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. 1c). 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.

\begin{equation}
  i(G) = \frac{1}{|G|} \sum_{g \in G} \chi(g) \neq 0,
\end{equation}

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 \( G \).}

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.

---

**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, Portuguese, 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-communications 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 "e-10"

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. 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 "fl" 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 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 ã or an Icelandic "thorn" should be expressed as backslash sequences. Thus the French pattern which allows a hyphen to occur before any ñ would be encoded as 1\c c and \accenthyphcodes preceding the inputting of these patterns would include \c c{"Æ} for the ñ 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.

\begin{itemize}
  \item 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
\end{itemize}
Extended TeX Font Encoding Scheme — Latin

Cork, September 12, 1990

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>'00x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'01x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'02x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'03x</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'04x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'05x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'06x</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>'07x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'10x</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
</tr>
<tr>
<td>'11x</td>
<td>H</td>
<td>I</td>
<td>J</td>
<td>K</td>
<td>L</td>
<td>M</td>
<td>N</td>
</tr>
<tr>
<td>'12x</td>
<td>P</td>
<td>Q</td>
<td>R</td>
<td>S</td>
<td>T</td>
<td>U</td>
<td>V</td>
</tr>
<tr>
<td>'13x</td>
<td>X</td>
<td>Y</td>
<td>Z</td>
<td>[</td>
<td>\</td>
<td>]</td>
<td>~</td>
</tr>
<tr>
<td>'14x</td>
<td>a</td>
<td>b</td>
<td>c</td>
<td>d</td>
<td>e</td>
<td>f</td>
<td>g</td>
</tr>
<tr>
<td>'15x</td>
<td>h</td>
<td>i</td>
<td>j</td>
<td>k</td>
<td>l</td>
<td>m</td>
<td>n</td>
</tr>
<tr>
<td>'16x</td>
<td>p</td>
<td>q</td>
<td>r</td>
<td>s</td>
<td>t</td>
<td>u</td>
<td>v</td>
</tr>
</tbody>
</table>
| '17x| x   | y   | z   | {   | }   | (   | hyph.char |)

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 (%0) or per million (%00) 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.