



ucc

University College Cork Computer Centre
Electronic Publishing Unit
Research Document Server



Sorry, Professor, the dog ate my thesis

How to expect the unexpected when using L^AT_EX

Peter Flynn

Living and Working with L^AT_EX
London Mathematical Society
20 October 2006

Abstract

The last thing you want when submitting a major thesis is for unexpected problems to crop up at the last minute. Even if they do, you need to have a plan for dealing with them which doesn't involve exporting the text into a wordprocessor to finish the job.

This paper describes some of the most commonly-underused techniques for providing the safety-net which can rescue success from the jaws of defeat. It is based on the author's experience of supporting L^AT_EX in an academic environment as well as in commercial publishing.

Contents

1	A little bit of background	2
2	Trips and traps	3
2.1	The FAQ	3
2.2	CTAN	3
2.3	Usenet News	4
2.4	'My equations won't...'	4
2.5	Get equipped!	4
2.6	Bibliography	5
2.7	Keep a copy online	6
2.8	Learn to write simple macros	6
2.9	The best form of attack is defence	7
2.10	Backup	8
3	Expectans expectavi	9

This document is available online at <http://epu.ucc.ie/articles/lms06>

1 A little bit of background

T_EX — the language — was made available by a computer scientist and mathematician principally for other computer scientists — and for those mathematicians willing to learn some computing using a dumb terminal. At the time (1978–1982) no other interface was immediately envisaged, either logical (syntax) or physical (editor). Knuth has on several occasions expressed moderate surprise (and pleasure) that T_EX has been used so widely, especially by people well outside the original implicit user community (for example in the Humanities, in schools, and domestically).

The original T_EX language is no longer required knowledge for any user writing a conventional document, and is now of interest only to the developer, the computer scientist, and the typographic programmer. Systems such as L^AT_EX and ConT_EXt provide a logic-based interface designed to relieve the normal user engaged on conventional document work from the need to know or understand anything of the internal workings of T_EX. Even in mathematics, which is closest to the core of T_EX's original designs, the need to know and understand the syntax and operation of the language *as applied to mathematics* no longer implies a need to understand everything else about T_EX.

In the physical interface — editors — progress has been less successful. *Textures* (Apple Mac), *Scientific Word* (Microsoft Windows), and LyX (all platforms) have amply demonstrated that a synchronous typographic interface to a L^AT_EX document is possible; and systems like the T_EX daemon¹ (Fine, 2001) and *preview-latex*² (Kastrup, 2002) provide an alternative near-synchronous display capability. However, none of these systems address anything remotely like the interface required by the new (non-programmer) user, where *all* trace of the markup and processes is invisible, and the interaction design meets the expectations of the user. It is perhaps instructive to note that in the parallel field of XML editors, despite the extensive investment capital available, a similar position still obtains (Flynn, 2006).

For the postgraduate student using L^AT_EX to prepare a thesis, there is thus the advantage of not needing an in-depth knowledge of plain T_EX, and the disadvantage that a usable markup-free synchronous typographical editor does not exist. For the mathematical thesis author, whether in mathematics *per se* or in another field making use of mathematics, the suitability or otherwise of graphical interfaces to formulae and expressions outside the T_EX field has become a matter of choice. Until recently, mathematicians absorbed T_EX syntax from an early age, but the prevailing hegemony of the wordprocessor now poses a significant threat to T_EX's survival, despite the obvious disadvantages of the wordprocessor paradigm.

Apart from the interface, however, there are numerous other hidden traps awaiting the thesis author using L^AT_EX, and the remainder of this paper deals with some of the most common in this author's experience.

¹<http://www.pytex.org/texd/>

²<http://www.gnu.org/software/auctex/preview-latex.html>

2 Trips and traps

This collection is based on the author's experiences of supporting $\text{T}_{\text{E}}\text{X}$ and $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ among postgraduate authors and academic staff (faculty). It is wholly unscientific, but it is factual, not anecdotal. In recommending $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ to this community the author has found that conscientious users can save about two months formatting time off the average 3–4 year PhD writing cycle, in addition to having a more reliable, better-looking, and more long-lasting *opus*.

2.1 The FAQ

Find the $\text{T}_{\text{E}}\text{X}$ FAQ³ and bookmark it. This is your first port of call for questions about 'How do I...?' and it enables you to solve most of the common queries that users come up with.

Only look for additional sources of information once you're certain your question isn't answered here. The Internet is full of friendly and helpful people, but they are volunteers, and it's courteous not to waste their time by asking questions that are answered elsewhere.

2.2 CTAN

The Comprehensive $\text{T}_{\text{E}}\text{X}$ Archive Network is a collection of servers around the world hosting all the free software relating to $\text{T}_{\text{E}}\text{X}$ systems. It's rooted at the $\text{T}_{\text{E}}\text{X}$ Users Group⁴ (TUG), DANTE⁵ (the German $\text{T}_{\text{E}}\text{X}$ Users Group), and UKTUG⁶ (the UK $\text{T}_{\text{E}}\text{X}$ Users Group).

From here you can download everything from complete systems (ISO CD and DVD images for installation) right down to the smallest component of an individual style file (package) or font.

Packages are the plug-in components of $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$. There are hundreds of them, providing refinements, formatting solutions, and fonts. CTAN has an index of them you can search.

Most institutional thesis layouts come with a specific set of packages designed to provide the features the institution mandates, so you should stick to these unless you have permission to diverge.

For other purposes you should learn how to download and install packages unless your $\text{T}_{\text{E}}\text{X}$ installation provides automated download-and-install-on-demand (Mik $\text{T}_{\text{E}}\text{X}$, for example).

³<http://www.tex.ac.uk/faq>

⁴<http://tug.ctan.org>

⁵<http://www.dante.de/software/ctan>

⁶<http://www.tex.ac.uk>

2.3 Usenet News

If you're not already connected, go and learn how to read Usenet News. You'll need a newsreader (a Usenet News reader, not a blog/RSS newsreader) — see the Newsreaders site⁷ for lists — but at a pinch you can use a good mailer with built-in newsreading capability like *Thunderbird*⁸, or a web interface like *Google Groups*⁹, or even Microsoft *Outlook*.

The group for discussions of \TeX is `comp.text.tex`, but you don't have to plough through all 100+ daily posts: just get the hang of how to ask questions and comport yourself — see one of the many guides to 'Netiquette' such as Griffith's¹⁰ (Barry *et al.*, 2000).

2.4 'My equations won't...'

...line up, number themselves, use the right font, fit the width, etc. The finer points of mathematical typesetting are addressed in numerous documents, both online and in print.

If you are using maths heavily, you need to have read at least some of them: there is a short list¹¹ in Flynn (2002, p. ix).

2.5 Get equipped!

Despite my rant about editors, there are plenty of usable interfaces around. I'm not going to mention any of them by name, except that *AUCTEX*¹² appears to be the most favoured by the professional mathematician.

Don't even *think* of using a non- \TeX editor like *Notepad* unless you are fond of self-inflicted pain.

The other tools you will almost certainly need are:

A bibliographic manager — See the discussion on \BIBTeX in section 2.6 on p. 5.

A knowledge of \LaTeX is required

A graphics editor — If you're including photographs or other continuous-tone artwork, use *GIMP*¹³, Adobe *PhotoShop* or *PaintShop Pro*, or similar for any preparatory work, and save them as JPG, PNG, or PDF files. You must use

⁷<http://www.newsreaders.com/>

⁸<http://www.getthunderbird.com>

⁹<http://groups.google.com>

¹⁰<http://www.cit.gu.edu.au/guides/netiquette.html>

¹¹<http://research.silmaril.ie/latex/intro.html#A17>

¹²<http://www.gnu.org/software/auctex/>

¹³<http://www.gimp.org/>

pdflatex to process documents using these formats: standard L^AT_EX with *dvips* only works with EPS files, which are unreasonably huge when used with bitmap formats like these.

If you're including diagrams of any kind, make sure they are done as vector drawings (PDF or EPS), *not* bitmaps. EPS only works with standard L^AT_EX using *dvips*; PDF only works with *pdflatex*.¹⁴

A spellchecker and/or thesaurus — *Ispell*¹⁵ and *Aspell*¹⁶ work well, but with limitations, although it doesn't take too long to train them to your specialist vocabulary. Beware the generic dictionary-based products which may be too general.

Thesauruses are all commercial products, and I'm not aware of any which work specifically with T_EX editors.

Another useful tool (*Emacs* only at the moment) is the Remembrance Agent¹⁷, which indexes all your text files (L^AT_EX, HTML, email, etc) and provides a context-sensitive window showing documents related to the topic around the cursor position ('point'). This enables you to check and avoid (or deliberately re-use) phrases used elsewhere.

Learning to spell and punctuate is good, too.

2.6 Bibliography

If you're not already using B_IB_T_EX¹⁸, start now. There is no excuse *at all* for manually formatting every entry and trying to maintain them in that format.

If you can't stomach the thought of typing the B_IB_T_EX syntax, use a graphical interface like *tkbibtex*¹⁹ or *JabRef*²⁰. The latter is preferred, as it will accept a large number of other bibliographic reference formats such as RIS and SilverPlatter, which are in common use on web reference sites and in libraries.

Find out what format your references must be in for submission (if this is not already mandated by an institutional thesis document class) and make sure you have the relevant packages installed.

In the extreme case of your professor requiring an obtuse or degenerate reference format (which is not unknown), ask the very helpful people on `comp.text.tex`.

¹⁴This is a gross simplification, but that's what I'm trying to do: simplify.

¹⁵<http://www.gnu.org/software/ispell/ispell.html>

¹⁶<http://aspell.sourceforge.net/>

¹⁷<http://www.remem.org/>

¹⁸<http://research.silmaril.ie/latex/chapter7.html#bib>

¹⁹<http://www.cat.csiro.au/ict/staff/pic/tkbibtex.html>

²⁰<http://jabref.sourceforge.net/>

2.7 Keep a copy online

As an additional form of backup, copy your thesis-in-development to your web site. If necessary, don't link it, just put it there, if you have concerns about people browsing a partial document.

When you're away at a conference or you meet someone useful or interested, you can then show them or just give them the URI.

Better still, create a page about your research where you can publicise what you're doing — and in the process, effectively stake your claim to be *the* researcher in this field. Some universities *require* all postgraduates to maintain a web site about their research. Others doubtless run a mile.

2.8 Learn to write simple macros

Having said in section 1 on p.2 that you don't need to know any T_EX source code, I'm now going to break the rule.

If your thesis contains very frequently-repeated objects, it will be much easier and more maintainable in the long-term to have a macro which can act as both label and formatter, and be re-used *ad infinitum*.

amo (1)

Person	Singular	Plural
1	amo <i>I love</i>	amamus <i>we love</i>
2	amas <i>you love</i>	amatis <i>ye love</i>
3	amat <i>s/he loves</i>	amant <i>they love</i>

```
\newenvironment{verbal}[1]{%
  \begin{center}\sffamily\noindent\textit{#1}\par\medskip
  \begin{tabular}{c>{\bfseries}l>{\itshape}l>{\bfseries}l>{\itshape}l}
  {\end{tabular}\end{center}}
\newcommand{\dbl}[1]{\multicolumn2l{#1}}
...
\begin{verbal}{amo (1)}
Person&\dbl{Singular}&\dbl{Plural}\\\[6pt]
1&amo&I love&amamus&we love\\
2&amas&you love&amatis&ye love\\
3&amat&s/he loves&amant&they love\\
\end{verbal}
```

2.9 The best form of attack is defence

Your defence. As of submission, *you* are the domain expert. No-one else has researched your particular aspect of the topic in as much depth (or you wouldn't have been allowed to select that topic), so it's your responsibility to put it forward as effectively as possible.

Don't let other things get in the way: as with any publication, the objective is communication, not raising barriers. Your viva is your chance to convince a sceptical extern of your worth, not a chance to show off your flowers of eloquence or your skill with fonts.

2.9.1 Formatting style

Stick to the plain and simple. This ought to be mandated by your institution's thesis standard, if there is one. A thesis is not the place for quirks and cunning stunts of typography — keep that for when you turn it into a book.

Pick a common and readable typeface and stick to it (another reason for avoiding wordprocessors, which have a nasty habit of changing font for no good reason), and don't mix typefaces.

Adhere to the conventions of your discipline, and *be consistent* (the mantra of every publisher since Gutenberg, and one of the principal advantages of L^AT_EX).

2.9.2 Structural style

Most readers can hold three levels of sectional structure in their head while reading. Below that, they lose track of where they are, especially if you have subsubsubsub-sections more than two pages long, so that there is no visible indication of what part of the tree you're in. Keep away from the very finely-nested structures unless absolutely unavoidable.

Avoid nested lists as well: they are part of the structure. Unless mandated otherwise, use a simple decimal notation for sections, and arabic or alphabetic labelling for lists.

2.9.3 Writing style

Eschew obfuscation: while you are the domain expert for the duration, write for the intelligent, informed, and expert reader from a closely-related domain. You're among equals, so don't explain the obvious, but cite anything that would otherwise look like a leap of faith — extraordinary claims require extraordinary proof

Short sentences are good, but don't be afraid to use a long one where there is a complex point to be made. Make sure you write complete sentences, too, without missing important parts of speech.

There is some debate about the merits of writing in the first or the third person ('I designed an experiment. . . ' as opposed to 'An experiment was designed. . . ' or 'We designed an experiment. . . '). Formal usage dictates the third person, but the first person is becoming more widespread.

2.10 Backup

This shouldn't need mentioning, and it is arguably out of scope anyway, as it applies to all work on a computer, whether L^AT_EX or not. If you don't have a recent backup of your entire thesis — document, diagrams, photographs, research data, background information, supporting documents, etc — go and make one immediately.

Any competently-designed editor will automatically make a backup file each time you start modifying a document, by default, silently, and without question. This is part of taking a pragmatic approach to computing. I would submit that if your present editor does not do this (or some equivalent) then you are using the wrong editor, and you should replace it as soon as possible.

In theory the same should apply to editing other types of file, such as images, but the large size of images compared with the size of a text document may preclude this on some systems

Having made sure your system is creating editorial backup copies, you then need to ensure that you make a full duplicate of all your files onto some other medium. Tape is still the reliable standby of all computer centres, but it is expensive for a user's workstation. Unless your institution creates backups for you over the network automatically (rare), you should write your files to CD, DVD, or memory device regularly and frequently, and keep the copies at home, out of your office, lab, or locker.

You may be doing this anyway, if you bring your thesis with you between home, campus, and anywhere else you work. Just make sure that you label the medium with the correct date.

A final word of caution about USB memory devices: these are not disks, and cannot be re-written *ad infinitum* as disks can. Depending on the make, model, and capacity, there is a large but finite limit to the number of rewrites they will tolerate, after which they are read-only devices. Under normal use (copying a few files) this limit will scarcely be reached within the physical lifetime of the device, but when regularly copying very large volumes or very high numbers of files, the limit may be reached within 2–3 years.

3 Expectans expectavi

Writing a thesis takes a long time. Both the research and the writing require infinite patience as well as a capacity for handling the unexpected. If the hints in the foregoing section are to be of use, they need to be accompanied by a core set of computing skills which are not nowadays taught to undergraduates.

The European Computer Driving Licence (ECDL) was supposed to provide a minimal set of computing skills, but for practical reasons it is largely restricted to office procedure and software.²¹ The specific skills (with their ECDL paragraph number) which will serve you well are:

- how to use a good plain-text editor (not a wordprocessor and not a toy like Microsoft *Notepad*); 2.1.3
- where to find all 94 of the printable ASCII characters on your keyboard and what they mean, and how to type accents and symbols if you use them; 3.2.1.2
- how to create, open, save, close, rename, move, and delete files and folders (directories); 2.3
- how to use a Web browser and/or File Transfer Protocol (FTP) program to download and save files from the Internet; 7.3.1.6
- how to uncompress and unwrap (unzip or detar) downloaded files. 2.3.7

This may seem like a trivial list, but the inability to perform simple housekeeping and installation is the cause of more helpdesk calls than any other single category of request.

Armed with knowledge, you should be prepared to handle most of the twists of fate which come your way. The rest is in your capable hands.

²¹In fact most of the additional skills *are* mentioned, but rarely taught with any enthusiasm.

References

- Sarah Barry, Anthony Thyssen and Kim Taylor:** Electronic Mail/News Etiquette. Brisbane, Australia, Jul 2000 – Technical report (URL: <http://www.cit.gu.edu.au/guides/netiquette.html>).
- Jonathan Fine:** Instant Preview and the T_EX daemon. TUGboat, 22 Dec 2001, Nr. 4, 292–298.
- Peter Flynn:** Formatting Information: A beginner’s introduction to typesetting with L^AT_EX. TUGboat, 23 2002, Nr. 2 (URL: <http://research.silmaril.ie/latex>).
- Peter Flynn:** If XML is so easy, how come it1’s so hard?: The usability of editing software for structured documents. In Proc. Extreme Markup Conference. Montréal, QC, Aug 2006 (URL: <http://epu.ucc.ie/articles/extreme06>).
- David Kastrup:** Revisiting WYSIWYG Paradigms for Authoring L^AT_EX. TUGboat, 23 2002, Nr. 1, 57–64.