[16. mars 1997]

EuroTEX

Fra og med versjon 3.0 tillater TEX bruk av åttebits fonter i manuskriptet. I EuroTEX er dette utnyttet, slik at ikke-engelske bokstaver i størst mulig grad behandles på samme måte som engelske bokstaver. Denne beskrivelse gjelder EuroTEX-utvidelsene som bygger på «plain» TEX, versjon 3.141, med bruk av dc-fonter.

- Ikke-engelske bokstaver tillates i manuskriptets tekstdel (i ligninger bare ved bruk av `\hbox`). Man kan benytte alle bokstaver i fonten ISO Latin 1 unntatt bokstaven `ÿ`. Bokstaver som `æ ø å ä ö ü ß à â ç è é ê î ï ô û ù` kan altså skrives direkte inn i `tex`-filen. Omskrivninger som `\ae \o \aa \"a \"o \"u \ss \`a \^a \c c \`e \'e \^e \^\i \"\i \^o \^u \`u` blir imidlertid fortsatt forstått. Bokstavene i fonten ISO Latin 1 kan også benyttes i styreord. Tegn i ISO Latin 1 som ikke er bokstaver, blir i regelen ikke forstått. TEX forstår imidlertid tegnene `§ ¡ ¿ £ « »` i tekstmodus og tegnene ± · i ligningsmodus.
- TEX kan dele ord som inneholder bokstavene i dc-fonten. Behandlingen blir den samme, enten bokstavene skrives direkte eller ved hjelp av makroer.
- Språk kan velges med kommandoen `\language\`⟨språk⟩, hvor ⟨språk⟩ er et av ordene `english german french italian polish danish norwegian swedish icelandic finnish`. Ved valg av språk settes regler for orddeling, anførselstegn og setningsmellomrom. For hvert av de ti språkene er det lest inn en stavelsesliste, med regler for hvordan ord kan deles. Riktige anførselstegn i hvert enkelt språk fremkommer ved å skrive `\qt{`⟨tekst⟩`}` eller — for enkle anførselstegn — `\siqt{`⟨tekst⟩`}`. De ulike anførselstegn er behandlet nedenfor. Setningsmellomrom er for alle språk unntatt fransk satt til `\nonfrenchspacing`; for fransk er den satt til `\frenchspacing`. Man kan veksle mellom språk vilkårlig ofte innenfor et dokument. Språkvalget gjelder inntil et nytt språk velges. Ved utelatelse av kommandoen `\language\`⟨språk⟩ velges engelsk.
- Test på det valgte språk kan gjennomføres med kommandoen `\ifnum\language=\englishhyphenation` og tilsvarende for de andre språkene. Andre `\if`-tester kan også benyttes.
- Anførselstegn skrives tradisjonelt forskjellig i de ulike språk. Den vanligste skrivemåten i de ti språkene er: “engelsk” „tysk“ «fransk» «italiensk» „polsk“ „dansk“ «norsk» ”svensk” „islandsk“ ”finsk”. Ved å skrive `\qt{`⟨tekst⟩`}` fåes de nevnte anførselstegn i de ti språkene. Enkle anførselstegn fåes ved å skrive `\siqt{`⟨tekst⟩`}`. Med dette styreordet fåes altså: ‘engelsk’ ‚tysk‘ ‹fransk› ‹italiensk› ‚polsk‘ ‚dansk‘ ‹norsk› ’svensk’ ‚islandsk‘ ’finsk’. Uavhengig av språkvalget er fortsatt ` `` ` og `''` ligaturer for “ og ”. Videre er `<<` og `>>` ligaturer for « og ». Disse kan også frembringes ved å skrive dem direkte, altså `«` og `»`. Likeledes kan enkelttegnene ‘ og ’ naturligvis skrives direkte med `‘` og `’`. Også de øvrige anførselstegnene kan velges uavhengig av det valgte språk. Anførselstegnene kan frembringes som vist i nedenstående tabell. Tabellen viser samtidig i hvilke språk et anførselstegn tradisjonelt brukes.

`\lohi{nede-oppe}`	gir	„nede-oppe“	tysk polsk dansk islandsk
`\hihi{oppe-oppe}`	gir	“oppe-oppe”	engelsk spansk italiensk
`\hiqq{oppe-like}`	gir	”oppe-like”	svensk finsk norsk
`\ltgt{utover}`	gir	«utover»	fransk italiensk spansk norsk
`\gtlt{innover}`	gir	»innover«	tysk dansk
`\gtgt{høyrevendt}`	gir	»høyrevendt»	svensk finsk

  De enkle anførselstegnene fremkommer ved å sette `si` foran, altså med makroene `\silohi \sihihi \sihiqq \siltgt \sigtlt \sigtgt`. For foranstilte gåseøyne gir makroen `\hiqq` litt bedre plassering enn ligaturen `''`:

`\hiqq{svensk}`	gir	”svensk”
`''svensk''`	gir	”svensk”.

  Kan du se forskjellen?
- Makroer for bokstaver og diakritiske tegn. I tillegg til de bokstaver som er definert i TEX-boken, har følgende bokstaver egne navn:

`\dh \DH`	gir	ð Ð	(stungen d (edd))
`\th \TH`	gir	þ Þ	(thorn)
`\ng \NG`	gir	ŋ Ŋ	(samisk eng)
`\y \Y`	gir	ij IJ	(nederlandsk ij)
`\dj \DJ`	gir	đ Đ	(kroatisk og samisk gjennomstrøken d)
`\lp \Lp`	gir	l· L·	(katalansk l prikk, som i «paral·lel»).

  (`\lp` er egentlig ikke en bokstav, men en makro som setter en prikk bak l-en hvis ordet ikke deles der.) Diakritiske tegn (tødler, aksenter o.l.) kan frembringes som omtalt i TEX-boken. Tre tilleggsmakroer er definert:

`\k o`	gir	ǫ	(kvist, ogonek, høyrejustert for Ą, Ę, ą, ų)
`\p L`	gir	Ľ	(merket, bare for tsjekkisk og slovakisk ď, ť, ľ, Ľ og latvisk ģ)
`\r u`	gir	ů	(ring).

Med ovenstående makroer kan man skrive navn som Lech Wałęsa, Pelėda Lazdynų, Piešťany, Þingvellir, Morgunblaðið og Báđašoaivi. Korrekte skrivemåter for slike navn kan ikke frembringes med de tradisjonelle cm-fontene.

- Styresymbolet `\|` benyttes til å skille tegn. Det kan brukes til å skille et ord fra en bindestrek (-) dersom TEX ikke klarer å dele ordet. Hvis for eksempel TEX ikke klarer å dele `$n$-komponentsystem`, kan man isteden skrive `$n$-\|komponentsystem`. Tilsvarende, hvis TEX ikke klarer å dele ordet foran bindestreken i et uttrykk som `Tsjebysjov-polynomer`, kan man isteden skrive `Tsjebysjov\|-polynomer`. Med en slik bruk av `\|` vil TEX kunne dele de aktuelle ordene. Bak en apostrof (') kan `\|` brukes på samme måte. For å få TEX til å dele ordet «d'Alembert» kan man skrive `d'\|Alembert`. Bak en skråstrek (/) kan `\|` også brukes på denne måten, men makroen `\hskip0pt` vil der være bedre siden denne tillater ombrekk umiddelbart bak skråstreken. Styresymbolet `\|` kan også brukes til å skille tegnene i en ligatur. Således gir `fl` ligaturen «ﬂ», mens `f\|l` gir «fl». En slik splitting av ligaturer er nesten aldri viktig, men for kyrilliske fonter kan den være påkrevet. I ligningsmodus gir `\|` normtegnet $\|$.
- For tysk orddeling er to makroer definert. Makroen `\SS` er stor ß (eszett). Den gir utskriften «SS», men fremkaller en annen orddeling enn den man får ved å skrive `SS`. Skrivemåten `\SS` gir alltid riktig orddeling for stor ß. Dette er vist i nedenstående tabell:

`GRÖSSE`	gir	GRÖS-SE	(galt)
`GRÖ\SS E`	gir	GRÖ-SSE	(riktig, jevnfør Grö-ße)
`MASSTHEORIE`	gir	MASSTHEO-RIE	(mangler en bindestrek)
`MA\SS THEORIE`	gir	MASS-THEO-RIE	(fullstendig, som i Maß-theo-rie).

Makroen `\ck` (og den tilhørende for store bokstaver `\CK`) gir riktig orddeling i gammel rettskrivning for bokstavkombinasjonen «ck». Med skrivemåten `Verrü\ck ungen` fåes således orddelingsmulighetene «Ver-rückungen», «Verrük-kungen» og «Verrückun-gen». Makroen må imidlertid ikke benyttes i ord hvor «ck» ikke kan deles, slik som i ordet «rückwärts». Skrivemåten `rü\ck wärts` ville gi orddelingen «rük-kwärts» (galt), mens skrivemåten `rückwärts` gir orddelingen «rück-wärts» (korrekt).

- For skandinavisk og tysk før rettskrivningsreformen er definert makroer for tre like konsonanter som er trukket sammen til to. Slike konsonantgrupper må deles slik at man får 2+1 like konsonanter. Eksempler på slike ord er «toppunkt» og «tallinje» (norsk) og «Nullösung» og «Programmodul» (tysk) som ved deling på det aktuelle sted heter «topp-punkt», «tall-linje», «Null-lösung» og «Programm-modul». Ved å benytte skrivemåtene `to\ppp unkt`, `ta\lll inje`, `Nu\lll ösung` og `Progra\mmm odul` fåes alltid riktig utskrift av disse ordene. Følgende makroer gir korrekt behandling av slike konsonantgrupper: `\bbb`, `\ddd`, `\fff`, `\ggg`, `\kkk`, `\lll`, `\mmm`, `\nnn`, `\ppp`, `\rrr`, `\sss`, `\ttt`. (Tilsvarende for store bokstaver: `\BBB`, `\DDD`, osv.)
- Åttebitsfontene inneholder enkelte tegn som har egne navn. Disse er vist i følgende tabell:

`\pound`	gir	£	
`\promille`	gir	‰	
`\?`	gir	¿	(omvendt spørsmålstegn)
`\!`	gir	¡	(omvendt utropstegn)
`\caret`	gir	ˆ	(innskuddstegn (litt forskjellig fra cirkumfleks ˆ))
`\wave`	gir	˜	(gjentagelsestegn (litt forskjellig fra tilde ˜))
`\underscore`	gir	_	(understrek (forskjellig fra makroen `\_` som gir _))
`\visiblespace`	gir	␣	(synlig mellomrom).

Tegnfølgene `?'` og `!'` var tidligere ligaturer for ¿ og ¡, men disse er ikke lenger definert. I ligningsmodus er styresymbolet `\!` fremdeles definert som «negativt tynt mellomrom». Skrivemåten `\it\$` gav tidligere *£*, men gir nå *$*. Tegnene `<` `>` `|` (og `"`) i tekstmodus var tidligere bare brukbare i skrivemaskinskrift, men kan nå anvendes i alle tekstfonter. Makroene `\{`, `\}` og `\backslash` var tidligere bare definert i ligningsmodus, men gjelder nå også i tekstmodus.

- I makroen `\hyphenation` kan både bokstavene i ISO Latin 1 og makroer for bokstavene i dc-fontene benyttes. Man kan altså skrive `\hyphenation{ång-ström}` eller `\hyphenation{\aa ng-str\"om}`. I begge tilfeller vil «Ångström» alltid bli riktig delt. Klammer kan imidlertid ikke benyttes i `\hyphenation`. Således gir skrivemåten `\hyphenation{{\aa}ng-str\"om}` feilmelding i TEX.
- I makroene `\uppercase` og `\lowercase` kan bokstavene i ISO Latin 1 og makroer for diakritiske tegn benyttes. Makroer for egne bokstaver slik som `\aa \oe \l` får imidlertid ikke riktig behandling. Heller ikke bokstaver som skifter diakritisk tegn (som ď, Ď og ť, Ť) får riktig behandling. Således gir kommandoen `\uppercase{\"o\oe\ss\l\k a\"\i\p t}` resultatet «ÖœßłĄÏŤ». For å avhjelpe dette,

er det innført to nye makroer `\upcase` og `\lowcase`. Disse gir riktig behandling for alle bokstaver i dc-fontene, hva enten de skrives direkte eller som makroer, samt de fleste bokstaver med diakritiske tegn. Således gir `\upcase{\"o\oe\ss\l\k a\"\i\p t}` resultatet «ÖŒSSŁĄÏŤ». Et unntak danner de litauiske vokaler med ogonek (į, ų) og de latviske konsonanter med cédille (ģ, ķ, ļ, ņ). TEX gir feilmelding hvis man forsøker å behandle disse med `\upcase` eller `\lowcase`. Makroene `\upcase` og `\lowcase` virker dessuten ikke over avsnitt. De aktive tegnene `§ ¡ £ « »` kan ikke benyttes med `\uppercase`, og dermed heller ikke med `\upcase`. TEX gir feilmelding ved slik bruk. De tilsvarende makroene `\S \! \pound \qt` kan imidlertid brukes.

- Utskrift til fil (med `\write`) — f.eks. i forbindelse med utarbeidelse av et register — var tidligere problematisk i TEX for ikke-engelske bokstaver. Alle bokstaver i ISO Latin 1 blir nå skrevet ut som sådanne, enten de er skrevet direkte eller som makroer. Andre dc-bokstaver blir skrevet ut som makroer på formen ``\char `^^xy `` . Eksempelvis blir `\"o` skrevet ut som `ö`, `\ss` blir skrevet ut som `ß`, `\oe` blir skrevet ut som ``\char `^^f7 `` , mens `\k a` blir skrevet ut som ``\char `^^a1 `` . Siden makroen ``\char `^^xy`` sluker etterfølgende mellomrom, innebærer det at mellomrom vil forsvinne bak et ord som slutter med en slik makro. Bokstaver som ikke inngår i dc-fontene, men som kan dannes med TEX, vil normalt bli skrevet ut som makroer. Et unntak danner de litauiske vokaler med ogonek (į, ų) og de latviske konsonanter med cédille (ģ, ķ, ļ, ņ). Disse vil ikke kunne bli viderebehandlet med TEX når de skrives ut til fil. Problemene med vanskelig lesbare eller ikke behandelbare makroer reduseres ved å bruke `\string`. Eksempelvis gir `\string\k\space a` utskriften `\k a`. Dersom imidlertid de ordene som er skrevet ut til en fil, skal sorteres, er det i praksis bare de bokstaver som inngår i ISO Latin 1, som er hensiktsmessige i slik utskrift.
- dc-fontene benyttes overalt i tekstmodus. I ligningsmodus brukes cm-fontene som omtalt i TEX-boken. Nedenfor er vist fonten dcr10 (tipunkts antikva). Tegnene har samme plassering i alle dc-fonter. dc-fontene inneholder bokstavene i de viktigste europeiske språk som skrives med latinsk skrift. Dessverre er det ikke funnet plass for de katalanske bokstavene ŀ, Ŀ, de litauiske bokstavene ė, Ė, į, Į, ų, Ų, ū, Ū og de latviske bokstavene ā, Ā, ē, Ē, ī, Ī, ū, Ū, ģ, Ģ, ķ, Ķ, ļ, Ļ, ņ, Ņ. Fra nordisk synsvinkel er det også av interesse at de samiske bokstavene ŧ, Ŧ mangler. På den annen side inngår bokstavene ÿ, Ÿ visstnok ikke i noe europeisk språk. De nederlandske bokstavkombinasjonene ij, IJ synes også å ha svært begrenset nytteverdi.

	´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	