Typography

Typographers' Inn

Peter Flynn

1   Return to the fold

I'm delighted that the plan to incorporate parts
of the former *TEX and TUG News* has now come
to fruition, and credit is due to the editors for
their hard work on this. I know that when it was
first suggested there were objections from several
people that it would in some way dilute the TEXnical
excellence of *TUGboat* to include articles about
anything except internal TEXnicalities, but I am
happy to see from recent issues that this has not
been the case.

On the contrary, I think *TUGboat* has ben-
efited from the cross-fertilization that the change
has involved. Given that computer users have easy
access to other (lesser!) typesetting systems, I be-
lieve it is important that users, companies, libraries,
universities, and the printing and publishing indus-
tries see that the TEX community is continuing to
research and publish better ways of doing the job, in
all fields, not just the mathematical and scientific.

2   H&J

No, it's not a fetish like S&M (or even a genre
like S&F), but the compositor's abbreviation for
hyphenation and justification. Those who have had
to produce publication-quality typesetting will be
aware of the problems raised by setting in short mea-
sures like narrow newspaper or magazine columns,
especially when you have to include lengthy names
from a computing environment which include a
'path' (the hierarchy of folder names which pro-
vides the route by which you can locate a file),
for example: `Hard Disk:Clients:Acme Cookies
Inc:Projects:Bakery:Progress Report January
1998`.

While names with embedded spaces can be bro-
ken at a space, it may not always be clear to the
reader that there was originally a space at what has
become a line-break. Many publications still seem
to use typesetting systems which fail grotesquely to
break at sensible places, especially in names with
no embedded spaces, such as URLs. Achieving a
smooth and even finish is admittedly difficult in
texts with many reluctant hyphenations, but the
reason I bring this up again is in answer to those of
you who have written to me asking for ammunition

to use against well-meaning but ignorant publishers who insist on inserting hyphens or breaking URLs *before* punctuation (putting periods *etc* at the beginning of lines!). This looks terrible but most of them seem oblivious of it. If you want to hit them with chapter and verse, therefore, it's the `path` algorithm and macro by Nelson Beebe and Phil Taylor that I have referred to before. It works like `\verb`, it's in `eplain` and there's a `path` package for LaTeX as well, and it also forms part of the `url` package used for *TUGboat*. It makes *un*hyphenated breaks, and `path` allows you to specify the characters (usually punctuation) after which breaking is to be allowed, for example:

`\discretionaries +\:/.$%&;=@_?#|~+`

However, all such sets of characters have some drawbacks, and there are a few hidden traps for the unwary. The rules for `url` are documented in the style file and were chosen for the *TUGboat* production environment.

- For `path`, you may want to restrict the set of characters for each document, as there may be some that you want to exclude from being breakpoints. I use the example above for some of my own work, but your mileage will vary: plus, exclamation mark, and double-quote do occur in email addresses, and you may want angle brackets or curly braces if you need to illustrate markup.

- Bear in mind that folder path-names should never, never hyphenate at a genuine embedded hyphen, such as in `emacs-19.27.tar.gz`, because of the risk of confusing the reader, who may not then be able to work out if the hyphen belongs to the name or if it has been inserted by the typesetting process.

- None of the implementations appears to let you distinguish between *a*) characters which must be treated *verbatim* simply because they are literals which you do not want TeX to interpret on some occasion, and *b*) those characters (perhaps including some of those required *verbatim*) which are being specified as potential breakpoints. There thus appears to be no way to say 'break at a period or a comma or a slash but not at a backslash or percent sign or hash mark (pound sign); but still allow me to include backslashes and percent signs and hash marks as verbatim characters.' If you want to use `\path` for URLs, for example, you must include the tilde, because it occurs in some URLs as a 'personal directory' flag character for usernames, even though it would be a semantic er-

ror to break a URL after such a tilde. Personally I'd also rather not have breaks occurring after the question mark, hash sign, or vertical bar (XML) separating a URL path from its query or locator string, but if you don't include them, the macros will gag when you feed it a URL containing one.

- An unfortunate by-product is that the start of a URL falling close to the end of a line may break somewhere within the hyphen and the two slashes, whereas the ideal place would be after the second slash; or that TeX control sequences may break *after* the backslash!

It would be nice if some byte-level guru could beef up the selectivity of the algorithm to let it cater for string exceptions, and also add some kind of interpretive shield which will let it gobble *all* characters *verbatim* except the sentinel (the character enclosing the definition), but never break after them (but this would then mean no documenting both TeX control sequences and PC folders with the same setting...)

3 Quote unquote

Has anyone any idea where the English use of the curious mirror-quote comes from? English quotation marks are 66s and 99s, as all TeX users learn, but I see more and more books using open-quotes which are lateral mirrors of the 99s, **ccs** as it were, 'like this'. I have hundreds of fonts here, but I can't find any which contain it. It's a perfectly normal quote in some languages, and it has its own Unicode (ISO 10646) code point (`u+201d`, 'single high-reversed-9 quotation mark' or 'single reversed-comma quotation mark') but how it migrated into use in English is a mystery, but there is one of them in the `wsuipa` fonts at `\char'163`.

I first noticed it a long time ago, in *'Rommel?'—'Gunner Who?'*, the second volume of Spike Milligan's WWII autobiography, published in London in 1974 by Michael Joseph and printed by Hollen Street Press in Slough in what looks like Century Schoolbook. These open-quotes are used throughout, except for the 'Rommel?' on the cover and title page, which is in Fraktur, and has the open-quotes the right way round (as do the sans-serif sidebars). It may be too long ago now to find out, but I've seen it dozens of times since, in books and other documents. Typographically it's a nonsense in English, and if this was still the days of filmsetting I'd say someone put the mat in back to front, which ought to be impossible, but as it has persisted into DTP I'd be interested to know who's doing it and why.

4 Encoding variations

Some time ago I bought the MathTime fonts from Y&Y, to do Times math setting. It was a bit of a struggle to install them, as I use UNIX most of the time, and their installation assumes *a*) a PC, and *b*) Adobe Type Manager. I finally did it, though, and I'm going to document it for posterity soon, but it raised some additional questions about the use of PostScript font encodings in TeX.

I use a lot of PostScript fonts from sources such as the Monotype and other CD-ROMs. Most of my work has been in English, using plain TeX, so I've been content with my homebrew `psfonts.tex` macros to handle the occasional accent, and the rare time or two I needed LaTeX, I stuck with the OT1 ('old TeX') default encoding. (For those of you not using PostScript fonts, 're-encoding' is needed to reconcile the fact that Adobe uses a different font layout from the TeX one.) Having to load the Y&Y fonts for a LaTeX job, however, meant choosing a more comprehensive solution. Their preference is to use their own 'TeX'n'ANSI' encoding, and as I had no personal axe to grind, I picked this in order to minimize any installation problems.

I've also been re-encoding a few other PostScript fonts. I'm happy to report that it's quite painless, and I've only hit one or two small kinks (my æ vanished for some reason). So much so that I've revised my script which copies a font off a CD-ROM, passes a dead chicken over the keyboard, and mutters the incantations about `afm2tfm` and `vptovf` necessary to make it TeXable. This is now at `ftp://ftp.ucc.ie/pub/tex/mkvf` (with a PC version, `cdvf.bat`, in the same directory). If you haven't started using PostScript fonts, it's well worth it, especially as the CM fonts are now available in that format, saving you acres of disk space in bitmaps. Now that I think I've grasped the way you specify additional fonts with PSNFSS, it has freed me from the use of CM in LaTeX (and there was much rejoicing!).

5 TeX and TeXability

I have long resisted using LaTeX simply because of the pain of having to undo almost everything it does in order to make it work the way that real-world publishing needs. Plain TeX with Karl Berry's eplain provides many of the necessary textual features (cross-referencing, indexing, citation, *etc*) in a more usable manner without the heavily idiosyncratic default layout styles of LaTeX. However, the range of packages now available for LaTeX 2ε, and my re-reading the LaTeX 2ε version of Lamport's *LaTeX: A Documentation Preparation System* as well as

The LaTeX Companion over the summer, persuaded me to try one more time to come to terms with it, and it has not been unsuccessful.

The *Companion* is crucial: it's the only reliable source of explanatory information in one volume about all the major packages as well as much of the guts of LaTeX 2ε. A copy of the new edition of Lamport's book is also required, as it covers some of the extremely basic things I didn't know, some of which are not covered in the *Companion*. In the past I have found that even for relatively short jobs like articles for a journal, you can need a surprisingly large number of packages, and for a book you may need a dozen or more. What did please me was that most of the packages I have used so far are pretty much bug-free, well-documented, and perform as advertised, even in combination. Well done, the authors.

The stumbling-blocks are a mixed bunch. LaTeX 2ε has roughly the same default behaviour and appearance as LaTeX 2.09, so anything even mildly different needs fairly verbose recoding. I only came across four serious annoyances (and I've no doubt that I will be severely treated for airing them if I've missed something):

- As any reader of `comp.text.tex` knows, it's the simplest changes which are the most difficult to find out about for the beginner: try looking up how to shift the left margin inwards by the value of `\parindent` for the duration of a verbatim environment. It's probably simple, but it doesn't seem to be obvious. Hanging footnote indentation is another quite common requirement: it *is* in the *Companion*, but as with Lamport's book you have to hunt for it (p.73, called a 'complicated variant'!).

- Documentation of float control is much improved in LaTeX 2ε, but it's still relatively opaque: it's sometimes hard to understand *why* your figure has wandered off by three pages (`\begin{tracing}[floats]...\end{tracing}`?). One problem is still that a figure has to fit on a single page, max. Long examples of source code, even set `\tiny`, can run to several pages, but for some reason LaTeX does not seem able to break them onto multiple pages.

- Contextually-sensitive hyperlinks are not a luxury. When I'm writing, I really don't want to bother having to remember if the phrase I want to refer to is in a chapter, or a section, or a subsection, or a figure, or a table, or a sidebar, or an example, or whatever. I know the label I gave it, and I expect a cross-reference system to know the type of object for me, and

to insert the relevant phrase or symbol accordingly; and if I move the target from a figure to an example, I expect the point of reference to reflect this. In other words, \ref{foo} must produce 'Figure 7.4' when \label{foo} occurs inside a figure, and 'section 3.5' or 'chapter 9' or whatever when it is located in a section or a chapter. In a sidebar, I would expect it to retrieve the title and page number: 'see the box "Cattle rustling" on page 173'. In complex technical or analytic texts this can save days of work: the facility is built into eplain. LaTeX already knows the environment for any \label, so it should be possible.

- Lastly, there must still be a way to get better parameterisation of the typographic dimensions that standard compositors' specifications require. Cutting and pasting large chunks of code (whole macros) just to change a value from 3pt to 4pt is not the way to go. The control over formatting changes such as the appearance of sectional headings is already excellent, and many packages extend this level of control, but more are needed.

The work done by the *Companion* is magnificent, but there is still a tendency for some other documentation to concentrate on the 'borderline cases and special parameters' rather than the daily necessities: as I said earlier, some of the most elementary requirements are still entirely non-obvious, even to someone familiar with TeX. Anyway, I'm almost a convert. I'm going to use LaTeX 2$_\varepsilon$ for my forthcoming book on SGML (plug, plug) and we'll see how it copes.

6   Humanities citations

The one big thing I did have to hack together just before Yuletide was a new citation style. I was setting a thesis for a colleague in the Humanities, and this is still a neglected area for TeX and LaTeX. The edmac system does a great job for the annotation of *varia* and the *apparatus criticus*, but there are still just as many citation styles as in the natural sciences. One of the most common is the footnote, and of the two packages available, camel and footbib, the first doesn't yet do what I needed, and the second tries to do too much and requires the use of additional control sequences: I wanted something transparent. The style is so prevalent that historians actually use the term 'footnote' to mean 'citation' (I haven't yet found out what term they use when they want to talk about 'normal' footnotes.[1]) In this

style, you would refer for example to the \nutshape control sequence,[2] as well as giving the full citation in the list of bibliographic references at the end.

The requirement, therefore, was to let the author type \cite{foo} in the normal way, and have the bibliographic style retrieve the short citation in author-title-year format to be put into a standard footnote with superior numbering, rather than put author and year or a reference number in brackets. But that wasn't all: if the immediately following citation is to the same work by the same author, then all the footnote should say is [3]. And if the same work was cited again later in the document,[4] the footnote should just say [5].

In both cases you also need to be able to add an optional comment or page number, but this is adequately handled by \cite[options]{foo} (although putting square brackets themselves in the optional text caused some interesting error messages). I used the chicago style, renamed it human and hacked the .bst file to output the new citation instead of the traditional author and year. This last bit was the worst: the documentation on BibTeX styles is thin, most styles themselves are uncommented, and the BibTeX language itself, while ingenious and efficient, is obtuse and poorly documented. However, the fix worked, to the accompaniment of reams of BibTeX error messages about the stack, which I don't grok,[6] but which don't seem to affect the result. Once I've hammered out the error messages I'll make it available and perhaps some reverse Polish gureaux can find my errors in time for TUG's Polish sojourn.

⋄ Peter Flynn
Computer Centre,
University College Cork,
Ireland
pflynn@imbolc.ucc.ie
URL: http://imbolc.ucc.ie/
   ~pflynn/

---

[1] Like this one

[2] Goossens M, Mittelbach F, and Samarin A. *The LaTeX Companion*, 1994, p.56

[3] *ibid.*

[4] But with an intervening reference or footnote about something else.

[5] Goossens M *et al*, 1994, *op. cit.*

[6] I mean I know about stacks, but I can't work out why what I've done should affect it.