The text companion encoding*
(draft)

Jörg Knappen
Institut für Kernphysik
D-55099 Mainz

31–Mar–1995

This document lists the glyphs to be included in the first incarnation of the text companion font encoding. There are many free slots reserved for further standardisation.

About the text companion fonts

The text companion fonts (TS encoding) shall complete the T1-encoded text fonts with custom glyphs which should agree in their style with the text typeface. It should also remove the necessity of 'hidden math' in text mode (e.g. for the footnote marks). There will be two sets of fonts, their encodings are termed 'TS1' and 'TSA' informally.

The T1 encoding covers the *letters* from the ISO standard latin alphabets latin-1 and latin-2, but some symbols are left unsupported. To allow 'latin-1-verbatim' and 'latin-2-verbatim', those symbols are provided by the TS1 encoding.

The TS1 encoded fonts will be named `tc*`, I have checked that there is currently no conflict with other font names. The tc fonts will be implemented in METAFONT and freely available. They will be released together with dc-fonts version 1.2 at the beginning of september 1995. At about the same time, virtual fonts created with fontinst shall come out.

Another goals is to support the full Adobe standard and expert sets of glyphs. For this purpose a second set of fonts (TSA) will be created as virtual fonts. The computer-modern style version could be called '`td...`'

Reasons for introducing TS

With LaTeX 2_ε, text and math fonts are separated from each other. However, there are several places, where math creeps into plain text (bullets in the itemize environment, footnotemarks, some custom signs). To remove this, the necessary symbols have to be taken out of the math fonts into the text companion font.

In addition, there are currently some text glyphs (such as ¥ or ®) sitting in the math fonts. Some of these (for example ¥) are sufficiently rare in mathematics as to warrant removing from the math fonts and placing in a text symbol encoding. They will still be accessable in math mode, either by loading a TS-encoded font as a math family, or by clever tricks with `\mathchoice`.

Three glyphs which occurred in those cm fonts, where `ligs = 0`, are missing in the Cork (T1) encoding. They are single quote (‚), arrow up and arrow down (‸, ˬ).

A good typeface should have different style accents for capitals and lowercase letters, the accents on the lowers being more 'steep' and the ones on capitals more 'flat'. At this time, the cm fonts only have 'steep' accents, where as the dc fonts version 1.1 only have 'flat' ones. Since TeX can do cross-font accenting, I

*I acknowledge intense e-mail exchange with Alan Jeffrey, without which this document had quite another shape.

propose to move the capital accents to the companion font.

There are a number of symbols in the Adobe standard and Expert encoding (such as florin or paragraph) that aren't in the Cork encoding. These can be placed in the TS encoding.

The following shall be in the first TS font (TS1)

The first TS font shall complete the Cork encoding with obviously missing glyphs. They are new 'primitives' needed to have their own METAFONT description.

All accents in the style appropriate for capitals (13 glyphs).

Single quote (‚), arrow up and arrow down (‸, ˬ) (3 glyphs)

The two troublesome glyphs $ and £ (2 glyphs).

The ¥ glyph currently in the ams fonts (1 glyph).

More currency signs. For the first version the following are considered: Colon, cents, dollaroldstyle, centoldstyle (4 glyphs)

The florin, though rejected by the dutch authority Karel Treebus, will be included for those who really want to use it. In the METAFONT implementation, probably just the appropriate letter 'f' will be inserted. (1 glyph)

The footnote symbols (∗, §, ¶ , †, ‡, ‖) (6 glyphs).

A ready-made per thousand sign (1 glyph).

The 'sputnik' or 'international currency sign' (1 glyph)

The bullet for the itemize environment (1 glyph)

Copyright and circled R (◯ , ®) (Copyright is currently a macro, circled R is part of the ams symbol fonts) (2 glyphs).

A dash with a length between en-dash and em-dash. Needed for polish. Identical to Adobe's 'threequartersemdash' (1glyph)

'Tick' and 'backtick' (ASCII-acute and ASCII-grave). Found in the Malvern documentation (2 glyphs)

Custom signs occuring in the iso-8859 latin alphabets (broken vert, logical not, macron, male and female ordinal indicator (looking like raised o and a), plus-minus, superscript 1, 2, and 3, centered dot, vulgar fractions $^1/_4$, $^1/_2$, $^3/_4$, times and div (latin 1); 'ASCII'-style breve, hachek and hungarian long umlaut (latin 2)). (18 glyphs)

A degree sign and eventually a ready-made celsius sign (2 glyphs)

A micron sign (1 glyph)

Oldstyle digits. Thus it will be possible to build dc-based virtual fonts with oldstyle digits (10 glyphs)

Fraction (a slash to build frations). In the METAFONT implementation probably identical to the usual slash (1 glyph)

Sum: 70 out of 256 possible glyphs.

Glyphs to be included in the second TS font (TSA)

The second TS font provides a bunch of glyphs supporting the Adobe standard and expert sets. Most of them are hard to do in METAFONT, but not necessary, since METAFONT supports the notion of a design size. They are however needed to complement so-called scalable fonts.

In the METAFONT world, the second TS font can be done either as a virtual font or by doing some METAFONT hacks.

Superior digits (0–9). This makes 10 glyphs. Allthough 1, 2, and 3 are already in TS1, they are repeated here for consistency. (10 glyphs)

Inferior digits (0–9) (10 glyphs)

Small versions of some punctuation: !, &, ?, !‘, ?‘ (5 glyphs)

Superior letters and symbols (The Adobe selection is quite arbitrary and incomplete): (,), $, period, comma, a, b, cent, d, e, i, l, m, n, o, r, s, t, hyphen, trademark (20 glyphs)

Inferior symbols: (,), hyphen, $, cent, hyphen (6 glyphs)

Fractions: $^1/_8$, $^3/_8$, $^5/_8$, $^7/_8$, $^1/_3$, $^2/_3$ (6 glyphs)

Misc. Symbols: onefitted, twodotenleader, onedotenleader, ellipsis (4 glyphs)

Sum: 61 out of another 256

I need some feedback on the following

More currency signs (Guarani, Lira (different to pounds e. g. in the HP character set), Naira, Peseta, Rupee, Won) (6+ glyphs)

Humanist editions often use more special glyphs, like angle brackets and 'blackbord bold' square brackets. Can someone provide more information on these? (? glyphs)

Recipe: Capital R with crossed right leg (1 glyph)

Fortran-style 'blank' symbol (b with slash through the ascender) (1 glyph)

A low tilde (1 glyph)

Interrobang (overstruck ? and !) (1 glyph, evt. 2 with spanish inverted 'Gnaborretni').

Shortequals: A special dash or hyphen used sometimes in german. (1 glyph)

Warenzeichen: Circled Wz. German usage. Do we really want this? (1 glyph)

Alan proposed: 'La' and 'TeX'. Anyone who thinks that these aren't needed should read ltnews.cls, of which 80% is hacks to set up the logos for Palatino, Helvetica, CMSS, URW Grotesk and Utopia. (2 glyphs).

The following shall *not* occur in the TS encoding

Additional letters to support more languages. They belong to T*n* text font encodings.

Any mathematical symbols. They should be added to new math font encodings.

Dingbats. They aren't needed in all the different styles a text font can have, and are better placed in special dingbat fonts.

Playing cards. They count as dingbats.

Code points for the TS1 font

Here are the assigned Code points for the TS1 fonts. Several points are choosen because they are used either in the Cork encoding (the accents) or the latin-1 alphabet. Glyphs which could not be placed this way are put in positions 128 ff. in consecutive order. I have assigned positions to some glyphs which may not be in the first release of TS1 (e. g. 'Recipe'), but may be added later.

code (oct)	name	why
Accents for capital letters		
000	grave	Cork
001	acute	Cork
002	circumflex	Cork
003	tilde	Cork
004	umlaut	Cork
005	hungarian	Cork
006	ring	Cork
007	hachek	Cork
010	breve	Cork
011	macron	Cork
012	dot above	Cork
013	cedilla	Cork
014	ogonek	Cork
Miscellaneous		
026	three quarters emdash	Cork —
044	dollar	ASCII
047	straight quote	ASCII
052	centered star	ASCII
057	fraction	ASCII-/
Oldstyle digits		
060	oldstyle digit 0	ASCII
061	oldstyle digit 1	ASCII
062	oldstyle digit 2	ASCII
063	oldstyle digit 3	ASCII
064	oldstyle digit 4	ASCII
065	oldstyle digit 5	ASCII
066	oldstyle digit 6	ASCII
067	oldstyle digit 7	ASCII
070	oldstyle digit 8	ASCII
071	oldstyle digit 9	ASCII
Miscellaneous		
136	arrow up	ASCII-^
137	arrow down	ASCII-_
140	backtick	ASCII-`
176	low tilde	ASCII-~
177	short equals (hyphenchar)	Cork
TS1-symbols		
200	ASCII-style breve	TS1
201	ASCII-style hachek	TS1
202	double tick (ASCII-style hungarian)	TS1
203	double backtick	TS1
204	dagger	TS1
205	ddager	TS1
206	double vert	TS1

code (oct)	name	why
207	per thousand	TS1
210	bullet	TS1
211	centigrade	TS1
212	dollaroldstyle	TS1
213	centoldstyle	TS1
214	florin	TS1
215	colon	TS1
216	won	TS1
217	naira	TS1
220	guarani	TS1
221	peso	TS1
222	lira	TS1
223	recipe	TS1
224	interrobang	TS1
225	gnaborretni	TS1
symbols from ISO-8859-1 (latin-1)		
242	cent	latin-1
243	sterling	latin-1
244	currency sign	latin-1
245	yen	latin-1
246	broken vertical bar	latin-1
247	section sign	latin-1
251	copyright	latin-1
252	feminine ordinal indicator	latin-1
254	logical not	latin-1
256	circled R	latin-1
257	high bar (ASCII-style macron)	latin-1
260	degree sign	latin-1
261	plus-minus sign	latin-1
262	superscript 2	latin-1
263	superscript 3	latin-1
264	tick (ASCII-style acute)	latin-1
265	micro sign	latin-1
266	pilcrow sign	latin-1
267	centered dot	latin-1
271	superscript 1	latin-1
272	masculine ordinal indicator	latin-1
274	fraction one quarter	latin-1
275	fraction one half	latin-1
276	fraction three quarters	latin-1
326	times (multiplication sign)	latin-1
366	division sign	latin-1

Code points for the TSA font

This section is not fully worked out yet. Every position of a lowercase letter in the Cork fonts is reserved for the respective superscript lowercase letter. Any of them can in principle occur in an abbreviature (e. g. in $2^{\text{ième}}$). The codes of the digits 0–9 are reserved for the superscript digits, the remaining glyphs need to be assigned points later.