Philology

Philological facilities for the Coptic script

Claudio Beccari and Cristiano Pulone

Abstract

Since 1995 some Coptic fonts by Serge Rosmorduc have been available on CTAN, along with minimal support for using them with LaTeX. We have extended them a little bit and added some support for philological typesetting, including hyphenation patterns and a small collection of macros for the ease of the philologists.

1 Introduction

Thanks to Serge Rosmorduc, since 1995 one Coptic font has been available on CTAN together with the font description file necessary for its use with LaTeX.

Rosmorduc provided the METAFONT description with file `copte.mf`; apparently he obtained the contour descriptions by tracing some fonts very similar to those that appear in a hieroglyphic dictionary [2]; one line of that source METAFONT file says:[1]

```
{limn output Sep 24 17:59:49 1995 from
imageto output Sep 24 16:54:15 1995}
```

but nothing else is said about the source images he operated on. He put his fonts in the public domain with a generic sentence, but an auxiliary file of his bundle contains the whole specification of the Free Software Foundation Licence.

We therefore felt free to add and modify Rosmorduc's files, by changing names and giving him full credit for the original work he had done. His work continues to be at least 80% of the new files. In particular his approach has been almost completely maintained in this sense: his tracing algorithm gave him the Bezier nodes and control points, therefore his METAFONT description is explicit, not a parameterized algorithmic one as we are accustomed to from the Computer Modern font source files; on this point you may see [5] for further information.

On the other side the fonts whose pictures have been traced by Rosmorduc have a very interesting appearance, since they give the impression of being stroked with a quill pen or some other old handwriting instrument of that sort.

Since Coptic fonts appear mostly in Christian liturgical texts we added some symbols that frequently occur in such texts. Nevertheless, philolo-

[1] `limn` and `imageto` are two early programs for tracing font contours; they are part of the GNU Font Utilities.

gists today dedicate most of their attention to para-Christian texts, especially Gnostic ones, as those in the renowned Nag Hammadi Library. Such signs as ⲻ ⲽ ⸗ ⁕ are specific to ritual texts, not necessarily strictly Christian ones; ⳿ and ☥ are of evident Christian origin, but they frequently appear in pagan ritual texts, Gnostic cosmologies, etc. For the philologist's ease we also added other glyphs that are in use in their texts.

We decided also to collect all the signs on the first page of the font table; in other words, we used only the first 128 slots of the font table. We did not care much about the encoding; Rosmorduc himself had in mind a philologist writing critical texts, not a theologian writing a whole text in Coptic, thus requiring numbers, punctuation, extra signs, etc. We are aware that there is an effort among the clergy in the Coptic Church trying to define a common encoding scheme (see [3], for example). On the other hand, the Unicode standard allocates the unique Coptic signs in slots `0x03E2` through `0x03EF`, while apparently the other symbols derived from Greek share the same positions of the Greek letters.

For reasons of compatibility we retained the ligatures defined by Rosmorduc, so that a text originally written to be typeset with the `copte` fonts can be processed with our fonts `copto` (ordinary upright font) and `copti` (inclined font), obtaining the same output except for a possible inclination.

In the end, the ordinary Coptic font turned out as shown in Table 1.

We also provide the font definition files, and, most important of all, we provide Type 1 versions of the fonts. So there should be no difficulty in producing fine documents in PostScript or PDF format.

2 The Coptic font

As we said in the introduction, Rosmorduc's fonts were obtained by tracing the images of Coptic glyphs taken from some source where they had suitable dimensions. The METAFONT specifications are in terms of contour Bezier nodes and control points, although the latter are specified indirectly.

The font design size is declared to be 10pt and everything inside the METAFONT files is specified in terms of a unit u declared to be one tenth of the design size, therefore nominally $u = 1\,\mathrm{pt}$. In practice the heights, widths and depths of all characters were as the tracing algorithm had determined; virtually all of them were distinct. No *x_height*, *quad_width*, etc., were defined, but it turned out that the average *x_height* of the various glyphs was significantly larger than the corresponding dimension of comparable fonts; in particular, it was almost 25% larger than the corresponding height of the CM fonts.

We decided to modify the font metric dimensions and to specify the other missing font dimensions that define the font-dependent TEX units.

We modified several glyphs; the most significant modification was the one concerning the letter "shima" whose upper stroke protrudes far out of the glyph bounding box, so that it would not butt against the ascenders or the next right character, specifically with "lauda", as in the word ϭⲁⲟϭ.

We added some special symbols, such as

ⲻ ⲽ ⳿ ☥ ‐ ⲓⲉ ⲛ̈ ⲓ̈ ⲩ̈ ⳾ ⸗

and redefined the ligature table to cope with new glyphs and their ligatures, while preserving Rosmorduc's ligatures, for compatibility reasons. Probably in this respect more work should be done, but the result appears acceptable.

3 The font keyboard mapping

We use the word *transliteration* instead of encoding, because we had the philologist in mind; that is, a person who is writing critical texts in Coptic, but using a "Latin" keyboard. In fact, we made all our experiments with the Italian keyboard which lacks some important ASCII characters, but does have various others that are missing from a "normal" US keyboard.

For example, we had to define the grave accent. The Italian keyboard lacks the key with the grave accent (or back tick); we used the apostrophe key instead, on the assumption that the acute accent is seldom if ever used in Coptic. On the other hand, we defined a macro `\°` because the `°` sign is on the Italian keyboard and the macro appears less obtrusive when inserted into the source file. We provided the `\0` alias for those who use keyboards without the `°` sign.

The correspondence between the Latin and the Coptic signs is as shown in Table 2. It is based on the correspondence of the "sounds" or of the "shapes" or ... on the availability of a free key!

Notice that in Table 2 we make no attempt to spell out the names of the Coptic letters; there are several naming conventions that depend on the Western language of those who named them. At the web site `http://www.copticchurch.net` there is a short outline regarding Coptic fonts with their names. The point is that the Coptic letters should carry a phonetic value, but those who know how to read and write this language do not agree on their pronunciation, therefore the phonetic transliteration is not unique. Nevertheless, those who master the Coptic language can perfectly understand Table 2.

	’00	’01	’02	’03	’04	’05	’06	’07	’10	’11	’12	’13	’14	’15	’16	’17	
’000	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	"00
’020	16	17	18	19	20	¨ 21	– 22	23	24	25	26	27	28	29	30	31	"10
’040	32	·: 33	:· 34	ψ 35	Ψ 36	37	38	39	Θ 40	Ξ 41	Ξ 42	Φ 43	Φ 44	- 45	· 46	ⳉ 47	"20
’060	ⳉ 48	49	ⳉ 50	ⳇ 51	ⳇ 52	ⳋ 53	ⳉ 54	55	ⲟ 56	ⳅ 57	⸗ 58	ⲭ 59	Ⲭ 60	⸗ 61	ⳍ 62	ⳏ 63	"30
’100	64	Ⲗ 65	Ⲃ 66	Ⲥ 67	Ⲗ 68	Ⲉ 69	Ⲏ 70	Ⲓ 71	Ⲏ 72	Ⲓ 73	Ⲋ 74	Ⲕ 75	Ⲗ 76	Ⲩ 77	Ⲏ 78	Ⲟ 79	"40
’120	Ⲡ 80	Ⲅ 81	Ⲣ 82	83	Ⲧ 84	Ⲩ 85	ⳁ 86	Ⲙ 87	Ⲭ 88	Ⲫ 89	Ⲍ 90	[91	92	] 93	94	95	"50
’140	` 96	ⲁ 97	ⲃ 98	ⲥ 99	ⲍ 100	ⲉ 101	ⲏ 102	ⲓ 103	ⲕ 104	ⲓ 105	ⳇ 106	ⲕ 107	ⲗ 108	ⳍ 109	ⲛ 110	ⲟ 111	"60
’160	ⲛ 112	ⳅ 113	ⲣ 114	115	ⲧ 116	ⲩ 117	ⳉ 118	ⲱ 119	ⲭ 120	ⲫ 121	ⲍ 122	123	124	125	126	127	"70
	"00	"01	"02	"03	"04	"05	"06	"07	"08	"09	"0A	"0B	"0C	"0D	"0E	"0F	

Table 1: The ordinary upright Coptic font

4 The Coptic fonts in Type 1 format

As mentioned previously, we produced our Coptic fonts in Type 1 PostScript format.

We followed the near-standard procedure of tracing large-scale raster fonts made with METAFONT by means of `mftrace` [7] and pipelining the output to `pfaedit` (now `fontforge` [8]).

We did not try to create an encoding vector with any kind of names for the Coptic letters, not even trying to “copy” them from existing encoding vectors. Our fonts are rather non-standard, so their use is confined to (LA)TEX use only.

5 The Coptic font description file

Since we produced two basic shapes for the Coptic font family, we had to produce a new LATEX font definition file, `lcopcoptic.fd`, substantially different from the single-shape one by Rosmorduc.

In addition to the obvious point of declaring two shapes, instead of just one, the main changes are the following:

1. The font encoding was named `LCOP` in accordance with the recommendation of the LATEX3 team that any non-standard encoding should start with the letter `L`, for “local encoding”.
2. The font is loaded with a default magnification of 0.83. This scales the Coptic font to have a closer match between its x-height and the one of the surrounding text in Latin characters.

As with the original font by Rosmorduc, our fonts come in one size, although the METAFONT source files may be invoked on the fly by modern TEX systems so as to generate the raster files at the desired magnification. The heavy strokes of these fonts cope well with shrinking, but we think they become too black when enlarged too much, as seen in titles. In any case, we produced the PostScript Type 1 versions of these fonts so that a single source is used at all magnifications and PostScript or PDF documents can be well typeset, easily readable on screen and perfectly printable on paper.

6 Coptic hyphenation

We produced also a pattern file for the hyphenation of the Coptic language; it was “hand made” since we were not able to find either a word list or a Coptic dictionary specifying hyphenation points.

We worked on and implemented grammar rules [6] based on open and closed syllables, similar to the ancient Greek and Italian rules. While typesetting a master’s thesis on some ancient Coptic texts, we entered the words into a word list without hyphenation points and checked the hyphenations by means of a little LATEX program implementing V. Eijkhout’s `\printhyphens` macro [4].

The present patterns appear to hyphenate correctly all the words on the word list (a few hundred). Some possible hyphenation points may have been missed, but this is not a real inconvenience. (Every time the `patgen` program is used to create new patterns or analyze existing ones, the statistics of the missed hyphens are output. This is useful information when `patgen` is used to create or to modify hyphenation patterns from a hyphenated word list, but is of minor importance when analyzing existing patterns.)

In fact, typographical hyphenation does not necessarily coincide with grammatical hyphenation, even though, of course, it must not violate grammatical rules. Typographical hyphenation fulfills two main purposes: (a) allowing the typeset text to be broken into lines and justified without ugly white gaps, and (b) keeping the reader comfortable in reading broken lines. For the second purpose it may be desirable to refrain from certain hyphen-

Latin	Coptic	Latin	Coptic
a	ⲁ	A	Ⲁ
b	ⲃ	B	Ⲃ
c	ⲥ	C	Ⲥ
d	ⲇ	D	Ⲇ
e	ⲉ	E	Ⲉ
f	ⲫ	F	Ⲫ
g	ⲅ	G	Ⲅ
h	ⲏ	H	Ⲏ
i	ⲓ	I	Ⲓ
j	ⲱ	J	Ⲱ
k	ⲕ	K	Ⲕ
l	ⲗ	L	Ⲗ
m	ⲙ	M	Ⲙ
n	ⲛ	N	Ⲛ
o	ⲟ	O	Ⲟ
p	ⲡ	P	Ⲡ
q	ϭ	Q	Ϭ
r	ⲣ	R	Ⲣ
t	ⲧ	T	Ⲧ
u	ⲩ	U	Ⲩ
v	ⲯ	V	ⲯ
w	ⲱ	W	Ⲱ
x	ⲭ	X	Ⲭ
y	ⲩ	Y	Ⲩ
z	ⲍ	Z	Ⲍ
8	ⲑ	81	Ⲑ
ks	ⲝ	KS, Ks	Ⲝ
ps	ⲯ	PS, Ps	Ⲯ
dj	ⲋ	DJ, Dj	Ⲋ
hj	ⲱ	HJ, Hj	Ⲱ
tj	ⲧ	TJ, Tj	ⲧ
h1	ⲋ	H1	Ⲋ
h2	ⲱ	H2	Ⲱ

Table 2: Correspondence between Coptic and Latin signs or sign sequences on a Latin keyboard

command	example	output
\H	\H	ⲋ
\h	\h	ⲋ
\=	\={me}	ⲙⲉ
\"	\"	¨
\"i	\"i	ï
\"u	\"u	ⲩ̈
\’	\’e	è
\’m	\’m	ⲙ̀
\’n	\’n	ǹ
\°[a]	\°	-
\0	\0	-
\+[b]	\+	≠
\pont	\pont{c}	ⲥ̇
\trepun	\trepun	⁖
\threedots	\threedots	⁖
\trepund	\trepund	⁖
\sic	\sic e.\=nk	sic ⲉ.ⲛ̄ⲕ
\dubious	\dubious{anokpe}	ⲁⲛⲟⲕⲡⲉ
\barretta	\barretta{dj}	ⲋ̶
\Asterisk	\Asterisk	*
\Crux	\Crux	☩
\crucicula	\crucicula	☩
\iesus	\iesus	ⲓ̄ⲥ̄
\djois	\djois	ⲭ̄ⲥ̄
\xcr	\xcr	ⲭ̄ⲥⲣ̄
\xc	\xc	ⲭ̄ⲥ̄

[a] Besides the sign - this command and its alias \0 introduce a discretionary break after the short hyphen.

[b] This double inclined dash mark inserts an unbreakable point within a word, but it does not inhibit hyphenation in the remaining part of the word.

Table 3: New commands with the Coptic fonts

ations; this is certainly so in Italian, and therefore we followed our Italian tradition.

7 Macros for Coptic philologists

We completed the Coptic bundle with a `coptic.sty` file containing some useful macros in order to easily typeset Coptic source `.tex` files for critical texts.

The style file obviously provides a command and an environment, `\textcoptic` and `coptic` respectively, for typesetting the marked text using the Coptic alphabet and hyphenation. There is also a command `\textlatin` for inserting some text in the Latin alphabet. For compatibility reasons the code fragment "`coptic`" may be substituted by "`copte`" or "`copto`", so old documents, the source files of which were composed with the original Rosmorduc files or with our alpha versions of the package, are still usable, with no need to correct the commands and the environments.

It must be noted that the loading order of the packages `fontenc` and `coptic` makes a difference. If the former precedes the latter, `coptic` remembers the correct Latin encoding; if the latter precedes the former, `coptic` remembers the default encoding, presumably `OT1`, before the subsequent `fontenc` package changes the Latin encoding. Therefore a little care should be exercised when loading packages, or the `\textlatin` might yield unexpected results.

The specific Coptic language macros that are introduced with the package are collected in Table 3. Some of them operate on arguments, others are freestanding. The diaeresis accent may be set on every letter, but special commands are defined for

the cases when the letters are ⲓ and ⲩ, to use the special accented glyphs already present in the font. Similarly, the grave accent has special glyphs when the letter is ⲉⲉ or ⲛ.

It is a well-known problem that accent macros interrupt what TEX considers to be a word; in general, they inhibit subsequent hyphenation in the word. By resorting to special characters and the advanced LATEX 2_ε composite symbol command definitions it is possible to address the special symbols directly, thus allowing hyphenation and letting TEX work with the possible ligature and/or kerning properties of the characters involved.

Finally, we note that the `coptic` package is compatible with the `teubner` package, so that some synergy can be exploited between the two. In practice, typesetting critical texts in Coptic almost always implies citing numerous Greek references and text samples, possibly from the same ancient periods, so that all the facilities available with `teubner`, that are not directly connected with the Greek language, can be quite useful.

8 Conclusion

In preparing the modified Coptic fonts, from the original work of Serge Rosmorduc, and all the related files for typesetting critical texts in Coptic, we think we have made a first attempt to extend the present situation; but the actual approval of this work may come only from those scholars and Coptic clergy that use this alphabet and this language.

We did not experiment with `ledmac` [9] simply for lack of time; we suppose that the `coptic` and the `ledmac` packages should be compatible. If there are any, they may in fact be between `teubner` and `ledmac` and these possible bugs will be examined for the next release of `teubner`.

We are very grateful to the TEX users who have been so patient to use our material and submit constructive criticism. We will continue working on the refinement of this bundle in order to make it more useful to the Coptic experts.

As a concluding display, see in Figure 1 a sample text typeset with the Coptic fonts of this article; the versicle marks were obtained by means of the `teubner` package facilities. If upper and/or lower philological marks are present it is advisable to spread out the typeset lines a little bit; in Fig. 1 they were processed with a spread factor of 1.5.

Wien K 8304 (Rainer, AN 201)

ϯⲱⲣⲕⲉ ⲉⲣⲟⲕ ⲉⲉⲡⲟⲟⲩ ⲱ ⲥⲣⲟⲩⲫⲟⲥ ⲡⲁⲩⲩⲉⲗⲟⲥ ⲉⲧⲧⲏⲩ ϯ ⲉⲭⲉⲛ ⲧⲉⲭⲱⲣⲁ ⲛⲩⲏⲉⲉⲉ ⲭⲉⲕⲁ̣ⲥ ⲉⲕⲉⲡⲱⲣⲩ ⲛⲛⲉⲕ- ϯ ⲧⲉⲛⲥ ⲉⲭⲉⲛ ⲉⲉⲁ ⲉⲉⲁ (sic) ⲛⲓⲉⲉ ⲉⲧⲉⲣⲉⲡⲓⲥⲏⲁⲩ ⲛⲃⲁⲣⲱⲧ ϯ ⲛⲥ̣ⲏⲧⲩ ⲩⲁⲛⲧⲉⲩⲕⲧⲟⲩ ⲡⲉⲉⲁ ⲛⲧⲁⲩⲉⲓ ⲉⲃⲟⲗ ⲛⲥ̣ⲏⲧⲩ ⲛⲓ ϯ ⲉⲓⲃ̣̈ⲧ ⲉⲉⲡⲡⲉⲉⲉⲛⲧ ⲡⲉⲉⲉⲥⲓⲧ ⲉⲉⲛⲑⲁⲗⲁⲥⲥⲁ ⲉⲩⲱⲡⲉ ⲉⲩ- ϯ ⲧⲟⲉⲉⲥ̣ ⲥⲁⲛⲕⲁⲥ ⲉⲕⲉⲟⲧⲱⲛⲥⲩ ⲉⲃⲟⲗ ⲉⲩⲱⲡⲉ ⲉⲩⲥⲏⲛ ϯ ⲥⲉⲉⲉⲁ ⲛⲧⲟⲉⲉ ⲉⲕⲉⲕⲧⲟⲩ ⲉⲡⲉⲓⲉⲉⲁ ⲡⲡⲧⲁⲩⲭⲓⲧⲩ ⲉⲉⲡⲉⲣ- ϯ ⲧⲣⲉⲡⲕⲁⲥ̣ ⲧⲁⲭⲣⲟⲩ ⲥⲁⲣⲟⲩ ⲉⲉⲡⲉⲣⲧⲣⲉⲧⲡⲉ ⲉⲣⲥⲁⲓⲃⲥ ⲉⲣⲟⲩ ϯ ⲁⲧⲱ ⲉⲉⲡⲉⲣⲧⲣⲉⲗⲁⲁⲧ ⲛⲉⲉⲉⲧⲟⲛ ⲩⲱⲡⲉ ⲛⲁⲩ ⲡⲧⲏⲣⲩ ϯ ⲁ̈ⲓⲟ ⲁ̈ⲓⲟ ⲧⲁⲭⲏ ⲧⲁⲭⲏ ⲧⲁⲭⲏ

Figure 1: A sample typeset with the `copto` font

[1] Budge, Wallis, *Egyptian Hieroglyphic Dictionary: With an index of Egyptian words, king list and geographical list with indexes, list of hieroglyphic characters, Coptic and Semitic alphabets, etc.*, New York, Dover Publications, 1978.

[2] *The Coptic Standard Character Code (CSCC)*, `http://www.copticchurch.net/coptic_fonts/`

[3] Eijkhout, Victor, "The bag of tricks", *TUGboat* 14(4), 1993, p. 424.

[4] Hall, Timothy, "The METAFONT approach: Implicit, relative, and analytical font design", *TUGboat* 24(2), 2003, pp. 200-205.

[5] Mallon, Alexis, *Grammaire copte*, Beirut 1904, 1926.

[6] Nienhuys, Han-Wen, *mftrace — Scalable fonts for Metafont*, `http://www.xs4all.nl/~hanwen/mftrace/index.html`

[7] Williams, George, *fontforge*, `http://fontforge.sourceforge.net/`. Originally named pfaedit.

[8] Wilson, Peter, *ledmac — A presumptuous attempt to port EDMAC, TABMAC and EDSTANZA to* LATEX, present in any distribution of the TEX system when the ledmac package is installed.

◇ Claudio Beccari
Politecnico di Torino, Turin, Italy
`claudio.beccari@polito.it`

◇ Cristiano Pulone
Università di Bologna, Bologna, Italy
`c_pulone@tin.it`