The release 1.2 of the dc-fonts: Improvements to the european letters and first release of text companion symbols

Jörg Knappen
Barbarossaring 43
D-55118 Mainz
R. F. A.
knappen@vkpmzd.kph.uni-mainz.de

Abstract

I describe the improvements made to the dc fonts in release 1.2, and the text companion font added. The ec fonts will finally replace the present 128-character cm fonts as the default fonts of LaTeX.

1 Introduction

In 1990 at the TUG meeting at Cork, Ireland, the european TeX user groups agreed on a 256 character incoding supporting many european languages with latin writing. This encoding is both an *internal encoding* for TeX and a *font encoding*. This double nature is a consequence of the fact, that both kind of encodings cannot be entirely separated within TeX.

The design goals of the Cork encoding are to allow as many languages as possible to be hyphenated correctly and to guarantee correct kerning for those languages. Therefore it includes many ready-made accented letters.

It also includes some innovative features, which have not become very popular yet, though they deserve to become so. First to mention is a special, zero width invisible character, the compound word mark (cwm). The second is the separation of the two characters <hyphen> and <hyphenchar>. By appropriate design of the hyphenchar glyph, hanging hyphenation can be achieved.

The final version of the Cork encoded fonts will be called ec (European Computer Modern or Extended Computer Modern) fonts. The current version, called dc fonts, is an intermediate step towards the final version. Note, that in the cause of bug fixes and improvements, the metrics may change.

The need for a text companion font was first articulated in the discussion of new 256 character mathematical fonts in 1993. In order to achieve a better orthogonality between text and math, some text symbols stored in the math fonts should be moved to the text companion fonts[1]. The text companion fonts are also the ideal place to store some new characters, like currency symbols.

[1] The archives of the math-font-discuss mailing list are available for ftp on ftp.cogs.susx.ac.uk in directory pub/tex/mathfont.

2 Supported languages

The following languages are supported by the Cork encoding: Afrikaans, Albanian, Breton, Croatian, Czech, Danish, Dutch, English, Estonian, Faroese, Finnish, French, Frisian, Gaelic, Galician, German, Greenlandic, Hungarian, Icelandic, Irish (modern orthography), Italian, Letzeburgish, Lusatian (Sorbian), Norwegian, Polish, Portuguese, Rhaetian (Rumantsch), Rumanian, Slovak, Slovene, Spanish, Swedish, Turkish. Many non-european languages using the standard latin alphabet (e. g. Bahasa Indonesia, Suaheli) are also supported.

In europe, the following languages aren't supported: Azeri, Basque, Catalan, Esperanto, Irish (old orthography), Latvian, Lithuanian, Maltese, Sami, Welsh. Of course, Greek and all languages with cyrillic writing are outside the scope of the Cork encoding.

3 Improvements to the dc fonts

3.1 Accents

In good typography, the accent marks should look different for capitals and lowercase letters respectively. The accent over a capital should be of a 'flat' design, while the accent on a lowercase letter should be 'steep'. The Computer Modern fonts by D. E. Knuth only have steep accents, well made for lowercase letters.

óx́ Óx́

Fig. 1: Letters with acute accent in the cmr font

There are no readily accented letters provided which leads to problems with proper hyphenation and kerning. However, the floating accent approach guarantees the consistency of all accented letters.

With the (now out-of date) version 1.1 of the dc fonts, the situation is different. We have readily designed accented letters for all languages included in the ISO standards 8859-1 and 8859-2. If an 'exotic' accented letter is needed, it does not fit to the provided ones.

óx́ Óx́

Fig. 2: Letters with acute accent in the dcr font (v 1.1)

Note that the floating acute accent is the same for capitals and lowers, but different from both, being even steeper than the lowercase one.

With the version 1.2 of the dc fonts, all inconsistencies have gone. The accents are different between capitals and lowers as they should be, and floating accents can be applied in a consistent manner.

Fig. 3: Letters with acute accent in the dcr font (v 1.2)

Since the Cork encoding does only provide one slot for each accent, the capital acute accent is taken from the *text companion* font tcr. This is possible, because TEX allows cross-font accenting.

The acute accent and the readily accented letters were taken with kind permission of the authors from the polish pl fonts, which provide the highest available quality for these shapes. The hungarian double acute accent and the grave accent follow the design of the acute accent.

3.2 Quotation marks

The design of quotation marks provides another challenge to the ec fonts. In the Computer Modern fonts, they are optimised to english usage.

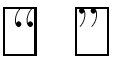

Fig. 4: Quotation marks in the cmr font

They lie asymmetrically in their boxen, which makes a wider space before and after a quotation. However, this kind of design produces a desaster, if the same english opening quotation mark is used as a german or polish closing quotation mark. Currently, macros have to compensate for this.

In the dc fonts 1.2, the quotation marks are lying symmetrically in a tighter box, the additional space is created by kerning against the `boundarychar`.

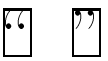

Fig. 5: Quotation marks in the dcr font (1.2)

The boundarychar feature was introduced with TEX3 and METAFONT2, it is reasonable to assume that nowadays every TEX user has access to these or later versions[2].

3.3 Miscellaneous

The shapes for polish letters are now taken from the polish pl font, leading to improved shapes on the ogoneked letters and the crossed l.

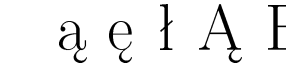

Fig. 6: Polish special letters in the dcr font (1.2)

With the help of the czechoslovak TEX users' goup, the shapes of czech and slovak special letter have been improved, too.

[2] Maybe it was not a reasonable assumption in 1990, when the Cork encoding was born and the above mentioned versions were brand new

Fig. 7: Some czech and slovak special letters in the dcr font (1.2)

The height of umlaut dots has been adjusted to the value contributed by the czechoslovak group (ä occurs in slovak), the value used in the version 1.1 of the dc fonts was considered too low even by german users.

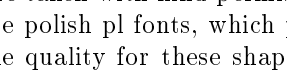

Fig. 8: The letter ä in cmr, dcr v1.1, and dcr v1.2

The `hyphenchar` is now designed to hang out of its bounding box, thus allowing for hanging hyphenation.

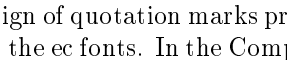

Fig. 9: Hyphen and hyphenchar with their bounding box

The release 1.2 also contains a new shape, a classical serif italic font. It was already prepared in version 1.1, but no parameter and driver files were present for it. It is an italic with upper serifs instead of ingoing hooks. This paragraph is typeset the dcci font to show its appearance.

4 The tc fonts

4.1 A text symbol encoding

Over the years, many reasons have accumulated for a new text symbol encoding. There are some text symbols currently stashed in the math fonts, the footnote marks, and the bullet (•) are among them.

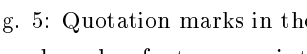

Fig. 10: Footnote symbols from cm math fonts

In 256-character math fonts they should not be preserverved, but moved to a text symbol encoding.

Fig. 11: Footnote symbols from the tcr font

I have added serifs to the paragraph sign (¶) in the serif typefaces, and I have added another one having only one vertical stroke. The design of the section sign (§) was improved significantly, as can be seen from the boldface glyphs.

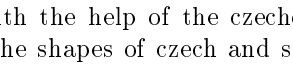

Fig. 12: Old and new design of the section mark

The ISO standards 8859-1 (Latin 1), 8859-2 (Latin 2), and 6937 contain several custom signs. It will be easier to typeset text encoded according to those standards if the neccessary symbols were easily accesible through a text symbol font.

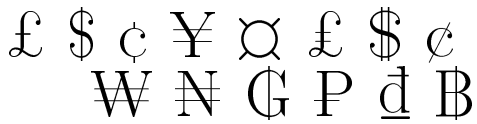

Fig. 13: Currency signs from the tcr font

Finally, I wanted to have different style accents for capitals and lowercase leters. Since the Cork encoding does not have the space for another fourteen accent glyphs, I decided to have the lowercase accents which are far more often needed in the dc fonts, and to put the capital accents into the text companion fonts.

The users of commercial font also want to access all glyphs stored in those fonts. Since most of those glyphs are textual, they all should be included into a text symbol font encoding.

4.2 The font encodings TS1 and TSA

For mainly technical reasons, I think the candidates for a text symbol encoding should be distributed over two fonts, their encoding named TS1 and TSA respectively. There are important differences between the technologie supported by METAFONT and TEX compared to the path most commercial font suppliers choose.

The computer modern family of fonts supports the notion of a designsize, i. e. there are subtle differences between the shapes at different point sizes as illustrated in the next section. TEX is able to raise and lower letters, thus it does not need a readily raised digit to produce a superscript. It can also produce nice fraction using a macro from the TEXbook, exercise 11.6, like $^1/_2$, $^1/_4$, and $^1/_6$.

Most commercial vendors took the easier path, their fonts come in only one size and are scaled up and down to the other sizes. Thus, a small superscript does not look right, and to compensate this a readily designed superscript is added to the fonts. A subscript, too, because earlier text processors weren't able to raise or lower letters. For similar reasons, fraction glyphs were provided, or fraction were contructed out of a sequence <superscript digit> <fraction> <subscript digit>, where fraction is a special slash to construct fractions.

On the other hand, it is almost impossible to follow this path with TEX and METAFONT: The size of the superscripts can be influenced by TEX macros, and therefore there is no unique 'virtual designsize' for ready-made superscripts.

The selection of superscripts offered by commercial vendors is at the moment rather sparse, many often needed ones are lacking.

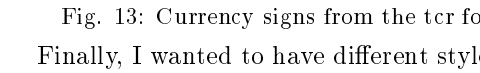

Fig. 14: Some superscript letter lacking in expert fonts

Therefore the rule of thumb for the distribution of glyphs is the following: Put all glyphs wich can be conveniently made with METAFONT and are needed with TEX into the encoding TS1, and put the remaining glyphs, mainly superscripts and subscripts, into the encoding TSA. There are some duplications and derivations from this rule of thumb, e. g. superscript 1, 2, and 3 are part of ISO 8859-1, thus they occur in TS1 as well as in TSA.

5 Standard Control Sequences

The following standard control sequences are assigned with LATEX's T1 encoding for the dc fonts:
\r Ring accent (\r u gives ů)
\k Ogonek (\k e gives ę)
\dh, \DH Icelandic letter edh (ð, Ð)
\dj, \DJ Letter d with stroke (đ, Đ)
\ng, \NG Letter eng (ŋ, Ŋ)
\th, \TH Icelandic letter thorn (þ, Þ).

The package textcomp by Sebastian Rahtz assigns standard names to all text companion symbols. The documentation prints a nice table.

6 Ligatures

In the proportional fonts, the following ligatures are implemented:

--	– (en dash)
---	— (em dash)
‘‘	“ (english opening quotes, german closing quotes)
’’	” (english and polish closing quotes)
,,	„ (german and polish opening quotes)
<<	« (french opening quotes)
>>	» (french closing quotes)
!‘	¡ (spanish opening exclamation mark)
?‘	¿ (spanish opening question mark)
fi	fi
ff	ff
fl	fl
ffi	ffi
ffl	ffl

There is another important ligature, not shown above:
<hyphen><hyphenchar> gives <hyphenchar>.
This allows the implementation of new hyphenation patterns with hyphenchar *´127* allowing hyphenation of words containing explicits hyphens. This ligature

was missing by accident in the september 1995 release of dc 1.2, but has been added in patchlevel 1 released in december 1995.

7 New names of the font files

Currently, the extended computer modern font have the prefix dc. This prefix will changew to ec with the final release after another round of bug fixing. I hope to make the transition from dc

to ec in about one year. The text companion fonts have the prefix tc, which is not subject to change. However, later releases may included more characters and therefore have different checksums. No characters shall be removed from the tc fonts.

Most of the dc fonts can be generated at any size one want in the range from 5pt to 100pt. For each size, a unique name is needed.

With the release 1.2 of the dc fonts, a new, more precise naming scheme is in effect. Since there are widely used operating sytems limiting the file name to 8 character (plus an extension of 3 characters) the following scheme is used:

- The first two letters (either dc or tc denote the encoding and the general design of the font.
- The one or two following letters denotes the family, shape, and series attributes of the font. E. g. r for roman, bx for bold extended, it for italic, ot bi for bold extended italic. A complete overview is given at the end of this section.
- The following four digits give the design size in TEX's points multiplied with 100. E. g 1000 denotes tex point, 1440 denotes magstep 2, i. e. 14.4 point, and 0500 denotes five point.

Here are the implemented styles:

Roman family: r roman, b bold, bx bold extended, sl slanted, bl bold extended slanted, cc caps and small caps, ti (text) italic, bi bold extended italic, u unslanted italic, ci classical serif italic (new design).

Sans serif family: ss sans serif, si sans serif inclined (slanted), sx sans serif bold extended, so sans serif bold extended oblique (slanted).

Typewriter family: tt typewriter, tc typewriter caps and small caps, st slanted typewriter, it italic typewriter, vt variable width typewriter.

Various other fonts: bm variant bold roman, dh dunhill, fb Fibonacci parameters, ff funny, fi funny italic.

Here are some examples:

dcr1000	European computer modern roman at 10pt
tcr1000	Text companion symbols roman at 10pt
dcss1728	European computer modern sans serif at 17.28pt
dcbx0900	European computer modern roman bold extended at 9pt

Some remaining fonts come at one size only, those are

dcssdc10	sans serif demi-bold condensed
dcsq8	sans serif quotation
dcqi8	sans serif quotation inclined
dclq8	latex sans serif quotation
dcli8	latex sans serif quotation inclined
idclq8	invisible latex sans serif quotation
idcli8	invisible latex sans serif quotation inclined.

The last four fonts are for the slides document class, which replaces old SLiTEX. They contain a special version of the capital letter 'I'.

8 Upgrading to ec

There is the following non-official time schedule for the upgrade from the dc to ec fonts: One intermediate release (1.3) shall come out in spring 1996, the final release of the ec fonts shall be made in autumn 1996. Afterwards, the fonts will be frozen and only necessary bug fixes will be applied.

A The Cork Encoding

position (octal)	description
Accents for lowercase letters	
000	grave
001	acute
002	circumflex
003	tilde
004	umlaut
005	hungarian
006	ring
007	hachek
010	breve
011	macron
012	dot above
013	cedilla
014	ogonek
Miscellaneous	
015	single base quote

016 single opening guillemet
017 single closing guillemet
020 english opening quotes
021 english closing quotes
022 base quotes
023 opening guillemets
024 closing guillemets
025 en dash
026 em dash
027 compound word mark (invisible)
030 perthousandzero
031 dotless i
032 dotless j
033 ligature ff
034 ligature fi
035 ligature fl
036 ligature ffi
037 ligature ffl
040 visible space

ASCII

041 exclamation mark
042 straight quotes
043 hash mark
044 dollar sign
045 percent sign
046 ampersand
047 apostroph
050 opening parentheses
051 closing parentheses
052 asterisk
053 plus sign
054 comma
055 hyphen (note: not minus sign)
056 full stop
057 solidus
060 digit 0
...
071 digit 9
072 colon
073 semicolon
074 less than sign
075 equals sign
076 greater than sign
077 question mark
080 commercial at
081 capital letter A
...
132 capital letter Z
133 opening square bracket
134 backslash
135 closing square bracket
136 ASCII circumflex
137 underscore
140 opening quote (not ASCII grave!)
141 lowercase letter a
...
172 lowercase letter z
173 opening curly brace
174 vertical bar
175 closing curly brace
176 ASCII tilde
177 hyphenchar (hanging)

Letters for eastern european languages (from latin-2)

200 capital letter A with breve
201 capital letter A eith ogonek
202 capital letter C with acute
203 capital letter C with hachek
204 capital letter D with hachek
205 capital letter E with hachek
206 capital letter E with ogonek
207 capital letter G with breve
210 capital letter L with acute
211 capital letter L with hachek
212 capital letter crossed L
213 capital letter N with acute
214 capital letter N with hachek
215 capital letter Eng
216 capital letter O with hungarian double acute
217 capital letter R with acute
220 capital letter R with hachek
221 capital letter S with acute
222 capital letter S with hachek
223 capital letter S with cedilla
224 capital letter T with hachek
225 capital letter T with cedilla
226 capital letter U with hungarian double acute
227 capital letter U with ring
230 capital letter Y with diaeresis
231 capital letter Z with acute
232 capital letter Z with hachek
233 capital letter Z with dot
234 capital letter IJ
235 capital letter I with dot
236 lowercase letter d with bar
237 section sign
240 lowercase letter a with breve
241 lowercase letter a with ogonek
242 lowercase letter c with acute
243 lowercase letter c with hachek
244 lowercase letter d with hachek
245 lowercase letter e with hachek
246 lowercase letter e with ogonek
247 lowercase letter g with breve
250 lowercase letter l with acute

251	lowercase letter l with hachek
252	lowercase letter crossed l
253	lowercase letter n with acute
254	lowercase letter n with hachek
255	lowercase letter eng
256	lowercase letter o with hungarian double acute
257	lowercase letter r with acute
260	lowercase letter r with hachek
261	lowercase letter s with acute
262	lowercase letter s with hachek
263	lowercase letter s with cedilla
264	lowercase letter t with hachek
265	lowercase letter t with cedilla
266	lowercase letter u with hungarain double acute
267	lowercase letter u with ring
270	lowercase letter y with diaeresis
271	lowercase letter z with acute
272	lowercase letter z with hachek
273	lowercase letter z with dot
274	lowercase letter ij
275	spanish inverted exclamation mark
276	spanish inverted question mark
277	pound sign

Letters for western european languages (from latin-1)

300	capital letter A with grave
301	capital letter A with acute
302	capital letter A with circumflex
303	capital letter A with tilde
304	capital letter A with diaeresis
305	capital letter A with ring
306	capital letter AE
307	capital letter C with cedilla
310	capital letter E with grave
311	capital letter E with acute
312	capital letter E with circumflex
313	capital letter E with diaeresis
314	capital letter I with grave
315	capital letter I with acute
316	capital letter I with circumflex
317	capital letter I with diaeresis
320	capital letter Edh (D with bar)
321	capital letter N with tilde
322	capital letter O with grave
323	capital letter O with acute
324	capital letter O with circumflex
325	capital letter O with tilde
326	capital letter O with diaeresis
327	capital letter OE
330	capital letter O with slash
331	capital letter U with grave
332	capital letter U with acute
333	capital letter U with circumflex
334	capital letter U with diaeresis
335	capital letter Y with acute
336	capital letter Thorn
337	capital letter Sharp S (deviating from latin-1)
340	lowercase letter a with grave
341	lowercase letter a with acute
342	lowercase letter a with circumflex
343	lowercase letter a with tilde
344	lowercase letter a with diaeresis
345	lowercase letter a with ring
346	lowercase letter ae
347	lowercase letter c with cedilla
350	lowercase letter e with grave
351	lowercase letter e with acute
352	lowercase letter e with circumflex
353	lowercase letter e with diaeresis
354	lowercase letter i with grave
355	lowercase letter i with acute
356	lowercase letter i with circumflex
357	lowercase letter i with diaeresis
360	lowercase letter edh
361	lowercase letter n with tilde
362	lowercase letter o with grave
363	lowercase letter o with acute
364	lowercase letter o with circumflex
365	lowercase letter o with tilde
366	lowercase letter o with diaeresis
367	lowercase letter oe
370	lowercase letter o with slash
371	lowercase letter u with grave
372	lowercase letter u with acute
373	lowercase letter u with circumflex
374	lowercase letter u with diaeresis
375	lowercase letter y with acute
376	lowercase letter thorn
377	lowercase letter sharp s (deviating from latin-1)

B The Text Companion Encoding

position description
(octal)

Accents for capital letters

000	grave
001	acute
002	circumflex
003	tilde
004	umlaut
005	hungarian
006	ring
007	hachek
010	breve
011	macron
012	dot above

preliminary draft, December 9, 1995 2:35

013	cedilla
014	ogonek

Miscellaneous

015	base single straight quote
022	base double straight quotes
025	twelve u dash
026	three quarters emdash
030	left pointing arrow
031	right pointing arrow
032	tie accent (lowercase)
033	tie accent (capital)
040	blank symbol
044	dollar sign
047	straight quote
052	centered star
057	fraction

Oldstyle digits

060	oldstyle digit 0
061	oldstyle digit 1
062	oldstyle digit 2
063	oldstyle digit 3
064	oldstyle digit 4
065	oldstyle digit 5
066	oldstyle digit 6
067	oldstyle digit 7
070	oldstyle digit 8
071	oldstyle digit 9

Miscellaneous

115	mho sign
117	big circle
127	ohm sign
136	arrow up
137	arrow down
140	backtick (ASCII grave)
142	born
144	died
154	leaf
155	married
156	musical note
176	low tilde
177	short equals

TS1-symbols

200	ASCII-style breve
201	ASCII-style hachek
202	double tick (ASCII double acute)
203	double backtick
204	dagger
205	ddager
206	double vert
207	perthousand
210	bullet
211	centigrade
212	dollaroldstyle
213	centoldstyle
214	florin
215	colon
216	won
217	naira
220	guarani
221	peso
222	lira
223	recipe
224	interrobang
225	gnaborretni
226	dong sign
227	trademark

Symbols from ISO-8859-1 (latin-1)

242	cent
243	sterling
244	currency sign
245	yen
246	broken vertical bar
247	section sign
250	high dieresis
251	copyright
252	feminine ordinal indicator
254	logical not
256	circled R
257	macron
260	degree sign
261	plus-minus sign
262	superscript 2
263	superscript 3
264	tick (ASCII-style acute)
265	micro sign
266	pilcrow sign
267	centered dot
271	superscript 1
272	masculine ordinal indicator
274	fraction one quarter
275	fraction one half
276	fraction three quarters
326	multiplication sign (times)
366	division sign