[March 16, 1997]

EuroTEX

From version 3.0 onwards, TEX allows the use of eight-bit fonts in the manuscript. In EuroTEX this is utilized, so that non-English letters as far as possible are treated as English letters. This description concerns the EuroTEX extensions founded on plain TEX, version 3.141, with the use of dc fonts.

- Non-English letters are allowed in the text part of the manuscript (in equations only by using `\hbox`). All letters in the font ISO Latin 1 except ÿ may be used. Letters like ä ö ü ß à â ç è é ê î ï ô û ù æ ø å can hence be typed directly into the `tex` file. Paraphrases like `\"a \"o \"u \ss \`a \^a \c c \`e \'e \^e \^\i \"\i \^o \^u \`u \ae \o \aa` are still understood, however. The letters in the font ISO Latin 1 may also be used in control words. Characters in ISO Latin 1 which are not letters, will as a rule not be understood. However, TEX understands the characters § ¡ ¿ £ « » in text mode and the characters ± · in math mode.
- TEX can hyphenate words containing the letters in the dc font. The treatment is the same, whether the letters are typed directly or as macros.
- Languages can be selected with the command `\language\`⟨language⟩, where ⟨language⟩ is one of the words `english german french italian polish danish norwegian swedish icelandic finnish`. By the selection of language, rules for hyphenation, quotes and sentence space are set. For each of the ten languages, a list of syllables is read, with rules for hyphenation of words. Correct quotes in each language appear by typing `\qt{`⟨text⟩`}` or — for single quotes — `\siqt{`⟨text⟩`}`. The different quotes are treated below. Sentence spaces are for all languages, except French, set to `\nonfrenchspacing`; for French it is set to `\frenchspacing`. One may switch between languages arbitrary often within a document. The selected language is valid until a new language is selected. The default language is English.
- Test on the selected language can be done with the command `\ifnum\language=\englishhyphenation` and equivalently for the other languages. Other `\if` tests may also be used.
- Quotes are traditionally written differently in the different languages. The customary notation in the ten languages is: “English” „German“ «French» «Italian» „Polish“ „Danish“ «Norwegian» ”Swedish” „Icelandic“ ”Finnish”. By typing `\qt{`⟨text⟩`}`, one gets the mentioned quotes in the ten languages. Single quotes appear by typing `\siqt{`⟨text⟩`}`. Hence, with this control word one gets: ‘English’ ‚German‘ ‹French› ‹Italian› ‚Polish‘ ‚Danish‘ ‹Norwegian› ’Swedish’ ‚Icelandic‘ ’Finnish’. Independently of the selected language, ``` `` ``` and `''` are still ligatures for “ and ”. Further, `<<` and `>>` are ligatures for « and ». They may also be produced by typing them directly, i.e., `«` and `»`. Of course, also the single characters ‘ and ’ can be typed directly with `` ` `` and `'`. The other quotes may also be selected independently of the selected language. The quotes can be produced as shown in the table below. The table also shows in which languages a quote traditionally is used.

`\lohi{low-high}`	yields	„low-high“	German Polish Danish Icelandic
`\hihi{high-high}`	yields	“high-high”	English Spanish Italian
`\hiqq{high-equal}`	yields	”high-equal”	Swedish Finnish Norwegian
`\ltgt{outwards}`	yields	«outwards»	French Italian Spanish Norwegian
`\gtlt{inwards}`	yields	»inwards«	German Danish
`\gtgt{right}`	yields	»right»	Swedish Finnish

  Single quotes appear by putting `si` in front, i.e., with the macros `\silohi \sihihi \sihiqq \siltgt \sigtlt \sigtgt`. For “Gänsefüßchen” in front, the macro `\hiqq` yields a better positioning than the ligature `''`:

`\hiqq{Swedish}`	yields	”Swedish”
`''Swedish''`	yields	”Swedish”.

  Do you see the difference?
- Macros for letters and diacritical marks. In addition to the letters defined in the TEX book, the following letters have names of their own:

`\dh \DH`	yields	ð Ð	(edh)
`\th \TH`	yields	þ Þ	(thorn)
`\ng \NG`	yields	ŋ Ŋ	(Lappish eng)
`\y \Y`	yields	ij IJ	(Dutch ij)
`\dj \DJ`	yields	đ Đ	(Croatian and Lappish d bar)
`\lp \Lp`	yields	l· L·	(Catalan l dot, as in “paral·lel”).

  (`\lp` is actually not a letter, but a macro that puts a dot after the l, if the word is not hyphenated there.) Diacritical marks (umlaut, accents, etc.) can be produced as mentioned in the TEX book. Three additional macros are defined:

`\k o`	yields	ǫ	(ogonek, right-adjusted for Ą, Ę, ą, ų)
`\p L`	yields	Ľ	(prime, only for Czech and Slovak ď, ť, ľ, Ľ and Latvian ģ)
`\r u`	yields	ů	(ring).

With the above macros, names like Lech Wałęsa, Pelėda Lazdynų, Piešťany, Þingvellir, Morgunblaðið and Báđašoaivi can be written. Correct spellings of such names cannot be produced with the traditional cm fonts.

- The control symbol `\|` is used to split characters. It can be used to split a word from a hyphen (-), if TEX does not manage to hyphenate the word. As an example, if TEX does not manage to hyphenate the word `$W$-orthogonal`, one may instead type `$W$-\|orthogonal`. Correspondingly, if TEX does not manage to hyphenate the word before the hyphen in an expression like `Lebesgue-integrable`, one may instead type `Lebesgue\|-integrable`. With such a use of `\|`, TEX will be able to hyphenate the words in question. Behind an apostrophe ('), `\|` may be used in the same way. To make TEX hyphenate the word "d'Alembert" one may type `d'\|Alembert`. Behind a slash (/), `\|` can be used in the same way, but the macro `\hskip0pt` will there be better, since it allows a line break just behind the slash. The control symbol `\|` can also be used to split the characters in a ligature. Thus, `fl` yields the ligature "ﬂ", whereas `f\|l` yields "fl". Such a splitting of ligatures is almost never important, but for Cyrillic fonts it may be necessary. In math mode, `\|` produces the norm character $\|$.
- For German hyphenation, two macros are defined. The macro `\SS` is capital ß (eszett). It is printed out as "SS", but produces another hyphenation than the input `SS`. The notation `\SS` always gives a correct hyphenation for words with capital ß. This is shown in the table below:

`GRÖSSE`	yields	GRÖS-SE	(wrong)
`GRÖ\SS E`	yields	GRÖ-SSE	(correct, compare Grö-ße)
`MASSTHEORIE`	yields	MASSTHEO-RIE	(one hyphen is missing)
`MA\SS THEORIE`	yields	MASS-THEO-RIE	(complete, as in Maß-theo-rie).

The macro `\ck` (and the corresponding for capital letters `\CK`) gives correct hyphenation in the old spelling for the letter group "ck". The notation `Verrü\ck ungen` thus yields the hyphenation possibilities "Ver-rückungen", "Verrük-kungen" and "Verrückun-gen". However, the macro must not be used in words where "ck" cannot be hyphenated, as in the word "rückwärts". The notation `rü\ck wärts` would give the hyphenation "rük-kwärts" (wrong), whereas the notation `rückwärts` gives the hyphenation "rück-wärts" (correct).

- For Scandinavian and German before the spelling reform, macros for three equal consonants that are contracted to two, are defined. Such consonant groups must be hyphenated, such that $2 + 1$ equal consonants appear. Examples of such words are "Nullösung" and "Programmodul" (German) and "toppunkt" and "tallinje" (Norwegian), which on hyphenation on the spot in question are written "Null-lösung", "Programm-modul", "topp-punkt" and "tall-linje". By typing `Nu\lll ösung`, `Progra\mmm odul`, `to\ppp unkt` and `ta\lll inje`, one always gets a correct spelling of these words. The following macros give correct treatment of such consonant groups: `\bbb`, `\ddd`, `\fff`, `\ggg`, `\kkk`, `\lll`, `\mmm`, `\nnn`, `\ppp`, `\rrr`, `\sss`, `\ttt`. (Correspondingly for capital letters: `\BBB`, `\DDD`, etc.)
- The eight-bit fonts contain some characters which have names of their own. They are shown in the following table:

`\pound`	yields	£	
`\promille`	yields	‰	
`\?`	yields	¿	(inverse question mark)
`\!`	yields	¡	(inverse exclamation mark)
`\caret`	yields	ˆ	(caret (a little different from circumflex ˆ))
`\wave`	yields	˜	(repeat mark (a little different from tilde ˜))
`\underscore`	yields	_	(underscore (different from the macro `\_` which yields _))
`\visiblespace`	yields	␣	(visible space).

The character sequences `?'` and `!'` were earlier ligatures for ¿ and ¡, but they are no longer defined. In math mode, the control symbol `\!` is still defined as "negative thin space". The notation `\it\$` gave earlier *£*, but now gives *\$*. The characters < > | (and ") in text mode were earlier only usable in typewriter fonts, but may now be used in all text fonts. The macros `\{`, `\}` and `\backslash` were earlier only defined in math mode, but are now valid also in text mode.

- In the macro `\hyphenation`, the letters in ISO Latin 1 as well as macros for the letters in the dc fonts may be used. Hence, one may type `\hyphenation{ång-ström}` or `\hyphenation{\aa ng-str\"om}`. In both cases "Ångström" will be correctly hyphenated. However, braces can not be used within `\hyphenation`. Thus, the notation `\hyphenation{{\aa}ng-str\"om}` yields an error message in TEX.

- In the macros `\uppercase` and `\lowercase`, the letters in ISO Latin 1 and macros for diacritical marks may be used. Macros for letters of their own, like `\aa \oe \l`, however, will not be treated correctly. Also, the letters which change diacritical marks (like ď, Ď and ť, Ť) will not be treated correctly. Thus, the command `\uppercase{\"o\oe\ss\l\k a\"\i\p t}` produces "Öœßł Ą̨ïT". In remedy of this, two new macros `\upcase` and `\lowcase` are introduced. They give correct treatment for all letters in the dc fonts, whether they are typed directly or as macros, together with most letters with diacritical marks. Thus, `\upcase{\"o\oe\ss\l\k a\"\i\p t}` yields the result "ÖŒSSŁĄÏŤ". An exception constitute the Lithuanian vowels with ogonek (į, ų) and the Latvian consonants with cedilla (ģ, ķ, ļ, ņ). TEX gives an error message, if one tries to treat them with `\upcase` or `\lowcase`. Also, the macros `\upcase` and `\lowcase` do not work across paragraphs. The active characters `§ ¡ £ « »` cannot be used with `\uppercase`, and hence not with `\upcase`. TEX gives an error message by such a use. The corresponding macros `\S \! \pound \qt` may, however, be used.
- Print to file (with `\write`) — e.g. in connection with preparation of an index — was earlier problematic in TEX for non-English letters. All letters in ISO Latin 1 are now printed as such, whether they are typed directly or as macros. Other dc letters are printed as macros on the form `\char '^^xy` . As an example, `\"o` is printed as `ö`, `\ss` is printed as `ß`, `\oe` is printed as `\char '^^f7` , whereas `\k a` is printed as `\char '^^a1` . Since the macro `\char '^^xy` gobbles up succeeding space, this means that space will disappear behind a word which ends with such a macro. Letters which are not in the dc fonts, but which can be created with TEX, will normally be printed as macros. An exception constitute the Lithuanian vowels with ogonek (į, ų) and the Latvian consonants with cedilla (ģ, ķ, ļ, ņ). These letters cannot be further treated with TEX when they are printed to file. The problems with difficultly readable or not treatable macros are reduced by using `\string`. As an example, `\string\k\space a` produces `\k a`. However, if the words printed to file are to be sorted, in practice only the letters of ISO Latin 1 are appropriate in such a print.
- The dc fonts are used everywhere in text mode. In math mode the cm fonts are used as mentioned in the TEX book. Below, the font dcr10 (ten-point Roman) is shown. The characters have the same position in all dc fonts. The dc fonts contain the letters of the most important European languages written with Latin letters. Unfortunately, it was impossible to find room for the Catalan letters l·, L·, the Lithuanian letters ė, Ė, į, Į, ų, Ų, ū, Ū and the Latvian letters ā, Ā, ē, Ē, ī, Ī, ū, Ū, ģ, Ģ, ķ, Ķ, ļ, Ļ, ņ, Ņ. From a Nordic point of view, it is also of interest, that the Lappish letters ŧ, Ŧ are missing. On the other hand, the letters ÿ, Ÿ do not enter into any European language. The Dutch letter groups ij, IJ also seem to have a very limited value of use.

	'0	'1	'2	'3	'4	'5	'6	'7	
'00x	`	´	ˆ	˜	¨	˝	˚	ˇ	"0x
'01x	˘	¯	˙	¸	˛	‚	‹	›	
'02x	“	”	„	«	»	–	—		"1x
'03x	₀	ı	ȷ	ﬀ	ﬁ	ﬂ	ﬃ	ﬄ	
'04x	␣	!	"	#	$	%	&	’	"2x
'05x	(	)	*	+	,	-	.	/	
'06x	0	1	2	3	4	5	6	7	"3x
'07x	8	9	:	;	<	=	>	?	
'10x	@	A	B	C	D	E	F	G	"4x
'11x	H	I	J	K	L	M	N	O	
'12x	P	Q	R	S	T	U	V	W	"5x
'13x	X	Y	Z	[	\	]	^	_	
'14x	‘	a	b	c	d	e	f	g	"6x
'15x	h	i	j	k	l	m	n	o	
'16x	p	q	r	s	t	u	v	w	"7x
'17x	x	y	z	{	|	}	~	-	
'20x	Ă	Ą	Ć	Č	Ď	Ě	Ę	Ğ	"8x
'21x	Ĺ	Ľ	Ł	Ń	Ň	Ŋ	Ő	Ŕ	
'22x	Ř	Ś	Š	Ş	Ť	Ţ	Ű	Ů	"9x
'23x	Ÿ	Ź	Ž	Ż	Ĳ	İ	đ	§	
'24x	ă	ą	ć	č	ď	ě	ę	ğ	"Ax
'25x	ĺ	ľ	ł	ń	ň	ŋ	ő	ŕ	
'26x	ř	ś	š	ş	ť	ţ	ű	ů	"Bx
'27x	ÿ	ź	ž	ż	ĳ	¡	¿	£	
'30x	À	Á	Â	Ã	Ä	Å	Æ	Ç	"Cx
'31x	È	É	Ê	Ë	Ì	Í	Î	Ï	
'32x	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	Œ	"Dx
'33x	Ø	Ù	Ú	Û	Ü	Ý	Þ	SS	
'34x	à	á	â	ã	ä	å	æ	ç	"Ex
'35x	è	é	ê	ë	ì	í	î	ï	
'36x	ð	ñ	ò	ó	ô	õ	ö	œ	"Fx
'37x	ø	ù	ú	û	ü	ý	þ	ß	
	"8	"9	"A	"B	"C	"D	"E	"F	