

The Experimental Unicode ‘EUx’ Encodings

Will Robertson Elie Roux Khaled Hosny

2010/05/27 v0.1h

Abstract

This file documents some experimental Unicode font encodings, used for X_YTeX and LuaTeX. Includes Latin Modern font definitions for version 1.6 and later of those fonts.

Contents

1 Introduction	1	4 The Latin Modern OpenType fonts	3
2 The EU1 encoding	2	4.1 Conversion scripts	3
3 The EU2 encoding	2	5 Latin Modern test file	5

1 Introduction

L^AT_EX’s font encoding system provides a method for ensuring that characters in a document can be rendered properly by glyphs in the output. It is also a sensible system for typesetting multilingual documents while remaining within T_EX’s restrictions of 8-bit fonts.

When X_YTeX came along, no work had been done on using this infrastructure for Unicode fonts. For one thing, the notion of a well-defined font encoding — which neatly pigeonholes a font into a category of containing a specific set of glyphs — doesn’t work so well when designers have an essentially unrestricted selection of glyphs they may add to their fonts.

Both technical advantages of L^AT_EX’s font encodings — error checking, and multilingual typesetting — were no longer relevant with X_YTeX, due to impracticality for the former and redundancy in the latter; X_YTeX’s support of OpenType fonts obviates the need for font switching (for glyph coverage only) by necessity when changing languages, although multiple fonts will often be used for stylistic purposes.

Nonetheless, font encodings are tied inextricably into L^AT_EX’s font selection mechanisms. For the first couple of years of X_YTeX’s life, defined font families were

placed in the U encoding; appropriate, given the Unicode initial, but *inappropriate* within the framework of L^AT_EX itself. The U encoding stands for ‘undefined’, to be used for fonts that contain glyphs that are essentially arbitrary; for example, a Dingbats font.

Since X_YL^AT_EX documents didn’t interact with the ‘regular’ world so much, this overloading of the encoding didn’t really matter in the long run. But eventually, it became apparent that it would be appropriate to use font encodings to represent Unicode subsets and/or different methods of font loading. With the advent of LuaT_EX, this latter use has justified the approach originally taken here.

2 The EU1 encoding

The EU1 encoding represents the first attempt of providing a standard L^AT_EX encoding for Unicode fonts. At present, it loads and defines nothing in terms of declaring text glyphs and accents. That rôle is played for now by Ross Moore’s xunicode package; in the future I hope some of that package will migrate here.

Because we can assume that all current T_EX distributions will contain the Latin Modern OpenType fonts (fonts/opentype/public/lm), these are selected as the default fonts. More details on this later.

I am now defining the convention that EU1 encoded fonts have no necessary subsets of Unicode. Perhaps in the future this will be restricted to fonts that support Unicode accent characters.

Furthermore, no restrictions are made on the method used to load the font. X_YL^AT_EX allows fonts of two types: system installed fonts (also accessible via regular applications or programs); and ‘local’ fonts, which are specified by filename to be found either in a texmf tree of the local directory.

```
1 \DeclareFontEncoding{EU1}{}{}
2 \DeclareErrorFont{EU1}{lmr}{m}{n}{10}
3 \DeclareFontSubstitution{EU1}{lmr}{m}{n}
```

The : character needs to be sanitised in case babel makes it active:

```
4 \g@addto@macro\nfss@catcodes{\@makeother:}
```

3 The EU2 encoding

The EU2 encoding is to be used with LuaT_EX. It is the same as EU1 for now; the actual difference is in the font definition files.

```
5 \DeclareFontEncoding{EU2}{}{}
6 \DeclareErrorFont{EU2}{lmr}{m}{n}{10}
7 \DeclareFontSubstitution{EU2}{lmr}{m}{n}
```

The : character needs to be sanitised in case babel makes it active:

```
8 \g@addto@macro\nfss@catcodes{\@makeother:}
```

4 The Latin Modern OpenType fonts

Version 0.1g of euenc supports version 1.6 of the LM fonts, released in October 2009.

4.1 Conversion scripts

The .fd files are converted from the T1 encoding with the following shell script. Only the package maintainers need to be able to run this script and those following.

4.1.1 Encoding conversion shell script

```
9 sed -f sed-eu1lm.sed 'kpsewhich t1lmr.fd' eu1lmr.fd
10 sed -f sed-eu1lm.sed 'kpsewhich t1lmdh.fd' eu1lmdh.fd
11 sed -f sed-eu1lm.sed 'kpsewhich t1lmss.fd' eu1lmss.fd
12 sed -f sed-eu1lm.sed 'kpsewhich t1lmssq.fd' eu1lmssq.fd
13 sed -f sed-eu1lm.sed 'kpsewhich t1lmvtt.fd' eu1lmvtt.fd
14
15 sed -f sed-eu2lm.sed 'kpsewhich t1lmr.fd' eu2lmr.fd
16 sed -f sed-eu2lm.sed 'kpsewhich t1lmdh.fd' eu2lmdh.fd
17 sed -f sed-eu2lm.sed 'kpsewhich t1lmss.fd' eu2lmss.fd
18 sed -f sed-eu2lm.sed 'kpsewhich t1lmssq.fd' eu2lmssq.fd
19 sed -f sed-eu2lm.sed 'kpsewhich t1lmvtt.fd' eu2lmvtt.fd
20
21 sed -f sed-eu1lm.sed -e \
22 's/:mapping=tex-text//g' 'kpsewhich t1lmtt.fd' eu1lmtt.fd
23 sed -f sed-eu2lm.sed -e \
24 's/;.trep;.tlig;//g' 'kpsewhich t1lmtt.fd' eu2lmtt.fd
```

4.1.2 Encoding conversion sed scripts

The shell script above requires the following sed scripts.

Sed scripts

```
25 /^%.*$/d
26 <eu1sedlm>s/{T1}/{EU1}/g
27 <eu1sedlm>s/t1lm/eu1lm/g
28 <eu2sedlm>s/{T1}/{EU2}/g
29 <eu2sedlm>s/t1lm/eu2lm/g
30
31 s/ec-lmri\([0-9]*\) /lmroman\1-italic/g
32 s/ec-lmro\([0-9]*\) /lmromanslant\1-regular/g
33 s/ec-lmr\([0-9]*\) /lmroman\1-regular/g
34 s/ec-lmcsc\([0-9]*\) /lmromancaps\1-oblique/g
35 s/ec-lmcsc\([0-9]*\) /lmromancaps\1-regular/g
36
```

37 s/ec-lmbxo\([0-9]*\)/lmromanslant\1-bold/g
38 s/ec-lmbxi\([0-9]*\)/lmroman\1-bolditalic/g
39 s/ec-lmbx\([0-9]*\)/lmroman\1-bold/g
40 s/ec-lmbo\([0-9]*\)/lmromandemi\1-oblique/g
41 s/ec-lmb\([0-9]*\)/lmromandemi\1-regular/g
42
43 s/ec-lmdunh\([0-9]*\)/lmromandunh\1-regular/g
44 s/ec-lmduno\([0-9]*\)/lmromandunh\1-oblique/g
45
46 s/ec-lmss\([0-9]\{1,\}\)/lmsans\1-regular/g
47 s/ec-lmss\([0-9]\{1,\}\)/lmsans\1-oblique/g
48
49 s/ec-lmssdc\([0-9]\{1,\}\)/lmsansdemicond\1-regular/g
50 s/ec-lmssdo\([0-9]\{1,\}\)/lmsansdemicond\1-oblique/g
51
52 s/ec-lmssbx\([0-9]\{1,\}\)/lmsans\1-bold/g
53 s/ec-lmssbo\([0-9]\{1,\}\)/lmsans\1-boldoblique/g
54
55 s/ec-lmssq\([0-9]\{1,\}\)/lmsansquot\1-regular/g
56 s/ec-lmssqo\([0-9]\{1,\}\)/lmsansquot\1-oblique/g
57 s/ec-lmssqbx\([0-9]\{1,\}\)/lmsansquot\1-bold/g
58 s/ec-lmssqbo\([0-9]\{1,\}\)/lmsansquot\1-boldoblique/g
59
60 s/ec-lmtt\([0-9]\{1,\}\)/lmmono\1-regular/g
61 s/ec-lmtto\([0-9]\{1,\}\)/lmmonoslant\1-regular/g
62 s/ec-lmtti\([0-9]\{1,\}\)/lmmono\1-italic/g
63
64 s/ec-lmtcsc\([0-9]\{1,\}\)/lmmonocaps\1-regular/g
65 s/ec-lmtcso\([0-9]\{1,\}\)/lmmonocaps\1-oblique/g
66
67 s/ec-lmtl\([0-9]\{1,\}\)/lmmonolt\1-regular/g
68 s/ec-lmtlo\([0-9]\{1,\}\)/lmmonolt\1-oblique/g
69 s/ec-lmtk\([0-9]\{1,\}\)/lmmonolt\1-bold/g
70 s/ec-lmtko\([0-9]\{1,\}\)/lmmonolt\1-boldoblique/g
71
72 s/ec-lmtlc\([0-9]\{1,\}\)/lmmonoltcond\1-regular/g
73 s/ec-lmtlco\([0-9]\{1,\}\)/lmmonoltcond\1-oblique/g
74
75 s/ec-lmvt\([0-9]\{1,\}\)/lmmonoprop\1-regular/g
76 s/ec-lmvto\([0-9]\{1,\}\)/lmmonoprop\1-oblique/g
77
78 s/ec-lmvtl\([0-9]\{1,\}\)/lmmonoprop\1-regular/g
79 s/ec-lmvtlo\([0-9]\{1,\}\)/lmmonoprop\1-oblique/g
80 s/ec-lmvtk\([0-9]\{1,\}\)/lmmonoprop\1-bold/g
81 s/ec-lmvtko\([0-9]\{1,\}\)/lmmonoprop\1-boldoblique/g
82
83 <euIsedlm>s/\(. []*\)\(lm[a-z0-9-]\{1,\}\)/\1"[2]:mapping=tex-text"/g

```
84 <eu2sedlm>s/\(. [ ]*\)\(lm[a-z0-9-]\{1,\}\)/\1file:\2:script=latn;+trep;+tlig;/g
```

5 Latin Modern test file

Adapted and improved from my PracT_EX Journal article on the Latin Modern fonts.

```
85 \documentclass[12pt]{article}
86 \usepackage[margin=2.5cm]{geometry}
87 \newenvironment{vrb}
88     {\begin{tabular}{@{}p{7cm}l@{}}}
89     {\end{tabular}}
90
91 \usepackage{fancyvrb}
92 \DefineShortVerb{\|}
93
94 \usepackage{slantsc}
95 \usepackage{ifluatex}
96 \usepackage{ifxetex}
97 \ifxetex
98   \def\EU{EU1}
99 \else
100  \ifluatex
101    \usepackage{luaotfload}
102    \def\EU{EU2}
103  \else
104    \def\EU{T1}
105  \fi
106 \fi
107 \usepackage[\EU]{fontenc}
108 \usepackage{lmodern}
109 \begin{document}
110
111 \title{Testing the Latin Modern \texttt{\EU} encoding}
112 \author{Will Robertson}
113 \maketitle
114 \thispagestyle{empty}
115 \pagestyle{empty}
116
117 \section*{Roman}
118
119 \fontfamily{lmr}\selectfont
120 \begin{vrb}
121 |\rmdefault| & {\selectfont Latin Modern Roman} \\
122 | \itshape| & {\itshape Latin Modern Roman Italic} \\
123 | \slshape| & {\slshape Latin Modern Roman Oblique} \\
124 | \scshape| & {\scshape Latin Modern Roman Small Caps} \\
```

```

125 | \slshape| & {\scshape\slshape Latin Modern Roman Oblique Small Caps} \\
126 | \bfseries| & {\bfseries Latin Modern Roman Bold Extended} \\
127 | \itshape| & {\bfseries\itshape Latin Modern Roman Bold Italic Extended} \\
128 | \slshape| & {\bfseries\slshape Latin Modern Roman Bold Oblique Extended} \\
129 | \fontseries{b}| & {\fontseries{b}\selectfont Latin Modern Roman Bold} \\
130 | \fontseries{b}\slshape| & {\fontseries{b}\slshape Latin Modern Roman Bold Oblique} \\
131 \end{vrb}
132
133 \section*{Dunhill}
134
135 \fontfamily{lmdh}\selectfont
136 \begin{vrb}
137 |\fontfamily{lmdh}\selectfont| & {Latin Modern Dunhill} \\
138 | \slshape| & {\slshape Latin Modern Dunhill Oblique} \\
139 \end{vrb}
140
141 \section*{Sans}
142
143 \fontfamily{lmss}\selectfont
144 \begin{vrb}
145 |\sffamily| & {\selectfont Latin Modern Sans} \\
146 | \slshape| & {\slshape Latin Modern Sans Oblique} \\
147 | \bfseries| & {\bfseries Latin Modern Sans Bold} \\
148 | \slshape| & {\bfseries\slshape Latin Modern Sans Bold Oblique} \\
149 | \fontseries{sbc}\selectfont| & {\fontseries{sbc}\selectfont Latin Modern Sans Demi Cond} \\
150 | \slshape| & {\fontseries{sbc}\selectfont\slshape Latin Modern Sans Demi Condensed Ob} \\
151 \end{vrb}
152
153 \section*{Sans Quotation}
154
155 \fontfamily{lmssq}\selectfont
156 \begin{vrb}
157 |\renewcommand\sfddefault{lmssq}| \\
158 |\sffamily| & {\selectfont Latin Modern Sans Extended} \\
159 | \slshape| & {\slshape Latin Modern Sans Extended Oblique} \\
160 | \bfseries| & {\bfseries Latin Modern Sans Bold Extended} \\
161 | \slshape| & {\bfseries\slshape Latin Modern Sans Bold Extended Oblique} \\
162 \end{vrb}
163
164 \section*{Typewriter}
165
166 \fontfamily{lmtt}\selectfont
167 \begin{vrb}
168 |\ttfamily| & {\selectfont Latin Modern Typewriter} \\
169 | \itshape| & {\itshape Latin Modern Typewriter Italic} \\
170 | \slshape| & {\slshape Latin Modern Typewriter Oblique} \\
171 | \scshape| & {\scshape Latin Modern Typewriter Small Caps}

```

```

172 | \scshape\slshape | & {\scshape\slshape Latin Modern Typewriter Oblique Small Caps} \\
173 | \bfseries      | & {\bfseries Latin Modern Typewriter Dark} \\
174 | \slshape      | & {\bfseries\slshape Latin Modern Typewriter Dark Oblique} \\
175 | \fontseries{b}\selectfont | & {\fontfamily{lmtt}\fontseries{b}\selectfont Latin Modern
176 | \fontseries{b}\slshape   | & {\fontfamily{lmtt}\fontseries{b}\slshape Latin Modern Typ
177 | \fontseries{l}\selectfont | & {\fontfamily{lmtt}\fontseries{l}\selectfont Latin Modern
178 | \fontseries{l}\slshape   | & {\fontfamily{lmtt}\fontseries{l}\slshape Latin Modern Typ
179 | \fontseries{lc}\selectfont | & {\fontfamily{lmtt}\fontseries{lc}\selectfont Latin Moder
180 | \fontseries{lc}\slshape   | & {\fontfamily{lmtt}\fontseries{lc}\slshape Latin Modern T
181 \end{vrb}
182
183 \section*{Variable-width Typewriter}
184
185 \fontfamily{lmvtt}\selectfont
186 \def\fixedspacing{\fontdimen3\font0pt\fontdimen4\font0pt}
187 \begin{vrb}
188 |\renewcommand\ttdefault{lmvtt}| \\
189 |\ttfamily          | & {\selectfont \fixedspacing Latin Modern Typewriter Propo
190 | \slshape           | & {\slshape \fixedspacing Latin Modern Typewriter Proporti
191 | \fontseries{l}\selectfont | & {\fontseries{l}\selectfont \fixedspacing Latin Modern Typ
192 | \fontseries{l}\slshape   | & \mbox{\fontseries{l}\slshape \fixedspacing Latin Modern T
193 | \fontseries{b}\selectfont | & {\fontseries{b}\selectfont \fixedspacing Latin Modern Typ
194 | \fontseries{b}\slshape   | & \mbox{\fontseries{b}\slshape \fixedspacing Latin Modern T
195 \end{vrb}
196
197 \end{document}

```