[16. März 1997]

EuroTEX

Ab Version 3.0 erlaubt TEX Eingabe von Achtbitfonts. In EuroTEX hat man das ausgenutzt, so daß nicht-englische Buchstaben in höchtsmöglichem Grade wie englische Buchstaben behandelt werden. Diese Beschreibung betrifft die EuroTEX-Erweiterungen, aufbauend auf „plain“ TEX, Version 3.141, mit Verwendung von dc-Fonts.

- Nichtenglische Buchstaben sind im Textteil des Manuskripts erlaubt (in Gleichungen nur durch Gebrauch von `\hbox`). Man kann alle Buchstaben im Font ISO Latin 1 außer dem Buchstaben `ÿ` verwenden. Buchstaben wie `ä ö ü ß à â ç è é ê î ï ô û ù æ ø å` lassen sich also direkt in die `tex`-Datei eintasten. Umschreibungen wie `\"a \"o \"u \ss \`a \^a \c c \`e \'e \^e \^\i \"\i \^o \^u \`u \ae \o \aa` werden aber weiterhin verstanden. Die Buchstaben im Font ISO Latin 1 können auch in Namensbefehlen verwendet werden. Zeichen in ISO Latin 1, die nicht Buchstaben sind, werden in der Regel nicht verstanden. TEX versteht aber die Zeichen `§ ¡ ¿ £ « »` in Textmodus und die Zeichen ± · in Mathematikmodus.
- TEX kann Wörter trennen, die die Buchstaben im dc-Font enthalten. Die Behandlung wird dieselbe, unabhängig davon, ob die Buchstaben direkt oder mittels Makros eingegeben werden.
- Sprachen lassen sich durch den Befehl `\language\`⟨Sprache⟩ wählen, wobei ⟨Sprache⟩ eines der Wörter `english german french italian polish danish norwegian swedish icelandic finnish` ist. Bei der Wahl einer Sprache werden Regeln für Silbentrennung, Anführungszeichen und Satzleerraum gesetzt. Für jede der zehn Sprachen ist eine Trenntabelle mit Silbentrennungsregeln eingelesen. Richtige Anführungszeichen in jeder Sprache werden durch den Befehl `\qt{`⟨Text⟩`}` oder — für einfache Anführungszeichen — `\siqt{`⟨Text⟩`}` erzeugt. Die unterschiedlichen Anführungszeichen werden unten behandelt. Satzleerraum ist für alle Sprachen außer Französisch zu `\nonfrenchspacing` definiert; für Französisch ist er `\frenchspacing`. Innerhalb eines Dokuments kann man beliebig oft zwischen den Sprachen wechseln. Die Sprachwahl gilt, bis eine neue Sprache gewählt wird. Beim Auslaß des Befehls `\language\`⟨Sprache⟩ wird Englisch gewählt.
- Test auf die gewählte Sprache läßt sich mittels des Befehls `\ifnum\language=\englishhyphenation` durchführen. Entsprechend für die anderen Sprachen. Andere `\if`-Tests lassen sich auch benutzen.
- Anführungszeichen werden traditionell unterschiedlich in den einzelnen Sprachen geschrieben. Die üblichsten Schreibweisen in den zehn Sprachen sind: “Englisch” „Deutsch“ «Französisch» «Italienisch» „Polnisch“ „Dänisch“ «Norwegisch» ”Schwedisch” „Isländisch“ ”Finnisch”. Durch den Befehl `\qt{`⟨Text⟩`}` werden die erwähnten Anführungszeichen in den zehn Sprachen erzeugt. Einfache Anführungszeichen erhält man durch den Befehl `\siqt{`⟨Text⟩`}`. Mit diesem Befehl ergibt sich also: ‘Englisch’ ‚Deutsch‘ ‹Französisch› ‹Italienisch› ‚Polnisch‘ ‚Dänisch‘ ‹Norwegisch› ’Schwedisch’ ‚Isländisch‘ ’Finnisch’. Unabhängig von der Sprachwahl sind weiterhin ``` `` ``` und `''` Ligaturen für “ und ”. Ferner sind `<<` und `>>` Ligaturen für « und ». Diese erhält man auch durch direkte Eingabe von `«` und `»`. Ebenfalls werden selbstverständlich die Einzelzeichen ‘ und ’ durch die Eingaben `` ` `` und `'` erzeugt. Auch die übrigen Anführungszeichen lassen sich unabhängig von der gewählten Sprache erzeugen. Die Anführungszeichen erscheinen, wie in der untenstehenden Tabelle gezeigt. Die Tabelle zeigt gleichzeitig, in welcher Sprache ein Anführungszeichen traditionell gebraucht wird.

`\lohi{unten-oben}`	ergibt	„unten-oben“	Deutsch Polnisch Dänisch Isländisch
`\hihi{oben-oben}`	ergibt	“oben-oben”	Englisch Spanisch Italienisch
`\hiqq{oben-gleiche}`	ergibt	”oben-gleiche”	Schwedisch Finnisch Norwegisch
`\ltgt{außen}`	ergibt	«außen»	Französisch Italienisch Spanisch Norwegisch
`\gtlt{innen}`	ergibt	»innen«	Deutsch Dänisch
`\gtgt{rechts}`	ergibt	»rechts»	Schwedisch Finnisch

  Einfache Anführungszeichen erscheinen durch ein vorangestelltes `si`, d.h. mittels der Makros `\silohi \sihihi \sihiqq \siltgt \sigtlt \sigtgt`. Bei vorangestellten Gänsefüßchen ergibt der Makro `\hiqq` eine bessere Lage als die Ligatur `''`:

`\hiqq{Schwedisch}`	ergibt	”Schwedisch”
`''Schwedisch''`	ergibt	”Schwedisch”.

  Sehen Sie den Unterschied?
- Makros für Buchstaben und diakritische Zeichen. Zusätzlich zu den Buchstaben, die im TEX-Buch definiert sind, haben folgende Zeichen eigene Namen:

`\dh \DH`	ergibt	ð Ð	(Edh)
`\th \TH`	ergibt	þ Þ	(Thorn)
`\ng \NG`	ergibt	ŋ Ŋ	(lappisches Äng)
`\y \Y`	ergibt	ij IJ	(niederländisches ij)

`\dj \DJ`	ergibt	đ Đ	(kroatisches und lappisches durchstrichenes d)
`\lp \Lp`	ergibt	l· L·	(katalanisches l Punkt, wie in „paral·lel“).

(`\lp` ist eigentlich kein Buchstabe, sondern ein Makro, der ein Punkt hinter das l anbringt, falls das Wort dort nicht getrennt wird.) Diakritische Zeichen (Umlaute, Akzente u.ä.) lassen sich erzeugen, wie im TEX-Buch beschrieben. Drei zusätzliche Makros sind definiert:

`\k o`	ergibt	ǫ	(Krummhaken, Ogonek, rechtsverschoben für Ą, Ę, ą, ų)
`\p L`	ergibt	Ľ	(Strich, nur für tschechisches und slowakisches ď, ť, ľ, Ľ und lettisches ģ)
`\r u`	ergibt	ů	(Kringel).

Mit den obigen Makros kann man Namen wie Lech Wałęsa, Pelėda Lazdynų, Piešťany, Þingvellir, Morgunblaðið und Báďašoaivi schreiben. Richtige Schreibungen solcher Namen lassen sich nicht anhand der traditionellen cm-Fonts erzeugen.

- Der Kurzbefehl `\|` wird benutzt, um Zeichen zu scheiden. Er wird benutzt, um ein Wort von einem Bindestrich (-) zu scheiden, falls TEX es nicht schafft, den Ausdruck zu trennen. Falls zum Beispiel TEX den Ausdruck **`$n$-Komponentensystem`** nicht trennen kann, sollte man stattdessen schreiben: **`$n$-\|Komponentensystem`**. Ebenfalls, wenn TEX es nicht schafft, das Wort vor dem Bindestrich in einem Ausdruck wie **`Tschebyschow-Polynome`** zu trennen, sollte man stattdessen schreiben: **`Tschebyschow\|-Polynome`**. Mit einem solchen Gebrauch von `\|` wird TEX die aktuellen Wörter trennen können. Hinter einem Apostroph (') läßt sich `\|` auf dieselbe Weise benutzen. Um TEX dazu zu bringen, das Wort „d'Alembert“ zu trennen, schreibe man **`d'\|Alembert`**. Hinter einem Schrägstrich (/) läßt sich auch `\|` auf diese Weise benutzen, aber der Makro **`\hskip0pt`** ist dort besser, weil dieser einen Umbruch unmittelbar hinter dem Schrägstrich erlaubt. Der Kurzbefehl `\|` läßt sich auch dazu benutzen, die Zeichen einer Ligatur zu scheiden. Somit ergibt **`fl`** die Ligatur „fl“, aber **`f\|l`** ergibt „fl“. Eine solche Scheidung der Ligaturen ist kaum wichtig, aber für kyrillische Fonts kann sie notwendig sein. In Mathematikmodus ergibt `\|` das Normzeichen $\|$.
- Für deutsche Silbentrennung sind zwei Makros definiert. Der Makro **`\SS`** ist großes ß (Eszett). Er ergibt die Ausgabe „SS“, ruft aber eine andere Silbentrennung hervor als diejenige, die man durch die Schreibweise **`SS`** erhält. Der Befehl **`\SS`** ergibt immer richtige Silbentrennung für großes ß. Dies ist in der untenstehenden Tabelle gezeigt:

`GRÖSSE`	ergibt	GRÖS-SE	(falsch)
`GRÖ\SS E`	ergibt	GRÖ-SSE	(richtig, vergleiche Grö-ße)
`MASSTHEORIE`	ergibt	MASSTHEO-RIE	(ein Bindestrich fehlt)
`MA\SS THEORIE`	ergibt	MASS-THEO-RIE	(vollständig, wie in Maß-theo-rie).

Der Makro **`\ck`** (und der für große Buchstaben zugehörige **`\CK`**) ergibt richtige Silbentrennung nach der alten Rechtschreibung bei der Buchstabenverbindung „ck“. Mit der Schreibweise **`Verrü\ck ungen`** erhält man somit die Trennmöglichkeiten „Ver-rückungen“, „Verrük-kungen“ und „Verrückun-gen“. Der Makro muß aber nicht in Wörtern wie „rückwärts“ benutzt werden, in denen „ck“ nicht getrennt werden kann. Die Schreibweise **`rü\ck wärts`** ergäbe die Trennung „rük-kwärts“ (falsch), während die Schreibweise **`rückwärts`** die Trennung „rück-wärts“ ergibt (richtig).

- Für Skandinavisch und Deutsch vor der Rechtschreibungsreform sind Makros für drei gleiche Konsonanten, die zu zwei Konsonanten zusammengezogen sind, definiert. Solche Konsonantengruppen müssen derart getrennt werden, daß $2+1$ gleiche Konsonanten entstehen. Beispiele solcher Wörter sind „Nullösung“ und „Programmodul“ (Deutsch) und „toppunkt“ und „tallinje“ (Norwegisch), die durch Trennung an der aktuellen Stelle „Null-lösung“, „Programm-modul“, „topp-punkt“ und „tall-linje“ heißen. Durch die Schreibweisen **`Nu\lll ösung`**, **`Progra\mmm odul`**, **`to\ppp unkt`** und **`ta\lll inje`** wird immer die richtige Ausgabe dieser Wörter erzeugt. Folgende Makros ergeben richtige Behandlung solcher Konsonantengruppen: **`\bbb`**, **`\ddd`**, **`\fff`**, **`\ggg`**, **`\kkk`**, **`\lll`**, **`\mmm`**, **`\nnn`**, **`\ppp`**, **`\rrr`**, **`\sss`**, **`\ttt`**. (Entsprechend für große Buchstaben: **`\BBB`**, **`\DDD`**, usw.)
- Die Achtbitfonts enthalten einige Zeichen, die eigene Namen haben. Diese sind in der untenstehenden Tabelle gezeigt:

`\pound`	ergibt	£	
`\promille`	ergibt	‰	
`\?`	ergibt	¿	(umgekehrtes Fragezeichen)
`\!`	ergibt	¡	(umgekehrtes Ausrufezeichen)
`\caret`	ergibt	ˆ	(Einschaltungszeichen (etwas verschieden von Zirkumflex ˆ))
`\wave`	ergibt	˜	(Wiederholungszeichen (etwas verschieden von Tilde ˜))
`\underscore`	ergibt	_	(Unterstrich (verschieden vom Makro **`\_`**, der _ ergibt))
`\visiblespace`	ergibt	␣	(sichtbares Leerzeichen).

Die Zeichenfolgen `?‘` und `!‘` waren früher Ligaturen für ¿ und ¡, diese sind aber nicht mehr definiert.

In Mathematikmodus ist der Kurzbefehl `\!` weiterhin als „negativer kleiner Leerraum“ definiert. Die Schreibweise `\it\$` ergab früher *£*, ergibt aber jetzt *$*. Die Zeichen `<` `>` `|` (und `"`) in Textmodus waren früher nur in Schreibmaschinenschrift verwendbar, können aber jetzt in allen Textfonts benutzt werden. Die Makros `\{`, `\}` und `\backslash` waren früher nur in Mathematikmodus definiert, gelten aber jetzt auch in Textmodus.

- Im Makro `\hyphenation` lassen sich sowohl die Buchstaben in ISO Latin 1 wie Makros für die Buchstaben in den dc-Fonts benutzen. Man kann also `\hyphenation{ång-ström}` oder `\hyphenation{\aang-str\"om}` schreiben. In beiden Fällen wird „Ångström“ immer richtig getrennt. Geschweifte Klammern lassen sich aber nicht in `\hyphenation` benutzen. Somit ergibt `\hyphenation{{\aa}ng-str\"om}` eine Fehlermeldung in TEX.
- In den Makros `\uppercase` und `\lowercase` können die Buchstaben in ISO Latin 1 und Makros für diakritische Zeichen benutzt werden. Makros für eigene Buchstaben wie `\aa \oe \l` werden aber nicht richtig behandelt. Auch nicht Buchstaben, die ihr diakritisches Zeichen wechseln (wie ď, Ď und ť, Ť), werden richtig behandelt. Somit ergibt `\uppercase{\"o\oe\ss\l\k a\"\i\p t}` die Ausgabe „ÖœßłĄïT“. Um diesem abzuhelfen, sind zwei neue Makros `\upcase` und `\lowcase` eingeführt. Diese ergeben richtige Behandlung für alle Buchstaben in den dc-Fonts, ob sie direkt oder mittels Makros geschrieben sind, samt die meisten Buchstaben mit diakritischen Zeichen. Somit ergibt der Befehl `\upcase{\"o\oe\ss\l\k a\"\i\p t}` die Ausgabe „ÖŒSSŁĄÏŤ“. Eine Ausnahme bilden die litauischen Vokale mit Ogonek (į, ų) und die lettischen Konsonanten mit Cedille (ģ, ķ, ļ, ņ). TEX gibt eine Fehlermeldung, wenn man versucht, diese mit `\upcase` oder `\lowcase` zu behandeln. Die Makros `\upcase` und `\lowcase` wirken außerdem nicht über Abschnitte. Die aktiven Zeichen `§ ¡ £ « »` dürfen nicht mit `\uppercase`, und damit auch nicht mit `\upcase`, benutzt werden. TEX gibt eine Fehlermeldung bei solchem Gebrauch. Die entsprechenden Makros `\S \! \pound \qt` lassen sich aber benutzen.
- Ausgabe in eine Datei (mit `\write`) — z.B. in Verbindung mit Erarbeitung eines Registers — war früher für nichtenglische Buchstaben problematisch in TEX. Alle Buchstaben in ISO Latin 1 werden jetzt als solche ausgegeben, ob sie direkt oder mit Makros eingeben sind. Andere dc-Buchstaben werden als Makros auf die Gestalt `\char '^^xy` ausgegeben. Beispielsweise wird `\"o` als `ö` ausgegeben, `\ss` wird als `ß` ausgegeben, `\oe` wird als `\char '^^f7` ausgegeben, während `\k a` als `\char '^^a1` ausgegeben wird. Weil der Makro `\char '^^xy` nachfolgende Leerzeichen verschluckt, bedeutet dies, daß ein Zwischenraum hinter einem Wort, das mit einem solchen Makro endet, verschwinden wird. Buchstaben, die nicht in die dc-Fonts eingehen, die aber mit TEX gebildet werden können, werden normalerweise als Makros ausgegeben. Eine Ausnahme bilden die litauischen Vokale mit Ogonek (į, ų) und die lettischen Konsonanten mit Cedille (ģ, ķ, ļ, ņ). Diese lassen sich nicht mit TEX weiter verarbeiten, wenn sie in eine Datei geschrieben werden. Die Probleme mit schwer lesbaren oder nicht verarbeitbaren Makros werden durch Verwendung von `\string` reduziert. Beispielsweise ergibt der Befehl `\string\k\space a` die Ausgabe `\k a`. Falls aber die Wörter, die in eine Datei geschrieben sind, sortiert werden sollen, sind in der Praxis nur die Buchstaben, die in ISO Latin 1 eingehen, für eine solche Ausgabe zweckmäßig.
- Die dc-Fonts werden überall in Textmodus benutzt. In Mathematikmodus werden die cm-Fonts benutzt, wie im TEX-Buch beschrieben. Unten ist der Font dcr10 (Zehnpunkt-Antiqua) gezeigt. Die Zeichen stehen in allen dc-Fonts auf derselben Stelle. Die dc-Fonts enthalten die Buchstaben der wichtigsten europäischen Sprachen, die mit lateinischer Schrift geschrieben werden. Leider wurde kein Platz für die katalanischen Buchstaben l·, L·, die litauischen Buchstaben ė, Ė, į, Į, ų, Ų, ū, Ū und die lettischen Buchstaben ā, Ā, ē, Ē, ī, Ī, ū, Ū, ģ, Ģ, ķ, Ķ, ļ, Ļ, ņ, Ņ gefunden. Aus nordischer Sicht ist es auch von Interesse, daß die lappischen Buchstaben ŧ, Ŧ fehlen. Auf der anderen Seite gehen angeblich die Buchstaben ÿ, Ÿ in keine europäische Sprache ein. Die niederländischen Buchstabenverbindungen ij, IJ scheinen auch, einen sehr beschränkten Nutzwert zu haben.

	´0	´1	´2	´3	´4	´5	´6	´7	
´00x	`	´	ˆ	˜	¨	˝	˚	ˇ	″0x
´01x	˘	¯	˙	¸	˛	‚	‹	›	
´02x	“	”	„	«	»	–	—		″1x
´03x	o	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	[	\	]	ˆ	_	
´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	