

German T_EX

Hubert Partl
EDP Center of the
Technical University Vienna

Although T_EX and L^AT_EX have been designed for American standards only, they are being used all over the world and with a lot of different languages. This article is intended to show an example of the problems that arise when modifying T_EX or L^AT_EX for easier application with a language other than English.

One of the great advantages of T_EX and L^AT_EX is the portability of document files among all T_EX installations. In order to prevent users from each inventing their own incompatible modifications, which would destroy that portability, the first step should be to standardize the user interface — i.e. the control sequences and commands to be used in the T_EX input files. Together with this “standard”, a “quick and dirty” or “prototype” solution should be provided, so that users can start to apply the new features. Then, usually in several steps, better and more complete and finally even optimized solutions should be developed in such a way, that the users’ (authors’) input files need not be changed, but only the style files, font files, hyphenation patterns and other files that comprise a T_EX implementation are replaced or improved by the installation’s T_EX guru.

This is the way that has been adopted for “German T_EX”. As Joachim Lammarsch reported in TUGboat No. 8/3, the German T_EX Users Group has agreed on a standard for a “Minimal Subset of German T_EX Commands” at its 6th meeting in Münster (Germany) last October. These commands will make it easier to set German texts — both with Plain T_EX and with the commonly used macro packages like L^AT_EX, *A_MS*-T_EX etc. It is recommended that all T_EX installations in the German speaking countries (Germany, Austria, Switzerland) implement at least these commands on all their mainframes and Personal Computers. Then, all T_EX and L^AT_EX input files that use these commands can be exchanged freely among all participating sites.

1 The User Interface

The German T_EX commands fall into two categories:

- commands that provide additional features which are needed to typeset German texts, e.g. the German „Anführungszeichen“,
- “shorthand” commands that are easier to type than the corresponding original T_EX commands, e.g. "s instead of \ss{ }, or "ck in-

stead of the corresponding \discretionary command.

The standardized “Minimal Subset of German T_EX Commands” consists of the following control sequences and commands:

- "a for the umlaut a (ä, short for \a) — also for the other vowels,
- "s for the sharp s (ß, short for \ss{ }),
- "ck for ‘ck’ that is to be hyphenated as ‘k-k’,
- "ff for ‘ff’ that is to be hyphenated as ‘ff-f’ — also for certain other consonants,
- "‘ or \glqq for German left double quotes and "’ or \grqq for German right double quotes, to produce „deutsche Anführungszeichen“, also known as „Gänsefüßchen“,
- \glq for German left single quotes and \grq for German right single quotes, to produce ‚einfache Anführungszeichen‘,
- "< or \flqq for French left double quotes and "> or \frqq for French right double quotes, to produce «French quotes», also known as «guillemets». These quotes are also used in certain German text styles, sometimes pointing >in< rather than <out>.
- \flq for French left single quotes and \frq for French right single quotes <like this>,
- "|" to disable forbidden ligatures in words that consist of several parts — e.g. to produce ‘Auflage’ (not ‘Auf~~l~~age’) for the word meaning ‘Auf-Lage’,
- "- to mark a hyphenation exception within a long word (like \-, but without disabling automatic hyphenation in the rest of the word),
- "" for an analogous hyphenation exception, where *no* hyphen sign is added in the case of hyphenation (e.g. in a hyphenated word like ‘Eingabe-File’),
- \dq to print the quote character ("),
- \setlanguage{\xxx} to switch to the language ‘xxx’. Arguments to this command are predefined command names like \german, \austrian, \english, \USenglish, \french etc. This command will switch everything that is language specific, e.g. the format of today’s date, and the texts of the captions used with chapters, tables, figures and the like. In a more complete implementation, this should also include language specific hyphenation patterns and exceptions, special fonts or ligatures, different enumeration conventions

and so on. However, the German \TeX commands remain available, regardless of the language specified. This is useful for multi-lingual documents, e.g. an article that is written in English but contains German citations (like this one).

- $\backslash\text{originalTeX}$ to reset everything to its original meaning in \TeX or \LaTeX . This is needed to generate environments that are completely compatible with the rest of the \TeX world.
- $\backslash\text{germanTeX}$ to switch on the German \TeX commands (modifications) again.

The last three commands are usually applied locally for different parts of a multi-lingual document. They have been designed in such a way that they can be easily extended to other languages in the obvious way. The author expresses his hope that other national \TeX Users Groups will adopt similar or perhaps even compatible conventions for their language specific \TeX modifications.

Both the shorthand forms and the original forms of the commands for the Umlaute and for the sharp s will be modified so that automatic hyphenation remains in effect — either for the whole word (with Umlaute and sharp s included in the hyphenation patterns) or at least for the rest of the word, which can be accomplished by using constructs like

```
 $\backslash\text{nobreak}\backslash\text{hskip}\backslash\text{z@skip}$ 
```

which make \TeX “think” that a new word is started after the umlaut. (This trick has been found and reported by Norbert Schwarz from Bochum.)

The standard does *not* include layout conventions. On the contrary, a variety of document layouts is encouraged. As with conventional typesetting methods, all authors, editors, and institutions should be free to chose their individually preferred document styles and should not be forced to an unnatural uniformity.

Tables 1 and 2 show examples for typical applications of the German \TeX commands.

Table 1: Examples

sch"on	produces:	schön
Stra"se	produces:	Straße
"'Ja, bitte!"	produces:	„Ja, bitte!“
"<Merci bien!"	produces:	<Merci bien!>
Dru"cker	produces:	Drucker or Druk-ker
Ro"lladen	produces:	Rolladen or Roll-laden
Auf"lage	produces:	Auflage

Table 2: Date Formats

$\backslash\text{setlanguage}$	$\backslash\text{today}$
$\backslash\text{german}$	31. Januar 1988
$\backslash\text{austrian}$	31. Jänner 1988
$\backslash\text{english}$	31st January 1988
$\backslash\text{USenglish}$	January 31, 1988
$\backslash\text{french}$	31 janvier 1988

With Plain \TeX , the German commands are made available by an input command like

```
 $\backslash\text{input german}$ 
```

With \LaTeX , they are made available by specifying the document style option `german`, e.g. with

```
 $\backslash\text{documentstyle}[11pt,german]\{\text{article}\}$ 
```

In addition, the user should take care that the correct hyphenation patterns for his language are used — usually by specifying the appropriate format file when calling the \TeX program. Among the German \TeX users, the German hyphenation patterns generated by Jost Krieger and Norbert Schwarz at the University of Bochum are the preferred ones.

2 The Present Solution

A “quick and dirty” realization of these German \TeX commands has been compiled by the author with the help of several other \TeX users in Basle, Bonn, Bochum, Darmstadt, Stuttgart, and Vienna. The file, known as `GERMAN.TEX` or `GERMAN.STY`, is public domain. Mainframe installations can obtain it via Electronic Mail from several file servers: ArpaNet users can FTP it from the Rochester \LaTeX Style File Collection, and BitNet users can GET it from `NETSERV AT AEARN` in Linz (Austria) or from `LISTSERV AT DHDURZ1` in Heidelberg (Germany). PC users can obtain it on floppy disk from the German PCTeX distributor.

Besides being quick and dirty, this solution has the advantage that it can be used with the original versions of \TeX and \LaTeX and with the fonts and hyphenation patterns as they are available now. Everything is defined and re-defined using \TeX commands only, and it is just one \TeX input file that can easily be ported to every computer (including Personal Computers) and is independent of the output devices used.

Care has been taken to make the same file usable both with Plain \TeX and with \LaTeX and other macro packages. This has been accomplished by us-

ing Plain $\text{T}_{\text{E}}\text{X}$ commands only, with the only exception of the $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ command `\protect` which is defined to `\relax` within this file if it has not been defined before.

The umlaut accent is redefined such that with the letters A, a, O, o, U, and u, the dots are positioned a bit lower than in the original version, and that the commands `\nobreak` and `\hskip` are added to enable automatic hyphenation in the rest of the word, as mentioned above.

For the sharp s (β), the command `\lccode` is used to enable automatic hyphenation in words containing this letter.

The German left double quotes („) are formed by taking the English right double quotes (") and lowering them by the height difference between quotes and comma, with some extra kerning. The German right double quotes (") are the same as the English left double quotes except for the kerning. The German single quotes are formed in a similar way.

For the French quotes, the appropriate math symbols are used.

The quotes character (") is made an active character and is defined as a control sequence that takes the following character as its parameter and, depending on the value of this character, does the appropriate actions, i.e. it prints the corresponding umlaut or sharp s or quotes character, or it performs the required combination of `\discretionary`, `\nobreak`, and `\hskip` commands.

The quotes character is added to the `\dospecials` command which is used in the `verbatim` environments.

The different versions of today's date are obtained by re-definitions of the `\today` command in analogy to the original definition by Leslie Lamport.

The different versions of the chapter and table titles are obtained in the following way: The language changing commands re-define command names like `\contentsname` to contain the appropriate texts (e.g. 'Inhalt' for German texts and 'Contents' for English texts). With Plain $\text{T}_{\text{E}}\text{X}$ or other macro packages, this will have the desired effect only if these command names are actually used to print the respective title lines. With $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$, it means that the original document style files have to be modified in the following way: The hard coded English words (like 'Contents') have to be replaced by the corresponding command names (e.g. `\contentsname`), and these command names have to be defined to contain the original words, e.g. with

```
\def\contentsname{Contents}
```

Leslie Lamport's comments in the DOC-Files provide help in finding all places where such modifications are necessary. There has been some discussion recently, whether these modified style files should be available from one official central "clearinghouse".

The three language switching commands are defined to switch on and off all the appropriate modifications. Finally, the command `\germanTeX` is executed, which switches on everything that is appropriate for typesetting German texts.

3 Future Work

For the future, a better realization of the German $\text{T}_{\text{E}}\text{X}$ commands is planned by a team of advanced $\text{T}_{\text{E}}\text{X}$ experts in Germany. This solution will include the following features:

- The Umlaute and special quotes will be designed with METAFONT as separate characters in the text fonts, and they will be accessed as ligatures.
- New hyphenation patterns will be generated that include the unlauded characters, the sharp s, the special ck, and the special double consonants that hyphenate as triple consonants.
- The multi-lingual $\text{T}_{\text{E}}\text{X}$ software will be used to enable the switching of hyphenation patterns for the different languages.

Due to the complexity of this project, it will take some time until this solution becomes available for all $\text{T}_{\text{E}}\text{X}$ installations (i.e. all computer types and all fonts for all output devices). However, the user interface (i.e. the $\text{T}_{\text{E}}\text{X}$ commands described above) will remain unchanged with this new solution. Therefore, users who start using them now will not have to change their $\text{T}_{\text{E}}\text{X}$ input files later, and they will still be able to exchange their $\text{T}_{\text{E}}\text{X}$ files with all installations where either the present or the future version of the German $\text{T}_{\text{E}}\text{X}$ commands is installed.