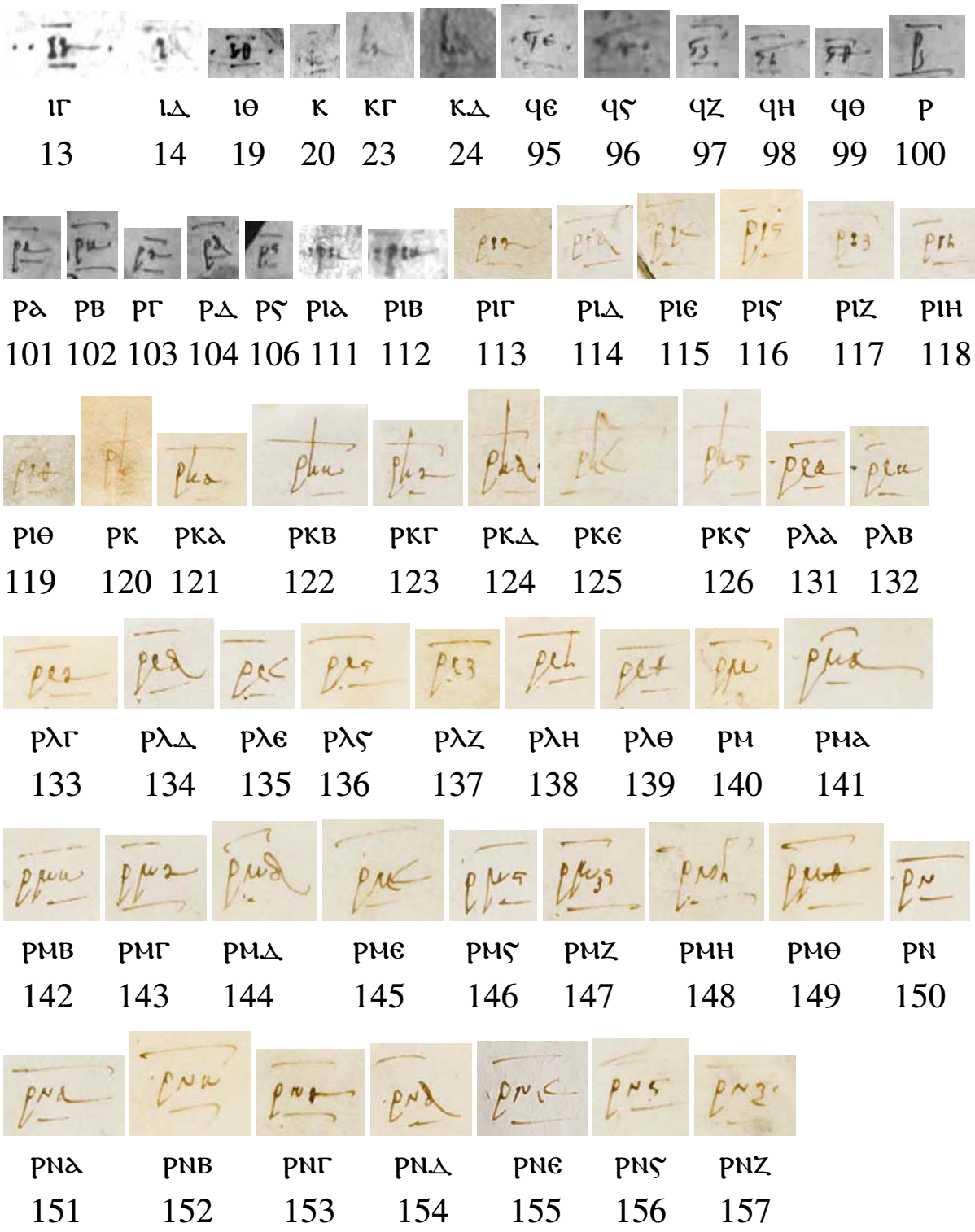

Figure 4. The "Epact ('imported' [Greek ἐπακτός]) Numerals" used as page numbers in White Monastery codex GI. Note: Not all the numerals in these page numbers are "epact" in form, some being only more or less slightly cursive forms of the normal uncial Coptic alphabetic characters. The horizontal lines above and below the page numbers are merely decorative and are not an essential part of the numbers themselves. For a sample complete page of codex GI, with page number 122 in the upper left-hand corner, see the following page. (Information provided by Stephen Emmel, Westfälische Wilhelms-Universität Münster, July 2010.)

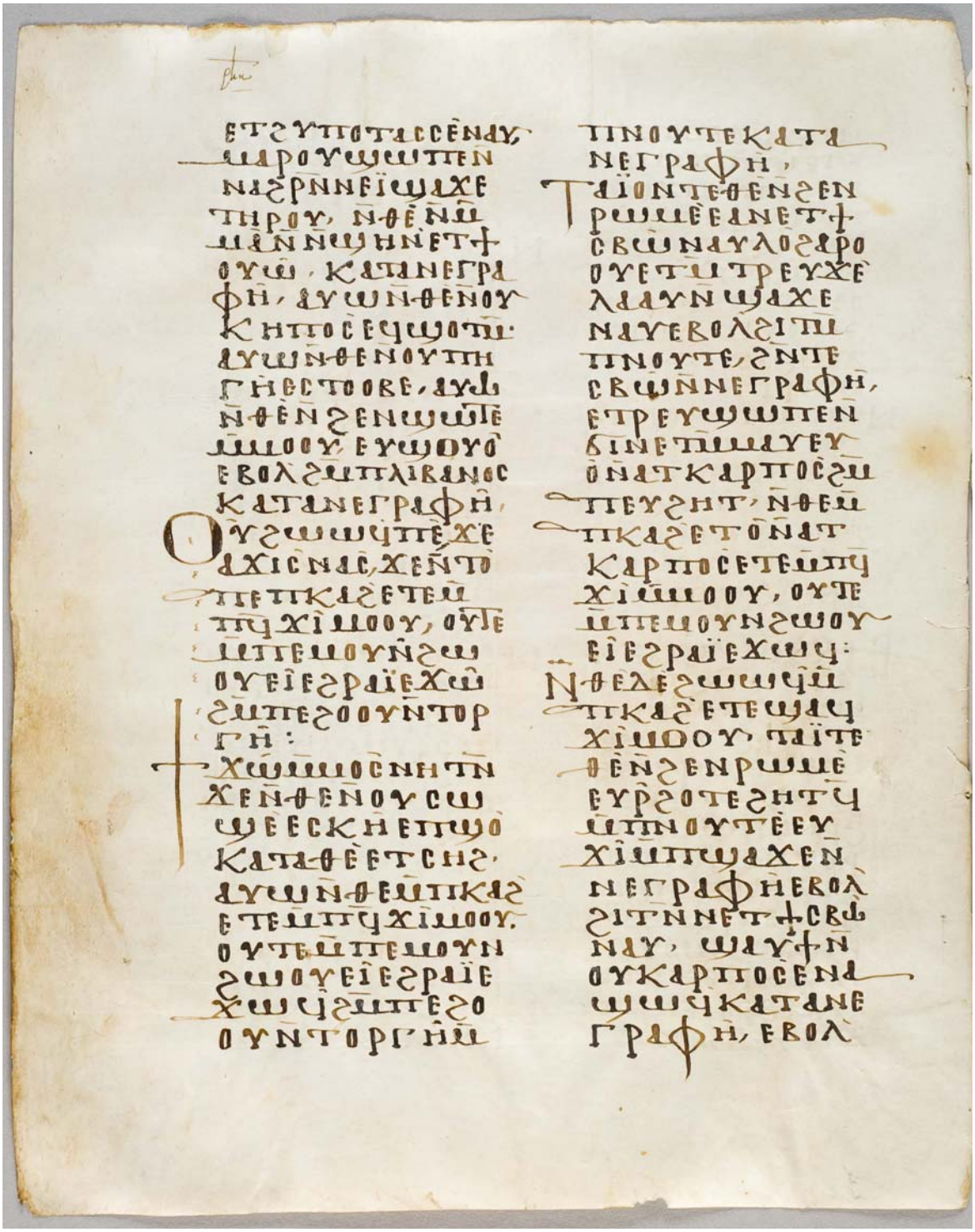

Figure 5. White Monastery codex GI, p. 122 (Coptic PKB)

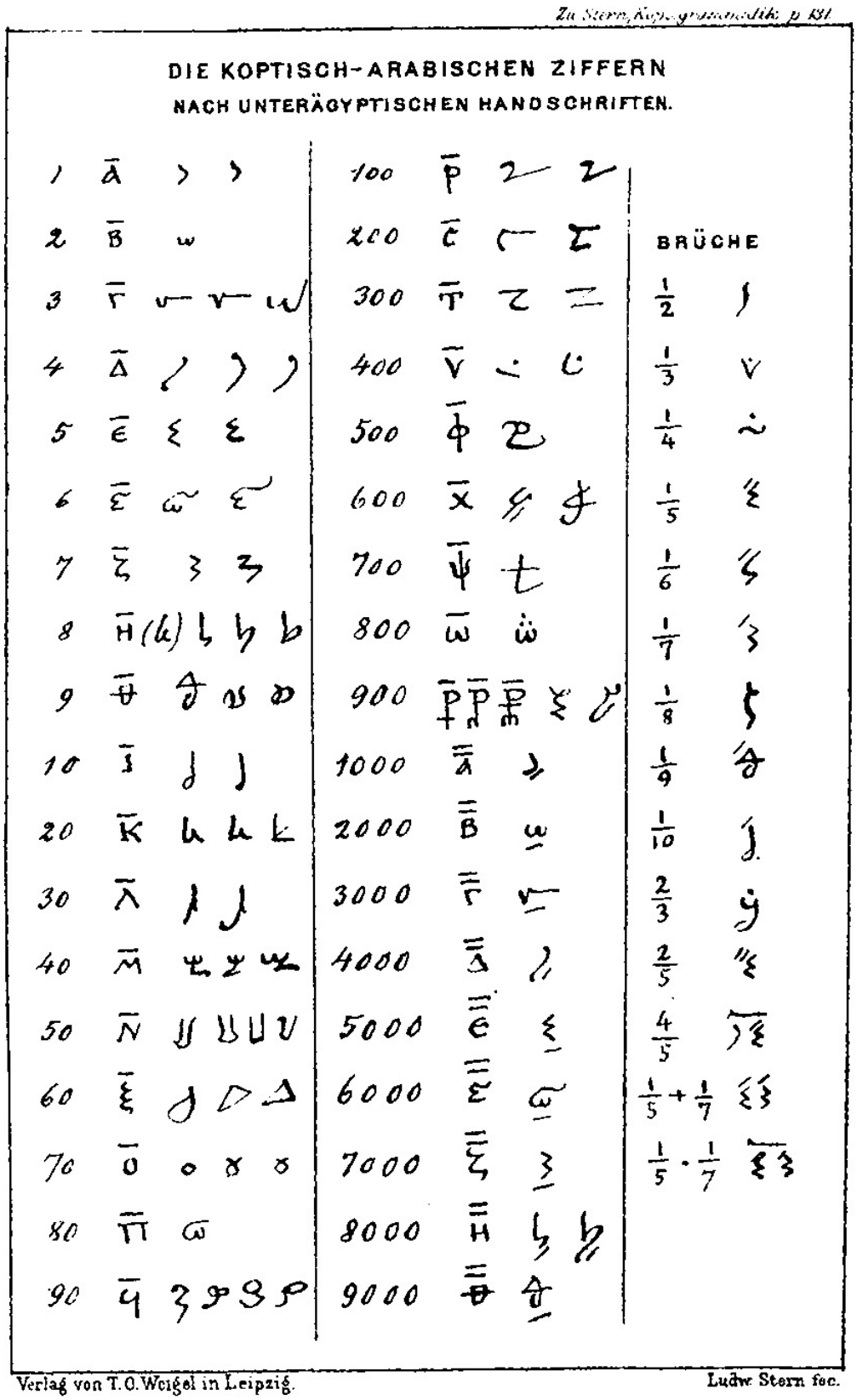

Figure 6. Table from Stern 1880 showing various forms of epact characters. Of particular concern with regard to the completeness of Pandey's proposed encoding are the fractions: (col. 3 "Brüche"). Note that the fraction ½ is already encoded in the main Coptic block at U+2CFD COPTIC FRACTION ONE HALF.