We solved this problem by designing a new set of fonts. We call them "stick-figure" fonts. The T is displayed as two line segments (Fig. 1*c*). In Fig. 2 we show a sample of some text, including a displayed equation. (The actual clarity will depend on the resolution of the graphics screen used.) Using the stick-figure font, a page containing 3000 characters can usually be displayed on the screen in less than a minute. Some users have even used this previewer from home over a modem.

(1) *Subduction condition.* The irrep D of G_0 must subduce into the indentity irrep of G, that is,

$$i(G) = \frac{1}{|G|} \sum_{g \in G} \chi(g) \neq 0, \qquad (1)$$

where $i(G)$ is the subduction frequency, and $\chi(g)$ is the character of D for the element g of G. The summation is taken over all elements g of G.

(1) *Subduction condition.* The irrep D of G_0 must subduce into the indentity irrep of G, that is,

$$i(G) = \frac{1}{|G|} \sum_{g \in G} \chi(g) \neq 0, \qquad (1)$$

where $i(G)$ is the subduction frequency, and $\chi(g)$ is the character of D for the element g of G. The summation is taken over all elements g of G.

Fig. 2. Sample TEX output (*a*) from our laser printer and (*b*) how it appears using our previewer with stick-figure fonts.

Our previewer has one major fault. We cannot distinguish between some of the different fonts. But the speed gained more than compensates for this disadvantage.

Our previewer is written in a portable C computer language.

◇ Harold T. Stokes
Department of Physics and Astronomy
Brigham Young University
Provo, Utah 84602
Phone: (801) 378-2215
Bitnet: stokesh@byuvax

Philology

Report on Multilingual Activities

Michael J. Ferguson

We now have an Extended TEX Font Encoding which should be officially approved by TUG and available on the archives by the time you read this report. This new standard is the first step in exploiting the enhanced capabilities of TEX 3.0 in a multilingual environment. The use of this standard will enable the direct use of accented characters, hyphenation, and will create, for the first time, a consistent interface to all text fonts in TEX systems. The new standard is intended to be used for all of TEX's text fonts. This means that accessing a particular code in any font will produce similar results. To this end, a **tt** font will have a **ffi** sequence of monospaced characters at the same location as the **ffi** ligature in the bold roman font. It should also serve as the encoding standard for non-Greek variable names in math mode.

The new standard extends the alphabetic content of ISO Latin 1 by including all of the linguistic characters in ISO Latin 2 along with ligatures and punctuation relevant to these languages. This standard includes the national characters, without the necessity of explicit accenting, for Albanian, Czech, Danish, Dutch, English, Faeroese, Finnish, French, German, Hungarian, Icelandic, Irish, Italian, Norwegian, Polish, Portugese, Rumanian, Serbocroatian, Slovak, Slovene, Spanish, Swedish and Turkish. It also includes a block of diacritics, so that these "basic" characters may be extended by use of TEX's accenting mechanism. This capability should allow for the extension to other languages, for example Lappish, at the possible penalty of not allowing hyphenation of words containing these explicitly accented characters.

The standard also includes some important innovative extensions to increase flexibility of use and enhance lexical semantic consistency.

- It includes both a "dash" and an explicit "hyphen char". This capability allows font designers the option of replacing the "-" with an "=" without losing the dash. Since the "-" is no longer the hyphen char, it allows words with explicit dashes, such as INRS-Télécommunications to be hyphenated. It also allows for the Serbocroatian hyphenation rule which says that a word broken at a "-" should

have the "-" both at the end of the line and the beginning of the next.

- It includes codes for opening quotes ", closing quotes ", and the ASCII double quote character ". TEX normally accesses the opening and closing quotes as ligatures, and tends to use the ASCII double quote for special purposes such as designating hex numbers. The ASCII double quote code in a font should correspond to straight quotes, much as they are in the `tt` fonts. This makes a new character available in the Roman fonts and removes an irritant in the use of double quotes in these fonts. The font designer can decide whether the opening double quotes, closing double quotes, and the ASCII double quotes are distinguishable. The standard also includes < > in the normal ASCII location.
- It includes a small "o" to match the "%" to enable allow for a "milli" or "milli.milli...", a "visible space", and both diacritic and non-diacritic versions of such characters as the " ^ ~ ".
- Perhaps the most innovative feature of the new standard, and potentially the most powerful, is the inclusion of a "`Compound Word Mark -- <cwm>`". The `<cwm>`, whose image is invisible, is effectively a character of zero width and depth. This character would be used to break up ligatures on subword boundaries in German. For example, the German word "auflage" should not have the "fl" ligature because it occurs on the subword boundary between "auf" and "lage". However a word such as "flach" should have the ligature. There are many ways to enable input of the `<cwm>`. One might be to define the "`\fl `" as meaning the `f<cwm>l`. Then the input sequence "`au\fl age`" would be rendered as "auflage". The `<cwm>` can also be used to selectively enable ligatures. For instance, if one wanted the "ff" ligature in the font to be off most of the time, the ligature access sequence in the font would be "`f<cwm>f`". Explicitly including the `<cwm>` at input would enable the ligature. The `<cwm>` can also serve as an invisible hyphen char. It is our belief that this is just the beginning of the possible uses for this character.

Although we have an excellent extended text font definition, there is much work to be done. Perhaps the most urgent is the definition of an extended math symbol font. A key decision in this particular exercise is whether to include Greek symbols in an extended math symbol font or to access Greek symbols used in math from a Greek text font or from an extended math font. In the interim the Greek characters may be accessed from the current cm fonts.

Now that we have an extended font encoding standard, we are in a position to create complete TEX 3.0 compatible hyphenation patterns. These patterns should be expressed using only the normal printable ASCII character set so that they can be transmitted by electronic mail. All non-printable ASCII characters such as an ç or an Icelandic "thorn" should be expressed as backslash sequences. Thus the French pattern which allows a hyphen to occur before any ç would be encoded as `1\c c` and `\accenthyphcodes` preceding the inputting of these patterns would include `\c c{"e7}` for the ç definition necessary within the patterns. Finally, the new patterns need to account for use of the `<cwm>`. It may be as simple as including the pattern equivalent to `1<cwm>`, or it may be more elaborate. I would like to review these patterns for syntactic consistency but would like the originators to make the arrangements for the actual distribution. I could act as the coordinator of the list of "responsables".

Finally, this new font encoding standard is important for all TEX ports because it defines TEX's internal character codes. Thus this encoding will be the basis of the `xchr[..]` and `xord[..]` translation arrays.

I should especially like to thank Jan Michael Rynning for his work in collecting the detailed information needed for creating this standard. I should also like to thank Jan Michael and Norbert Schwarz for their leadership, hard work, and attention to detail during the Cork meeting that made this important standard possible.

◇ Michael J. Ferguson
Coordinator for Multilingual Activities
INRS - Télécommunications
Université du Québec
3 Place du Commerce
Verdun H3E 1H6
Québec, Canada
mike@inrs-telecom.uquebec.ca

	'0	*'1*	*'2*	*'3*	*'4*	*'5*	*'6*	*'7*	
'00x	`	´	^	~	¨	˝	°	ˇ	"0x
'01x	˘	¯	˙	¸	˛	¸	<	>	
'02x	“	”	„	≪	≫	–	—	<cwm>	"1x
'03x	0	ı	j	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	{	|	}	~	(hyph.char) -	
'20x	Ă	Ą	Ć	Č	Ď	Ě	Ę	Ğ	"8x
'21x	Ĺ	Ľ	Ł	Ń	Ň	Eng/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	í	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	Đ	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	

A few words of explanation:

'000–'014 are accents. *'014* is an ogonek. *'015–'024* are quotation marks. *'030* is a small 0 to put after the per cent sign, to turn it into a per thousand (‰) or per million (‱) sign. *'027* (cwm) is a compound word mark (a zero-width invisible character) used e.g. for avoiding ligatures. *'040* is a visible space. *'042* is a *straight* double quotation mark.

'041–'176 is like the 7 bit ASCII code. Some characters that — at first glance — appear duplicated as accent characters usually have a different shape.

'177 is the hyphen character (that may be different from the dash (*'055*)). *'202*, *'210*, *'242*, and *'250* are A's and E's with ogonek accents.

The table has been sorted to reflect `\uppercase` `\lowercase` mechanism for all characters.

This table shows the character codes' positions, but the shapes are only approximations.