



HAL
open science

Tralics, a LaTeX to XML translator; Part I

José Grimm

► **To cite this version:**

| José Grimm. Tralics, a LaTeX to XML translator; Part I. [Technical Report] 2005. inria-00000198v1

HAL Id: inria-00000198

<https://inria.hal.science/inria-00000198v1>

Submitted on 7 Sep 2005 (v1), last revised 25 Apr 2008 (v4)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Tralics, a \LaTeX to XML translator
Part I

José Grimm

N° 309

September 2005

Thème NUM



*rapport
technique*

Tralics, a \LaTeX to XML translator

Part I

José Grimm*

Thème NUM — Systèmes numériques
Projet Apics

Rapport technique n° 309 — September 2005 — 232 pages

Abstract: In this paper we describe *Tralics*, a \LaTeX to XML translator. A previous version of the software (written in Perl) was used to obtain the Pdf version of Inria’s “Rapport d’Activité” for year 2001. The current version of the software (written in C++) is used for both the HTML and Pdf version for the year 2002: the XML generated by *Tralics* is conforming to a local DTD, similar to the TEI; it was converted to Pdf via pdf\TeX and the `xmltex` package, and the HTML via a `xslt` processor.

We explain here the philosophy of the software, its usage, its limitations, and how to customize it. All commands defined by *Tralics* are explained in this document, for most of them, we give an example of use. The index lists all commmands, environments, and options.

Key-words: Latex, XML, HTML, MathML, Perl, PostScript, Pdf

* Email: Jose.Grimm@sophia.inria.fr

Tralics, un traducteur de \LaTeX vers XML

Partie I

Résumé : Dans cet article nous décrivons le logiciel Tralics, un traducteur de \LaTeX vers XML. Une version antérieure de ce logiciel, écrite en Perl, a été utilisée pour générer la version Pdf du Rapport d'activité de l'Inria en 2001. La version actuelle du logiciel, écrite en C++, a été utilisée pour obtenir à la fois le HTML et le Pdf de la version 2002 : nous avons utilisé une DTD locale, similaire à la TEI, et pdf\LaTeX plus $\text{xm\textit{t}ex}$ pour obtenir le Pdf.

Nous expliquons ici la philosophie de Tralics, son usage, ses limitations, et comment paramétrer le logiciel. Toutes les commandes définies par Tralics, sont expliquées, pour la plupart d'entre elles on donne un exemple d'utilisation. Un index regroupe l'ensemble des commandes, environnements et options.

Mots-clés : Latex, XML, HTML, MathML, Perl, PostScript, Pdf

Chapter 1

Introduction

The Tralics software was designed as a tool for the Raweb. In this chapter, we explain some of our motivations. In the next chapter, we study some T_EX commands and explain how they are handled in the same fashion by Tralics. Following chapters explain some differences, merely because XML is not dvi. In a final chapter, we explain how to configurate Tralics. There is a second part, that explains how the XML files can be used, converted into Pdf or HTML; it describes also the Raweb DTD.

1.1 A short history of the Raweb

A short history of the Raweb may be found on the Inria internal web site¹. The question concerns Inria’s Annual Activity Report, also known as “Rapport d’activité”, or “Annexe technique” to the RA or “annexes scientifiques” to the RA. This is a document, written by the research teams, at the end of the year N (October, November), and published in March of year $N + 1$.

Until the 1993 edition (published in 1994), only a paper version existed. A L^AT_EX model was used since 1987, designed by Jacques André then Martin Jourdan. See the reference [5], by Louarn in the first Cahiers Gutenberg.

In 1993, contacts were made with the Grif S.A. society, for the design of a SGML DTD and a L^AT_EX-to-SGML converter². As a result, Philippe Louarn was able to put on the web the RA (year 1994) in its HTML version³. But this converter was judged too complicated (rules were too strict) and for several years, the HTML was directly produced from the L^AT_EX source, using `latex2html`.

In 1996, a working group (conducted by Albert Benveniste) gave new specifications: independent modules, grouped into ten sections, etc. A technical group was created (conducted by Gérard Paget), whose objective was to find a company that could sell a software (maybe using XML as intermediate language). None was found, but the design of modules (in L^AT_EX syntax) was well-defined by Laurent Pierron and José Grimm with the aid of Marie-Pierre Durolet and Jean-Claude Le Moal. For the Ra98, a Perl script did some preprocessing, splitting the L^AT_EX source into modules (one module per HTML page). The author wishes to thank all these people (including A. Quadrant), who gave him the idea to work on L^AT_EX and write a translator.

In 1999, the Scientific Annexes to Inria’s Annual Report were renamed RAWEB, to emphasize the role played by the Web (it is available as a CD-Rom, but no more printed by Inria).

¹See <http://www.inria.fr/interne/disc/apropos/chantiers/raweb-xml/histoire.html>

²Grif stands for ‘GRenoble Interactive Formatter’; further developments of this editor by the Opera team led to the Amaya software.

³See <http://www.inria.fr/rapportsactivite/RA94/RA94.html>

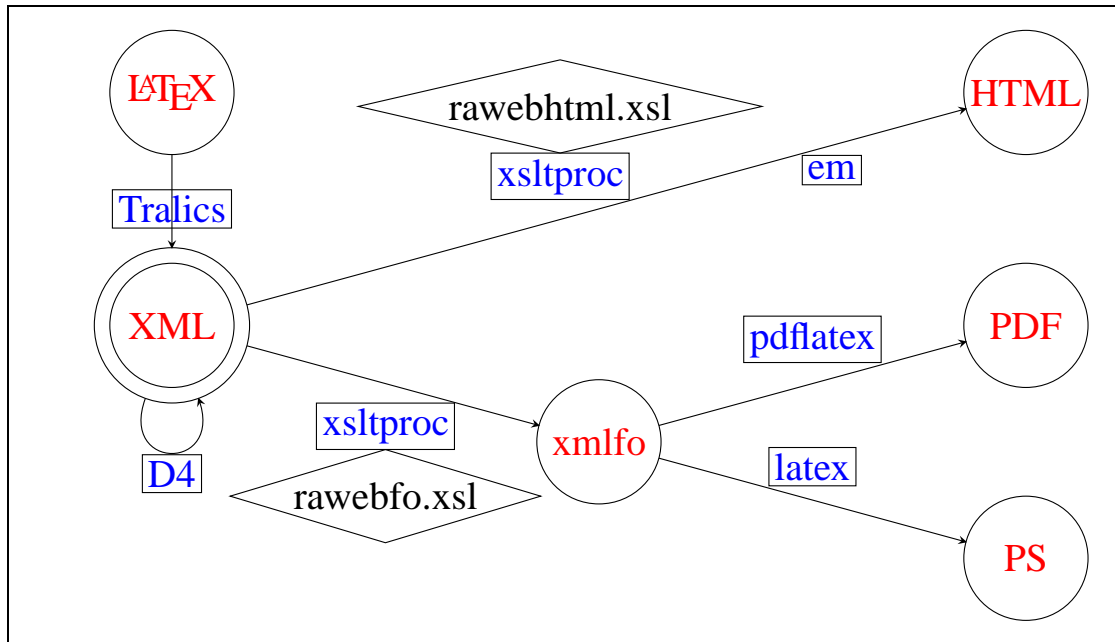


Figure 1.1: A diagram that explains how the Raweb operates. Rectangular boxes contain tools, diamond-shape boxes are style sheets, and circles contain language names. The name XML is in a double circle, it is the central object; the arrow labeled ‘D4’ that connects it to itself indicates conversion from one DTD to the other, used in 2004. The box containing ‘em’ represents the Perl script `extract-math.pl` that handles the math formulas; it uses tools borrowed from `latex2html`. The diagram is written in ‘pgf’, a format that `Tralics` cannot interpret yet.

In 2001, the Perl scripts mentioned above evolved into a $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ -to-XML converter (some ideas were borrowed from `latex2html`, which is also a Perl script). The main trouble was conversion from XML to Pdf, and we used tools from the $\text{T}_{\text{E}}\text{X}$ community (by S. Rahtz and D. Carlisle [1]) and `pdfLATEX`. On the other hand, the images in the HTML files were converted by `latex`. Additional software (for creation on a global index, etc.) was written/used by Marie-Pierre Durolet.

This gave a complicated object: a Perl script, that converts a $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ file into XML (using Omega as a subprocess for the math) followed by another Perl script that extracts the math, calls `latex`, then `dvips`, then `pstoimg` (a Perl script that calls `ppmquant`, `pnmcrop`, and so forth, whose job is to produce a `png` image for the math formula), and finally an XSLT processor for the effective conversion. This became even more complicated in 2004, where a new DTD was introduced (designed by Bruno Marmol and people mentioned above), hence a XML-to-XML translator. See Figure 1.1.

1.2 Birth of Tralics

The big Perl script was rewritten as a C++ translator, renamed `Tralics`, and got (for version 1.6) a first IDD^N number in December 2002. This software was still able to produce a $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ preview of the Raweb: The source is read, the syntax is tested, a $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ file is written for each module, `latex` and `bibtex` are called, the resulting dvi is converted to PostScript. But the same `Tralics` can be used in a different way: the source is read, the syntax is tested, an XML file is created, an

⁴InterDepositDigitalNumber = IDD^N.FR.001.510030.000.S.P.2002.000.31235

XSLT processor is called to generate the XSL/Format, and `pdflatex` generates the Pdf (you can also generate an HTML version).

Since 2003, there are a few people writing their RA directly in XML. As a consequence, the new `Tralics`, that is used in 2004, does only the bare minimum: it converts the L^AT_EX code into XML code. There is a Perl script that does everything else (calling external programs like `latex`, `xsltproc`, etc). A non-obvious point concerns the math and the images. For the math, see above; for the images, a Perl script (the same as above) is used for conversion from PostScript to `png`. In some cases, the image already exists in `png` format and it is unwise to re-create it.

1.3 Main objectives

The main objectives of `Tralics` are described in [3]. Recall that we want a program that converts from one language to another, such that

- the result is easily analyzable,
- the structure is preserved.
- the translation is univocal.

By the very choice of the XML language as target, the first objective is automatically satisfied. The only little problem concerns spaces: in which cases are spaces used as delimiter or as text? The translation of `\par_x{x}\par` is `<p>x</p>`, and this is often interpreted in the same fashion as a single space.

The third item has the following meaning: the translation of `\foo` and `\bar` should be different in case these objects have different meanings in L^AT_EX, but should be the same otherwise. One could argue that some tokens could be translated more than once in a different context (remember the `\texorpdfstring` command). We could imagine special rules for the bibliography; in fact, in math mode you must use a different function to add attributes to an element. On the other hand, we have an application where `\tt` and `\bf` are treated alike: in fact, all font information is removed, by redefining element names.

The important point is preservation of the structure. This means of course that the initial document has some structure and the target has one also. Clearly the structure of an XML document is given by its DTD, but this is completely ignored by `Tralics`. The first version of the program was very strict; nowadays, in your document, you can omit the `\documentclass`, the `\begin{document}`, or the `\end{document}`.⁵

There are some implicit rules, for instance that `\section` terminates a `\subsection`, the occurrence of a character in vertical mode triggers the start of a `<p>` element, etc.

In [3], we wrote that the document should satisfy three kinds of assumptions: some technical assumptions (like: there is no `\0` in the document), some validity assumptions (the document can be compiled by L^AT_EX without errors, it respects the rules of the Raweb as given on the web pages), and more general, unwritten, rules. The current translator is more flexible: there is no restriction on the syntax of commands, except that null characters are not allowed, plain T_EX documents are translated (for instance the file `xii.tex` by D. Carlisle that starts with `\let~\catcode~'76` and ends with `Yer.W.:jbye`), and we have an application where only one part of the document is translated: only code before `\maketitle` is used. In this document we shall explain these things.

⁵Put 'Hello, word!' in a tex file, compile, and admire the result!

1.4 An example

Let's consider the following piece of code.

```
\left [1=2\right\]$ est une \emph{formule de mathématiques}~!
```

You would expect this to be understood by \LaTeX as $[1 = 2]$ est une *formule de mathématiques* ! but the result is ‘Hello, word!’, since we have preceded it by this (curious) list of commands

```
\catcode'\$=\active\def$#1~{\catcode'\$=3 Hello, word} %$emacs
```

In \TeX , translation depends on a great number of tables, that assign values to numeric or symbolic quantities. For instance, each character (a number between 0 and 255) in each font (for instance \OT1/cmr/m/it/10) is associated to its dimensions, ligature informations, etc. Our example (with a normal dollar sign) uses three fonts: a math font, an italic font and an upright font. Often, these tables are read only by \TeX , in some other cases a user might look at them, and sometimes they are designed to be modified. For instance, the slant value of the current font is in general ignored (except by commands like \emph). A very important table holds the meaning of commands: they may be predefined (like \left), defined in a format (like \emph), or defined by the user (there are a great number of mathematicians that define a command named \RR for the set of real numbers).

An important table is the table of category codes. Each character has a category code, an integer between 0 and 15. For instance, ‘e’ has category letter, ‘é’ has category active (so that it is equivalent to \'e), backslash, open brace, close brace and dollar sign are of category 0, 1, 2, and 3 respectively. A command like \left is formed by an introducer (any character of category 0, its value is irrelevant) and a sequence of letters (in the example, the space that follows is not part of the command, since its category code is 5; this space will be read again, and ignored, unless its category code changes; such a change can be triggered by the evaluation of the command); while a command like \ or $\text{\$}$ is formed of an introducer and a single non-letter character (when followed by spaces, these spaces do not disappear). Writing a \TeX scanner is easy, the only difficulty is that category codes can change (for instance, in verbatim mode). In the example, the tokens are the following: $\$$ ₃ \left ₁ 1 ₂ $=$ ₁₂ \right ₁ $\{$ ₁, etc, $\$$ ₁₁ $\}$ ₂ \sim ₁₃ $!$ ₁₂. Note that a closing brace, with default category code is shown as $\}$ ₂, while a command whose name is formed of a closing brace is shown as $\}$ ₁. If you say \let\foo\bar , and ask \Tralics for the meaning of \foo , as in \show\foo , it knows the name ‘foo’ of the token and its value (this is a command code) and gets the name ‘bar’ from a lookup table, and adds the current value of the escape character in front. If this is the plus sign, you will see +foo=+bar .

Parsing an expression means finding for each command its arguments. In the example, the two active characters é and ~ take no argument. The \active command takes no argument either. The \emph command takes one argument⁶: a token, or a token list delimited by braces (i.e. characters with category codes 1 and 2), these braces are not part of the argument. The two commands \left and \right are special: they want a delimiter, in reality a pointer into a special slot in the current math font; the argument cannot be delimited by braces. The syntax of \catcode is more complicated. In fact, it is an instance of $\langle\text{codename}\rangle$ ⁷, and $\langle\text{codename}\rangle\langle 8\text{-bit number}\rangle$ is something that can follow \the or be used as $\langle\text{internal integer}\rangle$, while a $\langle\text{code assignment}\rangle$ is $\langle\text{codename}\rangle\langle 8\text{-bit number}\rangle\langle\text{equals}\rangle\langle\text{number}\rangle$. We shall explain this in details later; the idea is that, depending on the context, the \catcode command returns a value stored in a table or modifies it. In the case of

```
\catcode'\$=\active
```

the $\langle 8\text{-bit number}\rangle$ is the internal internal code of the dollar sign, expressed in the form ‘ $\text{\$}$ ’,

⁶Since \emph is robust, it takes no argument, but the non-robust command associated to it reads the argument. In any case an argument is read.

⁷We use angle brackets for the formal syntax of \TeX as explained in the \TeX book, [4].

and the \langle number \rangle is another character code⁸. Expanding a command means (roughly speaking) reading its arguments, and replacing it with the body of the command (where special markers like ‘#1’ have been substituted by the value of the arguments). In some cases, internal tables may be consulted, but they are never modified. Evaluating a command implies modification of some internal state variables (for instance, a character can be added to the current character list, or a complete paragraph split into lines, or a register modified). In this example, the dollar character becomes magically active: a dollar character is no more read as $\$$ ₃ but as $\$$ ₁₃.

The syntax of `\def` and friends is \langle def \rangle \langle control sequence \rangle \langle definition text \rangle where \langle definition text \rangle is \langle parameter text \rangle \langle left brace \rangle \langle balanced text \rangle \langle right brace \rangle . In the case of

```
\def$#1~{\catcode'\$=3 Hello, word}
```

the \langle control sequence \rangle is the dollar sign (the object to be defined, a command or, as in this example, an active character), the \langle parameter text \rangle is everything before the open brace, here ‘#1’, and the \langle balanced text \rangle is everything between the braces. Evaluation of ‘`\def`’ consists in storing in a table the \langle parameter text \rangle and the \langle balanced text \rangle (T_EX stores also a special marker representing the \langle left brace \rangle). What happens now when T_EX sees a dollar sign? since this character is active, the definition given above applies. The \langle parameter text \rangle explains how to read arguments. In this case, ‘#1~’ means that there is one argument, everything up to (but not including) a tilde character. In the case of ‘`\left...$...}`!’⁹, all characters are read (except the exclamation point). They are replaced by the body (no substitution is needed). After that, T_EX sees `\catcode` and evaluates it as before, so that the dollar sign becomes a math shift character again. The space after the digit 3 disappears, and we are left with ‘Hello, word!’⁹. Note: the example is given in French, first in order to show how 8-bit characters can be used, and also because, in English, there is no tilde before an exclamation point. In general, when a macro reads a delimited argument and sees an empty line instead of the delimiter, it signals an error of the form *Runaway argument? Paragraph ended before \$ was complete*.

Translation difficulties. There are different kinds of L^AT_EX to HTML/XML translators. Some, like `gellmu`⁹ use a syntax of their own; others, like `tex4ht` use T_EX as preprocessor, thus understand the full syntax; there are translators like `tth` or `hévée` that use a fixed (and efficient parser), or like `latex2html` (written in Perl) that use pattern matching, and global substitutions instead of sequential evaluation. Neither of these is perfect. We explain in this paper how Tralics deals with a certain number of problems.

1.5 Some remarks on the Translation

In L^AT_EX, commands can be defined in five places:

1. In the Pascal source. In this case, the command is a primitive, for instance ‘`\def`’ or ‘`\left`’. No equivalent might exist in the target language (for instance, you cannot define anything in a XML document, the result of ‘`\left`’ is an attribute of the `<mfenced>` element that results of the translation of the group implied by `\left`. And what about `\dump`?)
2. In the format file. In this case, the command is defined in a file (for instance `latex.ltx`), analyzed by T_EX and stored on disk using a fast retrieval method. For instance, plain T_EX, L^AT_EX, ConT_EXt have their own format file. Both plain T_EX and L^AT_EX define a command named ‘`\item`’, in a different way, for the same purpose (the L^AT_EX command must be used inside a special environment, the plain T_EX one can be used everywhere).
3. In a class file. A class file defines markup commands for a generic purpose (a book, an article, a presentation). For a book, you have a frontmatter, for an article, you can have a title page,

⁸In fact, you should use `\active` only in the scope of a `\catcode`, so you can forget that ‘`\active`’ typesets as ‘!’

⁹Generalized Extensible L^AT_EX-Like Markup

- in both you have sections, etc. The class defines also the current font, together with a lot of dimensions.
4. In a package. The difference between a class and a package was introduced by L^AT_EX 2_ε. Each document uses a single class, and lots of packages (this document uses the ‘report’ class, and the ‘RR’ package for the look; it uses ‘hyperref’ for hyper-links, ‘amsmath’ for the mathematical examples, ‘fancyvrb’ for the verbatim examples, etc.). Note that the plain T_EX format provides a macro `\proclaim` for theorems, while the amsmath package provides a command `\newtheorem` for defining theorems.
 5. In the T_EX source, or files included via `\input`. The T_EX source may redefine commands defined earlier; a package may redefine commands from the class, but should not redefine commands from other packages (i.e., the order in which packages are loaded should have no importance).

If one is to design a translator, the question is: which commands to translate? and how? Our idea is that all T_EX primitives should be understood (the difference between ‘`\dump`’ and ‘`\mydump`’ is that you get either an *Unimplemented* or *Undefined* command error), as well as all standard L^AT_EX commands; of course all user-defined commands are expanded. Concerning classes and packages, our model (the Raweb) looks like a report. We are planning to add support for PhD thesis. We are also interested in a class for slides (like beamer, but this one seems complicated). Other classes are not handled. Some L^AT_EX packages are standard; these are taken into account. There are too many packages for implementing them all.

Let’s assume that `\foo` is a command defined in one of these files. Should it be translated by Tralics? and how? Notice first that a user command is always expandable, and has to be expanded; only the input stream is affected by this operation that consists in reading arguments and replacing them by an instantiated body. On the other hand, evaluation may modify or use internal tables, which can be implemented differently in our translator than in T_EX (for instance, in pdfT_EX, there is a way to re-use an image). Whenever a command is defined in a style file¹⁰, we could use the L^AT_EX source. In some cases, for efficiency reasons, we implemented them in C++ (for instance, the whole fp package has been re-written).

A typical example is the `\it` command. This is a non official L^AT_EX command (see [2], or [6, section 7.3.6]): “it is legitimate for you to redefine them in a package or in the preamble according to your personal taste”. In Tralics, there is no difference between ‘`\it`’ and ‘`\normalfont\itshape`’. What these commands do in L^AT_EX is rather complex: essentially, some variables are set and ‘`\selectfont`’ is called. These commands are robust: if you use ‘`\it`’ in a chapter title, the toc file will contain the name of the command, not the result of the expansion. Since Tralics does not write anything in a toc file, this feature is not implemented (using font changes in a title is not recommended; look at the table of contents for this section).

Consider now a command like `\motscle`. This is defined by the Raweb as an environment (you can use it in Tralics only as an environment) and it expands to something like ‘`{\bf mots clés}`’. We do not want Tralics to use this expansion. One reason can be that the post-processor might prefer something like ‘`\textbf{Mots clés}`’:’ (i.e., use an alternate spelling, use an initial capital, add punctuation, etc). Another reason is that, since 2003, the Raweb is in English, and the translation should be an English word. For this reason, the translation of `\motscles` is `<keywords>`. The raweb class file contains commands that should not be translated: for instance, there is a command `\ra@finpart` whose purpose is to add a period after the last keyword. This should not be done by the translator, but by the program that will typeset the XML result.

Some T_EX primitives are hard to translate. For instance the ‘`\`’ command is assumed to put an acute accent over a character. It is defined (via an indirection through font encoding tables) in

¹⁰Sometimes packages are called style files, because this was the original term, and the extension is `.sty`.

terms of the `\accent` primitive. We could translate `\'` into a Unicode combining character (U+301, to be placed after the character, see [7, paragraph 7.7]). But in general, we have a construction like `\'e` which is defined in `iso-latin1`, this is the Unicode character “latin small letter e with acute”. In the same fashion `\k a` translates to ‘a’, this could produce `ą`, but the actual translation by Tralics is `ą`. In a previous version of Tralics, a construction like `\'\^` was illegal. In the current version, we use a double indirection table. You can typeset the name Hàn Thê Thành of the author of pdfT_EX, the input is `Th\'\{\'e}`, the Tralics output is `Thế`. If you know the Unicode character value, you can enter it (in Tralics as `Th^^^1ebf` or as `Th\char1EBF`).

A construction like `\font\myfont=cmt10 at 13pt` defines a command `\myfont` that can be used as `{\myfont \char217w}`. The effect of this command is to pick up two characters from the font `cmt10`, scaled by some ratio (it depends on the “at” size of the font), and typeset them. Note that the font could specify a ligature, so that the result could be a single character. In any case, the result might look very different from ‘Üw’ (for instance L^AT_EX provides a font containing only lines or circles). In the current version of Tralics, translation does not depend on the current font. The only interest of changing the current font with `\font` is that you can access or modify the internal tables of the font (but no metric file is read at all by Tralics).

1.6 Category codes and characters

A category code is an integer between 0 and 16, as explained in the T_EXbook. The codes are defined according to the following table.

0. A character of category code 0 is used to create a command like `\foo` or `\;`; the associated token is `\foo` or `\;`; the value of the character is irrelevant. By default, there is only one character of category 0, the backslash.
1. A character of category code 1 serves as group delimiter (opening character), as well as delimiter for macros and token lists. It is a `{` as used in the rules explained above. By default, the left brace character `{` is the only one with category 1.
2. A character of category code 2 serves as group delimiter (closing character), as well as delimiter for macros and token lists. It is a `}` as used in the rules explained above. By default, the right brace character `}` is the only one with category 2. The end of verbatim environment is handled by the following piece of code, that shows how to change the category codes of the three characters mentioned above, and use alternate characters for the same purpose.

```

\begingroup \catcode '{=0 \catcode '['= 1
\catcode']=2 \catcode '\{=12 \catcode '\}=12
\catcode'\=12 |gdef|@xverbatim#1\end{verbatim} [#1 |end[verbatim]]
|endgroup

```
3. A character of category 3 serves a math shift character. By default the dollar character `$` is the only character in this category. The same character can be used to start and end a formula. Two consecutive character of category code 3 are used to start or end display math; the characters need not be the same.
4. A character of category code 4 is used as delimiter in arrays; it indicates the end of a cell. A typical row in a L^AT_EX array is `a&b&c\;`; in plain T_EX, it would finish with `\cr` or `\crcr`. The `&` character is the only one of category code 4.
5. A character of category 5 is a end-of-line character. When such a character is seen, all remaining characters on the current line are ignored. After that, the reader behaves as if

- it had seen nothing, a space, or a `\par` token. In some cases, the `\par` token is invalid. In `Tralics`, this space character is special, in that it may print as a new line character, in `TEX`, it is a normal space. By default, only the carriage return character is of category code 5 (this character is inserted at the end of every line, instead of the line feed, carriage return, or both, that marks the end-of-line on the file).
6. A character of category 6 can be used as delimiter in a command, or a table preamble. By default, there is one character of this category, the `#`. A typical array preamble is `'\indent#\hfil&\quad#\hfil\cr'` (this is the first example of `\halign` in the `TEXbook`. We shall not describe `TEX` arrays here).
 7. A character of category code 7 is a superscript character; by default, it is the hat character. Do not confuse it with `\^` that produces an accent. Such a character can be used only in math mode. There is also the double hat construction, explained later in section 5.1: two identical characters of category code 7 can be used to read any 8bit character as in `^^ab`; in `Tralics`, a 16 bit character can be read using 4 such characters.
 8. A character of category code 8 (by default the underscore character) can be used as subscript character; as in the case of superscript characters, you can use it only in math mode.
 9. A character of category code 9 is ignored. `Tralics` ignores no character by default. On the other hand, it cannot put a null character in a string, so that the null character will not appear in the XML output (note: the same holds for `\char0` and `^^00`).
 10. A character of category 10 behaves like a space. By default, space, tabulation, character 160 behave like this (note that the character 160 is non-break space, it should be equivalent to `~`).
 11. A character of category code 11 is a letter. By default, only characters in the range ‘a’ to ‘z’ and ‘A’ to ‘Z’ are of category 11.
 12. A character of category 12 is an ‘Other character’. All characters not listed here are of category 12, including all digits.
 13. A character of category 13 is an active character. Currently, there is only one active character, the `~`. An error is signaled in the case where an active character is used, and undefined. `Tralics` defines `_`, `#` and `&`, to be the same as `_`, `\#` and `\&`.
 14. A character of category code 14, is a comment character: all characters remaining on the current line are discarded. By default `%` is a comment character.
 15. A character of category code 15 is invalid. By default, all characters are valid.
 16. A character of category code 16 is a ‘short verb’ character. This is a feature that does not exist in `TEX`. If you use `\DefineShortVerb` to make it a short verb, you should undefine the character before changing its category code.

If a `TEX` file contains the following lines,

```
\show{ \show} \show$ \show& \show# \show^ \show_ %$
\expandafter\show\space \show a \show 1 %
\def\foo+{}
\foo{ \foo} \foo$ \foo& \foo# \foo^ \foo_ \foo a \foo1
```

then `Tralics` will print:

```
begin-group character {.
end-group character }.
math shift character $.
```

```

alignment tab character &.
macro parameter character #.
superscript character ^.
subscript character _.
blank space .
the letter a.
the character 1.

```

In any case, we have a prefix that depends on the category code, then the value of the character. All calls to the command `\foo` are wrong and signal an error. We show here the first error message, followed by the “got” part of the other error messages. It is important to remember that the command has to be followed by a the right character with the right category code.

```

Error signaled at line 20 of file txt6.tex:
Use of \foo doesn't match its definition;
got {Character { of catcode 1}
expected {Character + of catcode 12}.
got {Character } of catcode 2}
got {Character $ of catcode 3}
got {Character & of catcode 4}
got {Character # of catcode 6}
got {Character ^ of catcode 7}
got {Character _ of catcode 8}
got a
got {Character 1 of catcode 12}

```

1.7 Considerations about mathematics

Translating mathematical formulas is rather difficult: this is because mathematics are complex in both T_EX and MathML. Basically, you start with kernels, add some decoration, and connect these things. Kernels can be letters like x , y , z or A , B , C , but you see very often Greek letters like α , β , Γ , Δ , Hebraic characters like \aleph , old German, like φ . Decoration can be, like in x^y , a second kernel on the left, the right, above, below. People use also bars, dots, rings, arrows, etc. There are different types of connectors: for instance you can say $x = y$ or $x \parallel y$, using parallel bars, or $x \times y$, $x + y$ using crossing bars. Amstex was designed by Spivak for easy typesetting of tensors of the form $T_{j_1 j_2 \dots j_m}^{i_1 i_2 \dots i_n}$.

The first difficulty consists in representing all these symbols. In T_EX, you use commands like `\alpha`, defined as `\mathchar"10B`, whose effect is to chose a character from a font (depending on the packages used; a big difficulty is to chose math fonts that go well with the main text font of the document). Some characters exist in bold version, or in italics version. A non trivial question is how to put everything in bold: you could use commands like `\boldx \boldequals \boldalpha`. It would be simpler to say `\bold{\$x=\alpha$}`. One solution is to say ‘`\boldmath`’¹¹, or equivalently ‘`\mathversion{bold}`’ before starting a math formula; the effect is to use a bold version of everything (in fact, of all characters that use one of the math families, but the number of families is so small that not all symbols use this mechanism). You can say ‘`\mathbf{x}`’ or ‘`\bm{x}`’ in a math formula. Only `\mathbf` is implemented in Tralics, it provides a bold upright font; on the other hand, the `\bm` command is defined in the `bm` package; provided that fonts are available, it should use a bold version of everything (at least for all characters for which a bold typeface exists).

¹¹An example of bold math is when typesetting the L^AT_EX 2_ε logo, see page 20.

Traditionally, uppercase letters were upright, lowercase letters were italics. This is the default for Greek letters, but \TeX uses italics for Roman letters (digits are by default upright uppercase, but lower case digits. Also known as “oldstyle numbers” can also be used). There are some exceptions to these rules: an operator like \sin is typeset using upright font. As said above, $\mathbf{}$ produces upright characters. The so-called “black board” or “double-struck” characters obtained by $\mathbb{}$ are upright, for instance \mathbb{R} . An interesting point: lots of people prefer \mathbb{R} , using \mathbb{m} . Some people fake it as \mathcal{R} , using $\mathcal{I}\!\!\!\mathcal{R}$. You can always use $\mathrm{}$ for an upright character, as in $\mathrm{E}=\mathrm{mc}^2$. Some packages provide italic uppercase Greek letters, and upright lowercase. Then you can apply the Laplace operator to the Delta function like this: $\Delta\Delta$.

In MathML, characters should not be used directly. For instance, you should use $\langle\mathrm{mn}\rangle 125\langle/\mathrm{mn}\rangle$ for a number $\langle\mathrm{mi}\rangle\mathrm{Foo}\langle/\mathrm{mi}\rangle$ for the variable Foo and $\langle\mathrm{mo}\rangle\mathrm{sin}\langle/\mathrm{mo}\rangle$ for the operator sin. These elements have a `mathvariant` attribute, which indicates which variant to use. In the case $\langle\mathrm{mo}\rangle$, the default is upright; in the case $\langle\mathrm{mi}\rangle$, the MathML norms says: “The default `mathvariant` would (typically) be `normal` (non-slanted) unless the content is a single character, in which case it would be `italic`.” So we could translated \mathbb{R} as an $\langle\mathrm{mi}\rangle$ element containing a normal R with a `mathvariant` attribute of value `doublestruck`. An alternate solution consists in using a double struck character R, and no `mathvariant` attribute. Here are all the possible variants, and for the character A its Unicode value: normal (1D400), italic (1D434), bold-italic (1D468), double-struck (1D358), bold-fraktur (1D56C), script (1D49C), bold-script (1D4D0), fraktur (1D504), sans-serif (1D5A0), bold-sans-serif (1D5D4), sans-serif-italic (1D608), sans-serif-bold-italic (1D63C), monospace (1D670). In the case of Greek letters, Unicode knows the following variants: bold, italic, bold italic, sans-serif bold, and sans-serif bold italic. And in the case of digits: bold, double-struck, sans-serif, sans-serif bold, monospace.

Note: there are some holes in the table, starting at 1D400. For instance the Laplace symbol (U+2112) looks like \mathcal{L} , but the the translation of \mathcal{L} (that \TeX shows as \mathcal{L}), is not 1D49C+11 (this character does not exist). Another special case: there are some variants of the Greek letters ϵ and ε ; which one is the default is unclear.

There are two complementary views of MathML: presentation and content. This is how you would convert $a = b$ using content markup: $\langle\mathrm{reln}\rangle\langle\mathrm{eq}/\rangle\langle\mathrm{ci}\rangle a\langle/\mathrm{ci}\rangle\langle\mathrm{ci}\rangle b\langle/\mathrm{ci}\rangle\langle/\mathrm{reln}\rangle$. And this is how Tralics converts the same using presentation $\langle\mathrm{mrow}\rangle\langle\mathrm{mi}\rangle a\langle/\mathrm{mi}\rangle\langle\mathrm{mo}\rangle=\langle/\mathrm{mo}\rangle\langle\mathrm{mi}\rangle b\langle/\mathrm{mi}\rangle\langle/\mathrm{mrow}\rangle$. A more complex example, in presentation mode:

```

<apply>
  <int/>
  <bvar><ci>x</ci></bvar>
  <condition>
    <apply><in/><ci>x</ci><ci type="set">C</ci></apply>
  </condition>
  <apply><sin/><ci>x</ci></apply>
</apply>

```

You can clearly see that we apply the $\langle\mathrm{int}\rangle$ operation to some quantity (in which $\langle\mathrm{sin}\rangle$ is applied to x), subject to some condition (in which $\langle\mathrm{in}\rangle$ is applied to x and some set). If you consider the expression $\int_0^{\infty} \sin(x) dx$, it is translated by Tralics as

```

<msubsup><mo>&int;</mo> <mn>0</mn> <mi>&infin;</mi> </msubsup>
<mrow>
  <mo form='prefix'>sin</mo>
  <mo></mo>
  <mi>x</mi>
  <mo></mo>
  <mi>d</mi>

```

```
<mi>x</mi>
</mrow>
```

Forget about the `<mrow>` element, this is added by Tralics using some heuristic rules that do not always work; their effect is to isolate the parentheses from the integral sign: the height of the parentheses should be normal. In the expression, one can see that the integral sign is a character considered as an operator (but there no `<int>` element here); it has an exponent and an index. In the same fashion, `\sin` is translated as a `<mo>` element (with an attribute that says it precedes its argument). Nothing in the formula says that the argument is x .¹² Nothing says that that ‘ dx ’ is not the product of d and x , but a conventional way of indicating that x is the bound variable (the `<bvar>` above). In fact, Tralics cannot guess the *use* of the tokens, it knows only the layout: for instance, is Γ_{ij}^k a Christoffel symbol with three indices or a randomly Greek letter, with two indices raised to some power? And what about ³He or this footnote?¹³ Note that such expressions are not part of content markup: in both cases, these things look like tensors and are to be produced with `<mmultiscripts>`. In what follows, we shall speak only of the presentation part of MathML.

Some features are difficult to implement. For instance, it is possible to group some equations in a single mathematical formula, and put a label (with a reference like 17) to the whole, as well as a label for each subequations (referenced as 17.a, 17.b, etc). It is also possible to split an equation on more than one line, with a single number for it:

$$\begin{aligned}\zeta(s) &= 1/1^s + 1/2^s + 1/3^s + 1/4^s \dots \\ &= 1/1^s + 1/3^s + \dots + 1/2^s + 1/4^s + \dots\end{aligned}\tag{1.17}$$

It is possible to add some lines of text between two equations

$$X = 1/2^s + 1/4^s + 1/6^s + 1/8^s + 1/10^s + 1/12^s + 1/14^s + 1/16^s + \dots\tag{1.18}$$

as in this example (equals signs are aligned)

$$= \frac{1}{2^s}(1/1^s + 1/2^s + 1/3^s + \dots)\tag{1.19}$$

Horizontal spacing in math formulas is managed intelligently by T_EX. In the case of $a + b$ or $a = b$, there is some space on each side of the operator, and this space disappears when the formula becomes an exponent or an index. Here is an example.

$$\frac{a + b}{a = b}, \quad x_{a=b}^{a+b}$$

Each operator has a type, and the value of horizontal spacing depends on the type of the tokens on the left and the right. Traditionally, parentheses are removed around the arguments of sine and cosine, and you say: sine squared of X instead of sine of X, the whole squared. Example:

$$\sin^2 x + \sin(x)^2.$$

In T_EX, you can consider every expression (a simple atom, or a list in braces) as an operator, provided that it is preceded by its type. Types Over, Under, Acc, Rad, Vcent are obtained by construction (overline, underline, adding an accent, constructing a radical, vertical centering). Types Ord, Op, Bin, Rel, Open, Close, Punct, Inner are obtained by defining characters via `\mathchar`, or using the following commands `\mathord`, `\mathop`, `\mathbin`, `\mathrel`, `\mathopen`, `\mathclose`, `\mathpunct`, `\mathinner`. These commands are understood by Tralics, but in general ignored. In

¹²Apart from the kludge mentioned above, Tralics make no difference between the letter x and the parentheses around it; since parentheses are often omitted in the case of the sine function, the argument could as well be a parenthesized x.

¹³Guess: in the previous formula, was the H in or out of the math formula.

the case of an inline expression, a line break can only appear after an operator (like plus or equal), provided it is the main operator (in particular, if you put the whole expression in braces, no line break will occur). This is ignored by `Tralics`. Hence something like $f(x^{a+b})$ cannot be broken (but Netscape does not hesitate to break after the plus sign, in the case of `<sup>`).

Note that exponents and indices use a smaller font size, and exponents in exponents use an even smaller one: compare x^{y^ϵ} with $x^{y_\epsilon^\epsilon}$. The image was obtained by converting the math formula into XML, then in a png image (such kind of images will be put into HTML pages, because there are few HTML browsers that understand MathML. We used anti-aliasing, because this is supposed to increase readability). There are three sizes, and four styles (`\displaystyle`, `\textstyle`, `\scriptstyle`, and `\scriptscriptstyle`). The first two styles in the list have the same size (in MathML, there is an attribute `scriptlevel` that controls the size of the expression, and another one `displaystyle` that says if the expression is in displaystyle or not). In fact, \TeX has two substyles, cramped or not. Consider carefully the placement of exponents in the following example

$$\frac{a_2 + x^2}{a_2 + x^2} = \frac{x_3^2}{x_3^2}.$$

One problem (for a translator like `Tralics`) deals with the placement of arguments versus operators. For instance, if you want to put a dot over a letter, you can say `\dot{x}` (compare the math version \dot{x} and the text version \dot{x} ; in text mode, slants are taken into account). The MathML equivalent is an `<msup>` whose first element is the identifier `x`. If you want to put a prime after a letter you say `x'`, or `x^{\prime}`. If you want to put an arrow over `x` prime, you have the choice between \vec{x}' and \vec{x}' . With the fonts used in this example, the first solution looks horrible. You can say `x_2^3` or `x^3_2`: the result is the same: a math item, formed of a nucleus, a superscript and a subscript. If you want something like $\frac{2}{3}x$, you have to use two items, the first one has an empty nucleus, the second has no scripts. You can say $R_i^{j_{kl}}$, using three items. The MathML translation should consist in a single `<mmultiscripts>` element. It is possible to enclose a formula by braces, brackets, etc, provided that the font contains the machinery needed for it. You can either use `\big` and its variants, if you know the height of the formula, or `\left` and `\right`, or as in the example that follows, use an environment like `matrix` that uses whatever is best.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

The `\over` command (and friends) are discouraged by `amstex`: you get a message of the form *Package amsmath Warning: Foreign command \over; \frac or \genfrac should be used instead*. In fact the command takes two arguments, one before and one after. Example `x=\left(a+b \over c+d\right)^2+1` gives

$$x = \left(\frac{a+b}{c+d} \right)^2 + 1$$

The trouble is the following: Assume that we have a command `\myfrac` that typesets its arguments as a/b if the current style is script or scriptscript, and $\frac{a}{b}$ otherwise. In an expression as above, the current style cannot be known before the `\over` is seen. For this reason, \TeX introduced a command `\mathchoice` that takes four arguments, one for each style; after \TeX has completely read the math expression, it takes, in a second pass, the relevant token list. This is complicated. This mechanism is partially implemented in the current version of `Tralics`.

1.8 Some subtleties of T_EX

Assume that we have a number, say ‘1.3’, and want to convert this to a dimension, say ‘1.3pt’. If the number is in a command `\foo`, we can say `\dimen0=\foo pt`. On the other hand, how can we get 1.3 given that `\dimen0` holds 1.3pt? We may use `\the\dimen0`, and remove the last two characters. Hence we say `\expandafter\rem@pt\the\dimen0`. One trouble is that the category codes of the characters ‘pt’ are 12, not 11, so that a definition like

```
\def\rem@pt#1pt{#1}
```

does not work. You could try to change the category codes of p and t in the definition, but these letters are part of the name of the command. Some black magic has to be used. The code shown here converts also ‘12.0pt’ to ‘12’.

```
\begingroup
\catcode'P=12
\catcode'T=12
\lowercase{
\def\x{\def\rem@pt##1.##2PT{##1\ifnum##2>\z@.##2\fi}}
\expandafter\endgroup\x
```

Implementing such a construction in a Perl script is not obvious. Consider then the following example:

```
\def\foo#1{#1x#1}\def\xbar#1{ $#1$}
\expandafter\foo\xbar y
\foo\xbar y
```

Consider line 2. After `\expandafter` has read the two tokens `\foo` and `\xbar`, current state is S, and a space and the character ‘y’ are not yet read. Expansion of `\xbar` reads an argument y; the space before it is ignored. The expansion is then `\foo` \sqcup_5 $\$_3$ y₁₁ $\$_3$. Hence the argument of `\foo` is the dollar sign, expanding it gives $\$_3$ x₁₁ $\$_3$ y₁₁ $\$_3$. Note that this gives an odd number of dollar signs.

Consider now the third line. Here the argument of `\foo` is `\xbar`. The expansion is `\xbar` x₁₁ `\xbar`, and a space and ‘y’ have to be read. The expansion of `\xbar` is \sqcup_5 $\$_3$ x₁₁ $\$_3$. The second `\xbar` reads a space and ‘y’ as its argument. The expansion is \sqcup_5 $\$_3$ y₁₁ $\$_3$. Note that `latex2html` complains with: *Unknown commands: xbarx*, because the string is rescanned, and a space should have been added. In the second version of our translator, which was a Perl script as `latex2html`, we solved this problem as follows. The commands `\` and `!` are replaced by `\00!` and `\01!`, special characters are replaced by `\3#!`, `\3&!`, `\3<!`, `\3>!`, and `\3~!`. A command like `\foo` is replaced by `\1foo!`, spaces after `cmdfoo` are removed. Using such a mechanism, we can handle spaces correctly, as long as the category codes do not change. In the current version of `Tralics`, a token `\foo` is represented as `\foo`, i.e., a pointer into the hash table, or like $\$_3$, i.e., a pair of two numbers category code, character value.

The behavior of a token is determined by its command code: the command code of $\$_3$ is 3. The command code of `\foo` depends on the current value of `\foo`. For instance, the default value of `\count` is (1,90,0). Here the first integer is the definition level, (1 is the bottom level), 90 is the command code and 0 is the subcode. Remember that, in the case of `\the\count0`, the `\count` command is evaluated for a value; in all other cases it is evaluated for side effects. In particular for a case like `\global\count3=17`. Assume that you say `\count3=17`; in such a case a number is read, namely 3, and checked for out-of-range. After that, the subtype is looked at (the commands `\count`, `\dimen` `\skip`, or `\muskip` have the same command code, but different subtypes, they read something and store it somewhere). Note: if you say `\countdef\foo3`, then `\foo` has some command code, and its subcode is 3. The action is exactly the same as `\count`, after 3 has been read. Said otherwise, `\foo=17`, and `\count3=\17` read a number 17. This number will be stored somewhere in the eqtb table; the location is 3 slots after the start of the `\count` table. This slot

contains (L,V) , for instance $(4,23)$. The first number is the definition level, and the second is the value. In the case $L=0$, this means that the object is undefined (in the case of a counter, this means zero). Otherwise this is the level. In the case $A\{B\{C\}\}$, A is at level 1, B at level 2, and C at level 3. The quantity L is never greater than the current level. In the case where the assignment is global, (L,V) is instantiated to $(1,17)$. In the case where the current level is L , then V in (L,V) is replaced by 17. Otherwise, the old value is saved on the stack, and (L,V) is replaced by $(l,17)$ (where l is the current level). When the group is closed, the old value is restored.

Note the following trick. Assume that $\backslash A$ increases some counter and puts the value in $\backslash foo$, $\backslash B$ does the same, but changes globally $\backslash foo$. Assume that you say $\{\backslash A\backslash A\backslash B\backslash A\backslash A\dots\}$. The first $\backslash A$ sets $\backslash foo$ to 1 and saves $(0,undef)$. The second $\backslash A$ sets $\backslash foo$ to 2. The first $\backslash B$ sets $\backslash foo$ to 3 at level 1, the second $\backslash B$ sets $\backslash foo$ to 4 at level. The next $\backslash A$ saves the old value of $\backslash foo$ and sets $\backslash foo$ to 5, etc. As a consequence: every $\backslash A$ preceded by a $\backslash B$ will put an item on the save stack. When the stack is restored, the value to restore will be (L,V) and the current value (L',V') . If L' is 1, nothing happens. Otherwise, (L',V') is replaced by (L,V) . As a consequence the value after the group is the value of the last $\backslash B$. This results in a waste of the save stack. For this reason, Knuth says: all assignments to the scratch registers whose numbers are 1, 3, 5, 7 and 9 should be $\backslash global$; all assignments to the other scratch registers (0, 2, 4, 6, 8, 255) should be non- $\backslash global$.

A silly question is: what happens if you say $\{\backslash let\ endgraf\ par\ gdef\ par\}\ edef\ foo\{\endgraf\}\ Foo\ def\ endgraf\}\ Bar\}$. In the current version of $\backslash Tralics$, but this is also true for $\backslash T_{E}X$, when you say $\backslash def\ foo$, the command code of $\backslash foo$ is changed to be used-defined, and the subcode is an address to a table containing the token list of the body. On the other hand, $\backslash let\ foo\ bar$ will use the command code and subcode of $\backslash bar$, and copy this in $\backslash foo$. In the original version of $\backslash Tralics$, the Perl version, we had two tables: the list of predefined commands, with their internal number, and the list of user defined commands with their body. In the code above, when $\backslash Foo$ is executed, then $\backslash foo$ is a user defined command, whose body contains $\backslash endgraf$, whose meaning is the original $\backslash par$. When $\backslash Bar$ is seen, the meaning of this token has changed. The essential reason why $\backslash Tralics$ was re-written in C++ is to make this piece of code work.

1.9 Language options

Most $\backslash T_{E}X$ formats (plain, $\backslash L_{A}T_{E}X$, $\backslash amstex$) have been written by American people; nowadays, major developments are done in Europe (including the $\backslash conT_{E}Xt$ format, and the $\backslash hyperref$ package). However, lots of people use basic primitives for their French publications, and the situation is not simplified by the fact that there are two package for writing French documents (one by B. Gaulle, and one by D. Flipo).

In the original version of $\backslash T_{E}X$ you had to say $\backslash 'e'$ for an e acute, and you had to say $\backslash c\ c'$ or $\backslash c\{c\}$ for a c cedilla (which form being the best is in general unknown). For homogeneity reasons, Lamport recommends $\backslash \{e\}$. Because some accents are redefined by tables, or tabular environments, the solution that always works is $\backslash a\ \{e\}$. This is something strange, but a translator like $\backslash Tralics$ has to cope with it (for instance, some authors of the Raweb use $\backslash bibtex$ files that are generated automatically from a data-basis, and this software systematically produces $\backslash a$ for accents). This makes texts rather uneasy to read, and not every spell-checker understands this (Ispell for instance allows $\backslash Jos\acute{e}$ or $\backslash Jos\backslash 'e$, but not both). By default (i.e. on Linux machine) $\backslash L_{A}T_{E}X$ understands iso-8859-1. This means that all characters used in France are recognized (except, $\backslash \oe\ \backslash \Y$, these characters cause also problems in HTML). Recently, another character was introduced, namely €. Nobody knows how to use it (according to [6, paragraph 7.8.7], published in april 2004, $\backslash \text{texteuro}$ is the official $\backslash L_{A}T_{E}X$ way). In case of doubt, you should use 'euro'.

In order to emphasize words, you can underline them, use a different font, or mark them with quotes. In English, you would use quotes "like these" or 'like these', but never like "this".

In France, guillemets « are used like this ». Note that the spacing is different from English, but the package should take care of everything. The forever question is: how to enter these funny characters in my keyboard made in Mexico¹⁴. One solution consists of typing two < in a row and hope for the best (we have either an active character, or a ligature). Note that Tralics translates `\verb+<<-->>+` as `<hi rend='tt'><​<​-​-​>​>​</hi>` the funny characters have as only purpose to inhibit ligatures in the resulting XML. The result might also be : `ij` and `il`. Depending on the packages, you should perhaps use `\guillemotleft`, `\guillemotright` or `\og`, `\fg`. Which method is the best is still unclear to me.

¹⁴or Thailand; it has an international layout, with a W. logo which is a registered trademark of M.C.

Chapter 2

Expansion

One part of the work of T_EX is to replace all user defined tokens by primitives; this is the main objective of the ‘expansion’ process. In this respect, there is little difference between T_EX and Tralics. In this chapter, we review some constructions.

2.1 Defining new commands

A definition is typically of the form

```
\def\fooi{foo}
\def\fooi#1#2{#2#1}
\def\fooi#1.#2=#3#{Seen#1#2#3.}
```

You may wonder why the commands are not called ‘\foo1’, ‘\foo2’ and ‘\foo3’. The reason is that, if digits have standard category codes, they are not of type letter, so that ‘\2foo’ is the command `\2`, followed by the letters ‘foo’ (the tokens are `2` `f`₁₁ `o`₁₁ `o`₁₁) and ‘\foo2’ is the command `\foo` followed by the digit 2 (the tokens are `f``o``o` `2`₁₂). It is possible to create the token `f``o``o``2` via `\csname foo2\endcsname`, and it is also possible to change the category code of 2. This is in general a bad idea: If you say `\setlength{\parindent}{\foo2+2cm}`, it is impossible to design the `\setlength` command so that ‘\foo2’ is read as a command and ‘2cm’ as a dimension. On the other hand, if you say `\def\foo2#1#2{#2#1}`, T_EX expects, after the second #, the character 2 with category code 12; if not it complains with: *Parameters must be numbered consecutively*. In Tralics, the message is a bit different, it says *Error while scanning definition of \foo2 expecting #2; got #{Character 2 of catcode 11}*.) Note how `2`₁₁ is printed.

Before `\def`, you can put a prefix: it can be `\long`, indicating that the command accepts whole paragraphs as arguments; it can be `\outer`, indicating that the command cannot be the argument of another command; it can be `\global`. This last prefix can be put before any assignment, it says that the assignment is global (unless `\globaldefs` is non-zero). The order of the prefixes is irrelevant. After the `\def` comes the object to define (this is either an active character, or a command name), then what T_EX calls ⟨parameter text⟩, and this is followed by the body. The body starts with the first opening brace (any character of category code 1) and ends with the first closing brace (any character with category code 2) that makes the body balanced against braces. These braces are not part of the body. The parameter text is an arbitrary sequence of tokens, but cannot contain braces. If it contains a # (in fact, any character of category code 6), it has to be the final character of the sequence, or be followed by the digits 1, 2, 3, up to 9, in order. If there is some text between #3 and #4 (or between #3 and the start of the body), this imposes a constraint on the third argument. If there is some text before #1, this imposes a constraint on the command itself. In the body you can use ##, this will be replaced by a #; you can also use #1, #2, etc., this

will be replaced by the value of the first, second, etc., argument. As above, the # is any character of category 6, the digits are of category 12, you cannot access the second argument if only one is available. If you define `\foo2` as above, \TeX will signal a second error: *Illegal parameter number in definition of \foo2*.

Once you have defined the commands, you can use them. We give here an example, and the translation by `Tralics`

```
\fooi\fooi12\fooi+ok. {\itshape 3} =xyz{}!
foo21Seenok <hi rend='it'>3</hi> xyz.!
```

and also by \LaTeX `'foo21Seenok 3 xyz.!'` Some explanations. The first command takes no argument, thus is easy to understand. The second command takes two arguments, its body is `'#2#1'` so that the expansion is the token list formed by the tokens of the second argument followed by the tokens of the first argument. In the case of `'\foo12'`, the arguments are '1' and '2' (a list of length one). In the case of `'\fooi {AB} {CD}'`, the arguments are 'AB' and 'CD', a list of length two. This is because \TeX ignores initial spaces when reading undelimited arguments; in any case, an argument is well-balanced against braces (same definition as above for the body of a command). The shortest possible sequence of tokens is read (in the case of an undelimited argument, this sequence is never empty). If the result starts with an open brace and ends with a closing braces, these are removed, provided that the remaining token list is well-balanced; for instance, in the case `'\fooi{ }a'`, the first argument is empty. If the command is not `\long`, then `\par` token are forbidden in the argument. In any case tokens that are defined to be `\outer` are forbidden in a parameter.

In the case of `\fooi`, the situation is a bit more complicated. Fetching the argument is more involved than in the general case. The specification is: plus sign, argument, dot, argument, equals sign, argument, sharp sign. Note first that the '+' sign is not part of the command name, but is required after it whenever used. The first argument here is the shortest sequence (possibly empty) of tokens, that is a balanced list, and this is followed by the required token list (here, a single dot). Here it is `'_{\it_3}_'`; a pair of initial and final braces disappear, if possible. The '#{' after '#3' says that the third argument is delimited by an open brace. This brace is left unread. Such a construction is rare: it occurs only four times in the \LaTeX sources, an example will be given later.

Consider the following example: `'\def\opt[#1]{'`. If you say `'\opt[foo]'` or `'\opt[{}foo]'`, the argument is 'foo'. If you say `'\opt[{}[foo]]'`, it is '[foo]'. It is important to know that braces are required if you want a closing bracket in the argument. In the case of `'\item[{}it foo]'`, the braces are useless; the scope of the `\it` command is limited to 'foo' because an additional pair of braces is added somewhere in the body of the `\item` command. The following example is non-trivial:

```
\def\@car#1#2\@nil{#1}
\if b\expandafter\@car\f@series\@nil\boldmath\fi
```

The command `\@car` takes two arguments, the first one is undelimited, the second one is delimited by `\@nil`, it is ignored. Let's assume that `\f@series` expands to a non-empty list, for instance 'mc' (this means that the current font has medium weight and is condensed). Then `'\@car md\@nil'` expands to 'm'. The second line of our example uses `\@car` to get the first character of `\f@series`, and compares it to 'b' (the result is true if the current font is bold, extra bold, bold condensed, etc). This code is used for typesetting the $\LaTeX 2_{\epsilon}$ logo in bold version as $\LaTeX 2_{\epsilon}$. The commands `\if` and `\expandafter` will be explained later. For the case where `\f@series` could be empty, it could be wise to insert a character like 'B' before `\@nil` (very often, you will see `\@empty`, because it typesets to nothing; here we want the full expansion of everything between 'b' and `\boldmath` to be a single token; using `\@car` produces a single token; if this token expands to nothing we lose).

The following example is from the \TeX book:

```
\def\cs AB#1#2C$#3\${#3ab#1 c##\x #2}
\cs AB{\Look}C${And\${ }{look}}\$ 5
```

If you feed this to Tralics¹, you will get three errors (one because of the ‘##’, and two undefined commands). In verbose mode, the transcript file of Tralics will contain the following

```
\cs AB#1#2C$#3\$ ->#3ab#1 c##\x #2
#1<-\Look
#2<-
#3<-{And\$ }{look}
```

One question is: should arguments be in braces or not? As seen elsewhere, some commands have a special syntax, and cannot be followed by braces (for instance, in the case of ‘\catcode‘\\$’ the argument is the backtick followed by the dollar). In a case like $\$a \over b+c\$$, there are two arguments, one before and one after the command. An expression like $\$a\over b\over c\$$ is a error. The error message says to add some braces, but they are used only for grouping. A similar error message is issued if you say $\$a^b^c\$$. But compare ‘ $\$a^{\{b^c\}d\$}$ ’ and ‘ $\$a\over \{b\over c\}d\$$ ’: the translation is $a^{b^c}d$ and $\frac{a}{\frac{c}{d}}$.

It is sometimes important to know which braces disappear or remain. As an example, you can say ‘\def\ap{a}’ in order to get a' ; but if you say this ‘ $\$x_{\ap}\not=x_{\ap}\$$ ’, you get $x'_a \neq x_{a'}$. In fact, you cannot say that ‘\ap’ is the argument of the underscore command; this is because T_EX expands everything; in one case, it sees that the underscore is followed by the letter a, in the second case by a brace, hence a delimiter for a math list.

In general, you will be faced with the following problem: you say ‘\def\foo#1{\xbar#1} and ‘\def\xbar#1{\itshape #1}’. Note the double braces: the outer braces delimit the argument (of \def, i.e., the body of \xbar), the inner braces delimit the scope of \itshape. When you say ‘\foo{12}’ only the first letter is in italics, another level of braces is needed. This is what you can see in the transcript file of Tralics:

```
\foo #1->\xbar #1
#1<-12
\xbar #1->{\itshape #1}
#1<-1
{begin-group character {}
+stack: level + 3 for brace
{\itshape}
{font change \itshape}
...
```

In this example, braces are missing in \foo, a remedy is to add a pair of braces in the argument, like ‘\foo{{12}}’.

It is possible to define commands inside commands. For instance, you can say

```
\def\foov#1{\def\xbar##1{##1###1###1}}
```

When the scanner reads a token list, it handles ‘#’ signs (in fact, any character of category 6) in a special manner inside a definition. The token list of the previous line is $\boxed{\text{def}} \boxed{\text{foov}} \#_{23} \{ \boxed{\text{def}} \boxed{\text{xbar}} \#_6 1_{12} \{ 1_{25} \#_6 1_{12} \#_6 1_{25} \#_6 \#_6 1_{12} \}_2 \}_2$. As you can see, there are three possibilities for a sharp sign: before the brace that defines the body, it is $\#_{23}$, and the digit that follows is omitted², it is 1_{25} , 2_{25} , in the body when followed by 1, 2, etc³. It is $\#_6$ when followed by a sharp sign. Said otherwise, a double sharp sign in a definition is equivalent to a normal one outside. Note the following trick.

```
\catcode'\^6
\def\foo#1^2{#1^1## #^ # ^^}
\show\foo
```

¹The last character on the line is not read

²T_EX uses $\#_{13}$, because no confusion with an active character can occur.

³T_EX uses 1_5 , because no confusion with a character of category 5 can occur (See T_EXbook, exercise 7.3).

A quantity like 1_{25} is shown as $\hat{1}$, because the hat character appears as $\hat{2}$ (i.e., the token $\hat{2}_5$) in the \langle parameter text \rangle part of the definition. Hence \TeX prints $\backslash\text{foo}=\text{macro}: \#1^2\rightarrow\hat{1}\hat{1}\#\hat{\hat{\#}}\hat{\hat{\#}}\hat{\hat{\#}}$. On the other hand, \Tralics uses a different mechanism for macros: it remembers the number of arguments and the items between them, hence does not make the difference with a macro defined as $\backslash\text{def}\backslash\text{xbar}\hat{1}\#2\{\dots\}$ ⁴. The printed result is $\backslash\text{foo}=\text{macro}: \#1\#2\rightarrow\#1\#1\#\hat{\hat{\#}}\hat{\hat{\#}}\hat{\hat{\#}}$.

Assume now that you say $\backslash\text{foov}\{17\}$. The result of the expansion is the token list shown above, with 1_{25} replaced by $1_{12} 7_{12}$. When $\backslash\text{xbar}$ is defined the $\#_6$ will read the character that follows, in this case 1_{12} . The situation is as if you had said $\backslash\text{def}\backslash\text{xbar}\#1\{17\#1\#17\#\#1\}$. Evaluating $\backslash\text{xbar}$ may signal an error, because of the $\#\#$ (no error is signaled in case the argument of $\backslash\text{xbar}$ is $\backslash\text{gee}$, a command that ignores its first and third argument). If you call $\backslash\text{foo}$ with $\{25\}$ instead of $\{17\}$ as argument, you will get the following error *Illegal parameter number in definition of $\backslash\text{xbar}$* ⁵.

2.2 Defining commands in \LaTeX

You can say

```
\newcommand*\fooi}{foo}
\newcommand*\fooui}[2]{\#2\#1}
\newcommand\fooiiv[3][bar]{Seen\#1\#2\#3}
```

The first two lines define the same commands as in the start of section 2.1. It is not possible to define $\backslash\text{fooiiii}$. However, you can define $\backslash\text{fooiiv}$, a command that takes an optional argument. In fact, you call it like this $\backslash\text{fooiiv}[X]YZ$; the expansion will be $\text{\textit{Seen}XYZ}$. You can put a pair of braces around the arguments, like $\backslash\text{fooiiv}[\{X\}]\{Y\}\{Z\}$, the result is the same. Braces are needed for the first argument in case you want a closing bracket in it. If the first argument is $\text{\textit{bar}}$, you can omit the $[\text{\textit{bar}}]$: for this reason, the argument is called optional. In \LaTeX , $\backslash\text{fooiiv}$ expands to $\backslash\text{@protected@ttestopt}$, which is a command to make $\backslash\text{fooiiv}$ robust (i.e., in some cases, the test for an optional argument is delayed); it then expands to $\backslash\backslash\text{fooiiv}$, which is a command that takes three arguments. In \Tralics , no auxiliary command is used. If you say $\backslash\text{show}\backslash\text{fooiiv}$, \Tralics will print the following on the transcript file.

```
\fooiiv=opt \long macro: bar\#2\#3->Seen\#1\#2\#3.
```

Commands defined by $\backslash\text{newcommand}$ are $\backslash\text{long}$ unless a star is used (they accept paragraphs as arguments.) The $\text{\textit{opt}}$ before it shows that the command takes an optional argument. We show the value of this argument instead of $\#1$ before the \rightarrow . The following is printed by \LaTeX

```
> \fooiiv=macro:
->\@protected@ttestopt \fooiiv \\\fooiiv {bar}.
```

Since being $\backslash\text{long}$ deals with reading parameters, in \LaTeX , it is the auxiliary command $\backslash\backslash\text{fooiiv}$ which is $\backslash\text{long}$. This shows how to ask \LaTeX for the meaning of the auxiliary command and its answer:

```
\expandafter\show\csname\string\fooiiv\endcsname
> \\\fooiiv=\long macro:
[\#1]\#2\#3->Seen\#1\#2\#3.
```

The philosophy of \LaTeX is that a user should not randomly redefine commands. For this reason, you must use $\backslash\text{newcommand}$ (for an undefined command) or $\backslash\text{renewcommand}$ (for overwriting an existing command). In the same fashion, $\backslash\text{renewenvironment}$ is used to redefine an environment; we shall see later that an environment $\text{\textit{foo}}$ is defined by two commands: $\backslash\text{foo}$ and $\backslash\text{endfoo}$. You should never define $\backslash\text{endfoo}$. This explains error messages of the form: *LaTeX Error: Command*

⁴Said otherwise $\backslash\text{ifx}\backslash\text{foo}\backslash\text{xbar}$ produces different results in \TeX and in \Tralics ; these two commands take two arguments, and have the same body, the only difference is in the introducing character.

⁵This was accepted by \Tralics until version 2.3: if a command takes one argument, the value of the second argument is the empty list.

`\endfoobar` already defined. Or name `\end...` illegal, see p.192 of the manual. In Tralics, we do not check that the command starts with ‘end’; the error message is `\newcommand: cannot define \foo; token is already defined`. You can use `\providecommand`, the syntax is the same. In this case, the definition is silently ignored if the command already exists. You can use `\DeclareRobustCommand`, this is defined by Tralics to be the same as `\providecommand` although the L^AT_EX behavior is different. You can say ‘`\global\def\foo{}`’, this is the same as ‘`\gdef\foo{}`’, it defines `\foo` globally. You cannot use the `\global` prefix for L^AT_EX commands.

It is now time to explain that braces have two different purposes: as a delimiter for an argument list, and also for grouping: in the same fashion that the formula $z(x + y)$ can be considered as z applied to $x + y$ or the product of z and $x + y$. In the case of ‘`\textit{12}`’, the braces delimit the arguments, in the case of ‘`\itshape 12`’, the braces are used for grouping. In both cases, all characters up to the closing brace are in italics, but this property depends on the semantics of the operator, not the syntax. There is a big difference between these two use of braces: the tokenizer produces token lists that are always balanced (there are as many opening delimiters as closing delimiters, where delimiters are characters of category code 1 and 2). On the other hand, if you say ‘`\let\bggroup=}`’, the `\bggroup` has the same meaning as an opening brace, hence triggers the start of a new group; but it is not an explicit character (such things are called “implicit characters” in the T_EXbook). When you say ‘`\hbox...`’ the opening brace can be implicit or explicit (in this case, braces are used both as delimiters and for grouping). Groups can also be defined by math shift characters (if you like empty lines in the source of a math formula, you can say ‘`\$let\par\relax ..\$\$`’), or implicitly for a cell in a table, or via `\left` and `\right` in a math formula, or via `\begingroup` and `\endgroup` (they define a “semi-simple group”).

One difference between plain T_EX and L^AT_EX is the existence of named groups: instead of saying ‘`\beginfoo`’ and ‘`\endfoo`’, you say ‘`\begin{foo}`’ and ‘`\end{foo}`’. This is interpreted by L^AT_EX as

1. When `\begin{foo}` is seen,
 - (a) A test is made to see if ‘`\foo`’ exists. If it does not exist, an error is signaled and steps (1c) and (1d) are skipped (via a call to `\expandafter`).
 - (b) The command `\begingroup` is executed (with space hacking).
 - (c) The name ‘foo’ is stored in `\@currenenvir`.
 - (d) The command `\foo` is executed.
2. When `\end{foo}` is seen,
 - (a) The command `\endfoo` is executed.
 - (b) The name ‘foo’ is compared with `\@currenenvir`, an error is signaled in case of mismatch.
 - (c) The command `\endgroup` is executed (with more space hacking).

This mechanism is not symmetric. It is implemented in Tralics in a similar manner (but there are some differences that can be analyzed by a malicious user). The first remark is the following: on entry, you may get a message that says *LaTeX Error: Environment unknown undefined*, on exit you would get *LaTeX Error: \begin{document} ended by \end{unknown}*. The trick is that the `\endfoo` token (created by `\csname`) is never undefined (its default action is `\relax`). One important point is that the command used in step (1d) is `\foo`, not `\beginfoo`. In [6, example 7-3-1], there is an example of ‘`bfseries`’ as an environment; there is no command `\endbfseries`. Note that in step (1d), the token that comes after `\foo` is the token after ‘`\begin{foo}`’, and this means that `\foo` can grab its arguments; on the other hand the token after `\endfoo` in step (2a) is the start of the sequence that checks the environment name: thus `\endfoo` cannot read its argument (we shall see in a minute why steps (2a) and (2b) cannot be swapped). In the current version of Tralics, the “space hacking” is not implemented; we shall not discuss it here.

This is how you can define new environments:

```
\newenvironment{x}[2]{#1BY\begin{y}#2AY} {by\end{y}ay}
\newenvironment{y}{Z}{z}
\begin{x}a b c \end{x}
```

This typesets as aBYZbAY c byzay. The `\begin` part reads two arguments. The `\end` part takes no argument; it could use the first argument of the `\begin`, provided that this one saves it in a command. In verbose mode, the following is printed by Tralics in the transcript file. We have removed all lines with ‘Character sequence’ and ‘Text’.

```
1 [61] \begin{x}a b c \end{x}
2 {\begin}
3 {\begin x}
4 +stack: level + 3 for environment
5 \x #1#2->#1BY\begin {y}#2AY
6 #1<-a
7 #2<-b
8 {\begin}
9 {\begin y}
10 +stack: level + 4 for environment
11 \y ->Z
12 {\end}
13 {\end x}
14 \endx ->by\end {y}ay
15 {\end}
16 {\end y}
17 \endy ->z
18 {\endgroup (for env)}
19 +stack: ending environment y; resuming x.
20 +stack: level - 4 for environment
21 {\endgroup (for env)}
22 +stack: ending environment x; resuming document.
23 +stack: level - 3 for environment
```

At lines 4, 10, 20 and 23, you can see that the current “level” changes (this is what \TeX calls the “semantic level”). The default level is level one, our example was done at level two, the first environment is at level three, the second at level four. When you see ‘level + 4’, it is because the level has just incremented; if you see ‘level - 4’ it means that the level will decrease. At lines 18 and 21, you see that Tralics uses a special ‘`\endgroup`’ token. Look closely at lines 13 and 19: when Tralics sees ‘`\end{x}`’, the current environment is y, it is only after evaluation of `\endx` that the environment is ‘x’ again; this example shows that steps (2a) and (2b) cannot be swapped. In Tralics the name of the environment cannot be modified by the user.

Because of the `\begingroup` command, everything, until the `\endgroup`, is local to this group; in particular `\@currenenv` will be restored. If you say something like

```
\begin{zfoo}\renewcommand\endzfoo{A}\end{zfoo}
```

the command associated to `\end{zfoo}` is locally redefined. In some cases, this is a big mistake: in Tralics, the start command can assume that the corresponding end command is executed or an error is signaled. In fact, the meaning of `\endzfoo` is stored on a special stack, and restored by `\end{zfoo}`. There is a big hack in \LaTeX (and also in Tralics): since no text should follow the end of the document, there is no need to store on the stack every definition given between the start and end of the document; thus `\document` executes a `\endgroup`; logically, `\enddocument` should insert a `\begingroup` token; in \LaTeX , this is not needed because step (2c) is never executed. In

Tralics this cannot be done, because we have to typeset the bibliography. Moreover, action cannot be completely trivial, because we have to re-insert all tokens saved by `\AtEndDocument`. We show here the transcript file, assuming that only one token has been saved, namely `\empty`. You can see the stack increase and decrease; you can see the `\endinput` that closes the current file; you can also see a second `\enddocument` command whose action is to pop the XML stack.

```
[490] \end{document}
{\end}
{\end document}
+stack: level + 2 for environment
{\enddocument}
{\endinput}
\empty ->
{\enddocument}
{Pop (module) 0: document_v}
{\endgroup (for env)}
+stack: ending environment document; resuming document.
+stack: level - 2 for environment
```

Clearly, you cannot use a document environment in a document; if you try, L^AT_EX complains with *LaTeX Error: Can be used only in preamble* (the preamble is everything before `\begin{document}`). The error message of Tralics is a bit more explicit: *Two environments named document*. If you put `\begin{it}` before `\begin{document}`, L^AT_EX does not complain. The trouble is at the end: you will get an error of the form *LaTeX Error: \begin{it} on input line 9 ended by \end{document}*, followed by a T_EX warning : (*\end occurred inside a group at level 1*). In Tralics, an error is signaled at the start: *\begin{document} not at level 0*. On page 161, you see statistics of the form ‘Save stack +1582 -1582’; this means that the semantic stack pointer has increased 1582 times, it has decreased the same number of times, so that the end of the document has been seen at level zero, no warning is issued in the case the two numbers are not the same.

2.3 Some small examples

Remember that `\foo` and `\;` are two commands who differ only in the following behavior: when the tokenizer sees a backslash followed by a semi colon (whose category code is not letter), it constructs a command whose name is formed by that character (and sets the internal state to a mode in which spaces are not ignored). On the other hand, if backslash is followed by a letter, all letters are read (and the state is set such that following spaces will be ignored). By space, we mean here every character that has the category code of a space. A space after `\verb` is never ignored, but it is unwise to use this space as delimiter. In the case of `\foo`, the tokenizer allocates a slot on the hash table (unless `\foo` already exists). The possibility to change category codes dynamically is interesting (however, the implementation of `\verb` in Tralics uses no category code changes, and is more efficient). The two commands `\makeatletter` and `\makeatother` change the category codes of the at sign character `@`, to letter and other. For instance

```
\makeatletter
\def\foo@val{
\def\foo#1{\def\foo@val{#1}\check@foo}
\def\usefoo{\foo@val}
\makeatother
```

In this example, we have two user commands: `\foo` that defines a variable, and `\usefoo` that uses it. The variable `\foo@val` has a reserved name; there is a command `\check@foo` that makes sure that the argument is correct. The default category code of `@` is 12; in most of the examples, we

shall assume that it is 11, because these examples come from the \LaTeX kernel or style files where the default category code is 11.

As explained above, `\catcode'\$=3'` changes the category code of the dollar sign. What follows has to be a character code (a number between 0 and 255) followed by an optional equals sign, followed by a valid category code (an integer between 0 and 15). Assume that you say `\def\A{25}`, followed by `\catcode\A7.`⁶. In the case where standard category codes are in effect this is tokenised as `\catcode \A 712 .12`. But when a number is read, all tokens are expanded, until the end of the number is found (in the case where the number is formed by digits, one space character after the number will be read, if possible). In this case, \TeX reads the digits 2, 5 and 7. It stops reading at the dot. This is an error (signaled by \Tralics as *Bad character code replaced by 0: 257*). Then \TeX reads an optional equals sign (there is none) and an integer (there is none). Hence you get a second error (*Missing number, treated as zero*). The result is that you have changed the category code of the null character to zero (like backslash). If you want to put 7 in the category code of the character defined by the command `\A`, you should say `\catcode\A=7~`.⁶ It is possible to make `\A` a reference to the character number 25, by using `\chardef`. Thus you can say `\chardef\A25~` and `\catcode\A7~`. Note that, in the context of routines like `scanint`, a character number is a valid number; so that `\A` can be used as the number 25, wherever a number is required. In the sources of \LaTeX you can see `\chardef\active=13'`. You will also see `\mathchardef@cclvi=256'`; there is no difference between `\chardef` and `\mathchardef`, except that a character is in the range 0-255, while a math char can take larger values (less than 2^{15}). You can use `\countdef\B26` (this will make `\B` as a reference to count register number 26), `\dimendef\C27` (this will make `\C` as a reference to dimension register number 27), `\skipdef\D28` (this will make `\D` as a reference to skip register number 28), `\muskipdef\E29` (this will make `\E` as a reference to muskip register number 29, and `\toksdef\F30` (this will make `\F` as a reference to token register number 30). There is no `\boxdef`. The reason is that, if you want to copy the value of counter 1 into counter 0, you say `\count0=\count1`. If you say `\count@=\B` this will put the value of the counter 26 into `\count@` (this is the counter 255). However, you say `\setbox0=\copy1` if you want to copy the content of box 1 into box 0: the syntax is not the same. When you use `\chardef`, a line will be added to the transcript file, see section 6.12.

Commands can be defined via `\let`. You say `\let\A=\B`, where `\A` is a token that can be defined (active characters or commands; \TeX does not care if the token is defined or not). It is followed by `<equals><one optional space>`. This means that \TeX reads all space tokens; if the first unread token is an equals sign, it is read as well as the next token, provided that it is a space. If the equals sign is followed by two space tokens, only one is read. Instead of `\B`, you can put any token. After that, the current meaning of `\A` will be the current meaning of `\B`. For instance, if you say `\let\foo\bar\show\foo` you will get `\foo=macro:->\mathaccent "7016\relax`. In plain \TeX , you would see a space instead of `\relax` (both a space and a `\relax` indicate the end of the number). In \Tralics , you would see `\foo=\bar`, this is because `\bar` is a primitive, instead of a user defined command. If you let `\let\A=+`, then `\A` will behave like a `+` character (of category 12). In fact, this is called an implicit character, and sometimes an explicit character is required. For instance in the case `\parindent=-3.4pt`, the minus sign, the digits, the dot, and the two letters `pt` must be explicit characters. However, after

```
\let\bgroup={ \let\egroup=} \let\sp=^ \let\sb=_
```

there is no difference between `\$x\sp\bgroup a+b\egroup\$` and `\$x^{a+b}\$`. A token list has to be well balanced against explicit braces. For instance

```
\def\foo{\catcode'}=0\egroup}
```

satisfies the requirements. The body of the command consists in `{1 \catcode '12 }2 =12 012 \egroup`. If you evaluate `\foo`, the `\catcode` command will read the four tokens that follow; it

⁶Here, and in the lines that follow, you should replace the tilde by anything that stops scanning the number.

will modify the category code of the opening brace. All this happens inside a group opened by `{` and closed by `\egroup`, so that this is harmless. One use of `\let` is the following:

```
\def\fooA{a very long command}
\def\fooB{another very long command}
\def\xbar#1{\ifx 0#1\let\foo\fooA \else \let\foo\fooB\fi}
```

Here we use the fact that `\let` just moves a pointer.⁷ This is faster than copying a list. In particular, consider

```
\def\xbar#1{\ifx 0#1\fooA \else \fooB\fi}
\def\xbar#1{\ifx 0#1\let\foo\fooA \else \let\foo\fooB\fi\foo}
```

The first line executes conditionally one of `\fooA` and `\fooB`. However, this command cannot read an argument (because `\fooA` is followed by `\else` and `\fooB` by `\fi`). In the second case, we define `\foo` conditionally, and it can read its arguments without problem.

You can use the following construct

```
\def\addtofoo#1{\let\oldfoo\foo\def\foo{#1\oldfoo}}
% example of use
\def\foo{A}\foo
\addtofoo{B}\foo
```

This typesets as ABA. Beware: the `\addtofoo` command can be used only once (the old value of `\oldfoo` has to be saved...). We shall see later how to replace in the definition above the `\oldfoo` by its value, using either tokens lists or `\edef`, using a method where `\oldfoo` is a temporary. This is another example:

```
\def\double#1#2{\let#1#2\def#2{#1#1}}
% example
\def\B{C}\def\C{to}\double\tmp\B
```

Here ‘B’ typesets as ‘toto’. In fact `\B` is defined as ‘`\tmp\tmp`’, where `\tmp` is the old definition of `\B`, namely a command that expands to ‘C’. If you say `\def\C{ti}\B`, you will get ‘titi’. If in `\double` the `\let` is replaced by a `\def` as `\def#1{#2}`, the expansion of `\tmp` would have been `\B`, and `\B` would have been the same as `\B\B`. You see the problem? This could provoke a stack overflow, a parameter stack overflow, or even a program crash.

Let’s mention the existence of `\futurelet\A\B\C`. It is the same as `\let\A\C\B\C`. The usefulness of such a construct will be explained later.

You can say `\expandafter\A\B`. In such a case, T_EX reads the first token, saves it somewhere, calls `expand` if possible, re-inserts the saved token. Nothing special happens if the second token (here `\B`) cannot be expanded, because it is a non-active character, or a command like `\par` or `\relax`. But assume that `\A` is a command that uses one argument (for instance `\textit`) and `\B` expands to ‘foo’. If you use `\expandafter`, only the first letter will be in italics. Assume that `\foo` expands to a dollar sign. Then `$$\foo` is an empty math formula because `\foo` is not expanded, but `\expandafter$$\foo. $$` is a display math formula with a dot. The main reason why tokens are not expanded after a dollar sign (when T_EX looks for an other dollar sign) is that a test `$$\ifmmode true\fi$` should evaluate to true. You can use `\expandafter` if you want the test to be executed outside math mode. Note: if a table contains a template of the form ‘`$$$`’, if the cell starts with `\ifmmode`, then the test is expanded (i.e. evaluated) before math mode is entered, because T_EX is looking for an `\omit` token. As a consequence you should always put ‘`\relax`’ before a test (this is not needed if a command is made “Robust”).

Look carefully at the following lines:

```
1 \def\toto{\titi!}\def\titi{\tata}\def\tata{\tutu}
2 \expandafter\expandafter\expandafter\def\toto{5}
```

⁷The reference count of the token list of the body of `\fooA` or `\fooB` must also be increased.

```

3 \let\E\expandafter \E\E\def\toto{6}
4 \def\E{\expandafter} \E\E\def\toto{7}
5 \expandafter\def\toto{8}

```

On the first line we define three commands `\toto`, `\titi` and `\tata`. As we shall see, lines 2, 3 and 4 do not change the meaning of `\toto`, so that, on line 5, the expansion of ‘`\toto`’ is ‘`\titi!`’. In this case, the effect of the `\expandafter` is to replace ‘`\toto`’ by ‘`\titi!`’. Hence, line 5 defines a macro `\titi`, that has to be followed by an exclamation point, takes no argument, and expands to 8. Consider now line 2. The first `\expandafter` puts apart the `\expandafter` token; it expands the next token, which is `\expandafter`, and the expansion of this is: read the token that follows (here ‘`\def`’), and expand the token that follows. This is ‘`\toto`’, that expand to ‘`\titi!`’. If we pop back the two tokens, line 2 is equivalent to ‘`\expandafter\def\titi!{5}`’. This looks like line 5, so that it is the same as ‘`\def\tata!{5}`’. There is no difference between lines 2 and 3: the `\E` command behaves exactly like `\expandafter`. Consider now line 4. What `TEX` does is expand the first token. It is `\E`, it expands to ‘`\expandafter`’. Since the token can be expanded, it will. Thus `TEX` reads and remembers the token that follows. It expands the next token (the third ‘`\E`’). Its expansion is ‘`\expandafter`’. Hence, line 4 is equivalent to ‘`\E\expandafter\def\toto{7}`’. Now, the `\E` in this list has as effect to try to expand the second token; it is `\def`, which cannot be expanded. Hence this ‘`\E`’ is useless. Line 4 is equivalent to ‘`\expandafter\def\toto{7}`’. And this defines `\titi`. We give here the trace of `TEX` (it is a bit more complete than the trace of `TEX`):

```

\E ->\expandafter
{\expandafter \E \E}
\E ->\expandafter
\E ->\expandafter
{\expandafter \expandafter \def}
{\expandafter \def \toto}
\toto ->\titi !
{\def}
{\def \titi !->7}

```

A question is : how many commands with two characters can be defined in `TEX`? The answer is 255 squared (all characters but the null character are allowed). Of course, if you say ‘`\def{++}`’, this defines the ‘`+`’ command not the ‘`++`’. You could imagine to change category codes (but, in a construction like ‘`\def\{-}`’, it is impossible to give a different role to the first and second opening brace). The solution is given by `\csname`, you can use it like this ‘`\csname1+1=2\endcsname`’. Note that this typesets nothing: when `\csname` manufactures a new control sequence name, it defines it as being `\relax` (the control sequence will exist, until the end of the job). You can hide the `\csname` command, like this

```

\def\nameuse#1{\csname #1\endcsname}
\nameuse{1+1=2}

```

If you want to define such a beast, you must use `\expandafter`.

```

\def\namedef#1{\expandafter\def\csname #1\endcsname}
\namedef{1+1=2}{true}

```

You can also say `\namedef{++}#1{#1+#1}` followed by `\nameuse{++}{3}`. This should give 3+3. If you want a macro named `\{}`, you can say `\nameuse{\string\{\string\}}`, provided that `\escapechar=-1`. If you do not like this setting of `\escapechar`, you can define a command, say `\Lbra`, that expands to `{12}` (an inactive opening brace character) using whatever method seems best. For instance

```

{\escapechar=-1 \xdef\Lbra{\string\{\}\xdef\Rbra{\string\}}
\namedef{\Lbra\Rbra}{Hey}

```

We explained above what happens when three `\expandafter` come in a row. Thus, it should not surprise you that the following command defines `\foo`.

```
\expandafter\expandafter\expandafter\def\nameuse{foo}{12}
```

A more realistic example of `\csname` is

```
\def\allocate#1{...}
\def\newcount#1{\allocate{ctr}\countdef#1\allocationnumber}
\def\newcounter#1{\expandafter\newcount\csname c@#1\endcsname}
```

Let's assume that `\allocate` takes as argument a type (for counters, dimensions, skip registers, muskip registers, box registers, token registers, input registers, output register, math families, language codes, insertions, etc), associated to `\newcount`, `\newtoks`, `\newbox`, `\newdimen`, `\newskip` (plus some other commands, not implemented in Tralics). It allocates a unique number depending on the type, and puts it in `\allocationnumber`. For instance, if you say `\newcount\Foo`, the number could be 110, if you say `\newskip\Bar`, the number could be 46. In the first case, the result is as if you had said `\countdef\Foo110`. In the case of `\newcounter{foo}`, the result is as `\newcount\c@foo111`. Note that there are only 256 count registers available in T_EX. You can use registers zero to nine as scratch registers (Do not forget that `\count0` contains the current page number), L^AT_EX uses registers 10 to 21 for its allocation mechanism. In the current version, the first free counter is 79. Some other counters are allocated by the class, and the package (in the transcript file, one line is printed for every call to `\allocate`, for instance: `\c@chapter=\count80`; in Tralics, the line looks like `{\countdef \c@foo=\count18}`).

A very important point is that all tokens between `\csname` and `\endcsname` are fully expanded. It is an error if a non-character token remains. Thus it is important to know which commands are expanded, and those that cannot be expanded. The exact rules are in the T_EXbook, chapter 20. As a rule of thumb, commands that do no typesetting and modify no internal table can be expanded. More precisely: user defined commands, conversions like `\string`, `\number`, conditionals like `\fi`, marks, and some special commands like `\csname`, `\expandafter`, `\the` can be expanded. A construction like `\csname\char'A\endcsname` is invalid.

If you say `\noexpand\foo`, the result is `\foo`, unexpanded. Example:

```
1 \def\F00{12}
2 %\csname\noexpand\F00\endcsname %bad
3 \edef\xbar{\noexpand\F00}
4 \noexpand\F00
5 \expandafter\textit\F00
6 \expandafter\textit\noexpand\F00
7 \count0=1\F00
8 \count0=1\noexpand\F00
```

Line two is an error: the no-expanded `\F00` is not a character. On line 3, the body of `\xbar` is '`\F00`', it will be expanded later. The translation of line 4 is empty (the command `\F00` is temporarily seen as `\relax`, and `\relax` does nothing). Because of the `\expandafter`, the argument of `\textit` on line 5 is 1, on line 6 it is 12. On line 7, 112 is put in `\count0`, because `\F00` is expanded. On line 8, 1 is put in the register, and 12 is typeset. On lines 8 and 6, `\F00` is expanded twice, the first expansion being inhibited by the `\noexpand`.

Some quantities are never expanded, for instance `\lowercase` (this is black magic), `\def` (more generally all assignments), `\relax` (it does nothing, but stops scanning integers, dimensions, glue, etc), `\hbox`, `\par`⁸, `\left`, etc. Be careful with constructs like `\csname\endcsname`: L^AT_EX may signal an error involving `unhbox`.

A command can be defined via `\edef` instead of `\def` (`\xdef` is the same as `\edef`, with an implicit `\global` prefix). All tokens in the body of the definition are expanded. Example:

⁸Unless redefined by L^AT_EX


```

\def\A{\B\C} \def\C{1}
\def\Bgroup{\iffalse}\fi}\def\Egroup{\iffalse{\fi}}
{\let\B\relax \global\edef\D\bgroup{\A\noexpand\C\egroup}}
{\let\B\relax \global\edef\E\Bgroup{\A\noexpand\C\Egroup}}

```

In this example, we consider two groups, that define (locally) a command `\B` and (globally) two commands `\D` and `\E`. The difference between these two commands is that `\bgroup` is an implicit character: when evaluated, it behaves like an opening brace, but it cannot be expanded. On the other hand, `\Bgroup` expands to an open brace. The `\edef` expands tokens following an explicit opening brace. It stops reading after having found an explicit closing brace (resulting from the expansion of `\Egroup`, not `\egroup`). The expansion of `\A` is `\B\C`, this is expanded again. Since `\B` is `\relax`, it cannot be expanded, and is left unchanged. The expansion of `\C` is `1`, so that the full expansion of `\A` is `\B1`. The expansion of `\noexpand\C` is `\C`. Thus, the example is equivalent to

```

\global\def\D\bgroup{\B1\C\egroup}
\global\E\Bgroup{\B1\C}

```

You can put three `\noexpand` in a row followed by some token `X`. After the first expansion, the result is `\noexpand` followed by `X`, after the second expansion, the result is `X`. In the example that follows, the value of `\B` is `\xbar`.

```

\def\xbar{xbar}
\edef\A{\noexpand\noexpand\noexpand\xbar}
\edef\B{\A}

```

Consider a realistic example like this

```

\def\add#1#2{\edef#1{#1\do{#2}}}
\def\cons#1#2{\begingroup\let\@elt\relax\xdef#1{#1\@elt #2}\endgroup}

```

We can say something like

```

\def\A{\def\B{}} %init
\let\do\relax% just in case
\add\A x, \add\A y, \add\A z,
\cons\B{ab}, \cons\B{cd}, \cons\B{ef}.
\show\A\show\B

```

This gives two ways to add some tokens to a list. Because both commands use `\edef`, full expansion is in use; you have to be very careful if the tokens contain macros that can be expanded. For the case of `\add`, we assume that `\do` does nothing; for the case `\cons`, the command resets `\@elt` to `\relax`. The body of `\A` will be `\do{x}\do{y}\do{z}` and the body of `\B` will be `\@elt ab\@elt cd\@elt ef`. Note the absence of braces: if you really need them, you should add them to the major argument of the `\cons` command.

The major problem with `\edef` is that it is not aware of assignments. Assume that `\def\@A\B{}`, and `\def\C{\let\@A\D}`, `\def\E{C}` have been somehow evaluated. Consider now an `\edef` containing `\E`. This implies expansion of `\C`, hence of `\let\@A\D`. The `\let` command cannot be expanded. Hence `\@A` is expanded, and you get the following error: *Use of \@A doesn't match its definition* from inside `\C`. You have never heard of this command `\@A`, and never used `\C`! For this reason some commands are made robust: for instance `\foo` expands to `\protect\foo` (the second command here has a space at the end), and `\protect` is defined to be `\relax`, or `\noexpand`, and sometimes `\string`. This mechanism works only if you use `\protected@edef` instead of `\edef`. (Note: `\protect` behaves like `\string` inside `\protected@write`, which is a variant of `\write`).

2.4 Variables in T_EX

By variable, we mean everything that the user can modify or watch changing. For instance, the current hash table usage is not a variable (it varies, of course, but the value is available only at the end of the run, in the transcript file). The current vertical list is updated whenever a paragraph is split into lines; you cannot access this list, however the `\output` routine gets the part of it that should be typeset on the current page in the box register 255. There are general purpose variables, and specialised ones: for instance `\spacefactor` makes sense only in horizontal mode, and the height of the material on current page (`\pagetotal`) can be used only between paragraphs (in fact, it is updated by T_EX whenever a line is added to the page; you can consult, and even modify, this value at any time). There are variables that you cannot modify (the version number, for instance) or only once (the magnification), or in the preamble (i.e., L^AT_EX reads some variables at begin-document, changes done later to these variables are ignored).

Variables can be classified into two categories depending on their use: in some cases you need to put a prefix before `\foo` if you want to use it, in other cases, if you want to modify it. For instance, if `\foo` is a user-defined command, you say `\let\foo`, or `\def\foo`, if you want to change the value, and simply `\foo` if you want to use it. In the same fashion `\font\tenrm` defines a font, and `\tenrm` is a use. On the other hand, if you say `\pageno=3`, this will set the current page number to 3 (this is plain T_EX syntax, the L^AT_EX syntax will be explained later). If you say something like `\hskip-\fontdimen2\font`, the `\hskip` command is a prefix that says that the variable that follows will be used. In this case, this is some dimension from a font. Note that `\fontdimen` is a prefix so that `\font` does not define a new font, but refers to the current font. The meaning of the above piece of code is: insert horizontal space, whose amount is the opposite of the second parameter of the current font (i.e., normal interword space).

The value of a variable can be

- an integer (32bit, signed, with magnitude less than 2³¹). The value can be restricted in some cases (to 0-255 if it is an index in a table of registers, to 0-255 if it is a character, etc).
- a dimension, often expressed in pt, (an integer number of times the small unit sp). Normally, the maximum value is 2¹⁴pt, but T_EX does not always check for overflow.
- a glue. Called rubber length in L^AT_EX. It is like a dimension with a stretch part, and a shrink part.
- a muglue. Like glue, but only one unit of measure is allowed: mu (math unit).
- a token list. This is a list of tokens (as always, well balanced against explicit braces).
- a font.
- a box (a box contains characters, rules, boxes, penalties, glue, whatsit, etc, but no commands). In Tralics a box contains XML stuff.

You can say ‘`\afterassignment\foo\count0=3`’; in this case, the command `\foo` is pushed on a special stack, and popped after assignment is complete. There is only room for one token on this special stack. For instance, if you write the following:

```
\def\fooA{\relax}\def\fooB{\relax}\def\fooC{\relax}\def\fooD{\relax}
\afterassignment \fooA\afterassignment\fooB
\fooC\count0=1\fooD
```

the transcript file of Tralics will contain (in verbose mode)

```
[9] \afterassignment \fooA\afterassignment\fooB
{\afterassignment}
{\afterassignment: \fooA}
{\afterassignment}
{\afterassignment: \fooB}
```

At this point, the after assignment stack contains `\fooB`. The order of evaluation is now the following: `\fooD` is expanded; this gives `\relax`, which terminates scanning of the number; it will be read again, after evaluation of `\fooB`:

```
[10] \fooC\count0=1\fooD
\fooC ->\relax
{\relax}
{\count}
+scanint for \count->0
\fooD ->\relax
+scanint for \count->1
{after assignment: \fooB}
\fooB ->\relax
{\relax}
{\relax}
```

You can use the `\showbox` command for displaying the content of a box. This is a little example. It uses `\everyhbox` and `\afterassignment`. Note the order in which these tokens are inserted.

```
\everyhbox{3}
\def\foo{12}
\afterassignment\foo\setbox0=\hbox{4}
\showbox0
```

This is what \TeX prints in the log file:

```
> \box0=
\hbox(6.4151+0.0)x19.99512
.\T1/cmr/m/n/10 1
.\T1/cmr/m/n/10 2
.\T1/cmr/m/n/10 3
.\T1/cmr/m/n/10 4
```

The first line of the trace starts with `\hbox` or `\vbox`, followed by the dimensions (height, depth, width; the unit is ‘pt’ by default), optionally followed by ‘shifted 27.1’ if the the box is shifted, and by ‘glue set 0.19’ if the glue has to be stretched or shrunk. After that, you will see the content of the box, one line per item (no more than `\showboxbreadth` lines are printed per box), each item is preceded by a context (a sequence of N dots at depth N , tokens at depth greater than `\showboxdepth` are not shown). In the box, you can see things like ‘`\penalty -51`’ or ‘`\kern 28.45274`’ or ‘`\glue 3.0 plus 1.0`’ or ‘`\glue(\baselineskip) 2.28015`’ (this last glue is inserted automatically by \TeX , it knows where it comes from, so that the name can be printed), `\special{...}`, `\write4{\indexentry...}`. The interesting point in the last object is that we have a list of tokens that will be evaluated later (when the page is shipped out). \Tralics does not put `\kern`, `\penalty`, neither `\glue` in a box. The `\special` command is not implemented; finally `\write` is never delayed. In our example, the box contains four items, which are characters (\TeX shows a command that contains the name of the font; in our example, the font is something like ‘`ecrm1000`’).

In \Tralics , you would see the same characters, but no font and no size. On the other hand, you can say something like

```
\everyxbox{Test}
\setbox0=\xbox{foo}{1\xbox{bar}{2} %
\AddAttToLast{x}{1}\AddAttToCurrent{y}{2}3}
\showbox0
```

and you will see

```
<foo y='2'>Test1<bar x='1'>Test2</bar> 3</foo>
```

Note the two commands that were used to add attributes to the current XML elements, and the last constructed one. We have added another command, `\XMLaddatt` that takes as optional argument the id of the element to which the attribute value pair should be added. This is an integer; if omitted, the current element is used. You can use `\XMLlastid` or `\XMLcurrentid` (there are references to variables, you must use `\the` if you want the value). If you want to overwrite an existing attribute pair, you must use a star. The previous example can be written like this:

```
\everyxbox{Test}
\setbox0=\xbox{foo}{1\xbox{bar}{2} %
  \XMLaddatt[\the\xMLlastid]{x}{1}\XMLaddatt[\the\xMLcurrentid]{y}{22}%
  \XMLaddatt[\the\xMLlastid]{x}{11}\XMLaddatt*{y}{2}3}
\showbox0
```

If `\foo` is any command then `\show\foo` will show its value. Here are some examples

```
\def\Bar#1#{\#1} \show\Bar
\let\foo\par \show\foo
\renewcommand\foo[2][toto]{#1#2} \show\foo
\let\foo=1 \show\foo
\let\foo=_ \show\foo
\let\foo=\undef \show\foo
\show\bgroup
```

This is what Tralics prints (it differs slightly from the L^AT_EX output)

```
\Bar=macro: #1#->#1.
\foo=\par.
\foo=opt \long macro: toto#2->#1#2
\foo=the character 1.
\foo=subscript character _.
\foo=undefined.
\bgroup=begin-group character {.
```

In the case of a variable, you can say `\the\foo`, the result is a token list that represents the value of `\foo` (if `\foo` is a token list, `\the\foo` is the value of `\foo`, otherwise, it is a list of characters). The command `\showthe` will show the value, i.e. print on the terminal the token list returned by `\the`. Example

```
\def\Show#1{\the#1\showthe#1}
\widowpenalty=3 \Show\widowpenalty
\parindent1.5pt \Show\parindent
\leftskip = 1pt plus 2fil minus 4fill \Show\leftskip
\thinmuskip = 3mu plus -2fil minus 4fill \Show\thinmuskip
\count0=17 \Show{\count0}
\dimen0=17pt \Show{\dimen0}
\skip0=17pt plus 1 pt minus 2pt \Show{\skip0}
\muskip0=17mu plus 1 mu minus 2mu \Show{\muskip0}
\Show{\catcode'\A}
\Show{\lccode'\B}
\Show\inputlineno
\font\xa=cmr10 at 11truept
\fontdimen6\xa = 11pt \hyphenchar\xa='\-
\Show{\fontdimen6\xa}
\Show{\hyphenchar\xa}
\chardef\foo25
\Show\foo
\Show\xa
\toks0={\foo = \foo} \def\foo{foo}
\Show{\toks0}
```

This is what Tralics prints on the screen. In the case of muglue, the output should use mu instead of pt.

```

\show: 3
\show: 1.5pt
\show: 1.0pt plus 2.0fil minus 4.0fill
\show: 3.0pt plus -2.0fil minus 4.0fill
\show: 17
\show: 17.0pt
\show: 17.0pt plus 1.0pt minus 2.0pt
\show: 17.0pt plus 1.0pt minus 2.0pt
\show: 11
\show: 98
\show: 333
\show: 11.0pt
\show: 45
\show: 25
\show: cmr10
\show: \foo = \foo

```

The typeset result is: *31.5pt0.0pt0.0pt1717.0pt17.0pt plus 1.0pt minus 2.0pt17.0pt plus 1.0pt minus 2.0pt119833411.0pt 45 25cmr10 foo= foo.*

In the case of `\the\foo`, `\showthe\foo`, `\advance\foo`, `\multiply\foo`, `\divide\foo`, the token that follows the first command is fully expanded.

2.5 All the variables

All variables (exceptions will be given later) are in the table of equivalents: this table contains the current meaning of quantities that are saved/restored by the grouping mechanism of T_EX. In T_EX this table is divided into six parts; in Tralics, the layout is slightly different, for instance, because T_EX makes a heavy using of glue (each space character produces a glue item), while Tralics ignores them completely. This big table contains the following objects

1. the current equivalent of single character control sequences (for `~` as well as `\~`);
2. the hash table (in Tralics, there are two such tables, if the command `\foo` produces `<bar gee='true'>`, the three strings 'bar', 'gee' and 'true' are in a special table).
3. all glue parameters.
4. all quantities that fit on 16 bits.
5. all integers.
6. all dimensions.

The glue parameters are the following (unused by Tralics, initialised to 0, except `\fill`).

- `\lineskip`: interline glue if `\baselineskip` is infeasible.
- `\baselineskip`: desired glue between baselines.
- `\parskip`: extra glue just above a paragraph.
- `\abovedisplayskip`: extra glue just above displayed math.
- `\belowdisplayskip`: extra glue just below displayed math.
- `\abovedisplayshortskip`: glue above displayed math following short lines.
- `\belowdisplayshortskip`: glue below displayed math following short lines.
- `\leftskip`: glue at left of justified lines.

- `\rightskip`: glue at right of justified lines. L^AT_EX uses `\leftskip` and `\rightskip` for commands like `\centering`, `\raggedright` etc. Unused by Tralics. On the other hand there is `\nocentering`, whose effect is the same as setting both `\leftskip` and `\rightskip` to zero.
- `\topskip`: glue at top of main pages.
- `\splittopskip`: glue at top of split pages.
- `\tabskip`: glue between aligned entries.
- `\spaceskip`: glue between words.
- `\xspaceskip`: glue after sentences.
- `\parfillskip`: glue on last line of paragraph.
- `\thinmuskip`: thin space in math formula.
- `\medmuskip`: medium space in math formula.
- `\thickmuskip`: thick space in math formula.
- `\itemsep`: defined by L^AT_EX. Rubber space between successive items in a list.
- `\labelsep`: defined by L^AT_EX. The space between the end of the label box and the text of the item in a list.
- `\parsep`: defined by L^AT_EX. Rubber space between paragraphs within an item.
- `\fill`: defined by L^AT_EX. Holds 0pt plus 1fill. You should not modify it.

The token parameters are the following (initially empty; unused by Tralics unless stated otherwise):

- `\parshape`: for funny paragraphs (not really a token list).
- `\output`: user defined output routine.
- `\everypar`: tokens inserted by T_EX at start of every paragraph.
- `\everymath`: tokens inserted by T_EX and Tralics at the start of every non-display math formula.
- `\everydisplay`: tokens inserted by T_EX and Tralics at the start of every display math formula.
- `\everyhbox`: tokens inserted by T_EX and Tralics at the start of every `\hbox`.
- `\everyvbox`: tokens inserted by T_EX and Tralics at the start of every `\vbox`.
- `\everyxbox`: tokens inserted by Tralics at the start of every `\xbox`.
- `\everyjob`: tokens inserted by T_EX and Tralics at the start of every job. This must be defined by the format (for T_EX), otherwise it is useless; in Tralics, you must put `\everyjob = "\something{like this}"` in the configuration file.
- `\everycr`: tokens inserted by T_EX after every `\cr` or non redundant `\crrc`.
- `\errhelp`: tokens that will be printed by T_EX in case of user-error.

The integer parameters are the following. These parameters are zero, unless stated otherwise.

- `\pretolerance`: badness tolerance before hyphenation (initialised to 100 by Tralics).
- `\tolerance`: badness tolerance after hyphenation (initialised to 200 by Tralics).
- `\linepenalty`: amount added to the badness of every line in a paragraph.
- `\hyphenpenalty`: penalty for break after discretionary hyphen.
- `\exhyphenpenalty`: penalty for break after explicit hyphen.
- `\clubpenalty`: penalty for creating a club line at a bottom of a page.
- `\widowpenalty`: penalty for creating a widow line at top of page.
- `\displaywidowpenalty`: ditto, just before a display.
- `\brokenpenalty`: penalty for breaking a page at a broken line.
- `\binoppenalty`: penalty for breaking after a binary operation in a math formula.
- `\relpenalty`: penalty for breaking after a relation in a math formula.

- `\predisdisplaypenalty`: penalty for breaking just before a displayed formula.
- `\postdisplaypenalty`: penalty for breaking just after a displayed formula.
- `\interlinepenalty`: additional penalty for a page break between lines.
- `\doublehyphendemerits`: demerits for consecutive broken lines.
- `\finalhyphendemerits`: demerits for a penultimate broken line.
- `\adjdemerits`: demerits for adjacent incompatible lines.
- `\mag`: magnification ratio, times 1000.
- `\delimiterfactor`: ratio for variable-size delimiters.
- `\looseness`: change to the number of lines in a paragraph.
- `\time`: current time of day. Number of minutes since midnight, computed by Tralics at start of run.
- `\day`: current day of the month (between 1 and 31).
- `\month`: current month of the year (between 1 and 12).
- `\year`: current year of our Lord. The initial values of `\time`, `\day`, `\month`, `\year`, are printed in the transcript file by Tralics in the following format: 9/12/2004 19:03:56
- `\showboxbreadth`: maximum items per level when boxes are shown (when Tralics shows the content of a box, it always shows everything).
- `\showboxdepth`: maximum level when boxes are shown (when Tralics shows the content of a box, it always shows everything).
- `\hbadness`: hboxes exceeding this badness will be shown.
- `\vbadness`: vboxes exceeding this badness will be shown.
- `\pausing`: pause after each line is read from a file. In Tralics there is no interaction with the user.
- `\tracingonline`: show diagnostic output on terminal. In verbose mode, this variable, and some other ones are set to a non-zero value, as explained in section 6.5.
- `\tracingmacros`: show macros as they are being expanded.
- `\tracingstats`: show memory usage if T_EX knows it.
- `\tracingparagraphs`: show line-break calculations.
- `\tracingpages`: show page-break calculations.
- `\tracingoutput`: show boxes when they are shipped out.
- `\tracinglostchars`: show characters that aren't in the font.
- `\tracingcommands`: show command codes.
- `\tracingrestores`: show equivalents when they are restored.
- `\uchyph`: hyphenate words beginning with a capital letter.
- `\outputpenalty`: penalty found at current page break.
- `\maxdeadcycles`: bound on consecutive dead cycles of output.
- `\hangafter`: hanging indentation changes after this many lines.
- `\floatingpenalty`: penalty for insertions heldover after a split.
- `\globaldefs`: override `\global` specifications.
- `\fam`: current family.
- `\escapechar`: escape character for token output. Initialised by Tralics to backslash.
- `\defaultshyphenchar`: value of `\hyphenchar` when a font is loaded.
- `\defaultskewchar`: value of `\skewchar` when a font is loaded.
- `\endlinechar`: character placed at the right end of the buffer when reading a new line. Initialised by Tralics to CR (ascii 13).
- `\newlinechar`: character that prints as a LF. Initialised by Tralics to LF, but not used.
- `\language`: the current set of hyphenation rules. For Tralics, 0 means english, 1 means french, 2 means german, and 3 stands for any other language.

- `\lefthyphenmin`: minimum left hyphenation fragment size.
- `\righthyphenmin`: minimum right hyphenation fragment size.
- `\holdinginserts`: do not remove insertion nodes from `\box255`.
- `\errorcontextlines`: maximum intermediate line pairs shown. In Tralics, the context of an error is not shown.
- `\lastpenalty`: read only variable; it contains the value of the current `\penalty` (always zero in Tralics).
- `\badness`: read only variable; it contains the current badness (always zero in Tralics).
- `\inputlineno`: read only variable; it contains the current input line number.
- `\XMLlastid`: Read only variable. It contains the unique identifier of the most recently created XML element.
- `\XMLcurrentid`: Read only variable. It contains the unique identifier of the current XML element.
- `\FPseed`. This is defined only when the FP package is loaded.

The counters defined in Tralics are the following. The counters are not used, but you can say `\renewcommand\thepage{...}`, this is not an error.

- `page`. This is `\count0`. Tralics initialises it to 1.
- `enumi`, `enumii`, `enumiii`, `enumiv`: for enumerations.
- `part`
- `subsection`
- `subsubsection`
- `paragraph`
- `subparagraph`.

The dimension parameters are the following:

- `\parindent`: indentation of paragraphs.
- `\mathsurround`: space around math in text.
- `\lineskiplimit`: threshold where `\baselineskip` switches to `\lineskip`.
- `\hsize`: line width in horizontal mode.
- `\vsize`: page height in vertical mode.
- `\maxdepth`: maximum depth of boxes on main pages.
- `\splitmaxdepth`: maximum depth of boxes on split pages.
- `\boxmaxdepth`: maximum depth of explicit vboxes.
- `\hfuzz`: tolerance for overfull hbox messages.
- `\vfuzz`: tolerance for overfull vbox messages.
- `\delimitershortfall`: maximum amount uncovered by variable delimiters.
- `\nulldelimiterspace`: blank space in null delimiters.
- `\scriptspace`: extra space after subscript or superscript.
- `\prelplaysize`: length of text preceding a display.
- `\displaywidth`: length of line for displayed equation.
- `\displayindent`: indentation of line for displayed equation.
- `\overfullrule`: width of rule that identifies overfull hboxes.
- `\hangindent`: amount of hanging indentation.
- `\hoffset`: amount of horizontal offset when shipping pages out.
- `\voffset`: amount of vertical offset when shipping pages out.
- `\emergencystretch`: reduces badnesses on final pass of line-breaking.
- `\z@`: You should forget that this is a variable, and use it only as the constant zero.

- `\evensidemargin`: defined by L^AT_EX, left margin for even pages.
- `\oddsidemargin`: defined by L^AT_EX, left margin for odd pages.
- `\leftmargin`: defined by L^AT_EX. Space between the left margin of the page and the left margin of the text. Depends on the list level.
- `\rightmargin`: defined by L^AT_EX. Similar to `\leftmargin`, but for the right margin.
- `\leftmargini`: defined by L^AT_EX. Value of left margin for a list at level one.
- `\leftmarginii`: defined by L^AT_EX. Value of left margin for a list at level two.
- `\leftmarginiii`: defined by L^AT_EX. Value of left margin for a list at level three.
- `\leftmarginiv`: defined by L^AT_EX. Value of left margin for a list at level four.
- `\leftmarginv`: defined by L^AT_EX. Value of left margin for a list at level five.
- `\leftmarginvi`: defined by L^AT_EX. Value of left margin for a list at level six.
- `\itemindent`: defined by L^AT_EX. Extra indentation added to the horizontal indentation of the text part of the first line of an item in a list.
- `\labelwidth`: defined by L^AT_EX. Nominal width of the box containing the label of an item.
- `\fboxsep`: defined by L^AT_EX. Space left between the edge of the box and its content produced by `\fbox` or `\framebox`.
- `\fboxrule`: defined by L^AT_EX. Width of line produced by `\fbox` or `\framebox`.
- `\epsfxsize`. Used for including images.
- `\epsfysize`. Used for including images.
- `\unitlength`. Used for the picture environment. Initialised by Tralics to 1pt.
- `\textwidth`. Width of the text. Initialised by Tralics to 427pt. More or less 15cm.
- `\textheight`. Height of the text. Initialised by Tralics to 570pt. More or less 20cm.
- `\columnwidth`: this is defined by L^AT_EX as containing the current line width; in a two-column document this is the half of `\textwidth` minus `\columnsep`, otherwise it is `\textwidth`. L^AT_EX copies this value in `\hsize` in some cases (for instance, when switch between one column and two columns).
- `\linewidth`: this is defined by L^AT_EX as containing the current line width (typically, this is `\columnwidth` minus the margins introduced by list environment). In Tralics, these two commands are undefined, except that you can use it as a unit of measure inside the optional argument of `\includegraphics` for ‘height’ and ‘width’: the value is 15cm.

The registers are the following

- `\skip xxx`: table of 256 “skip” registers.
- `\muskip xxx`: table of 256 “muskip” registers.
- `\toks xxx`: table of 256 token lists.
- `\box xxx`: table of 256 box registers.
- `\wd xxx`: Width of box N. If you ask for the value of the width, you will get zero. If you modify the width, nothing happens.
- `\ht xxx`: Height of box N.
- `\dp xxx`: Depth of box N.
- `\delcode xxx`: table of 256 delimiter code mappings. Unused by Tralics.
- `\catcode xxx`: table of 256 category codes. In Tralics, `\{ } $ & # ^ _ ~ %` have standard category codes. All letters have category code 11. Space, tabulation, non-breaking space (character 160) have category 10. All other characters have category 12.
- `\lccode xxx`: table of 256 lowercase equivalents.
- `\uccode xxx`: table of 256 uppercase equivalents. In Tralics, iso-8859-1 is assumed. Thus the lc-code of \hat{I} is \hat{i} , the uc-code of \hat{i} is \hat{I} . More precisely, the lccode of 206 is 238, and vice-versa.
- `\sfcode xxx`: table of 256 spacefactor mappings. Not used by Tralics.
- `\mathcode xxx`: table of math mode mappings. Not used by Tralics.

2.6 Using the variables

There are three routines defined in Tralics, named `scanint`, `scandimen` and `scanglue` that read a integer, a dimension and glue. Assume that `\count0` is 1, `\parindent` is 3pt, and you say `\skip\count0=2pt plus \parindent\relax`. The transcript file of Tralics will contain

```
[346] \skip\count0=2pt plus \parindent \relax
{\skip}
+scanint for \count->0
+scanint for \skip->1
+scanint for \skip->2
+scandimen for \skip ->2.0pt
+scandimen for \skip ->3.0pt
{scanglue 2.0pt plus 3.0pt}
{\relax}
```

The exact rules will be given later. The following happens here: After `\skip` there is an integer, an optional equals sign, then glue. After `\count` there has to be an integer. Thus, `scanint` reads an integer for `\count`, and an other one for `\skip`. A glue item is formed of a dimension (the natural width), optionally followed by ‘plus’ and a dimension (the stretch part), optionally followed by ‘minus’ and a dimension (the shrink part). In this case, there is no stretch part, because of `\relax`. The second dimension comes from the variable `\parindent`; the first dimension is explicit: the integer part of the dimension is read by `scanint`.

An integer can be explicit or implicit: an implicit integer comes from a command (it can be a variable like `\date`, or a constant like `\active`). In all other cases, the number can be followed by one optional space. In general, the number will be given as a non-empty sequence of digits, like 01239; you can specify digits in base 16 as "FF, this is 255, in this case, letters between A and F (uppercase, category 11 or 12) are allowed. You can specify digits in base 8 as '177, this is 127. You can also specify a digit as a character: 'A is 65. You can say '\A, this is also 65; note that a backslash is needed in cases like '\%. Only one letter is allowed in the command name, digits and quotes must have category 12.

An integer or a dimension can be preceded by a sign. This is a sequence of arbitrary length, formed of spaces or `+12` or `-12` signs. If the number of minus signs is odd, this changes the sign of the result. Hence if you say `\count0=+-+'77` and `\count1=-\count0`, this will put 63 in `\count1`.

A dimension can be implicit or explicit. You can say `\count0=\dimen0`: in this case the value of the dimension in sp units is put in the count register. You can say `\dimen0=\skip0`: the shrink and stretch part of the glue is ignored. You can also say `\count0=\skip0` (guess what happens). It is not possible to convert (directly) an integer to a dimension or glue. An explicit dimension is formed of a factor and a unit of measure. The factor can be an integer (hence `-'77pt` is a valid dimension), or a decimal number (given in base ten, like 1.5, or 1,5). Units can be pt, pc, in, bp, cm, mm, dd, cc, sp. The case is irrelevant: `Pt`, `pt`, `PT` and `pT` are valid, the category code may be anything (it cannot be active, because everything is fully expanded). Units shown above can be preceded by `true` (note that Tralics ignores magnification, thus the ‘true’ prefix). Units can also be `em` or `ex`. These values depend on the current font. Tralics always assumes that the font is a ten point one. A unit of measure can also be an integer, a dimension, or glue. For instance `\dimen0=1\count0` will multiply the value of `\count0` by one. This is the dual to `\count0=\dimen0`. You can say `\parindent=1.2\parindent` if you want to increase it by 20%.

A glue is formed of three parts: a dimension, a stretch part, and a shrink part. The stretch part can be a dimension (it can use special units like ‘fil’, ‘fill’ and ‘filll’, these are called infinite, of first, second and third order). You can say `\skip0=0pt plus 1fil`. For some strange reasons, after `fil` you can put a second L, and a third one. As is the case with other units like `ex` or `em`, the case is irrelevant. Spaces are ignored after the L. Moreover, T_EX continues scanning for an L

after having found ‘filll’; if found, it signals the following error: *Illegal unit of measure (replaced by filll)*. In the case of `\skip0=2\skip0`, the equals sign is followed by a dimension: there is a factor 2, and a unit (the fixed part of `\skip0`). As a consequence, this multiplies by two the fixed part of the glue, and sets the shrink and stretch part to zero (unless the code above is followed by ‘plus’ or ‘minus’).

Note: if you say `\chardef\foo=123\foo`, then `\foo` is made equal to 123: the first thing that `\chardef` does it to make `\foo` the same as `\relax`, so that scanning stops after digit 3. On the other hand in the case of `\count0=3\ifnum...` the conditional is evaluated while reading the number, thus before the assignment is complete. In particular, if the test compares `\count0` with something else, the value might be different from three. Assume that `\count0` and `\count13` contain the value 7. What happens if you say: `\count0=2\ifnum\count0=\count13\fi4`? It will put 2 in `\count0` and typeset 4. In fact, after the digit 3 is sensed, the `\fi` token terminates the `\ifnum`. It does so by inserting a `\relax` token, and a second `\fi` token. The effect of `\relax` is to finish reading the number. Thus `\ifnum` can compare the two values. If these two values are different, the expansion of the conditional is empty, and 24 is put in `\count0`. But the test is true, and `TEX` reads again the inserted `\relax`: it has as effect to stop scanning of the number 2. After that the inserted `\fi` is read. The transcript file of `Tralics` might look like this (note the two `\fi3532`):

```
[3506] \count0=2\ifnum\count0=\count13\fi4
{\count}
+scanint for \count->0
+\ifnum3532
+scanint for \count->0
+scanint for \ifnum->7
+\fi3532
+scanint for \count->13
+scanint for \ifnum->7
+iftest3532 true
+scanint for \count->2
{\relax}
+\fi3532
Character sequence: 4 .
```

A token list is a like a command without arguments. You can say `\foo={ABC}` if you want to put something in it, and `\the\foo` if you want to use the list. The equals sign is optional. You can insert a `\relax` between the equals sign and the opening brace. In the example that follows, you can see that, after the optional equals sign, you can put as many spaces or `\relax` tokens as you like; tokens are expanded, as long as no brace is found. The last line of the example shows that the token that follows `\the` is expanded (if `\the` itself is expanded). Thus, the last line adds some tokens at the end of the list. Note the space in `\A`: without it, `TEX` would see something like `\the\toks0\the\toks0`, and the second `\the` is expanded by the scanint routine, so that this inserts in `\toks0` the content of `\toks01` followed by a sharp sign.

```
\def\myrelax{ \relax}
\def\A{\toks0 }
\A=\relax\myrelax{1#}
\A=\expandafter{\the\A \the\A}\showthe\toks0
```

The `\showthe` command prints ‘1##\the \A’, but only a single # is in the list.

We have seen on page 30 how to use `\cons` to add some tokens to a command via `\edef`. The code that follows adds tokens to a list. The command is call `\addto@hook` in `LATEX` and is long, but the body is the same.

```

\def\addtohook#1#2{#1\expandafter{\the#1#2}}
\newtoks\foo
\addtohook{\foo}{\do{A}}\addtohook{\foo}{\do{B}}\addtohook{\foo}{\do{C}}

```

The command `\newtoks` defines its argument as a reference to a token register, for instance `\toks23`. Whenever you use `\addtohook` with `\A` as first argument, it is like the assignment `\A=\expandafter{\the\A...}` shown in the previous example. Other example

```

\T\expandafter{\L}
% \xdef\L{... \the\T}
\xdef\L{\catcode\the\count@=\the\catcode\the\count@\relax\the\T}

```

Let's assume that `\L` is a parameterless command, and `\T` a reference to a token register. The first line puts the value of `\L` in `\T`. The second line explains what we do in the third one. Remember that `\xdef` expands everything in the body. All tokens are fully expanded (except that the result of `\the` is not expanded). As a result, this will put some tokens in front of `\L`. Let's explain which tokens. We assume that `\count@` is a reference to some counter, that the counter contains 65, this is the ASCII code of the upper case letter A, and we assume that the category code is 11. The first token is `\catcode`, it cannot be expanded, it will be left unchanged. The second token is `\the`. It can be expanded, the result is the value of the counter, the two characters 65. The equals sign cannot be expanded. Then comes `\the`; this expands what follows. The `\catcode` command reads a number. Because of `\the`, it reads two digits 6 and 5, and looks at the `\relax`. Note: this `\the` is useless, this example revealed a bug in Tralics. This is the log of Tralics. The last line indicates the value of `\L`:

```

[18] \xdef\L{\catcode\the\countx=\the\catcode\the\countx\relax\the\T}
{\xdef}
{\the}
{\the \countx}
\the->65.
{\the}
{\the \catcode}
{\the}
{\the \countx}
\the->65.
+scanint for \catcode->65
\the->11.
{\the}
{\the \T}
\the->\catcode 48=12\relax .
{\def \L ->\catcode 65=11\relax \catcode 48=12\relax }

```

There are some advantages in putting items in a box. For instance, if it takes a long time to translate a piece of text that will be used several times, it can save some time. A second possibility is to create a box in a given context and use it in another one (this can be used for instance to put verbatim material in a section title; not in the toc, because the toc is obtained by reading characters from a file, but the box can be used for page headings). Finally, one can put some text in a box, measure the size of the box, and do some action according to the size of the box; it is not possible to measure a box in Tralics because no typesetting is done. Note that there is a limited number of boxes (there is a limit on the number of token registers, but you can always put your token list in a macro; in the same fashion, it is always possible to store integers and dimensions into token lists, i.e., in commands). Note that, if you want to implement arithmetics on big numbers, if you represent a number $x = \sum x_k B^k$ as a sequence of commands, try to access to x_k via `\csname x\the\k\endcsname`, and parse this as an integer, then you get something inefficient. It is much more efficient to say `\fontdimen\k\x` (there is a T_EX file by Denis Roegel that computes thousands of digits of π using font tables as auxiliary memory).

2.7 Counters

The most useful registers are counters. Rather than saying ‘`\count16=0`’, at the risk of destroying variables used by other packages, you should use named counters, together with an allocation scheme. We have seen that ‘`\newcount\foo`’ does that. In \LaTeX , we can do more. If you say ‘`\newcounter{foo}[bar]`’ then a counter `foo` is defined that depends on `bar`. Let’s assume, for simplicity, that the allocation mechanism allocates count register 17. Then `\c@foo` is a reference to ‘`\count17`’. It is assumed that no package defines a command that starts with `c@`, or `p@` or `cl@`, so that `\c@foo`, `\cl@foo`, and `\p@foo` are reserved for the counter `foo`. In \LaTeX , there is a command `\value` that takes one argument and expands to `\csname c@#1\endcsname`. The same command exists in `Tralics`, but it signals an error in the case where `\c@foo` is not a reference to a count register. You can say ‘`\value{foo}=10`’, this will put 10 into the counter, you can say ‘`\the\value{foo}`’, this will typeset the value of the counter. You should not use this low-level \TeX syntax. In fact, if you say ‘`\value{foo}=10\the\value{foo}`’ this will put 103 into the counter (assuming that it contained 3). Compare this with `\parindent=10\parindent` where there is an implicit multiplication.

Assignment should be done via ‘`\setcounter{foo}{10}`’. This is the same as ‘`\global\value{foo}=10\relax`’ (plus a check that ‘`foo`’ is a counter). The `\relax` has as effect to stop scanning the number. The `\global` makes the assignment global. In the same fashion, ‘`\addtocounter{foo}{4}`’ is the same as ‘`\global\advance\value{foo}4\relax`’. You can say something like ‘`\parindent=\value{foo}\xbar`’, this puts in `\parindent` the value of `\xbar` (let’s assume it is a dimension) multiplied by the the value of the `foo` counter. If you want to typeset the value of the counter, you say ‘`\number\value{foo}`’. You can also use `\romannumeral` or `\Romannumeral` (this last command is not defined by \TeX) instead of `\number` (it has to be followed by a number, for instance `\value...`). The following commands take as argument the name of a counter, and typeset the value: `\arabic` (it gives 7), `\roman` (it gives vii), `\Roman` (it gives VII), `\alph` (it gives g), `\Alph` (it gives G), `\fnsymbol` (it gives **). The following commands: `\@arabic`, `\@roman`, `\@Roman`, `\@alph`, `\@Alph`, `\@fnsymbol` are used internally by \LaTeX . They are defined in `Tralics` for compatibility reasons. Hence ‘`\number\value{foo}`’ is the same as ‘`\@arabic\c@foo`’ and the same as ‘`\arabic{foo}`’; using `\arabic` is the best choice.

Three operations are defined: `\advance` that increments a counter (or a dimension, or a glue), `\multiply` that multiplies it by an integer, and `\divide` that divides it by an integer. In the case of integer division, \TeX divides the absolute values, and adds the required sign to the quotient (the remainder is not computed). The following piece of code puts in `\count0` the number of hours and in `\count2` the numbers of minutes (quotient of remainder of the division of `\time` by 60).

```
\count0\time
\divide\count0 60
\count2=-\count0
\multiply\count2 60
\advance\count2 \time
```

You can say `\newlength\foo`. This allocates a new skip register. You can use `\setlength` and `\addtolength`, in the same way as `\setcounter` and `\addtocounter`. However, assignments are local. Using plain \TeX syntax, you can say:

```
\dimen0=2mm\dimen1=0.2cm
\advance\dimen0 by-\dimen1
\count0=\dimen0
```

Using \LaTeX syntax

```
\newcounter{foo}\newlength\lenA\newlength\lenB
\setlength{\lenA}{2mm}
\setlength\lenB{0.2cm}
```

```

\addtolength\lenA{-\lenB}
\setcounter{foo}{\lenA}

```

After this operation, the counter foo contains 5. This means that the difference between 2mm and 0.2cm is 5sp (two thousands of a micrometer). Note: Tralics uses exactly the same algorithms as T_EX, hence produces the same results.

Appendix A.3.1 of [6] describes the calc package. It allows to write commands like that:

```

\newcommand{\printtime}{%
  \setcounter{hours}{\time/60}%
  \setcounter{minutes}{\time-(\value{hours}*60)}
  \thehours h \theminutes min}
\def\today{\ifcase\day\or
  1st\or 2nd\or 3rd\or 4th\or 5th\or
  6th\or 7th\or 8th\or 9th\or 10th\or
  11th\or 12th\or 13th\or 14th\or 15th\or
  16th\or 17th\or 18th\or 19th\or 20th\or
  21st\or 22nd\or 23rd\or 24th\or 25th\or
  26th\or 27th\or 28th\or 29th\or 30th\or
  31st\fi~\ifcase\month\or
  January\or February\or March\or April\or May\or June\or
  July\or August\or September\or October\or November\or
  December\fi\space \number\year}

```

The time is \printtime, \today.

In this case, the result of Tralics could be: ‘The time is 16h 37min, 7th December 2004.’

You can do operations on integers like this:

```

\newcounter{Ac}
\setcounter{Ac}{(1+2)*(3+4)-20}           %% \theAc=1
\addtocounter{Ac}{(1*2)+(3*-4)+(34/7)}    %% \theAc=-5

```

and on dimensions:

```

\newlength{\Bc}
\setlength{\Bc}{(1cm+2cm)*(3+4)-200mm}    %%\the\Bc=28.4526pt
% exact results should be 1.0pt
\setlength\Bc{\the\Bc*\ratio{25.4pt}{722.7pt}} %%\the\Bc=0.99985pt
\Bc=1in \setlength\Bc{\the\Bc * 100 / 7227} %%\the\Bc=0.99998pt
\Bc=1in \setlength\Bc{\the\Bc * \real{ 0.01383700013837}} %%\the\Bc=1.00018pt
\Bc=1cm \setlength\Bc{\the\Bc / \real{28.452755}} %%\the\Bc=0.99985pt
\Bc=1cm \setlength\Bc{\the\Bc * \ratio{254pt}{7227pt}} %%\the\Bc=0.99985pt
\Bc=1in \setlength\Bc{\the\Bc / \ratio{7227pt}{100pt}} %%\the\Bc=1.00018pt
\Bc=1IN \setlength\Bc{\the\Bc / \ratio{7227PT}{100PT}} %%\the\Bc=1.00018pt

```

In L^AT_EX, there is a command called \stepcounter. Its effect is to increment a counter, and reset all counters that depend on it (see example below). There is also \refstepcounter whose purpose is to define the current label. This is not implemented in Tralics (see later for how \label works). The idea is that, for a counter ‘foo’, the printed value of the label is defined by ‘\p@foo \thefoo’. Here \thefoo is normally ‘\arabic{foo}’, but the quantity can be redefined. For instance, the book class has \renewcommand \thesection {\thechapter.\@arabic\c@section} (the article class has no chapter, and does not redefine \thesection). Both book and article classes say: \renewcommand \thesubsection {\thesection.\@arabic\c@subsection}.

Here we define some counters, and make them depend on other counters.

```

\newcounter{toto}           \setcounter{toto}{10}
\newcounter{titi}[toto]    \setcounter{titi}{20}

```

```

\newcounter{tata}[titi]    \setcounter{tata}{30}
\newcounter{tutu}[toto]    \setcounter{tutu}{40}

```

Here we call `\stepcounter`. The typeset result should be 11101=11101.

```

\stepcounter{toto} % kills titi, tutu
\stepcounter{tata} %% \thetata=31,
\stepcounter{titi} %% \thetata=0 % titi=1
\stepcounter{tutu}
\thetoto\thetiti\thetata\thetutu=11101

```

The magic is accomplished by the following command:

```

\def\@addtoreset#1#2{\expandafter\@cons\csname cl@#2\endcsname {#1}}

```

The first argument is the counter to define (for instance ‘tutu’), and the second argument is the dependent counter (for instance ‘toto’). The `\@cons` command is defined like on page 30. It modifies the command `\cl@toto` by adding `\@elt{tutu}`. If you say `\stepcounter{toto}`, then L^AT_EX executes `\let\@elt\@stpelt \csname cl@#1\endcsname`. Here is a part of the transcript file of Tralics that shows what happens (you won’t see the `\csname`, because characters needed for `\c@toto` and `\cl@toto` are read and expanded only once by Tralics.)

```

[1198] \stepcounter{toto}
\stepcounter->\global \advance \c@toto 1\relax {\let \@elt \@stpelt \cl@toto }
{\global}
{\global\advance}
+scanint for \c@toto->1
{\relax}
{begin-group character {}}
+stack: level + 3 for brace
{\let}
{\let \@elt \@stpelt}
\cl@toto ->\@elt {titi}\@elt {tutu}
\@elt->\global \c@titi 0\relax
{\global}
{\global\c@titi}
+scanint for \c@titi->0
{\relax}
\@elt->\global \c@tutu 0\relax
{\global}
{\global\c@tutu}
+scanint for \c@tutu->0
{\relax}
{end-group character {}}
+stack: killing \@elt
+stack: level - 3 for brace
[1199] \stepcounter{tata}
\stepcounter->\global \advance \c@tata 1\relax {\let \@elt \@stpelt \cl@tata }
{\global}
{\global\advance}
+scanint for \c@tata->1
{\relax}
{begin-group character {}}
+stack: level + 3 for brace
{\let}
{\let \@elt \@stpelt}
\cl@tata ->
{end-group character {}}

```

```
+stack: killing \@elt
+stack: level - 3 for brace
```

2.8 Fonts

One of the question we can ask is: what does `\it` do? As explained above, this is an unofficial command, thus could be implemented to do anything. Let's assume that it is defined in L^AT_EX 2.09 compatibility mode. It is then possible to explain what happens, but it is harder to explain what Tralics should do. A software like `latex2html` (that we studied carefully when implementing the first version of Tralics in Perl) uses a lot of energy in order to translate font changes properly. It is however very difficult to tell it that `\french` is a similar command (in fact, what we wanted is more than just finding the scope of the `\french`, we also wanted French syntax rules to apply, we wanted dashes instead of bullets in lists, etc.). In this paragraph, we shall explain all the gory details concerning fonts (however, look at [6] for what is in a `.fd` file).

One big table in T_EX is the table of fonts: there are N fonts with N characters in them (currently $N = 256$, and this is a small limit, in Ω , this value is 2^{16} ; the dvi format specifies $N = 2^{32}$). A book like [6] uses lots of fonts indirectly, via inclusion of PostScript files. Note that metric files designed for Ω cannot be read by T_EX. The hyphenation algorithm considers as a word only sequences of characters from the same font (hence 256 characters per font is a hard limit). A metric file contains all that it needed for T_EX to typeset a character; it does not contain glyphs. Essentially, it contains three tables, indicating for each character its height, its depth and its width. There are two other tables, the lig/kern table, and the kern table, that indicate, for instance in the case VA that some negative space should be used to make the characters narrower, and in the case of fi to use a single glyph instead of two. There is another table (useful only for math mode) that explains how to construct, for instance, braces of various size. Finally, there are some parameters. One parameter is the design size (the design size of a ten point font is 10pt), other parameters are the slant, the width of a space (this is glue), the two values of `ex` and `em`, and extra space. Math fonts have extra parameters, see [4, appendix G]. A font has two integer parameters: hyphen char, and skew char. These values are not in the metric file: when the font is loaded T_EX uses the values of `\defaultshyphenchar` and `\defaultskewchar`. Note: Tralics does not read TFM files, it sets all parameters to zero.

You load a font by saying `\font\preloaded=cmr7 scaled \magstep4` or `\font\foo=cmr10 at 12pt`. Such a construction will read a file `cmr7.tfm` or `cmr10.tfm` and apply a scale factor (a factor 2 in the first case, and 1.2, in the second case). A font like `ecrm` exists in size 5, 6, 7, 8, 9, 10, 10.95 (`magstephalf`), 12, 14.4, 17.28, 20.74, 24.88 (`magstep 1, 2, 3, 4, and 5` respectively), 29.86 and 35.83. There are some slight differences between `cmr10` at 12pt and `cmr12` (see the T_EXbook for details). You can simply say `\font\tenrm=cmr10`. After that you use it like this `{\tenrm test}`. This gives: test. You can use `\fontdimen1\tenrm` like any dimension. For instance, using `\the` to typeset the value, we get 0.0pt for the slant, 3.33333pt plus 1.66666pt minus 1.11111pt for the interword space, 4.30554pt for the ex-height, 10.00002pt for the quad,⁹ 1.11111pt for the extra space. Parameters for the current font are: 0.0pt for the slant, 3.33333pt plus 1.66666pt minus 1.11111pt for the interword space, 4.3055pt for the ex-height, 10.0pt for the quad, 1.11111pt for the extra space. If you say

```
\fontencoding{T1}\fontfamily{cmr}\fontseries{m}%
\fontshape{n}\fontsize{10pt}{12pt}
\selectfont
```

you specify all font parameter, and you switch (from 'ec-lmr10') to the default ten point font with T1 encoding, namely 'ecrm1000'. The default font in this document uses 'lmr' as family.

⁹This is a bit more than ten points.

The parameters are now: 0.0pt for the slant, 3.33252pt plus 1.66626pt minus 1.11084pt for the interword space, 4.3045pt for the ex-height, 9.99756pt for the quad, 1.11084pt for the extra space. As you can see, they are not exactly the same. However, the glyphs are similar. The current font name can be printed via `\fontname\font`.

An important characteristic of a font is how glyphs are represented: For \TeX , this is irrelevant, since the dvi file contains only the metrics. However, the reader will see some black and white pixels (of ink on a sheet of paper, or dots on a screen, or points on a wall projected by a beamer). The metafont program produces bitmaps; if the resolution of these bitmaps is different from that required by the printing device, some interpolation, extrapolation is required (this is sometimes called ‘antialiasing’, it may involve colored pixels instead of black and white). In a PostScript file, a font can be specified via different formats, Type1, Type3, TrueType etc. The simplest format is Type3, namely bitmaps. Some software like Acrobat Reader prefer Type1 (a format in which characters are defined by small programs). There is no direct way to produce a Type1 file from a metafont file, so that not all \TeX fonts exist in Type1. For instance, the computer modern fonts (in version OT1) have been translated but not the T1 version (said otherwise, cmr10 exists in Type1 format, but not ecrm1000). In this document, we experiment the Latin Modern font family; it is very similar to Computer Modern.

In the case of a format like plain \TeX , fonts are used according to the following scheme. First you define fonts like `\tenrm`, in three sizes (thus, you define `\sevenrm`, `\fiverm`), and different variants (say `\teni`, `\tensy`, `\tenex`, `\tensl`, `\tentt`, etc). Then you say `\textfont0=\tenrm`, `\scriptfont0=...`, `\scriptscriptfont0=...`: this defines family zero. You do the same for family 1, 2, 3, etc. We shall see later how certain math symbols use a specified family, in other cases the family specified by the `\fam` variable is used (there are only 16 families available). The size of a symbol is defined by the current style (`displaystyle`, `textstyle`, `scriptstyle`, or `scriptscriptstyle`). Then you say `\def\it{\fam4\teni}`. Thus `\it` has two effects: one is to switch to `\teni`, the second one is to set `\fam` to 4. Now, you can define a command `\twelvepoint` that modifies all the fonts values, using larger values. Gusee what happens for a definition like `\def\it{\temit\fam4}`.

In the case of a format like \LaTeX , the situation is different. There are some high level commands like `\large`, that are defined like `\@setsize \large {14pt}\xiipt\xiipt` (note: infinite recursion may be possible), and the `\xiipt` command is like the `\twelvepoint` command mentioned above. This is rather complicated. The situation became worse when people tried to replace computer modern fonts by other fonts. We shall describe here only the user interface of the NFSS (new font selection scheme).

We already mentioned that a important characteristic of the font is the encoding: We met OT1 (Original encoding by Knuth) and T1 (“Cork” encoding, similar to latin 1). There is an obsolete OT2 encoding for cyrillic, and new ones: T2A, T2B, T2C. The companion mentions over twenty standard font encodings. In the example of `\showbox` above, \TeX told us that the current font was `\T1/cmr/m/n/10`. The first two letters indicate the encoding. There are different families of fonts. Assume that you use Computer Modern fonts (you do this by selecting a package; after that, your whole document will be in computer modern, unless you use fonts selected via `\font` or `\selectfont`). There are six sub-families: Roman, Sans, Typewriter, Fibonacci, Funny roman, and Dunhill. The name of these families are: `cmr`, `cmss`, `cmtt`, `cmfib`, `cmfr`, `cmdh`. The default family in this document is `cmr`. You can chose another family via the commands `\rmfamily`, `\ttfamily` and `\sffamily` (no command is provided for the other families). The commands `\textrm`, `\textsf` and `\texttt` take an argument and typeset it using the family. The commands `\rm`, `\sf`, `\tt` do the same, but they reset the series to medium, and the shape to normal. The series of a font can be: bold, bold extended, semibold, etc. In \LaTeX you have `\mdseries` and `\bfseries` (you have also `\textmd` and `\textbf`, which are commands that take an argument; you have also `\bf` that selects roman family, bold series, normal shape). The shape can be: normal, italic, slanted, upright italic, small caps, etc. In \LaTeX we have `\upshape`, `\itshape`, `\slshape`, and `\scshape` (and as, usual, `\textup`, `\textit`, `\textsl` and `\textsc`; there is also `\it`, `\sl`,

\sc). There are two commands `\em` (a declaration) and `\emph` (that takes an argument) that use upright shape if the current font has a slant, and italics shape otherwise. These rules explain the `cmr/m/n` part in the font. In fact, the ‘`cmr`’ part comes from the command `\rmdefault`, but these commands are not implemented in Tralics.

There are two parameters that define the size of the font. First, document class options indicate the size used by `\normalsize`. In our example it is 10pt. There are ten commands that change the font size. In increasing order they are `\tiny`, `\scriptsize` `\footnotesize`, `\small`, `\normalsize`, `\large`, `\Large`, `\LARGE`, `\huge`, and `\Huge`. There is a command `\selectfont`; its purpose is to combine everything the result will be `\T1/cmr/m/n/10`. There is another process that converts this to the font name `ecrm1000`, using font definition files.

In math formulas, you see things like α' and \ddot{e} , but never ‘á’ and ‘ë’. If you want an acute accent you use `\acute`, if you want a double dot accent you say `\ddot`. In fact, the textfont used for math is very often a 7bit font, without accented letters. If you want $x^{\grave{e}me}$ you should say `x$\{\grave{e}me\}`, or perhaps `x$\{\hbox{ème}\}` (this gives $x^{\grave{e}me}$, letters are too big). A good solution is `xième`, $x^{i\grave{e}me}$. In French, you say 1^{er} , 1^{re} , 1^{ers} , 1^{res} , 2^e , 3^{es} , etc., via `1\ier`, `1\iere`, `1\iers`, `1\ieres`, `2\ieme`, `3\iemes`. In English, you say 1^{st} , 2^{nd} , 3^{rd} , 4^{th} .

2.9 Spaces

In T_EX spaces are ignored after a command like `\foo`, and a sequence of spaces is treated as a single one. The exact rule is the following. There is a variable whose values can be N (start of line), or M (middle of line) or S (when spaces are skipped). Whenever a line is read, T_EX removes every space character at the end of the line. It inserts the value of `\endlinechar` (provided this is a valid character, an integer between 0 and 255). The state is N. Spaces are ignored if the state is S or N; if the state is M, a space produces a space token, and the state is changed to S; in this sentence a “space” is any character whose category code is 10. If T_EX sees an end-of-line character (category 5), it ignores all other characters on the current line. If the state is N (line was empty), the tokeniser returns a `\par` token, if the state is M it returns a space token, otherwise the character is ignored. Note: in Tralics, the space token produced by an end-of-line is a line-feed character, this is to keep line breaks in the XML translation. If T_EX sees a backslash (or any character of category code 0), it reads a command; the state will be S if the character is a letter or a space, it will be N otherwise. If T_EX sees anything else, the state will be M.

For instance, if you say ‘`x_{ }_{ }_{ }_{ }_{ }`’ the tokeniser sees 5 spaces. If you say `\def\A{ }_{ }` and `\def\B{ }_{ }_{ }_{ }`, then the body of `\A` contains a space as well as the body of `\B`. Full expansion of `\B` contains three spaces and `x_{ }_{ }_{ }_{ }` contains four spaces. The command `\space` is defined like `\A` above.

Spaces discarded by the tokeniser do not appear in the translation. However, spaces produced by the tokeniser can be ignored in some cases. A typical example: a command can take a space as argument, and ignore the argument. For instance `\` is a command that ignores spaces that follow it using explicit scanning (i.e. `\futurelet`). We already mentioned that spaces between arguments are generally ignored. Spaces can be ignored because you say `\ignorespaces`: the effect of this command is to expand what follows, until a non-expandable token is seen. If it is a space, it is ignored, and the process continues. A space can be ignored because of a syntax rule (for instance, before an equals sign in an assignment). In L^AT_EX you can see things like that `\end{x}` `\end{y}` `\end{z}`, each ‘`end(xxx)`’ being on a line by itself: this produces a space, and the L^AT_EX environment mechanism is clever enough to remove these spurious spaces. It is also possible to remove a space from typeset material via `\unskip`.

Spaces are ignored in math mode. The reason is that spaces are used to separate words, and there are no words in math formulas. There are operators, and these operators know how much

white space to use. In the case of $x + y = z$, on each side of the plus sign there is some glue, the value comes from `\medmuskip`, it is 2.22 plus 1.11 minus 2.22; on each side of the equals sign there is `\thickmuskip`, namely 2.77 plus 2.77 (the unit is pt).¹⁰ After the zed, there is a kern of value 0.4398. Note: the plus sign is followed by a penalty of 700, the equals sign by a penalty of 500. Plain TeX defines

```
\thinmuskip=3mu
\medmuskip=4mu plus 2mu minus 4mu
\thickmuskip=5mu plus 5mu
```

In Tralics, constant values are used (expressed in terms of em units; one em is 18mu, in the example above one em is 10pt). You can say `\:`, `\>` and `\;`. This produces a space (thin, medium, thick) using the values given above. You can also use `\!`, this is the negative of thin space. The translation of `$A\B\C\D\!E$` is:

```
<mrow><mi>A</mi><mspace width='0.166667em' />
<mi>B</mi><mspace width='0.222222em' />
<mi>C</mi><mspace width='0.277778em' />
<mi>D</mi><mspace width='-0.166667em' /><mi>E</mi></mrow>
```

The `\space` command expands to a single space token. It may disappear in all cases where the syntax says that a space is optional (because in general these rules imply expansion); in a case like `\let\foo\space`, tokens are not expanded, and `\foo` is made equivalent to the current value of `\space`. The `_` command cannot be expanded. It starts a paragraph (if used in vertical mode). It inserts some white space whose value is the same as if the current space factor where 1000. You can use it after an abbreviation like Mr. in order to indicate that the dot is not an end of sentence marker. You can also use it after a command like `\TeX` if you want to leave some space. In math mode, Tralics interprets it as a space of width 6pt. The `~` character is usually active, its expansion is `\nobreakspace`. This is defined in Tralics to translate to ` `. You can say `\quad` or `\qqquad`. This inserts some space (the width is one or two em)¹¹. If you say `\hskip 1cm`, this will append some glue (in Tralics, it will generate a sequence of `~` whose width is more or less 1cm). If you say `\kern1cm` this will append a kern (like glue, but the size is fixed). This is ignored by Tralics. A normal space produces glue (the value of the glue depends on some font parameters; it can also depend on the current space factor). A glue may disappear at a line break. Kerns will not. In L^ATeX, you use `\hspace` instead of `\hskip`. You can use `\hspace*`, in this case, spaces at start of line are not ignored. Note the syntax `\hspace{2cm}` vs `\hskip2cm\relax`.

When TeX wants to split a paragraph into lines of equal width, it will have to stretch and shrink the glue that appears on the line; it will remove interword glue at break points. An item of glue has the form $x + y - z$, where x , y and z are dimensions (y and z can be expressed in terms of fil, fill and filll), all three values can be positive or negative. We can express this as: we have a vector of size 9: x_0 is the regular part of the glue, x_1 , x_2 , x_3 and x_4 are the stretch component (in units of pt, fil, fill, and filll, only one of these components can be given), x_5 , x_6 , x_7 and x_8 are the shrink components (in units of pt, fil, fill, and filll, only one of these components can be given). When two pieces of glue are added, all components are added. The convention is that x_2 is much larger than x_1 , so that the sum of x_1 and x_2 is x_2 (said otherwise if we add `1pt plus 2pt` and `3pt plus 4fil`, the result is `4pt plus 4fil`). Such simplifications are not done when TeX computes the sum of all glue items in a paragraph (as a result, addition is associative). The command `\hfil` is equivalent to `\hskip0pt plus 1fil`, the command `\hfill` is equivalent to `\hskip 0pt plus 1fill`, the command `\hfilneg` is equivalent to `\hskip 0pt plus -1fil`, the command `\hss` is equivalent to `\hskip 0pt plus 1fil`. It is an error to use infinite shrinkage, like `\hss`, in a paragraph, TeX complains with: *Infinite glue shrinkage found in a paragraph*. However

¹⁰these values were rounded; exact values are 2, 4 and 5 times 10pt/18, and 10pt/18 is a mu.

¹¹The value of the tilde is 3.33pt in math mode, and a `\quad` is 4 ` `; outside math mode since version 2.5(pl6).

you can say `123\hbox to1cm{\hss xxxxxx\hss}456`, the result is `123xxxxx456`, said otherwise, the text is centered, no overful neither underful box is signaled.

The commands `\vfil`, `\vfill`, `\vfilneg`, `\vss`, behave in the same fashion, in vertical mode, adding vertical space. Tralics translates `\hfil`, `\hfill`, `\hfilneg`, and `\hss` as `\leavevmode` followed by an element `<hfil>`, that has the same name as the command. It translates `\vfil`, `\vfill`, `\vfilneg`, and `\vss` in the same fashion, by using `\par` instead of `\leavevmode`. The three commands `\bigskip`, `\medskip` and `\smallskip` are used to insert vertical space between paragraphs, of size 12pt, 6pt and 3pt respectively (in L^AT_EX, this is some glue that the use can modify, however, Tralics ignores the shrink and stretch parts of the glue inserted by `\hskip`, `\vskip`, `\hspace` and `\vspace`.) These four commands read an argument (in L^AT_EX, `\hspace` and `\vspace` accept an optional star, that translates to an empty vertical or horizontal rule, Tralics ignores the star). In the case of an horizontal space, `\leavevmode` is executed, then `~` are produced (one every 4 pt, a negative dimension produces nothing). In the case of a vertical space, the current paragraph is terminated; if after that the mode is vertical, a new paragraph is started, it has an attribute `spacebefore` with as value the dimension. In L^AT_EX, the behavior is different (see appendix A.1.5 of [6]).

In T_EX, there is no command that starts a paragraph. The `\leavevmode` command is implemented as `\unhbox\voidb@x`, where `\unhbox` starts a new paragraph if needed, and produces nothing, provided that its argument is the void box; the paragraph may contain the current indentation and the value of `\everypar`. This is a primitive in Tralics, the value of `\everypar` is unused. Both commands `\indent` and `\noindent` make sure the current mode is horizontal, the first one inserts the current indentation (an empty box with the width of `\parindent`). In T_EX, you can use `\indent` anywhere in a paragraph. In Tralics, the translation of

```
a\noindent b \indent c
{\centering a\noindent b \indent c\par d}
{\raggedright a\noindent b \par\indent c\par d}
```

is

```
<p>a</p>
<p noindent='true'>b</p>
<p rend='center' noindent='false'>c
a</p>
<p rend='center'>b</p>
<p rend='center'>c</p>
<p rend='center'>d
a</p>
<p noindent='true' rend='flushed-left'>b</p>
<p noindent='false' rend='flushed-left'>c</p>
<p rend='flushed-left'>d</p>
```

The rules are the following: if `\indent` or `\noindent` appear in an empty paragraph, that is not centered, and that has no `noindent` attribute, one is set. Otherwise a new paragraph is started. It will have a `noindent` attribute, unless the paragraph is centered. The value of `\parindent` is never considered.

The translation of `\par` is a bit complicated. Nothing happens inside a `\hbox`, in `\term`, or if the current mode is not horizontal. The current XML element should be a `<p>`. A final space is removed from it. It will be popped from the stack. This restores the mode to the value of the previous mode. It restores the current XML element to the parent of the `<p>`. A newline character is added to it. There is an exception: in cases like `\noindent\par`, or `\bigskip\par`, or `\\par`, the `\par` command was ignored until version 2.5 (pl7). The behavior is now: if the paragraph is empty, but there are attributes, then the `<p>` is removed, and attributes are added to the next `<p>` element.

The translation of `\` depends on the context. The command can be followed by an optional star, and an optional dimension. Inside a cell, this indicates the end of the cell as well as the the end of the row. You can say `\newline`, this is like `\` without optional argument and array test. In vertical mode, L^AT_EX complains with *There's no line here to end*, but Tralics ignores the command. Inside a title, the command is ignored. Otherwise, the behavior is like `\noindent`; if an optional argument is given, it behaves like `\vskip`. For instance, the translation of

```
a \b \[2cm] c \newline[3cm]d \noindent e \vskip 4cm f
```

is

```
<p>a</p>
<p noindent='true'>b</p>
<p noindent='true' spacebefore='56.9055pt'> c</p>
<p noindent='true'>[3cm]d</p>
<p noindent='true'>e</p>
<p spacebefore='113.81102pt'>f</p>
```

Many people do not know that `\` takes an optional argument, and try to use different tricks in order to avoid errors triggered by `\\`. We have seen for instance

```
\def\ligne{\protect{\mbox{}}\\mbox{\indent}}
```

Remember that `\protect` is like `\noexpand`, it is not a L^AT_EX command that takes an argument! More strange cases can be found in [3].

The commands `\nolinebreak`, `\nopagebreak`, `\pagebreak`, and `\linebreak` are defined by L^AT_EX to take an optional argument, an integer between 0 and 4. They insert some penalty, but depend on the mode, like `\hspace` and `\vspace`. They are ignored in Tralics. The command `\break`, `\nobreak`, and `\allowbreak`, are defined by L^AT_EX, they insert some penalty (zero, plus or minus infinity). They are ignored by Tralics. The commands `\fillbreak`, `\goodbreak`, `\eject`, `\smallbreak`, `\medbreak`, `\bigbreak` are defined by L^AT_EX to terminate a paragraph and insert some penalty. In Tralics, they behave like `\par`.

2.10 Conditional expansion

In the previous paragraphs we have shown how to define a macro `\foo` that expands to `\bar` and a macro `\bar` that expands to `gee`.¹² Can a translator replace all `\foo` by `\bar` and all `\bar` by `gee`? the answer is obviously no; first because, if you say `\something\bar`, the argument will be (after expansion) `gee`, while in the case of `\something gee` it will be `g`; there is a second problem, that occurs in latex2html: if you replace `\bar` by its value, you get `\somethinggee`, and this is wrong, if you reparse it¹³; some commands can be randomly redefined (for instance, at first use) like this:

```
\def\NFSS{NFSS (New Font Selection Scheme)\global\def\NFSS{NFSS}}
```

The last reason is conditional expansion. Our original translator (written in Perl) has some troubles in these cases.

In this paragraph, we shall consider cases where expansion depends on the context. We have already seen the commands `\noexpand` for delayed execution and `\expandafter` that changes the order of expansion. We shall analyze three commands: `\color`, `\map` and `\loop`. The `\loop` command has a weird syntax, it needs a T_EX conditional (like `\if` or `\ifnum`) and executes some

¹²It is unwise to say `\def\bar`, because `\bar` is already defined for use in math mode, and who knows which commands rely on this definition.

¹³The remedy is to add a space after the command name. The trouble is when the command results from partial expansion of an active character, case where spaces should be preserved. This is one reason why active characters cannot be made robust in L^AT_EX.

piece of code as long as some condition is true. We are also interested in constructions that take a list of items, and add commas between them and a period at the end.

The easiest of these commands is ‘\map’: we want to apply some command to a list of arguments, constructed via techniques defined on page 30. It can be defined as

```
\def\map{\let\do}
```

This is a bit unsatisfactory: in some cases the list delimiter is different from \do, an example is given above: at the start of a chapter, we want to reset all counters that depend on the chapter counter, in this case \@elt is used as delimiter. We could imagine a map-with-argument macro, that would take as argument the \do. But this is nothing else than \let! Another example is given in section 2.12, it is ‘\let\do\@makeother\dospecials’, where the list is the last token. The command we are designing should be used as ‘\map\cmd\list’, where \cmd is the command to apply and \list is the list (note: you should not redefine the \list command, because it is used by environments like ‘itemize’). Let’s assume that our list is defined by

```
\def\mylist{\do{A}\do{B}\do{C}}
```

We have already seen how such a list can be created dynamically. The map command can be implemented as

```
\def\map#1#2{\let\do#1#2}
```

This definition is exactly the same as above, with a small difference: since the command takes an argument, a pair of braces disappear, for instance when you say \map\foo{\do{A}}; a second difference is that, for the argumentless version, you cannot use braces for the first argument. In general, you say ‘\map\@makeother\dospecials’, this is irrelevant. An alternate definition could be

```
\def\map#1#2{\def\do{#1}#2}
```

In this case, the first argument #1 can consist in more than one token. For instance, if you say \def\foo#1#2{_\#1#2_\} then \map{\foo A}\mylist gives ‘AA AB AC’. Note that there are too many spaces in this example: the last space in \foo is spurious.

The example with ‘\color’ is the following. Assume that we have two commands ‘\enrouge’ and ‘\envert’ that take an argument and typeset it in red and green. We want to define a command that takes two arguments, a color and text; nothing special should happen if the color is neither red nor green. We could implement this using a test, but this means that we have to change the macro if we add a new color (for instance ‘\enbleu’ for blue). The following works

```
\def\color#1{\csname en#1\endcsname}
```

The only drawback with this method is that it might produce unexpected results in the case where the command defined by \csname already exists (try ‘\color{d}{document}’).

We have in mind another example. Consider

```
% \newcommand{\fooiv}[3][bar]{Seen #1 #2 #3}
\def\fooivaux[#1]#2#3{Seen #1 #2 #3}
\def\fooiv{\@ifnextchar[{\fooivaux}{\fooivaux[bar]}}
```

The commented line is interpreted by L^AT_EX in the same fashion as the two other lines (except that the internal name is a bit more complicated than ‘\fooivaux’). We shall explain later how ‘\@ifnextchar’ works¹⁴. We are interested here in how L^AT_EX converts the ‘[3]’ into ‘[#1]#2#3’. Since the number of arguments is between zero and nine, a short sequence of conditionals could be used. Instead, the following code is used by L^AT_EX:

```
1 \long \def \@yargdef #1#2#3{%
2   \ifx#2\tw@
3   \def\reserved@b##11{[###1]}%
```

¹⁴The actual code uses \kernel@ifnextchar, because amsmath redefines sometimes \@ifnextchar

```

4   \else
5     \let\reserved@b\@gobble
6   \fi
7   \expandafter
8     \@yargd@f \expandafter{\number #3}#1%
9 }
10 \long \def \@yargd@f#1#2{%
11   \def \reserved@a ##1#1##2##{%
12     \expandafter\def\expandafter#2\reserved@b ##1#1%
13   }%
14   \l@ngre1@x \reserved@a 0##1##2##3##4##5##6##7##8##9##1%
15 }

```

In the case of ‘`\newcommand\fooiiii[3]{foo}`’ the `\@yargdef` command is called with three arguments, the first is `\fooiiii`, the command to be defined, then comes ‘`\@ne`’ (some randomly chosen token), then ‘3’ (the number of arguments) and finally ‘`{foo}`’, the body of the command to be defined. This argument is not read, but the code relies on the fact that it starts with an opening brace. The objective is to produce ‘`#1#2#3`’. In the case of ‘`\newcommand\fooi{foo}`’, arguments are the same with 0 as third argument, the objective is to produce the empty string. In the case of `\fooiiv`, the second argument is ‘`\tw@`’, this is something different from ‘`\@ne`’, the objective is similar, but a bit different: we want ‘`[#1]#2#3`’

In order to make things easier to understand, we shall proceed to the following simplifications: let’s forget about the percent signs (their purpose is to suppress unwanted space). Let’s forget about ‘`\long`’ (is it really needed?) and ‘`\l@ngre1@x`’ (this is something that adds conditionally a ‘`\long`’ token before the definition). Let’s simplify the names: we write ‘`\Ra`’ and ‘`\Rb`’ instead of ‘`\reserved@a`’ and ‘`\reserved@b`’. We also write ‘`\ydef`’ and ‘`\yaux`’ instead of ‘`\@yargdef`’ and ‘`\@yargd@f`’. Finally, we replace the arguments by X, Y, Z, and ‘`##`’ by a simple ‘`#`’. Hence we get

```

\def \ydef XYZ{
  \ifx Y\tw@
    \def\Rb#11{[#1]}
  \else
    \let\Rb\@gobble
  \fi
  \expandafter \yaux \expandafter{\number Z}X
}
\def\yaux XY{
  \def\Ra #1X#2#{\expandafter\def\expandafter Y\Rb #1X}
  \Ra 0#1#2#3#4#5#6#7#8#9#X
}

```

Let’s start the analysis with the lines 7 and 8. Because of the two ‘`\expandafter`’ tokens, the first token to be expanded is ‘`\number`’. This means that Z is replaced by its numeric value. Said otherwise, the number of arguments can be ‘03’, or ‘`\^^C`’, or even ‘`\value{page}`’ if the page number is not too big. In Tralics, only explicit numbers are allowed (You will get a message like *Only one token allowed; I will assume that the command takes no argument.*) In general, lines 7 and 8 are equivalent to `\yaux{Z}X`.

Let’s now explain lines 2 to 6. We are in a simple case of a conditional (the commands `\@ne` and `\tw@` are normally equivalent to 1 and 2, they compare unequal), so that line 3 is executed in case of an optional argument, and line 5 otherwise. In the last case `\Rb` is a command that takes an argument and ignores it; otherwise `\Rb` is a command that takes an argument, delimited by the character ‘1’, ignores it, and the expansion is ‘`[#1]`’ (four tokens). Remember that we want ‘`[#1]#2#3`’, that is a good starting point.

Consider now lines 11 and 12. In order to simplify explanations, we replace X by Z and Y by X (i.e. use the argument names of the outer function). We shall denote by U and V the arguments of \Ra. Thus \Ra is

```
\def\Ra UZV{\expandafter\def\expandafter X\Rb UZ}
```

The question now is what are the values of U and V? In order to answer this question we shall write line 14 in a different way. Let $s(n)$ be the sequence #1#2...#n#, and $S(n)$ the sequence #n...#9. The content of line 14

```
\Ra 0#1#2#3#4#5#6#7#8#9#Z
```

can be interpreted as \Ra, 0, $s(n-1)$, n , $S(n+1)$, #Z, whenever n is a digit between 1 and 9. Said otherwise, whenever Z is a digit between 1 and 9, the first argument U of \Ra is $0s(n-1)$ (the second argument is ignored, it is everything up to the first brace, the one that delimits the body). Obviously, in the case where Z is the digit 0, U is empty. We leave it as an exercise to the reader to see what happens in the case where Z is a sharp sign¹⁵. In all other cases, U is the sequence $0s(9)$. The important point is that, whatever Z, T_EX will not read beyond the opening brace of the body.

Assume now that we want to construct a normal command (case \Rb is gobble). It always gobbles a zero (if Z is zero, U is empty, and Z is gobbled). Thus \Rb UZ expand to: nothing if Z is 0, $s(n-1)n$ if Z is a digit between 1 and 9, and #1#2#3#4#5#6#7#8#9#Z otherwise. This yields an error *You already have nine parameters* which is adequate in case Z is a number larger than nine. Consider now the case of an optional argument. Here \Rb is a bit different: it reads the '0#1' part and replaces it by '[#1]'. You will get a *Runaway argument?* error (or some other strange behavior) in case Z is '0' because pattern matching fails (of course, you should never try to make optional the first argument of a function that takes none).

We shall in the sequel study the following problem (we shall call it *the \sharp problem*): we want a sequence of the form \sharp1\sharp2... \sharpN, where N is an integer between 1 and 9. After that, we can evaluate the '\sharp' command¹⁶, replacing it by '#'. One solution is: consider as above a sequence of length 9, and use pattern matching to reduce the length. We can also use a loop: the original L^AT_EX code used a loop from N down to Y (with $Y = 1$ in the case of normal argument, $Y = 2$ otherwise¹⁷). Here is an example of how this could be done with pattern matching:

```
\def\ydef#1#2#3{%
  \def\tmp##1#2##20{##1#2}%
  \def\sharp{\noexpand\sharp}%
  \edef\etmp{\tmp\sharp1\sharp2\sharp3\sharp4\sharp5\sharp6\sharp7%
    \sharp8\sharp90}%
  \def\sharp{#####}% needs 8 #
  \edef\etmp{\etmp}%
  \expandafter\def\expandafter#1\etmp{#3}}%
```

```
\ydef{\acmd}{3}{Args#1and#2and#3}
```

A silly question is: can we do loops without conditionals? The answer will be given later. We assume here that our loop will be of the form: while N is not too big, do something and increment N. This mechanism needs modifying a table (the location of N) hence is not pure expansion. In our example, we will write '\sharp\the\count0', and hope that this will evaluate to '#3' later on, assuming that \count0 contains '3' now. How that can be implemented is left as an exercise to the reader. See also section 2.11. We shall explain later all the silly details concerning conditionals

¹⁵Since Z is the result of expansion of \number it starts with a digit or a minus sign, so that this case will not happen.

¹⁶There is a \sharp command, that typesets as ‡, hence be careful.

¹⁷This explains the names \@ne and \@tw@.

in T_EX, all we need to know is that you can test $a < b$ and $a > b$, but neither $a \leq b$ nor $a \geq b$. Here is our code:

```
\def\code{\advance\count0 by 1 \sharp\the\count0}
\def\loop{\ifnum\count0<\count1 \code\loop\fi}
```

Assume that `\count0` holds 0 and `\count1` holds 3. In this case the test is true, ‘`\code`’ is evaluated, then ‘`\loop`’. The effect of evaluating ‘`\code`’ is to increment the counter and produce ‘`\sharp1`’. The loop terminates after ‘`\sharp3`’ has been produced. Notice that recursion is not terminal (but it would be in most computer languages): when the test is found false, there are four ‘`\if`’ tokens not yet evaluated. This example is atypical, in that the counter is modified before its use; exchanging the ‘`\sharp...`’ and the ‘`\advance...`’ part implies changing initial and final value (1,4 instead of 0,3).

Our `\loop` command is not generic, in that the name of the counters are built-in. Thus Knuth proposes the following:

```
\def\loop#1\repeat{\def\body{#1}\iterate}
\def\iterate{\body \let\next\iterate \else\let\next\relax\fi \next}
\let\repeat\fi % this makes \loop...\if...\repeat skippable
```

```
\loop \ifnum \count0<\count1 \sharp \the\count0 \advance\count0 by 1\repeat
```

Note that the last line contains an ‘`\ifxx`’ where the associated ‘`\fi`’ is the ‘`\repeat`’ at the end of the line. Thus, in the case where the ‘`\loop`’ command is *not expanded* this line is well-balanced regarding conditionals. In the case where ‘`\loop`’ *is* expanded, the value of the ‘`\repeat`’ token is irrelevant, it just serves as delimiter, and the ‘`\fi`’ has to be found in `\iterate`. In order for `\iterate` to work, the ‘`\body`’ should expand to an incomplete conditional, without `\else` part. It conditionally sets `\next`, and evaluates it after `\fi`; this trick makes the recursion terminal.

An alternate version is given by L^AT_EX, as follows¹⁸

```
\def\loop#1\repeat{\def\iterate{#1\relax\expandafter\iterate\fi}%
\iterate \let\iterate\relax}
```

Adding ‘`\let\iterate\relax`’ at the end of the definition has no real importance; but it causes no harm either. Note the `\expandafter` trick: if the test in the loop is false, neither `\expandafter` nor `\iterate` are expanded, if the test is true, `\fi` is evaluated before `\iterate`. Thus recursion is terminal. One difference with the T_EX method is that the body of the loop is put in `\iterate` rather than in a auxiliary command. The interesting point is the ‘`\relax`’. Guess what happens in this case:

```
\def\bloop#1\repeat{\def\iterate{#1\expandafter\iterate\fi}\iterate}
\count0=0 \count1=4
\bloop \ifnum \count0<\count1 \the\count0 \advance\count0 by 1\repeat
```

If you use L^AT_EX in verbose mode, you can see that the test is true, true and false, where you expect it to be true four times. The printed result is ‘0’ (hence the question: what did the second iteration do?). Using Tralics, you will get more information.

```
1 +\ifnum6
2 +scanint for \count->0
3 +scanint for \ifnum->0
4 +scanint for \count->1
5 +scanint for \ifnum->4
6 +iftest6 true
7 {\the}
8 {\the \count}
9 +scanint for \count->0
```

¹⁸The kernel contains two versions, one of them being `\long`.

```

10 \the->0.
11 Character sequence: 0.
12 {\advance}
13 +scanint for \count->0
14 {\expandafter \iterate \fi}
15 +\fi6
16 \iterate ->\ifnum \count 0<\count 1 \the \count 0 \advance \count 0 by 1\expandafter \...
17 +\ifnum7
18 +scanint for \count->0
19 +scanint for \ifnum->0
20 +scanint for \count->1
21 +scanint for \ifnum->4
22 +iftest7 true
23 {\the}
24 {\the \count}
25 +scanint for \count->0
26 \the->0.
27 +scanint for \count->10
28 {\advance}
29 +scanint for \count->0
30 {\expandafter \iterate \fi}
31 +\fi7
32 \iterate ->\ifnum \count 0<\count 1 \the \count 0 \advance \count 0 by 1\expandafter \...
33 +\ifnum8
34 +scanint for \count->0
35 +scanint for \ifnum->10
36 +scanint for \count->1
37 +scanint for \ifnum->4
38 +iftest8 false
39 +\fi8

```

Lines 16 and 32 are a bit too long; there are two token `\iterate\fi` that are replaced by `\...`

As you can see, all tests have a serial number. On lines 2–5, you can see why the first test is true: it is because the numbers 0 and 4 are compared. On lines 18–21, you see why the second test is true, and on lines 34–37, you see why the last test is false; in fact, `\count0` contains ten. On line 27, you see something strange. Explanations: Assume that you say ‘`\advance \Foo4`’, where `\Foo` is a reference to some counter. In this case, the trace of Tralics will contain `+scanint for \Foo->4`, and everybody understands this. If you replace `\Foo` by `\count0`, the trace will contain `\count`; it will also contain a line for the zero in `\count0`. Hence, the number that appears in line 27 is the value read by the `\advance` in line 12. What happened is the following: after ‘by’ we have seen the digit ‘1’. In the case of `\loop`, the next token would be ‘`\relax`’, and this stops scanning of the number. But here, we have ‘`\expandafter`’, which is expandable and expanded, as a consequence, this finishes the first conditional. After that comes the test; is it true, because we did not increment our counter yet. Then comes ‘`\the`’ which is expandable. This reads ‘`\count0`’, as well as the space after it. The expansion of ‘`\the...`’ is the digit zero; so far, we have read 10, and continue reading. The next token is ‘`\advance`’ and this is not expandable. Hence `\advance` has read everything up to the next `\advance`. Is it needed to explain what happens next? Let’s just notice that, at line 39, Tralics (and also T_EX) are still reading tokens for the second `\advance`.

2.11 Conditionals in T_EX

We shall discuss in this paragraph the following commands

- `\if AB` (comparison of two characters, character code),

- `\ifcat AB` (comparison of two characters, category code),
- `\ifint A=B` (comparison of integers),
- `\ifdim A=B` (comparison of dimensions),
- `\ifvmode`, `\ihmode`, `\ifmmmode`, `\ifinner` (test of the current mode),
- `\ifvoid A`, `\ifhbox A`, `\ifvbox A` (test box register status),
- `\ifx AB` (comparison of two tokens),
- `\ifeof N` (test if characters can be read from a channel),
- `\iftrue`, `\iffalse` (constant tests),
- `\ifcase` (this is called ‘switch’ in some other languages),
- `\or`, `\else`, `\fi` (must be used after if).

A conditional has the form `\if test true-code \else false-code \fi`. The `\else` part is optional; conditionals can be nested, and this nesting is independent from anything else. Conditionals are expanded: this means that conditionals are evaluated inside a `\edef`, you can use `\noexpand` to delay evaluation, and `\expandafter` to change the order of expansion.

An important point is the following: if you define a command `\ifthenelse` with three argument, that evaluates the first argument as a boolean, and expands conditionally to the second or third argument, then these two arguments must be balanced, and category codes are fixed. In the case of `\if`, there is no such limitations: if the test is found false, then all tokens are read at high speed until finding a `\else`, and normal processing occurs, or until finding a `\fi`, that indicates the end of the conditional; if the test is true, and if there is an `\else` part, all tokens between `\else` and `\fi` are read at high speed. Consider for instance this piece of code

```
\ifnum \A=\B do-nothing \else {\let\fi\relax\C}\fi
```

Assume that the test is false; this means that the else part is evaluated. Locally `\fi` is redefined to do nothing, and `\C` is evaluated. Let’s assume that `\C` does nothing special (it could typeset ‘Hello, word!’). In this case the `\fi` after the brace terminates the conditional. Assume now that the test is true. Skipping over the `\else` part at high speed just means compare the actual value of a token with `\if` or `\fi`: in the first case, the if-counter is incremented, in the second case it is decremented, in all other cases the counter is left unchanged; reading stops when the counter is zero. Here, the conditional is terminated by the first `\fi`. This means that you have to be very careful: the end of the conditional can change, depending on whether the test is true or false. When we say: ‘compare the actual value of the token’, this means that the name is irrelevant, only the meaning is used, for instance `\repeat` has the same value as `\fi`, and `\loop... \if... \repeat` is well balanced.

All constructions indicated above have a then-part and an else-part, except `\ifcase`: this command reads a number and you can specify action for the case zero, the case one, the case two, etc. For instance, we can solve the problem of constructing `\sharp1... \sharp N` as follows (assuming `\N` holds the value of N)

```
\ifcase \N \error{You cannot use zero here}
\or \sharp1
\or \sharp1\sharp2
\or \sharp1\sharp2\sharp3
\or \sharp1\sharp2\sharp3\sharp4
\or \sharp1\sharp2\sharp3\sharp4\sharp5
\or \sharp1\sharp2\sharp3\sharp4\sharp5\sharp6
\or \sharp1\sharp2\sharp3\sharp4\sharp5\sharp6\sharp7
\or \sharp1\sharp2\sharp3\sharp4\sharp5\sharp6\sharp7\sharp8
\or \sharp1\sharp2\sharp3\sharp4\sharp5\sharp6\sharp7\sharp8\sharp9
```

```

\else \error{Argument must be non-negative, at most nine}
\fi

```

The simple conditional ‘`\if AB ... \else ... \fi`’ compares two characters A and B, it shares some features with `\ifcat`. It expands tokens, using the following rules

- In the case `\noexpand\cmd`, where `\cmd` is a command that be expanded, it is as if you had said `\relax`.
- In the case `\noexpand X`, where X is an active character, the token that will be considered is X, without expansion.
- In the case `\noexpand Y`, that does not match any of the previous cases, the `\noexpand` is discarded.
- In the case `\cmd`, where `\cmd` can be expanded, it will be expanded.
- In the case X, where X is an active character, it will be expanded.
- In the case `\prim`, where `\prim` is a primitive that cannot be expanded, it is as if you have given character of numeric code `-1`, and category code `-1`.
- In the case of an implicit character `\cmd` (one obtained via `\let\cmd Y`, where Y is a character), the numeric code and the category code of the character is considered.
- In the case of an explicit character Y, the numeric code and the category code of the character is considered.

The command `\if` compares the two numeric codes, and `\ifcat` compares the category codes. If you say something like

```

\catcode ‘\A=3
\def\fooi{A}
\catcode ‘\A=11
\def\fooi{A}
\if\fooi\fooi H\fi \ifcat\fooi\fooi\else e\fi
\if\bgroup{1\fi \ifcat\egroup}1\fi \if\relax\par o\fi
\if01\else,\fi \ifcat01 w\fi \if\par1\else o\fi
\if\noexpand\fooi\relax r\fi \if\fooi Ald\fi \if!!!\fi

```

this should typeset as ‘Hello, world!’.

You must be very careful using a construction like ‘`\if\A\B...`’, because of the following

- If ‘`\A`’ expands to ‘`foo`’, the test is false, whatever ‘`\B`’,
- If `\A` is defined by ‘`\def\A\B{00}`’, the test is true, whatever ‘`\B`’,
- If `\A` is a command that takes an argument, the argument will be ‘`\B`’, and what happens depends on what the command does with the argument,
- If ‘`\A`’ expands to ‘`xxyy`’, the test is true, whatever `\B`, and a part of `\A` is typeset before ‘`\B`’,
- If ‘`\A`’ expands to ‘`xyyx`’, the test is false, ‘`\B`’ is ignored,
- If ‘`\A`’ and ‘`\B`’ expand to nothing, tokens that follow will be examined,
- If ‘`\A`’ expands to nothing, ‘`\if\A\B...`’ behaves like ‘`\if\B...`’, see above.

It is possible to write a macro `\newif` that takes an argument `\iffoo`¹⁹, and makes it a new conditional; the `ifthen` package provides the more L^AT_EXish syntax `\newboolean{foo}`. You can

¹⁹The first two letters should be I and F.

say ‘`\footrue`’ and the condition becomes true, or ‘`\foofalse`’ and it becomes false (the \LaTeX syntax is `\setboolean{foo}{true}`; Tralics knows both). These commands could be implemented as

```
\def\footrue{\def\iffoo{\if00}}
\def\foofalse{\def\iffoo{\if01}}
```

The trouble with this definition is that, when ‘`\iffoo`’ is read at high speed, it is not recognized as a conditional (it is a user defined command), see discussion about ‘`\ifhph`’ in [4, Chapter 20]. For this reason, the commands `\iftrue` and `\iffalse` were added, they evaluate respectively to true and false, and the following lines work (because `\let` is used instead of `\def`):

```
\def\footrue{\let\iffoo\iftrue}
\def\foofalse{\let\iffoo\iffalse}
```

You can use ‘`\ifnum`’ or ‘`\ifdim`’: in both cases a numeric quantity, an operator, and another numerical quantity are read. In the case of ‘`\ifnum`’, both quantities have to be numbers, otherwise dimensions. Three operators are recognized: less than, greater than and equal to. Here is an example

```
\count0=0 \count1=1 \dimen0=1pc \dimen1=12pt
\skip0=1cm minus3fill \skip1=1mmplus 2fill
\tracingall
\ifnum \count0<\count1
  \ifdim \dimen0=\dimen1
    \ifdim \skip0>\skip1 ok \fi\fi\fi
```

This is the trace of Tralics. Note that for \LaTeX , all lengths allocated by `\newlength` are “rubber” length, i.e. associated to a `\skip` register. Such quantities are automatically converted into rigid length (however, if you replace in the example ‘`\skip1`’ by ‘`1mmplus 2fill`’, then only a rigid dimension is read, the “plus 2fill” is not part of the condition).

```
+ \ifnum26
+scanint for \count->0
+scanint for \ifnum->0
+scanint for \count->1
+scanint for \ifnum->1
+iftest26 true
+ \ifdim27
+scanint for \dimen->0
+scandimen for \ifdim ->12.0pt
+scanint for \dimen->1
+scandimen for \ifdim ->12.0pt
+iftest27 true
+ \ifdim28
+scanint for \skip->0
+scandimen for \ifdim ->28.45274pt
+scanint for \skip->1
+scandimen for \ifdim ->2.84526pt
+iftest28 true
+ \fi28
+ \fi27
+ \fi26
```

This is one solution to our problem of producing N sharp signs in a row:

```
\ifnum \N>0 \sharp1\fi\ifnum \N>1 \sharp2\fi\ifnum \N>2 \sharp3\fi
\ifnum \N>3 \sharp4\fi\ifnum \N>4 \sharp5\fi\ifnum \N>5 \sharp6\fi
\ifnum \N>6 \sharp7\fi\ifnum \N>7 \sharp8\fi\ifnum \N>8 \sharp9\fi
```

The following construction is a priori more efficient (on the average there are less tests) but it takes more memory.

```
\ifnum \N>0 \sharp1\ifnum \N>1 \sharp2\ifnum \N>2 \sharp3
\ifnum \N>3 \sharp4\ifnum \N>4 \sharp5\ifnum \N>5 \sharp6
\ifnum \N>6 \sharp7\ifnum \N>7 \sharp8\ifnum \N>8 \sharp9\fi\fi\fi\fi\fi\fi\fi\fi
```

You can test whether a character can be read from an input channel, via the `\ifeof` command. Here is an example from the Tralics torture file. The file ‘tortureaux’ has six lines, the first one contains `abc`, the second one is empty, the third one contains `\a \b {c}`, the fourth one contains `{} \d} \e`, the next one contains `123`, the last one is empty. The `\testeq` commands compares two commands: things should be equal here. (See T_EXbook, exercise 20-18, if you do not understand the setting of `\endlinechar`). Commands starting with ‘bad’ are not evaluated in this example. Details can be found in section 5.8

```
{
\openin 5=tortureaux
\endlinechar=-1
\ifeof5 \badifeofatentry\fi
\read 5 to \foo\testeq\foo{abc}
\read 5 to \foo\testeq\foo{ }
\read 5 to \foo\testeq\foo{\a\b{c} \d} \e}
\global\read 5 to \foo
\closein5\relax
\ifeof5\else\badifeofatexit\fi
}\testeq\foo{123}
\ifeof3\else \badifeofnonexists\fi
```

You can say `\ifvoid25`, `\ifhbox25` or `\ifvbox25`. In T_EX these command would test the content of box register 25: if empty, the `\ifvoid` is true, the other tests are false; if not empty, the box contains a horizontal list or a vertical list, and `\ifhbox` and `\ifvbox` are respectively true, the two other tests being false. Since Tralics puts elements in box registers, and since elements have no orientation, the test is not well-defined. What Tralics does is just: read a number and return false. In the example that follows, only the first equals sign is part of an assignment.

```
\count0=0
\ifvoid\count0=1\fi
\ifvbox\count0=2\fi
\ifhbox\count0=3\fi
```

You can say `\ifmmode`, `\ifvmode`, `\ifhmode` and `\ifinner`. These commands check the current mode. The first three evaluate to true if the mode is math mode, vertical mode, or horizontal mode. The last is true if the mode is inner (internal vertical mode, restricted horizontal mode, or (nondisplay) math mode). The following example shows these modes.

```
\def\wm{\edef\res{\ifinner i\else I\fi
\ifhmode h\else H\fi
\ifvmode v\else V\fi
\ifmmode m\else M\fi}\res}
\par \wm$$\wm \hbox{\wm $\wm$} \eqno \wm$$
```

The result is: ‘IHvM IHV_m ihVM iHV_m iHV_m’. If you remove the ‘`\edef`’, the trouble will be that typesetting the ‘I’ enters horizontal mode. This example fails if ‘`$$...$$`’ is replaced by ‘`\[...]`’, because `\eqno` switches to inner math mode, and ‘`\]`’ checks for outer math. The same test, done by Tralics gives IHvM then IHV_m (in fact, ‘`\ifinner`’ always returns false, ‘`\eqno`’ is not recognized, and ‘`\hbox`’ has no influence of the current mode.)

The last conditional to explain is `\ifx`. This reads two tokens and compares them. Two tokens are equal if they are character tokens (implicit or explicit) with same character value and category

code, or two \TeX primitives with the same meaning, or two user-defined commands with the same value (same arguments, same body, same `\long` and `\outer` flags)²⁰.

Using `\ifx` we can code our `\color` command properly, like that

```
\def\color#1#2{%
  \def\crouge{rouge}\def\cvert{vert}\def\cc{#1}%
  \ifx\cc\crouge\enrouge{#2}\else\ifx\cc\cvert\envert{#2}\else#2\fi\fi}
```

It is possible to avoid these assignments in the `\color` macro, provided that they are hidden elsewhere. For instance

```
\def\ifstringeq#1#2#3#4{%
  \def\tempa{#1}\def\tempb{#2}%
  \ifx\tempa\tempb#3\else#4\fi}

\def\couleur#1#2{%
  \ifstringeq{#1}{rouge}{\enrouge{#2}}
  {\ifstringeq{#1}{vert}{\envert{#2}}{#2}}}
```

Note that the ‘ifthen’ package provides the `\equal` command as helper for such a situation: you could say `\ifthenelse{\equal{A}{B}}{X}{Y}` instead of `\ifstringeq {A}{B}{X}{Y}`. Caveat: the `\equal` command fully expands its two arguments, our version expands nothing.

In any computer language, you would define a command that compares two strings and returns true or false; this is not possible in \TeX because commands return no value. All you can do is modify some variable (a command, a register, a token list, etc). This assignment can be done by the caller or the callee. Here is a solution where the token `\next` is set by the caller:

```
\def\color#1{%
  \ifstringeq{#1}{rouge}{\let\next\enrouge}
  {\ifstringeq{#1}{vert}{\let\next\envert}{\let\next\relax}}%
  \next}
```

Note that, if `\envert` accepts an optional argument, for instance if `\envert[clair]{text}` typesets the text using light green, you can say `\color{vert}[clair]{text}`. We consider now a case where the assignment is done by the callee (via `\equaltrue` or `\equalfalse`; there is a variant that uses `\setboolean`).

```
\newif\ifequal
\def\streq#1#2{%
  \def\tempa{#1}\def\tempb{#2}%
  %variant: \setboolean{equal}{\ifx\tempa\tempb true\else false\fi}
  \ifx\tempa\tempb\equaltrue\else\equalfalse\fi}

\def\color#1{%
  \streq{#1}{rouge}%
  \ifequal\let\next\enrouge\else
    \streq{#1}{vert}%
    \ifequal\let\next\envert\else \let\next\relax\fi\fi
  \next}
```

A subtlety of \TeX is that tokens are read only when needed. Said otherwise, if you say ‘`\if AB C\else D\fi`’, \TeX will evaluate the test; it will remember that a new conditional has started. If the test is false, it will skip at high speed until the `\else`, and resume normal evaluation; but if the test is true, it will resume normal evaluation right now. It is only when \TeX sees an

²⁰If you remember the code of `\fooiv`, you can see that two \LaTeX commands with optional arguments always compare unequal.

`\else` token (and this can be another one) that it will read all tokens at high speed until the `\end`. And, when T_EX sees the `\fi`, it will pop the conditional stack. Consider the following example:

```
\def\ifstringeq#1#2#3#4{%
  \def\tempa{#1}\def\tempb{#2}%
  \ifx\tempa\tempb\aux{#3}\else\aux{#4}\fi}
\def\aux#1#2\fi{\fi#1}
\def\color#1{%
  \ifstringeq{#1}{rouge}{\enrouge}{\ifstringeq{#1}{vert}{\envert}{\relax}}}
```

Assume that the test is true. Then `\aux` reads all tokens, up to `\fi`, provides a `\fi` to finish the conditional now, then expands to its first argument (which is argument 3 of `\ifstringeq`). In the case where the test is false, the same thing happens. This is nicer than the solution that consists in defining conditionally `\next` and evaluating it after the `\fi`, it avoids an assignment.

Let's consider now a variant of the color problem. We want to write a command with three arguments A, B and C, it is assumed to read a token, compare it with A, and expand to B or C. We need an auxiliary command that reads the token. Thus the solution

```
\def\ifnextchar#1#2#3{%
  \let\tempa=#1\def\tempb{#2}\def\tempc{#3}%
  \ifaux
}
\def\ifaux#1{%
  \let\lettoken=#1%
  \ifx\lettoken\tempa\let\tempd\tempb\else\let\tempd\tempc\fi
  \tempd
}
```

Note that we have put an equals sign after `\let\tempa` and `\let\lettoken` for the case where the token to match is an equals sign. If you want to catch spaces, a bit more complicated machinery must be used. There is a problem with this command, because, if the argument of `\ifaux` is not a single token, say `'foobar'`, then only `'f'` will be put in `\lettoken` and `'oobar'` will be typeset. On the other hand, if the argument is empty, then `\ifx` will be put in `\lettoken`; after that `\lettoken` will be expanded. Since this is `\ifx`, the following tokens will be compared (said otherwise `\tempa` and `\let`), this is not exactly what is required. In order to solve this problem, we first modify slightly our code:

```
\def\ifnextchar#1#2#3{%
  \let\tempa=#1\def\tempb{#2}\def\tempc{#3}%
  \ifaux
}
\def\ifaux#1{\let\lettoken=#1\ifnch}
\def\ifnch{%
  \ifx\lettoken\tempa\let\tempd\tempb\else\let\tempd\tempc\fi
  \tempd
}
```

The `\ifnch` command given above looks like the L^AT_EX version of the beast. In fact, spaces are ignored in L^AT_EX, so that there is an additional test. Moreover, some variables have a different name, nevertheless, here is the code:

```
\def\@ifnch{%
  \ifx\@let@token\@sptoken
  \let\reserved@c\@xifnch
\else
  \ifx\@let@token\reserved@d
  \let\reserved@c\reserved@a
```



```

\else
  \let\reserved@c\reserved@b
\fi
\fi
\reserved@c}

```

The problem is the `\ifaux` command. The question is: can we rewrite it in such a way as to read a single token, before calling `\ifnch`. Recall that we want to distinguish between ‘`{x}`’ and ‘`x`’. A very interesting question is the following: if we read the opening brace, how can we put it back in the input stream? we cannot do so by just expanding a macro (because the body is always well balanced). You could try something like `{\ifnum0=} \fi` (that leaves an unmatched brace after expansion), or something like ‘`{\iffalse} \fi`’. Our solution is much simpler. There is a \TeX primitive that gets the token without reading it. To be precise, `\futurelet` reads a token A, that has to be a command name or an active character, then a second token B, then a third token C. The value of the token is put in A, using the equivalent of `\let`, then C and B are pushed back in the input stream (in this order, the token to be read first is B). The code of `\ifnextchar` is hence the following:

```

\def\ifnextchar#1#2#3{%
  \let\tempa=#1\def\tempb{#2}\def\tempc{#3}%
  \futurelet\lettoken\ifnch}

```

What ‘`\futurelet\lettoken\ifnch`’ does is read a token. This could be a space character, an equal sign, an open brace, a closing brace, whatever. It puts it back in the input stream. It puts it also in `\lettoken`. After that, it evaluates `\ifnch` (which is a command that should take no argument, of course; it should consult `\lettoken` and depending on the value, call a command that, maybe, reads the token). In \LaTeX , `\ifnextchar` ignores spaces; `amsmath` has a version that does not. The difference is the comparison of the current token with `\sptoken`. The non-trivial question is how to construct a token equivalent to a space. We give here three solutions. The first one uses `\futurelet`. If the arguments are A, B and C, where A is the command to define, and C the space, then B has to be a command (if it is a character, it will be typeset); this cannot be `\foo`, since spaces after `\foo` disappear, it has to be something like ‘`\;`’. This command must read the space, otherwise it appears in the output. We provide two solutions: a command that is delimited by a space, and a command that takes an argument (remember that spaces disappear before undelimited arguments):

```

\def\; {} \futurelet\Sptoken\; % comment required
\def\;#1{} \futurelet\Sptoken\; 0

```

In both cases, the command `\;` cannot be used for typesetting (in the \LaTeX kernel, it is used for computing the `\Sptoken`, and correctly redefined after that). We give here an example, where the redefinition is temporary, inside the box. Ee can discard the content of the box.

```

\setbox0\hbox{\def\;{} \global\futurelet\Sptoken\; }

```

We give now a solution using `\let`. Remember the syntax, after `\let` and `\sptoken` (the token to be assigned), comes `<equals>` and `<one optional space>` and `<token>`, where the last token is our space token. Since `<equals>` reads an arbitrary number of spaces and an optional equals sign, an equals sign is required. Our optional space cannot be optional. So we must produce a double space. This is not completely trivial. We give here two solutions (the comment is necessary)

```

\def\makesptoken#1{\let\sptoken= #1}\makesptoken{ }
\def\:{\let\Sptoken= } \: % this makes \Sptoken a space token

```

And now, how can we define `\@xifnch`? this command is assumed to read a space, discard it, and check again for the next character. Thus the question is to design a macro that reads a space. This cannot be done via `\def\@xifnch#1...`, since spaces are ignored before undelimited arguments; we cannot use the technique of the command ‘`\;`’ above, because we cannot read what follows the space; the solution consists in a command that takes no argument, and that starts with a

compulsory token, like `\def\foo\bar{etc}`. The non trivial point is that we want `\bar` to be replaced by a space token, but spaces disappear after `\foo`. We give here two solutions.

```
\expandafter\def\expandafter\foo\space{etc}
\def\:{\Foo}\expandafter\def\:{ etc}
```

Let's consider the following variant of the `\map` command. If we have `\do{A}\do{B}\do{C}`, we want to separate arguments with a comma, and put a period after the last argument; we might as well do something with the argument, say, typeset it in italics. This is not always possible. In one of the style sheets used by the Raweb, a Perl postprocessor is used for replacing some commas by a period. We assume here that we know where the list ends. For instance, we assume that we can put a `\endl` token at the end of the list. Then we can write something like

```
\def\foo#1#2\endl{\textit{#1}\ifx#2\endl\endl.\else, \foo#2\endl\fi}
```

Then `\foo{A}{B and C}{D}\endl` produces `'A, B and C, D.'` as expected. Let's analyze the code and try to see why it is wrong. We assume that you never say `\foo\endl`, because the list is assumed non-empty. We also assume that the list does not contain the `\endl` token (in L^AT_EX, you should use the special marker `\@nil` only as list delimiter). In our case, the first argument is `'A'`, the second is `'{B and C}{D}'`. In the case where the second argument is empty, the test is true, because `\endl` is compared against itself. In our case, the test is false because the brace is compared with the letter B. If we put the second argument in a pair of braces, we get an error: *Too many }*'s, because the test is true, and a part of `#2\endl\endl` has been evaluated. This means that our test is wrong. The only safe way to check whether `#2` is empty is to put it in a command, and check whether this is the same as `\empty`. We shall give a second version of the code where the test is replaced by `\ifx\endl#2\endl`. In the case where `#2` is empty, the test evaluates to true, and if `#2` evaluates to some token list that does not start with `\endl`, the test will be false; this is better.

Note that, when `\foo` is called again, it compares `'D'` with `'\endl'`. Does this surprise you? In fact, if you say `\foo{A}{XY}{UV}\endl`, you get `'A, XY, U, V.'` The trouble is the following: when T_EX reads the arguments of a command, a pair of braces disappears, when possible. Thus arguments are `'A'` (without braces) and `'{XY}{UV}'` (it is not possible to remove the braces). When `\foo` is called again, arguments are `'XY'` and `'UV'`, without braces. This explains why the test compares U and V (by the way, if `'UV'` is replaced by `'UUVV'`, the test will be true, yielding an *Undefined control sequence* error). When `\foo` is called again, arguments are now `'U'` and `'V'`, an unwanted result. There is a simple way to avoid disappearance of braces: it suffices to put a token before each item, for instance like this

```
\def\foo\do#1#2\endl{\textit{#1}\ifx\endl#2\endl.\else, \foo#2\endl\fi}
\foo\do{A}\do{B}\do{C}\endl
```

A more elegant solution: notice that `#2` starts with `\do`, unless it is empty. There is no need to read the argument for seeing this, we can use the `\ifnextchar` command. With the solution proposed here, the token that marks the end of the list is evaluated: we use `\relax`, because this is harmless.

```
\def\foo{\def\do##1{\textit{##1}\@ifnextchar{\do}{, }{.}}
\foo\do{A}\do{B}\do{C}\relax
```

Note that we can replace `\relax` by something more useful, for instance a period:

```
\def\foo{\def\do##1{\textit{##1}\@ifnextchar{\do}{, }{.}}
\foo\do{A}\do{B}\do{C}.
```

An alternate solution could use `\ifprevchar` instead of `\ifnextchar`. There is no such command in L^AT_EX, but the idea is the following: instead of putting a comma after each argument but the last, we can put a comma *before* each argument but the first. All we need to do is to know if this argument is the first. In one application, we have coded this as: apply `\do-first` on the

first argument, and map `\do-other` on the rest of the list. If side effects are allowed, we can use a piece of code like this (note how the final period is typeset):

```
\newif\iffirst
\def\do#1{\iffirst\firstfalse\else , \fi\textit{#1}}
\firsttrue
\do{A}\do{B}\do{C}.
```

In fact, there is no need to use an auxiliary command, it suffices to modify `\do` itself:

```
\def\foo{\def\do##1{\textit{##1}\def\do###1{, \textit{###1}}}}
\foo\do{A}\do{B}\do{C}.
```

If you think that there are too many sharp signs, you can try

```
\newcommand\normaldo[1]{, \textit{#1}}
\newcommand\firstdo[1]{\textit{#1}\let\do\normaldo}
\newcommand\foo{\let\do\firstdo}
\foo\do{A}\do{B}\do{C}.
```

There are other possibilities implying conditional commands. We shall see later how to define a ‘comment’ environment that ignores the content of it. It is as if you said

```
\newenvironment{comment}{\iffalse}{\fi}
```

One can make the following strange construct `{\ifnum0='}\fi`. In this case, we compare two numbers, zero and the internal code of the brace (which is in general non-zero). The result of the test is false, but who cares? the body of the conditional as well as the else part is empty. Hence, the result is like `\bgroup`, there are some differences because $\text{T}_{\text{E}}\text{X}$ has two brace counters: the balance counter and the master counter; there is only one counter in $\text{T}_{\text{r}}\text{a}\text{l}\text{i}\text{c}\text{s}$. For details, see the $\text{T}_{\text{E}}\text{X}$ book and its appendix D, where it is said “If you understand [...] you’ll agree that the present appendix deserves its name.” (the name of the appendix is ‘Dirty Tricks’).

A piece of code like this causes trouble to $\text{T}_{\text{r}}\text{a}\text{l}\text{i}\text{c}\text{s}$

```
\def\foo#1{%
  \sbox\tempboxa{#1}%
  \ifdim \wd\tempboxa >\hsize
    #1\par
  \else \hbox to \hsize{\hfil\box\tempboxa\hfil}%
\fi}
```

It is a simplification of the `\@makecaption` command of the ‘article’ class. The idea is to center the caption of an image if it fits on a line (centering is achieved via `\hfil`). The argument is typeset in a temporary box, and the width of the box is compared against `\hsize`. Captions in the $\text{R}_{\text{a}}\text{w}_{\text{e}}\text{b}$ are always centered, but this is not esthetic.

In appendix D of the $\text{T}_{\text{E}}\text{X}$ book, there are some examples of how to produce N asterisks in a row. The question is: can we produce this using pure expansion? this is a solution given by D. Kastrop:

```
\def\nlines#1{\expandafter\nlineii\romannumeral\number\number #1 000\relax}
\def\nlineii#1{\if#1m\expandafter\theline\expandafter\nlineii\fi}
\def\theline{A}
\nlines{5}
```

This produces ‘AAAAA’. The idea is the following: ‘`\romannumeral3000`’ expands to ‘mmm’. It is then rather easy to convert this sequence of m into a sequence of A . The argument of the command can be ‘`\count0`’; the ‘`\number`’ has as effect to convert the value of this counter into a number, it gobbles a space. The argument of the command can be ‘`\count1`’; the second ‘`\number`’ will gobble the second space (I don’t know if there is some other reason for these two `\number` commands). Here is the same idea, *without tests*:

```

\def\recur#1{\csname rn#1\recur}
\def\rn#1{}
\def\rnm#1{\endcsname{#1}#1}
\def\replicate#1{\csname rn\expandafter\recur
  \romannumeral\number\number#1 000\endcsname\endcsname}

\dimen0=4sp \replicate{\dimen0}{P}

```

You may wonder how this works. Here is the transcript file of Tralics.

```

1 [216] \replicate{\dimen0}{P}
2 \replicate #1->\csname rn\expandafter \recur \romannumeral
3 \number \number #1 000\endcsname \endcsname
4 #1<-\dimen 0
5 {\csname}
6 {\expandafter \recur \romannumeral}
7 +scanint for \dimen->0
8 +scanint for \number->4
9 +scanint for \number->4000
10 +scanint for \romannumeral->4000
11 \recur #1->\csname rn#1\recur
12 #1<-m
13 {\csname}
14 \recur #1->\csname rn#1\recur
15 #1<-m
16 {\csname}
17 \recur #1->\csname rn#1\recur
18 #1<-m
19 {\csname}
20 \recur #1->\csname rn#1\recur
21 #1<-m
22 {\csname}
23 \recur #1->\csname rn#1\recur
24 #1<-\endcsname
25 {\csname}
26 {\csname->\rn}
27 \rn #1->
28 #1<-\recur
29 {\csname->\rnm}
30 \rnm #1->\endcsname {#1}#1
31 #1<-P
32 {\csname->\rnm}
33 \rnm #1->\endcsname {#1}#1
34 #1<-P
35 {\csname->\rnm}
36 \rnm #1->\endcsname {#1}#1
37 #1<-P
38 {\csname->\rnm}
39 \rnm #1->\endcsname {#1}#1
40 #1<-P
41 {\csname->\rn}
42 \rn #1->
43 #1<-P
44 Character sequence: PPPP .

```

This is now something else, it is part of a command defined in the RR style file:

```

\bgroup
\edef\foo{\ifnum 0<0#1x\else y\fi}\def\xbar{x}%
\ifx\foo\xbar
\global\compteurtheme=#1
\else \global\compteurtheme=0 \@latex@error{Pas un thème #1}\@eha\fi
\egroup

```

Assume that #1 contains a positive number, for instance 25. In this case, the test will be true, \foo will be defined as ‘x’, and will be equal to \xbar. In this case, our command puts 25 in \compteurtheme. Some other tests (not shown here) are done for instance, the value should be a number between 1 and 4, or a number with two digits, each one being between 1 and 4. Assume that the argument is not a number, say it is ‘gee’; then \ifnum will compare 0 and 0, the test will be false, \foo will be defined as ‘y’ hence is not equal to \xbar. Assume that the argument is ‘3a’; this is not a theme, but a theme and a subtheme. In this case, the test is true, but \foo expands to ‘3x’, and this is not equal to \xbar. Nowadays, themes are ‘com’, ‘cog’, etc, and this piece of code has become useless. It is replaced by something different, see end of section 6.8.

2.12 A nontrivial command \verb

The code that follows is a simplified version of a L^AT_EX command

```

1 \def\verb{%
2   \bgroup
3   \let\do\@makeother \dospecials
4   \verbatim@font\@noligs
5   \@vobeyspaces \frenchspacing\@sverb}
6
7 \def\verb@egroup{\global\let\VBG\@empty\egroup}
8 \let\VBG\@empty
9
10 \def\@sverb#1{%
11   \catcode'#1\active
12   \lccode'\~'#1%
13   \gdef\VBG{\verb@egroup\error{...}}%
14   \aftergroup\VBG
15   \lowercase{\let~\verb@egroup}}

```

Note first that this code contains two empty lines, that are read by T_EX as a \par token (it is ignored, provided that the definition is read in vertical mode). Lines 5, 7, and 15 are terminated by a brace and the end of line character produces a space token, that is ignored for the same reasons. Lines 1, 10, 12, and 15 are terminated by a % character, since otherwise, it would produce a space character (ignored in case the command is executed in vertical mode, and that is not always the case). In the case of lines 2, 3, 4, etc., the end of line is converted into a space character that disappears because it follows a command name.

This code defines a command \verb that starts a group via \bgroup. At line 3, \dospecials is executed, after redefining \do. This changes the category code of all special characters (included all characters made active by packages like babel²¹). Line 4 changes the current font to a typewriter one, and it executes a piece of code that inhibits ligatures (for instance the one that converts a double dash in an en-dash). Note that this document contains a great number of verbatim examples, either inline or as environments. In some cases, we use a smaller font; it is hence important to allow the user to parameterize commands like these. Line 5 contains three commands: The first makes an end-of-line character active (usually, it will behave like \par), the second enters so-called french spacing mode (a mode where the width of a space is constant), and the last

²¹A safe solution would be to change all category codes; but this is time consuming.

command `\@sverb` will be explained later. The ‘s’ in the name of this command comes from the ‘starred’ version of `\verb`: If you say ‘`\verb*+ +`’, you will get ‘`␣`’. We have omitted the test with the star character.

On lines 7 and 8, we define a command `\VBG` that does nothing (i.e. expands to the empty list) and a command that evaluates to `\egroup` preceded by a global assignment of `\VBG` to nothing. On line 13, `\VBG` is defined as calling `\verb@egroup` plus some error, whose text is not shown here. Thus `\VBG` is a command that 1) resets `\VBG` to a harmless command, 2) closes the current group, 3) signals an error.

Let’s consider lines 11 and 12. We assume that the argument of `\@sverb` is some character c (If you say `\def\foo{\verb\foo=\foo}` then `\foo`, you will get an error *Improper alphabetic constant*, and after that, you’re really in trouble. In the usual case, the character that follows `\verb` is read with category code 11 or 12, because of the code line 3.) Line 11 makes the character c active (of category 13); the category code will recover its old value at the end of the group, and line 13 changes the lc-code of the tilde character (the lc-code will recover its value at the end of the group). The lc-code of a character will be used for hyphenation, as well as conversion from upper case to lower case. We assume here, for the sake of simplicity, that hyphenation is inhibited by the use of a verbatim font. Note that Tralics does not care about subtleties like hyphenation. For this reason, when you say `\verb+foo+`, it will execute `\verbprefix {\verbatimfont foo}`. You can redefine both commands (the prefix is empty, the font defaults to `\tt`). Notice that Tralics grabs the argument, contrarily to L^AT_EX.

Let’s come back to the L^AT_EX implementation of `\verb`. So far, we have read a character, changed its category code, changed the lc-code of the tilde character, changed the font and other tables, redefined `\VBG`, aftergrouped it (code on line 13: the token is saved on a special stack, and popped of the end of the current, that was opened on line 2, and normally closed on line 7). Line 15 is a kludge: what `\lowercase` does is replace in its argument every character by its lower case equivalent (using the lc-code table). The result is evaluated again. Here the argument is formed of three tokens: `\let`, the tilde and `\verb@egroup`. Since `~` is a character that has a lower-case equivalent, it will be replaced by that, namely the character c . Note: category codes are left unchanged by this procedure. It is hence important that `~` be an active character (because `\let` modifies that value of `~`) and that c be active (otherwise, there is no meaning in changing the value of c).

Consider the case of `\verb+\toto+`. Here the character c is the plus sign. After line 15 has been executed, the situation is the following: all characters are of category other, ligatures are disabled, french spacing is active, current font is typewriter, a group is opened, and a token is waiting for the group to terminate. In such a situation, you cannot go outside L^AT_EX properly. In fact, the carriage return has been made active in order to help error recovery (this is not shown here), and the ‘+’ sign has been made active: this will help us. T_EX sees now the following tokens `_12 _11 _11 _11 _11 +_13`. The first five tokens are added to the current horizontal list as characters in the current font, while the last one is expanded. The expansion is that of `\verb@egroup`, see line 7. This defines globally `\VBG`, then closes the group, restoring everything. It does not restore `\VBG` (because the last assignment was global). After the group, the after-grouped token `\VBG` is evaluated but it does nothing.

So far, so good: the translation of ‘`\verb+foo+`’ is the same as ‘`\texttt{\char‘\foo’}`’. Note that the author could have entered the previous expression as ‘`\verb-\verb+foo+-`’, or using the fancyvrb package as ‘`\|verb+\toto+|`’, but he used `\quoted{BS verbBS foo+}`.

Consider now the following example:

```
\def\duplicate#1{#1#1} ‘\duplicate{\verb+x}++’
```

You would expect ‘`xx++`’ but you get ‘`x+x+`’. Explanations: the expansion of `\duplicate` is `\verb+_12 x_11+_12 \verb+_12 x_11+_12 +?+_?`. The last two plus signs have not been read, and their category code is still unassigned. The `\verb` command reads the `+_12` via `\@sverb`. It changes the

category code of the plus sign. The second `\verb` does the same. It reads the `+?` as a `+13`, this finishes evaluation of `\verb`. The second `\verb` command does the same. In the case where you replace `++` by `--`, the `\verb` command will see an end of line character before a plus character and complain with *LaTeX Error: \verb ended by end of line.*

Consider now the following example:

```
\def\braceme#1{#{#1}} ‘\braceme{\verb+x+}++’
```

You get the following error *LaTeX Error: \verb illegal in command argument.* Let's try to see how this is done. The expansion of `\braceme` produces the following tokens: $\{_1 \overline{\text{verb}} +_{12} x_{11} +_{12} \}_2$. After `\@sverb` has finished, the first non-inactive character is `\}_2`, this closes the current group. Hence, as above, this restores category code, fonts, lc-codes, etc. It does not restore `\VBG` because assignment is global (`\gdef` at line 13 is like `\global\def`). The trick is now that a `\VBG` token is popped from the aftergroup stack. This one calls `\verb@egroup` and signals an error. What `\verb@egroup` does is to close a group (the one opened by `\braceme`), and reset `\VBG` to something harmless. Note that `TEX` is in a clean mode when the error is signaled. `Tralics` has no such error handling mechanism (however, no category codes are changed when scanning for the end of the command, so nothing harmful can be done). What this example shows is that error recovery is not completely trivial; nevertheless nice things can be done.

There is a variant to `\verb`, it is the ‘verbatim’ environment. The classical exercise is: write a command that reads everything up to `\end{verbatim}` (backslash and braces are of category 12 in this token list). There are different packages that solve this problem; For instance ‘fancyvrb’ is one of them. A solution is also given in the first chapter. It does not allow an optional space after ‘`\end`’.

2.13 Expandable tokens

We give here the list of all tokens that can be expanded.

- All user defined command.
- All undefined commands.
- `\a`, `\’`, `\‘`, `\"`, `\^`, `\~`, `\k`, `\H`, `\v`, `\b`, `\d`, `\u`, `\C`, `\f`, `\c`, `\.`, `\=`, `\r`, `\T`, `\V`, `\D`, `\h`: these commands produce an accented character.
- `\(`, `\)`, `\[`, `\]`: commands for math mode.
- `\noexpand`: inhibits expansion of the next token.
- `\expandafter`: changes order of expansion.
- `\UseVerb`: restores a quantity saved by `\SaveVerb`.
- `\@stpelt`, `\stepcounter`, `\refstepcounter`, `\addtocounter` `\setcounter`, `\value`. These are user-defined commands in `LATEX`. The expansion in `Tralics` can depend on the loading of the `calc` package.
- `\setlength``\addtolength`. These are user-defined commands in `LATEX`. The expansion in `Tralics` can depend on the loading of the `calc` package.
- `\csname`. This manufactures a token.
- `\number`, `\@arabic`, `\romannumeral`, `\Romannumeral`, `\string`, `\meaning`, `\fontname`, `\jobname`, `\tralicsversion`. These commands convert some internal quantities into tokens; for instance `\jobname` is the name of the file that is translated (without an extension ‘.tex’), and `\tralicsversion` is the version number of `Tralics`; it could be ‘2.5 (pl7)’.

- `\the`. You say `\the\foo`, if you want to typeset the value of a variable `\foo`.
- `cmdindexarabic`, `\roman`, `\Roman`, `\alph`, `\Alph`, `\fnsymbol`, `\@alph` `\@Alph`, `\@fnsymbol`: commands to typeset a L^AT_EX counter.
- `\input`, `\include`, `\endinput`: Special macros for files.
- `\if`, `\ifcat`, `\ifnum`, `\ifdim`, `\ifodd`, `\ifvmode`, `\ifhmode`, `\ifmmode`, `\ifinner`, `\ifvoid`, `\ifhbox`, `\ifvbox`, `\ifx`, `\ifeof`, `\iftrue`, `\iffalse`, `\ifcase`. These start a conditional.
- `\fi`, `\or`, `\else`. These modify the conditional stack.
- `\topmark`, `\firstmark`, `\botmark`, `\splitfirstmark`, `\splitbotmark`. Marks are not yet implemented.
- `\loop`, `\@whilenum`, `\@whiledim`, `\@whilesw`: For loops.
- `\xspace`. This adds a space if needed.

Chapter 3

Mathematics

3.1 Introduction

Mathematics play a great role in T_EX and Tralics. For instance, T_EX has three modes: vertical mode, in which no typesetting is done, horizontal mode (where everything happens) and math mode, a mode in which special objects are handled; a two phase process converts these special objects in normal ones. Fonts to be used in math mode have special properties (see appendices F and G of the T_EXbook). Not all subtleties of T_EX math can be implemented in Tralics; on the other hand, the XML translation is conforming to MathML. This defines some entities, for instance in isoamsc.ent, there is a definition of `&rceil`; to `⌉`;. As a consequence, Tralics will translate `\rceil` to `<mo>⌉</mo>`. For most elements, there is a mechanism to change the name, this is not done for MathML.

T_EX has also a notion of “inner” mode. Inside an inner object, you cannot put an outer one. As a consequence, you cannot put a display math formula into a box or another formula; in fact, in a construction like `\hbox{$$$}` the first dollar sign enters math mode, while in general, a test is made to see if a second dollar sign follows, case where display math mode is entered. In Tralics, there is no distinction between inner and outer mode; if you put a `\hbox` as above inside a math formula, the T_EX behavior is followed; otherwise, there will be a problem if you put a display math formula in a box; you will get an error of the form *Error in pop; stack holds p; trying to pop hbox*. This is essentially because an end-of-paragraph is inserted before the display math formula, which is harmless, and a start-of-paragraph after the formula; when we see the end of the box, there is a problem because of this unexpected paragraph.

The essential difference between normal math and display math, is that a display formula is (in general) centered on a line; it can have an equation number (via `\eqno` and `\leqno`, these two commands are not implemented). The MathML documentation says “One of the important uses of `<mlabeledtr>` is for numbered equations. In a `<mlabeledtr>`, the label represents the equation number and the elements in the row are the equation being numbered. The `side` and `minlabelspacing` attributes of `<table>` determine the placement of the equation number.” Thus, the recommended way, for MathML, is to use a table, like this (replace ellipsis by an expression)

```
<table>
  <mlabeledtr id='e-is-m-c-square'>
    <td>
      <mtext> (2.1) </mtext>
    </td>
    <td>
      ...
    </td>
  </mlabeledtr>
</table>
```

```

    </mtd>
  </mlabeledtr>
</mtable>

```

This mechanism is not yet implemented. We do not know how to insert numbers automatically, so that proposed solution is: you can use `\label`, `\ref` for any display math formula. This will add an id attribute to the `<formula>` object, which is a wrapper for the `<math>`. The `\tag/\notag` mechanism of `amsmath` does not work.

The syntax of mathematics is often strange. Instead of

```
\math{E=\frac{1}{2} m\power{v}{2}}
```

you say

```
$E={1\over 2} mv^2$
```

If you want a dollar character, you can say `\$`, if you want `^` (a hat character), you cannot say `\^`, because this will put an accent over what follows. In \LaTeX , you can say `\textasciicircum`, provided that you can guess the name. The previous example gives a warning: *Package amsmath Warning: Foreign command \over; \frac or \genfrac should be used instead.* The essential reason for this message is that `\over` is the only non-prefix command of \TeX : it influences what is on its left. This makes interpretation of math formulas a bit difficult.

When you say `{\alpha^2}`, \TeX will enter math mode with an error of the form *Missing \$ inserted*. On the other hand, `Tralics` will signal two errors, the first is *Math only command \alpha*. *Missing dollar not inserted*, the second is *Missing dollar not inserted, token ignored: {Character ^ of catcode 7}*. If you want a command that works in math mode and outside math mode, you can say:

```
\def\foo{\ifmmode \alpha^2 \else $\alpha^2$\fi}
```

This can be generalised, using the following command

```

\DeclareRobustCommand{\ensuremath}{%
  \ifmmode
    \expandafter\@firstofone
  \else
    \expandafter\@ensuredmath
  \fi}
\long\def\@ensuredmath#1{${\relax#1$}

```

The purpose of the `\relax` on the last line is for the case of an empty argument: we do not want `\ensuremath{}` to expand to `$$`. Note that the argument is handled only once (i.e. `\ensuremath` does not read it, but calls a helper), because of subtle bugs, see latex bugs data base `amslatex/2104`. We shall say later ‘Mode independent commands are interpreted as usual’, this implies that the `\relax` will do nothing. Other commands, not listed in this chapter, may signal an error. For instance, `\par` is forbidden. The same is true for `\char`; note that `\mathchar` provokes an *Unimplemented command* error. If you want a random Unicode character, you should use commands like `\mathmi`.

Because a math expression translates as `<math>` inside a `<formula>`, and that the math has a long namespace attribute, examples will never fit on a single line. In order to make the result easier to read, we have inserted some newline characters, and reindented all these examples. Two consecutive newline characters are scanned by \TeX as space plus `\par`. This space is ignored by \TeX (see \TeX book, the text between exercises 14.12 and 14.13). Hence the general rule in `Tralics`: when a `<p>` element is ended, a trailing space or newline is removed from the content of the element, a newline character is added to the parent of the `<p>`. As a result, you will very often see `<p>` at the start of a line and `</p>` at the end of a line in a XML file generated by `Tralics`.

Consider the following simple example:

α and β `\label{foo}`

The translation is the following

```
<p>
  <formula type='inline'>
    <math xmlns='http://www.w3.org/1998/Math/MathML'>
      <mi>&alpha;</mi>
    </math>
  </formula> and</p>
<formula id='uid1' type='display'>
  <math xmlns='http://www.w3.org/1998/Math/MathML'>
    <mi>&beta;</mi>
  </math>
</formula>
```

You can also say

```
\(\alpha\) and \[\beta \label{foo}\]
```

The result is exactly the same. In L^AT_EX, the commands `\(`, `\)`, `\[` and `\]` test the current mode. No such test is done by Tralics. The L^AT_EX implementation of `\[` is a bit strange. If the formula is in vertical mode, it will be preceded by a box of width `.6\linewidth` containing nothing (except two `\hss` commands to fill it) preceded by the current paragraph indentation. The command `\]` executes `\ignorespaces`. As you can see, there is some difference between a single dollar and a double dollar. In the first case, we are in normal math mode, otherwise in display math mode. One difference is the initial style: it is `\textstyle` (for normal mode) and `\displaystyle` otherwise (this will be explained later). A second difference is that the `everyXXX` token list inserted when scanning the formula depends on the mode. The third difference is specific to Tralics. A display math formula is never ‘trivial’ (see section 3.5), it can have a label (not more than one): in this case, the `<formula>` element has an `id` attribute. In any case, the `<formula>` element has a `type` attribute that explains that the formula is inline or display. A non-display formula starts a paragraph; a display math formula cannot appear in a paragraph (the equivalent of `\par` is executed), if the first non-space token (after expansion) that follows the math formula is not `\par`, a `\noindent` token will be inserted (see line 34 of the transcript at page 77). Note that, in T_EX, a math formula does not end a paragraph, in the sense that a `\parshape` is valid across math formulas; however what precedes the formula is split into lines, according to parameters in force at the start of the formula. Tralics does not split paragraphs into lines, and does not implement `\parshape`.

3.2 The basic objects

The following environments are recognized outside math mode, and produce a math formula: `eqnarray*`, `align`, `split`, `equation*`, `math` and `displaymath`. When Tralics sees a dollar character, it looks at the next character (without expansion). If this is a dollar sign, it will be read, and display math mode is entered, otherwise, normal math mode is entered. All environments showed above start display math mode (except ‘math’, which enters normal math mode). The environments ‘math’ and ‘displaymath’ are equivalent to `\(...\)` and `\[...\]` respectively. The environments ‘eqnarray’ and ‘split’ are implemented as arrays. There is no difference between

```
\begin{eqnarray} a&b \\ c&d \end{eqnarray}
\begin{split} a&b \\ c&d \end{split}
```

and

```
\[\begin{array}{rcl} a&b \\ c&d \end{array}\]
\[\begin{array}{r1} a&b \\ c&d \end{array}\]
```

Environments ‘equation’ and ‘align’ are translated as normal math. A star after the environment name is ignored. In the case of normal math mode, the content of the token list `\everymath` is inserted before the formula, for `displaymath` it is `\everydisplay`. For instance, if you say

```
\everymath={ (N)\ }
\everydisplay={ (D)\ }
 $\alpha$  and  $\beta$ 
```

the translation will be

```
<p>
<formula type='inline'>
  <math xmlns='http://www.w3.org/1998/Math/MathML'>
    <mrow>
      <mo>(</mo><mi>N</mi><mo></mo><mspace width='6pt' />
      <mi>&alpha;</mi></mrow></math></formula> and</p>
<formula type='display'>
  <math xmlns='http://www.w3.org/1998/Math/MathML'>
    <mrow>
      <mo>(</mo><mi>D</mi><mo></mo><mspace width='6pt' />
      <mi>&beta;</mi></mrow>
    </mrow>
  </math>
</formula>
```

In \TeX , you can put anything inside a math formula, provided it is hidden in a box; this is not possible in \Tralics , because we want the XML result to be conforming to MathML. We shall list here all commands valid in math mode, and explain later on how they are translated.

The following environments are recognized: `array`, `matrix`, `pmatrix`, `bmatrix`, `Bmatrix`, `vmatrix`, `Vmatrix`. All these environments produce arrays. For the first, an argument is required, explaining how cells are aligned. For all other environments, cells are centered. Environments of the form ‘`Xmatrix`’ have fences, an implicit `\left` and `\right`. In order: parentheses, braces, brackets, simple bars, double bars. There is also an environment ‘`cases`’, with two columns, left aligned, that has an open brace as left delimiter, an empty right delimiter. Example

```
\begin{array}{lcr}a&b&c\end{array}
\begin{bmatrix}d&e\end{bmatrix}
```

The translation is the following.

```
<formula type='inline'>
  <math xmlns='http://www.w3.org/1998/Math/MathML'>
    <mrow>
      <table>
        <tr>
          <td columnalign='left'><mi>a</mi></td>
          <td><mi>b</mi></td>
          <td columnalign='right'><mi>c</mi></td>
        </tr>
      </table>
      <mfenced open='{ ' close='}'>
        <table>
          <tr>
            <td><mi>d</mi></td>
            <td><mi>e</mi></td>
          </tr>
        </table>
    </mrow>
```

```

        <mtd><mi>f</mi></mtd>
        <mtd><mi>g</mi></mtd>
    </mtr>
</mtable>
</mfenced>
</mrow>
</math>
</formula>

```

The following delimiters are recognized: <, >, ., (,), [,], |, \{, \}, \langle, \rangle, \lbrace, \rbrace, \lceil, \rceil, \lgroup, \rgroup, \lfloor, \rfloor, \lmoustache, \rmoustache, \vert, \Vert, \uparrow, \downarrow, \updownarrow, \Uparrow, \Downarrow, \Updownarrow. A delimiter is anything that can follow \left or \right. For MathML, this has to be a character. As the following example shows, we use in most cases a character entity.

```

    $\left\lceil \left\uparrow x\right.\right\right.$
    $\lceil \uparrow x\}$

```

The translation is

```

<formula type='inline'>
  <math xmlns='http://www.w3.org/1998/Math/MathML'>
    <mfenced open='&lceil;' close='.'>
      <mfenced open='&uparrow;' close='&rbrace;'>
        <mi>x</mi></mfenced></mfenced></math></formula>
<formula type='inline'>
  <math xmlns='http://www.w3.org/1998/Math/MathML'>
    <mrow><mo>&lceil;</mo><mo>&uparrow;</mo><mi>x</mi><mo>&rbrace;</mo></mrow></math></formula>

```

This is the list of commands allowed in math mode, as well as in text mode: \dots, \ldots, \quad, \qquad, \sqcup, \\$, %, \&, \!, \, , \{, \}, \i, \sharp, \natural, \flat, _ . The following commands produce space: \; , \: , \> . Note that \! produces a negative space in math mode, nothing outside math mode. Example of use:

```

    $\dots\ldots\%\&\{\}\i\sharp\natural\flat\_ \$
    \dots\ldots\%\&\{\}\i\sharp\natural\flat\_

```

This is the translation:

```

<p><formula type='inline'>
<math xmlns='http://www.w3.org/1998/Math/MathML'>
<mrow><mo>&ctdot;</mo><mo>...</mo><mo>%</mo><mo>&amp;</mo><mo>{</mo>
<mo>&inodot;</mo><mo>&sharp;</mo><mo>&natur;</mo><mo>&flat;</mo>
<mo>_</mo></mrow></math></formula>
.....%&{}&inodot;&sharp;&natur;&flat;_</p>

```

We give here the list of all symbols that have a translation of the form <mi>α</mi>. They are of type Ord (ordinary symbol). We start with the lower case Greek letters: \alpha, \beta, \gamma, \delta, \epsilon, \varepsilon, \zeta, \eta, \theta, \iota, \kappa, \lambda, \mu, \nu, \xi, \pi, \rho, \sigma, \tau, \upsilon, \phi, \chi, \psi, \omega, \varpi, \varrho, \varsigma, \varphi, \vartheta, \varkappa, then upper case Greek letters: \Gamma, \Delta, \Theta, \Lambda, \Xi, \Sigma, \Upsilon, \Phi, \Pi, \Psi, \Omega, then other symbols: \hbar, \ell, \wp, \Re, \Im, \partial, \infty, \emptyset, \nabla, \surd, \top, \bottom, \bot, \angle, \triangle. Example

```

    $\alpha\Gamma \surd$

```

This translates as

```
<formula type='inline'>
  <math xmlns='http://www.w3.org/1998/Math/MathML'>
    <mrow><mi>&alpha;</mi><mi>&Gamma;</mi><mi>&radic;</mi>
    </mrow></math></formula>
```

Next comes the list of all symbols whose translation is like log. There are of type Ord (ordinary symbol), though they should be Op (large operator). The list is divided in two parts: these have movable limits: `\det`, `\gcd`, `\inf`, `\injlim`, `\liminf`, `\limsup`, `\max`, `\min`, `\sup`, `\projlim`, and these have not: `\dim`, `\exp`, `\hom`, `\ker`, `\lg`, `\lim`, `\ln`, `\log`, `\Pr`, `\arccos`, `\arcsin`, `\arctan`, `\arg`, `\cos`, `\cosh`, `\cot`, `\coth`, `\csc`, `\deg`, `\sec`, `\sin`, `\@mod`, `\sinh`, `\tan`, `\tanh`. Example

```

$$\lim_a \liminf_a \sin_a \hom_a$$

```

The L^AT_EX translation is $\lim_a \liminf_a \sin_a \hom_a$, and the Tralics version is

```
<formula type='inline'>
<math xmlns='http://www.w3.org/1998/Math/MathML'>
<mstyle scriptlevel='0' displaystyle='true'>
<mrow>
  <msub><mo movablelimits='true' form='prefix'>lim</mo> <mi>a</mi> </msub>
  <msub><mo movablelimits='true' form='prefix'>lim inf</mo><mi>a</mi></msub>
  <msub><mo form='prefix'>sin</mo> <mi>a</mi> </msub>
  <msub><mo form='prefix'>hom</mo> <mi>a</mi> </msub>
</mrow></mstyle></math></formula>
```

From now on, all symbols translate into the form `<mo>...</mo>`. We start with symbols of type Ord. In reality, most of them they should be of type Op (large operator). `\mho`, `\clubsuit`, `\diamondsuit`, `\heartsuit`, `\spadesuit`, `\aleph`, `\backslash`, `\Box`, `\imath`, `\jmath`, `\square`, `\cong`, `\lnot`, `\neg`, `\forall`, `\exists`, `\coprod`, `\bigvee`, `\bigwedge`, `\biguplus`, `\bigcap`, `\bigcup`, `\int`, `\sum`, `\prod`, `\bigotimes`, `\bigoplus`, `\bigodot`, `\oint`, `\bigsqcup`, `\smallint`. Examples

```

$$\bigcap \int \oint$$

```

The translation is

```
<mrow><mo>&bigcap;</mo><mo>&int;</mo><mo>&oint;</mo></mrow>
```

These are of type Bin (binary operator). `\triangleleft`, `\triangleright`, `\bigtriangleup`, `\bigtriangledown`, `\wedge`, `\land`, `\vee`, `\lor`, `\cap`, `\cup`, `\multimap`, `\dagger`, `\ddagger`, `\sqcap`, `\sqcup`, `\amalg`, `\diamond`, `\Diamond`, `\bullet`, `\wr`, `\div`, `\odot`, `\oslash`, `\otimes`, `\ominus`, `\oplus`, `\uplus`, `\mp`, `\pm`, `\circ`, `\bigcirc`, `\setminus`, `\cdot`, `\ast`, `\times`, `\star`, `\in`. Example

```

$$\cap \cup \wr$$

```

The translation is

```
<formula type='inline'><math xmlns='http://www.w3.org/1998/Math/MathML'>
<mrow><mo>&cap;</mo><mo>&cup;</mo><mo>&wr;</mo></mrow></math></formula>
```

These are of type Rel (relation). `\propto`, `\sqsubseteq`, `\sqsupseteq`, `\sqsubset`, `\sqsupset`, `\parallel`, `\mid`, `\dashv`, `\vdash`, `\Vdash`, `\models`, `\nearrow`, `\searrow`, `\nrightarrow`, `\swarrow`, `\Leftrightarrow`, `\Leftarrow`, `\Rightarrow`, `\neq`, `\neq`, `\leq`, `\leq`, `\geq`, `\geq`, `\succ`, `\approx`, `\succeeds`, `\preceq`, `\prec`, `\doteq`, `\supseteq`, `\subset`, `\supseteq`, `\subseteq`, `\bindnasrepma`, `\ni`, `\gg`, `\ll`, `\gtrless`, `\geqslant`, `\leqslant`, `\not`, `\notin`, `\leftarrow`, `\leftarrow`, `\owns`, `\gets`, `\rightarrow`, `\to`, `\mapsto`, `\sim`, `\simeq`, `\perp`, `\equiv`, `\asymp`, `\smile`, `\leftharpoonup`, `\leftharpoondown`, `\rightharpoonup`, `\rightharpoondown`, `\hookrightarrow`, `\hookleftarrow`, `\iff`, `\Longrightarrow`, `\longrightarrow`, `\longleftarrow`, `\Longleftarrow`, `\longmapsto`, `\frown`, `\bowtie`, `\longleftarrow`, `\Longleftarrow`, `\Join`. Example

```

$$\approx \leftarrow \Longleftarrow$$

```

Translation:

```
<formula type='inline'><math xmlns='http://www.w3.org/1998/Math/MathML'>
  <mrow><mo>&approx;</mo><mo>&leftrightharrow;</mo>
  <mo>&Longleftrightharrow;</mo></mrow></math></formula>
```

These are of type Inner: `\cdots`, `\hdots`, `\vdots`, `\ddots`. These are of type Between (they are of type Ord in T_EX, but are used as opening or closing delimiters): `\Vert`, `\lvert`, `\rvert`, `\uparrow`, `\downarrow`, `\Uparrow`, `\Downarrow`, `\Updownarrow`, `\updownarrow`. These are of type Open and Close: `\rangle`, `\langle`, `\rmoustache`, `\lmoustache`, `\rgroup`, `\lgroup`, `\rbrace`, `\lbrace`, `\lceil`, `\rceil`, `\lfloor`, `\rfloor`.

The following characters are classified as ‘small’: `<`, `>`, `.`, `:`, `*`, `?`, `!`, `x`, these are classified as ‘small-l’ and ‘small-r’: `()` `[]`, the vertical bar is small-l, these are bin: `+` `/` and the equals sign is of type Rel. Note: what you see here as `x` is in reality the character 215. It cannot be printed in verbatim mode by L^AT_EX.

```
$<>, .:;*?!x () [] | + - / = $
```

Translation:

```
<formula type='inline'>
  <math xmlns='http://www.w3.org/1998/Math/MathML'>
    <mrow><mo>&lt;</mo><mo>&gt;</mo><mo>,</mo><mo>.</mo><mo>:</mo>
    <mo></mo><mo>*</mo><mo>?</mo><mo>!</mo><mi>&times;</mi><mo>( </mo>
    <mo></mo><mo>[ </mo><mo>] </mo><mo>| </mo><mo>+ </mo><mo>- </mo>
    <mo></mo><mo>= </mo>
  </mrow></math></formula>
```

The following commands are used for accents: `\acute`, `\grave`, `\mathring`, `\ddddot`, `\dddot`, `\ddot`, `\tilde`, `\widetilde`, `\bar`, `\breve`, `\check`, `\hat`, `\widehat`, `\vec`, `\overrightarrow`, `\overleftarrow`, `\underrightarrow`, `\underleftarrow`, `\dot`.

The following commands are special. They will be explained later: `\overline`, `\underline`, `\stackrel`, `\underset`, `\overset`, `\mathchoice`, `\frac`, `\overbrace`, `\underbrace`, `\genfrac`, `\dfrac`, `\tfrac`, `\sqrt`, `\root`.

3.3 Parsing a math formula

This is a non-trivial operation, for this reason in verbose mode, the math expression will be printed on the transcript file. For instance, given

```
\tracingall
  $\begin{cases} x & \&y \\ \a & \&b \end{cases} \kern2cm x^{\{2\}}!$
```

you will see

```
1 {math shift character $}
2 +stack: level + 2 for math
3 +stack: level + 3 for math
4 \cases ->\left \{\begin{array}{ll}
5 +stack: level + 4 for math
6 +stack: level + 5 for cell
7 +stack: level + 6 for math
8 +stack: level - 6 for math
9 +stack: level - 5 for cell
10 +stack: level + 5 for cell
11 +stack: level - 5 for cell
12 +stack: level + 5 for cell
```



```

13 +stack: level - 5 for cell
14 +stack: level + 5 for cell
15 \endcases ->\end {array}\right .
16 +stack: level - 5 for cell
17 +stack: level - 4 for math
18 +stack: level - 3 for math
19 +scanint for \kern->2
20 +scandimen for \kern ->56.9055pt
21 +stack: level + 3 for math
22 +stack: level - 3 for math
23 +stack: level + 3 for math
24 +stack: level + 4 for math
25 +stack: level - 4 for math
26 +stack: level - 3 for math
27 +stack: level - 2 for math
28 Math: $\begin {cases}{\left \{\begin {array}{ll} x & y\end {array} \end {cases}
29 \end {array}\right .} \kern (Changed to)\hspace {56.9055pt}x^{\ 2 }!$
30 +scanint for \hspace->56
31 +scandimen for \hspace ->56.9055pt
32 {scanglue 56.9055pt}
33 Realloc xml math table to 20
34 {Push p 1}

```

We shall explain for each line in the transcript file where it comes from. Math mode scanning is entered when the translator sees a math shift character (line 1). The scanner reads some tokens and puts them in a list. The list is printed at the end (lines 28-29). The start of the formula is a bit special, in that the token that follows the first dollar sign is considered unexpanded when we check for a double dollar sign. A new group is entered, before scanning the whole formula (line 2).

The loop is as follows:

- A token is read and expanded. Lines 4 and 15 show expansion of user commands. An error is signaled in the case of end of data.
- In the case of `\nobreakspace`, we insert a `~`.
- If we get a font command, we proceed as follows. First `\cal` is transformed into `\@mathcal`. The font can be `\mathtt`, `\mathcal`, `\mathbf`, `\mathrm`, `\mathit`, `\mathbb`, `\mathsf`. These are basic math fonts; they have an inner variant, of the form `\@mathtt` (same name with an `@` in front). There is also `\mathnormal`. We allow old fonts (like `\rm`, `\sf`, `\tt`, `\bf`, `\it`, `\sl`), fonts switches of the form `\rmfamily`, or font commands that take an argument like `\textrm`. These fonts has an inner variant, say `T`. If the font takes no argument, then the token `T` is inserted (as explained for `\cal` above). Otherwise, let `S` be the current math font. In this case, an argument is read, then `S`, `T` and this argument is pushed back, to be read again. For instance, if the current font is `'sf'`, then `\mathrm{foo}` produces `\@mathrm foo\@mathsf`: these are five tokens to be read again. Note: `\mathbbm` is an alias for `\mathbb`.
- In all other cases the current token is added to the list. In particular, this explains while the trace starts with a dollar and `\begin`.
- If the token is an open brace, in fact any character of category code 1, a new math group is read. You can see on lines 23 and 24 that the stack level increases (a new semantic level is entered, all assignments are local).

- If the token is a close brace, in fact any character of category code 2, this terminates the current math group (see lines 25 and 26). An error is signaled in case the current group should be closed differently (for instance with `\end`, or `\right`, etc.)
- If the token is a dollar sign, in fact any character of category code 3, then four alternatives can be chosen. This dollar sign can be the end of the math formula. If we are in display math, a token is read with expansion. An error is signaled *Display math should end with \$\$* if this is not a dollar sign (in fact, a character of category code 3). If the current group is defined by `\hbox`, this can be the start of a math formula (never display math). The token in `\everymath` are inserted, then a math formula is read. Otherwise, an error is signaled *Extra \$ ignored...*, parsing continues.
- In the case of `\label`, an argument is read. If we are not in display math, or if the formula already has a label, you get an error: *Some labels may be lost*. Wherever the location of the label, an attribute will be added to the `<formula>` element that contains the `<math>` element.
- In the case of `\ensuremath`, a token list is read, and pushed back, so that this command acts as `\@firstofone`.
- The case `\begin` or `\end` is considered next. We make the assumption that this is a user defined environment, or a math environment. In the case of the example, we have a user environment that expands to a math environment. For a user defined environment, the following is executed:

```
{\cases .... \endcases}
```

In the trace, lines 28-29, you will see both `\begin{cases}` and the result of the expansion of `\cases`, but not `\cases` or the brace. However, you can see on lines 4 and 15 the expansion of the user defined commands, and on lines 3 and 18 the braces. You can see on lines 6 and 16 that a group (named ‘cell’) is opened and closed, because the builtin math environment starts a cell. This allows `&` or `\` tokens. The group defined on lines 7-8 does not exist in T_EX. Let’s hope for the best: the argument of the array should contain only letters. Whether these characters should be expanded is unclear.

- In the case `\left` and `\right`, a delimiter is read. The rules are: `\relax` and space tokens are ignored. After full expansion, the result should be one of the tokens listed above as valid delimiters, otherwise an error of the form *Invalid character in \left or \right* can be signaled. These commands come in pairs; you might get errors like *Missing \right. inserted*, or *Unexpected \right*. These commands define a group, see lines 5 and 17.
- Case `&` and `\`. These are valid only inside a cell group. They terminate a cell group and start a new one. See lines 9 to 14.
- The `\of` token is ignored.
- The `\mathchoice` command reads 4 arguments, which are remembered.
- Case of `\frac`, etc. These commands read their arguments. The main token list will contain a special slot, with the name of the command and the arguments. In the case of `\sqrt`, the first argument is optional. You say `\root A \of B`.
- The syntax of `\genfrac` is special. It takes six arguments. The following commands are executed when Tralics bootstraps:

```
\def\binom{\genfrac() \z@{}}
\def\dbinom{\genfrac(){0pt}0}
\def\tbinom{\genfrac(){0pt}1}
```

This defines three commands that take two arguments with regular syntax. The first two arguments of `\genfrac` are delimiters or empty. The next one is a dimension or empty. If empty, a default dimension will be used. In the example, the first argument is an opening parenthesis, the second is a closing parenthesis, the third is zero. The next argument is empty or a number between 0 and 3. These numbers correspond to a style: `\displaystyle`, `\textstyle`, `\scriptstyle`, and `\scriptscriptstyle` respectively. Currently an explicit number is required; everything else is treated as an empty list. On the other hand, the dimension is scanned via the `scandimen` routine (the procedure that prints lines of the form 19-20).

- `\hbox` and friends. Currently, only `\hbox` is implemented. The current value of `\everyhbox` token list is inserted, and the argument is read. It is possible to insert a math formula in such a box.
- Case of `\mbox`, `\text`, `\makebox`. Like `\hbox`, but no `\everyXXX` token list is inserted.
- Case of a math font, for instance `\@mathcal`. The font command is inserted in the token list, but the variable holding the current font is (locally) updated.
- Case of a math command (like `\alpha` listed as above). The command is read. There is a special hack: `\not\in` is converted to `\notin`, and `\not=` to `\ne`.
- `\kern`, `\hskip`, `\vskip`. See the example: on line 19 and 20, there is the trace of the routines that read the argument. The result is converted into a `\hspace`, with argument delimited by braces. Look at the trace, line 29. This will be read again later, see lines 30 to 32.
- An apostrophe is handled in a special way, as explained in the `TEX` book. Essentially `x'` is `x^{\prime}` and `x'^2` is `x^{\prime2}`.
- Mode independent commands are interpreted as usual (this includes undefined commands). This should not typeset anything.
- A character is remembered, together with the current font.
- Everything else is inserted without evaluation.

We give here an example with some fonts.

```

 $\mathtt{Ab}\mathcal{Cd}\mathbf{Ef}\mathrm{Gh}\mathit{Ij}$ 
 $\mathbb{Kl}\mathsf{Mn}$ 

```

The translation is as follows. You can notice that some variants affect only uppercase letters.

```

<formula type='inline'>
<math xmlns='http://www.w3.org/1998/Math/MathML'>
<mrow>
  <mi mathvariant='monospace'>A</mi>
  <mi mathvariant='monospace'>b</mi>
  <mi>&Cscr;</mi>
  <mi>d</mi>
  <mi mathvariant='bold'>E</mi>
  <mi mathvariant='bold'>f</mi>
  <mi> G </mi>
  <mi> h </mi>
  <mi>I</mi>
  <mi>j</mi>
  <mi>&Kopf;</mi>
  <mi>l</mi>

```

```

    <mi mathvariant='sans-serif'>M</mi>
    <mi mathvariant='sans-serif'>n</mi>
  </mrow>
</math>
</formula>

```

3.4 Translation of arrays

Whenever we see an array (this can be a global environment like ‘eqnarray’ or a local one, like ‘array’), we translate all cells one after the other. The character & is the cell separator. The command \\ is the row separator. In the case where an array ends with a \\, this gives an empty row: it will be removed. Each cell has an alignment, left, right, or center. An attribute is added only if this is not center. The ‘array’ environment has an argument that explains the type of the columns (columns not indicated are centered). The default alignment is ‘rl’ for ‘split’ and ‘align’, ‘rc’ for ‘eqnarray’, centered for matrix. You can use `\multicolumn`. This command takes three arguments: the span which should be some integer, then the alignment (one of r, l or c) and the content of the cell. The program may signal errors in case of wrong syntax. Here is an example:

```


$$\begin{array}{rc}
a&b&c&d\\
A&\multicolumn{1}{r}{B}&C&D\\
\end{array}$$


```

This is the translation of the array.

```

<table>
  <mtr>
    <td columnalign='right'><mi>a</mi></td>
    <td><mi>b</mi></td>
    <td columnalign='left'><mi>c</mi></td>
    <td><mi>d</mi></td>
  </mtr>
  <mtr>
    <td columnalign='right'><mi>A</mi></td>
    <td columnalign='right' columnspan='1'><mi>B</mi></td>
    <td columnalign='left'><mi>C</mi></td>
    <td><mi>D</mi></td>
  </mtr>
</table>

```

3.5 Trivial math

If you say ‘ x ’ and ‘123’, the translation will be

```

<p><formula type='inline'><simplemath>x</simplemath></formula> and 123</p>

```

Initially, we found this a good idea; because this can easily be converted in HTML into `<i>x</i>`. Moreover ‘ $2^{\grave{e}}$ ’ gives

```

<temporary>2<hi rend='sup'>e</hi></temporary>

```

Here the `<temporary>` element will not show in the XML tree, but is printed on the terminal if Tralics is called in the ‘interactivemath’ switch. If you invoke Tralics with the ‘-notrivialmath’ switch, these hacks are not tried, and the formula translates into:

```

<formula type='inline'>
  <math xmlns='http://www.w3.org/1998/Math/MathML'>
    <msup>
      <mn>2</mn>
      <mrow>
        <mi>i</mi>
        <mover accent='true'><mi>e</mi> <mo>&grave;</mo></mover>
        <mi>m</mi>
        <mi>e</mi>
      </mrow>
    </msup>
  </math>
</formula>

```

There are three hacks: the first is when the formula contains only a letter, the second is when the formula contains only digits, and the last one is when people use a math formula instead of `\textsuperscript`. This hack is applied only if the math formula starts with digits (no digit at all is OK; braces are ignored) followed by an exponent marker, followed by a special exponent; this has to be a single token or a token list. In the case of a single token, the hack is applied only if this is e or o. Typically, it applies in cases like 2^e and N^o. In the case of more than one token, it applies when the exponent is 'th', 'st', 'rd' and 'nd', for cases like 1st, 2nd, 3rd, and 4th. There are four rules for French: 'e', 'eme', 'ieme' and 'ieme' convert to 'e', 'ier' and 'er' convert to 'er', 'iemes', 'iemes' and 'es' convert to 'es', 'ère' and 're' convert to 're'. The accented letter can be given as è, or `\'e` or `\{e}` or `\grave{e}` or `\grave e`. The hack is applied in a case like:

```

$2 ~{\text{\small\rm \grave ere}} $

```

Instead of `\text`, `\hbox` can be used. Instead of `\small` or `\rm` any font change or font size command can be used. Up to two commands can be given. The original Perl version had 30 exceptions, including `$$\Sigma\{~{\rm it}\}$` and `\dot{\rm o}`. Compare Σ^{it} with Σ^{it} and \ddot{o} with \ddot{o} .

3.6 Conversion to XML

We use a highly recursive procedure that converts a math list into a formula. The procedure takes as argument the current style. This is one of D, T, S, or SS (display, text, script, or script script style). It is D for a display math formula, T for a normal formula.

Consider first the case where the formula has an `\over`, or a variant, not hidden inside braces. This example has 6 subexpressions, each of them have such an operator.

```

${a\over b}{a\above2mm b}{a\atop b}
{a\overwithdelims[] b}{a\abovewithdelims[]2mm b}{a\atopwithdelims[] b}$

```

The translation is

```

<formula type='inline'>
  <math xmlns='http://www.w3.org/1998/Math/MathML'>
    <mrow>
      <mfrac><mi>a</mi> <mi>b</mi></mfrac>
      <mfrac linethickness='2mm'><mi>a</mi> <mi>b</mi></mfrac>
      <mfrac linethickness='0.0pt'><mi>a</mi> <mi>b</mi></mfrac>
      <mfenced open='[' close=']'>
        <mfrac><mi>a</mi> <mi>b</mi></mfrac></mfenced>
      <mfenced open='[' close=']'>
        <mfrac linethickness='2mm'><mi>a</mi><mi>b</mi></mfrac></mfenced>
    </mrow>
  </math>
</formula>

```

```

    <mfenced open='[' close=']'>
      <mfrac linethickness='0.0pt'><mi>a</mi> <mi>b</mi></mfrac></mfenced>
    </mrow>
  </math>
</formula>

```

It is an error if the formula has more than one such operators. Otherwise, we have two parts: what precedes the operator and what follows the operator. As the example shows, some operators need delimiters. Other operators read a dimension. This dimension must be given explicitly as a sequence of digits and a unit of measure (we could do better; if you want `\parindent` instead of `2mm`, you should use `\genfrac` instead). After splitting the formula into two parts, the same idea than `\genfrac` is used. If the current style is C, the next style in the list is used for both parts of the formula (if the style is D or T, the next style is S, otherwise it is SS).

We assume from now on that the formula contains no more operators like `\over`. This means that the current style can be used for the current object. Items are handled as follows:

- A space is ignored.
- If the current token is `\text`, `\hbox`, `\mbox`, this is a command with an argument, that is interpreted using special rules. A sequence of characters produces `\mtext`, a space produces a `\mspace`, and math formulas are allowed. Errors may be signaled if the content of the argument is too complicated. The translation of

```
$ x=0 \text{provided that } y=0$ or } a=1$
```

is

```

<formula type='inline'>
  <math xmlns='http://www.w3.org/1998/Math/MathML'>
    <mrow>
      <mi>x</mi> <mo>=</mo> <mn>0</mn>
      <mrow>
        <mtext>provided</mtext>
        <mspace width='0.5em' />
        <mtext>that</mtext>
        <mspace width='0.5em' />
      </mrow>
      <mi>y</mi><mo>=</mo><mn>0</mn> </mrow>
      <mspace width='0.5em' />
      <mtext>or</mtext><mspace width='0.5em' /></mrow>
      <mi>a</mi><mo>=</mo><mn>1</mn>
    </mrow>
  </math>
</formula>

```

- In the case of `\mathop{\rm sin}`, the translation is `<mo form='prefix'>sin</mo>`. Any sequence of characters is allowed instead if 'sin'. Instead of `\rm`, any font change command that switches to 'rm' can be used.
- In the case of `\hspace`, an argument is read, converted to a dimension (in fact, a glue is read via the `scanglue` routine, the shrink and stretch parts of the glue are discarded), and the result is a `<mspace>` element. For instance `\hspace{2cm plus 3pt}` produces `<mspace width='56.9055pt' />`.
- In the case of `\displaystyle`, `\textstyle`, `\scriptstyle`, `\scriptscriptstyle`, the current style is changed, to D, T, S and SS respectively.

- In the case of `\nonscript`, the token is discarded if the style is D or T, kept otherwise. We shall see later that space disappears after such a token, if it is not discarded. Example.

```
\def\foo{\nonscript~} \foo x^{y\foo}_{\textstyle z\foo}
```

The translation is

```
<formula type='inline'>
  <math xmlns='http://www.w3.org/1998/Math/MathML'>
    <mrow>
      <mspace width='1em' />
      <msubsup>
        <mi>x</mi>
        <mstyle scriptlevel='0' displaystyle='false'>
          <mrow><mi>z</mi><mspace width='1em' /></mrow>
        </mstyle>
      <mi>y</mi>
    </msubsup>
  </mrow>
</math>
</formula>
```

- If the token is a character of category code 7 or 8, it is left unchanged (Typically, case of `^` and `_`).
- If the token is `\limits`, `\nolimits`, `\mathord`, `\mathop`, `\mathbin`, `\mathrel`, `\mathopen`, `\mathclose`, `\mathpunct`, `\mathinner`, `\ensuremath`, `\nonumber`, `\nolinebreak`, it is ignored. However, we remember that the next object should be Rel, Bin (if `\mathrel` or `\mathbin` has been seen.)
- If the token is `\big`, `\bigl`, `\bigm`, `\bigrm`, `\bigg`, `\biggl`, `\biggm`, `\biggrm`, `\Big`, `\Bigl`, `\Bigm`, `\Bigrm`, `\Bigg`, `\Biggl`, `\Biggm`, `\Biggrm`, we remember that a big object is wanted, with subtype left, right, middle, or other.
- The current token or group is translated, according to the rules given below. After that, flags may be added (if the object is declared Bin or Rel or big). The current style is changed to the next style in the case we are in a group, and the group is preceded by `^` or `_`.
- If the current token is a character, its translation will be a `<mi>`, `<mn>` or `<mo>` element. In the case of a letter, this may depend on the font attribute associated to the character. For instance `{\bf x=1}` gives `<mi mathvariant='bold'>x</mi>` and `<mo>=</mo>` and `<mn>1</mn>`.
- If the current token is `\left`, `\right`, or already translated, it is left unchanged.
- If the token is a constant, like `\alpha`, see the big list at the start of the chapter, its XML value is inserted.
- If the token is a list, like `{...}`, it will be translated, using a copy of the current style.
- If the token is a list of the form `\left ... \right`, it will be translated. After that, fences will be added (using what follows the `\left` and `\right`).
- If the token is a list of the form `\begin{xxx}... \end{xxx}`, we assume that this is an array, or a matrix, we already explained how it can be translated.
- If the token is `\mathchoice`, with its four arguments, one of them is selected according to the mode. For instance

```
\def\foo{\mathchoice{1}{2}{3}{4}}
$\foo{\displaystyle \foo}^{\foo^{\foo}}$
```

translates to

```
<mrow>
  <mn>2</mn>
  <msup>
    <mstyle scriptlevel='0' displaystyle='true'><mn>1</mn></mstyle>
    <msup>
      <mn>3</mn>
      <mn>4</mn>
    </msup>
  </msup>
</mrow>
```

- It is an error if the current token is not a command of the form `\acute`, etc, or `\overline`, those listed at the end of the section ‘basic objects’ on page 77. As a general rule, arguments are translated using the next style (i.e., smaller), unless the style is indicated (for `\dfrac` and `\tfrac`, the style is T and S, the style may be indicated for `\genfrac`). If `\foo` is as above, the translation of

```
$\tfrac{\foo}{\foo} = \dfrac{\foo}{\foo} = \frac{\foo}{\foo}
{\displaystyle \frac{\foo}{\foo}}$
```

is

```
<mrow>
  <mstyle scriptlevel='0' displaystyle='false'>
    <mfrac><mn>3</mn> <mn>3</mn></mfrac>
  </mstyle>
  <mo>=</mo>
  <mstyle scriptlevel='0' displaystyle='true'>
    <mfrac><mn>2</mn> <mn>2</mn></mfrac>
  </mstyle>
  <mo>=</mo>
  <mfrac><mn>3</mn> <mn>3</mn></mfrac>
  <mstyle scriptlevel='0' displaystyle='true'>
    <mfrac><mn>2</mn> <mn>2</mn></mfrac>
  </mstyle>
</mrow>
```

The translation of

```
\def\xbar#1{\genfrac{}{}{#1}{\foo}{\foo}}
$\xbar{0}\xbar{1}\xbar{2}\xbar{3}\xbar{}$
```

is the following. You can notice that, if the argument of `\xbar` is 2 or 3, this does not change the translation of the fraction. In T_EX we get two formulas that have the same size but are not vertically aligned (why?).

```
<mstyle scriptlevel='0' displaystyle='true'>
  <mfrac><mn>2</mn> <mn>2</mn></mfrac>
</mstyle>
<mstyle scriptlevel='0' displaystyle='false'>
  <mfrac><mn>3</mn> <mn>3</mn></mfrac>
</mstyle>
<mstyle scriptlevel='1' displaystyle='false'>
  <mfrac><mn>4</mn> <mn>4</mn></mfrac>
```



```

</mstyle>
<mstyle scriptlevel='2' displaystyle='false'>
  <mfrac><mn>4</mn> <mn>4</mn></mfrac>
</mstyle>
<mfrac><mn>3</mn> <mn>3</mn></mfrac>

```

As a final example, the translation of

```

 $\overline{x}\grave{y}\underbrace{z}\stackrel{a}{b}\overset{a}{b}$ 

```

is

```

<mover accent='true'><mi>x</mi> <mo>&OverBar;</mo></mover>
<mover accent='true'><mi>y</mi> <mo>&grave;</mo></mover>
<munder accentunder='true'><mi>z</mi> <mo>&UnderBrace;</mo>
</munder><mover><mi>b</mi> <mi>a</mi></mover>
<mover><mi>b</mi> <mi>a</mi></mover>

```

3.7 Final math mode hacks

Before, we forget it: when the formula is completely translated, we have a list of XML elements. If the list is empty, the result is `<mrow/>`. For instance, in the case of $x^{\{}}$, then exponent is empty. If the list has a single XML token, this will be the result. Otherwise, everything is put in a `<mrow>`. If the current formula, or subformula contains a style change, it is put in a `<mstyle>` element. This is not always the good solution, because the same style is used, for everything, what precedes and what follows the style command. If you look at the `\genfrac` example above, you can see that styles are added by the `\genfrac` interpreter. (the single \TeX switch is associated with two MathML attributes).

If we have a formula, of the form ${}_x^{\{2\}}_{abc}$, the translation rules explained so far tell us that we have: an underscore character, an XML element for x , a hat character, an XML element for $\{2\}$, an underscore, and an XML element for $\{abc\}$. We may have `\nonscript` tokens; they will be removed, as well as a space that follows. We have to evaluate the commands that control subscripts and superscripts. A hat character gives `<msup>`, an underscore character gives `<msub>`, and both give `<msubsup>`. It is possible for a formula to start with an underscore or a hat: in this case, the kernel is empty. It is not possible for a formula to end with hat or underscore. A kernel can have at most one subscript and at most one superscript; hence the formula above is wrong: the letter x is the first subscript to the empty kernel. A valid formula is for instance ${}_yx^2$. It translates as

```

<mrow>
  <msub><mrow></mrow> <mi>y</mi> </msub>
  <msup><mi>x</mi> <mn>2</mn> </msup>
</mrow>

```

We have mentioned above that some operators can be flagged as left, right, and that adding `\bigr` may convert a left operator into a right operator. There is a magic that converts, in some cases, the `\big` operator into fences. For instance

```

 $\bigl [ A \big ( x^2 \big ) B \bigr [$ 

```

translates as

```

<mfenced open='[' close='['>
  <mi>A</mi>
  <mfenced open='(' close=')'><msup><mi>x</mi> <mn>2</mn> </msup></mfenced>
  <mi>B</mi>
</mfenced>

```

There is another trick, that works in some cases. Consider:

$$\int_0^{\infty} f(x) dx = \big[U \big]$$

the translation is

```
<mrow>
  <msubsup><mo>&int;</mo> <mn>0</mn> <mi>&infin;</mi> </msubsup>
  <mrow>
    <mi>f</mi><mo>(</mo><mi>x</mi><mo>)</mo><mi>d</mi><mi>x</mi>
  </mrow>
  <mo>=</mo>
  <mfenced open='[' close=']'><mi>U</mi></mfenced>
</mrow>
```

The interesting point here is the placement of the inner `\mrow`. The idea is that the parentheses should remain small (not larger than the `\mrow`). In particular, it should not be influenced by the integral that precedes and the fence that follows. In some cases, it works.

3.8 Extensions

In Tralics, you can use the following three commands `\mathmo`, `\mathmi`, and `\mathmn`. They take an argument and produce a `<mo>`, `<mi>`, or `<mn>`. There is file `tralics-iso.sty` that contains

```
\def\makecmd#1{\expandafter\newcommand\csname math#1\endcsname}
\def\makemo#1#2{\makecmd{#2}{\mathmo{\&\##1;}}}
\def\makemi#1#2{\makecmd{#2}{\mathmi{\&\##1;}}}
\def\makemn#1#2{\makecmd{#2}{\mathmn{\&\##1;}}}
```

Then you can say `\makemo{x02190}{slarr}`, and this will define a command `\mathslarr`, whose translation (in math mode only) is `<mo>←</mo>`. The file provides nearly 2000 such definitions, taken from the MathML entity files, with the MathML names. These commands can be used instead of T_EX commands like `\mathchar`: remember that a math-char is a 15bit integer, where 8 bits are used for the position in a font table, 3 bits for the type, and 4 bits for the family. Only three types are defined for Tralics, but the content of the element is arbitrary (most math symbols are between U+2100 and U+27FF, there are also letters between U+1D400 and U+1D7FF). There is a command `\mathattribute` that adds an attribute pair to the last created math element. You can say for instance

```
\providecommand\operatorname[1]{%
  \mathmo{#1}%
  \mathattribute{form}{prefix}%
  \mathattribute{movablelimits}{true}%
}
```

After that,

$$\min_{x} f(x) > \operatorname{min}_{x} f(x)$$

translates as

```
<formula type='inline'>
  <math xmlns='http://www.w3.org/1998/Math/MathML'>
    <mrow>
      <msub><mo movablelimits='true' form='prefix'>min</mo> <mi>x</mi> </msub>
      <mrow>
        <mi>f</mi> <mo>(</mo> <mi>x</mi> <mo>)</mo> <mo>&gt;</mo>
      </mrow>
      <msub><mo movablelimits='true' form='prefix'>min</mo> <mi>x</mi> </msub>
```

```

<mrow>
  <mi>f</mi> <mo>(</mo> <mi>x</mi> <mo>></mo>
</mrow>
</mrow>
</math>
</formula>

```

You can use the command `\mathchardef`. This is like `\chardef`, it reads a command and a number. The number should fit on 15 bits. Otherwise, you will see an error of the form: *Bad mathchar replaced by 0: 1234567*. The command defined by `\mathchardef` behaves like `\mathchar`; said otherwise, it is not implemented in Tralics, an error will be signaled that takes the form *Undefined command \foo; command code = 232*.

TeX has a special register called `\fam`. If you say something like

```
\fam3 ${\fam9 \the\fam}\ \the\fam$
```

then the second `\the` expands to `-1`. The first give 9, but L^ATeX complains with: *\textfont 9 is undefined (character 9)*. In Tralics, you would see

```
<mrow><mn>9</mn><mspace width='6pt' /><mn>3</mn></mrow>
```

As the example shows, the family is unused, and not correctly restored. Each character has a `\mathcode`. The following

```
\mathcode{\a="0941 $a\the \mathcode{\a$
```

is interpreted by Tralics as `$a2369$`. However TeX complains, with *\textfont 9 is undefined (character A)*, because you ask the lower case letter a to be printed like the upper case letter A with `\textfont 9`.

There is a command `\delimiter`, it reads a number, but you cannot use it. The syntax rules of TeX allow `\delimiter \relax 12`, but these subtleties are not implemented in Tralics. There is a command `\radical`, it reads a number, then signals an error. The `\mathaccent` command is similar.

There are commands `\raise` and `\lower`, as well as `\vcenter`. The last one is not implemented in Tralics. The translation of

```
a\raise2cm\boxed{foo}\bar\lower 2pt\boxed{xfoo}\xbar
```

is

```
<p>a<foo>bar</foo><xfoo>xbar</xfoo></p>
```

As you can see, the specification disappear. Maybe in a future version, we will add an attribute to the box. You cannot use these commands in math mode in Tralics. In TeX, you can get an error of the form: *You can't use '\raise' in vertical mode*, while `\vcenter` is a math only command. Currently `\indent` and `\noindent` are ignored in math mode (in TeX `$$\indent_b$` produces a kernel and an index; the kernel is an empty box of width `\parindent`, of type Ord).

Chapter 4

Translating a bibliography

4.1 Introduction

As said in [6], “citations are cross-references to bibliographical information outside the current document, such as to publications containing further information on a subject and source information about used quotations. [...] There are numerous ways to compile bibliographies and reference lists. They can be prepared manually, if necessary, but usually they are automatically generated from a database containing bibliographic information.”

There are different ways to cite an author, or a text or a specific part of a text. The easiest way (for an automated system) is to use numbers, as above; if you are reading an interactive version of this document, you can click on the number, and you will see the entry in the bibliography, at the end of the document (between the index and the table of contents). This is standard practice; recommendations for a book series say: *References are cited in the text simply as numbers in square brackets, e.g. [165], do not use the abbreviations “Ref./Refs/” in the middle of a sentence. Only at the beginning of a sentence should you write “Reference [165]”*. In some cases, you can see ‘[17, p23]’, as the result of `\cite[p23]{foo}`; this means page 23 of the reference numbered 17. A bit more sophisticated are references like ‘[GMS93]’ instead of ‘[2]’ for a book by Goosens, Mittelbach and Samarin published in 1993. Computing the key is not obvious, because, if you cite a book by, say, Goethe, Molière and Shakespeare in 1793, the key will be the same, and a post-processor has to add a suffix (typically, this is done by a couple of routines named `forward.pass` and `reverse.pass` in a `bst` file). Sometimes, a more explicit scheme is used, for instance ‘Knuth, The Art Of ..., Algorithm P’, in the text, and the full reference can be found in the bibliography. A text of R. Ridolfi can be cited as ‘*Vita di Girolamo Savonarola*, 5^e éd, Florence, 1974, t. II., p. 182-183’. Note that the name of the author appears before the citation, and is not repeated inside it. In some books, citations are given as footnotes, and you can often see ‘*ibid.*’, meaning the previous cited text. These kinds of things are generally hard to fully automate. For this reason, only a simple scheme is provided by Tralics: a link to a bibliography section via a key.

The problem is essentially the following: The L^AT_EX source file contains a given number of citations, introduced by the `\cite` command or a variant. Each command defines one or more references. For each reference, a key has to be computed and typeset, an item added to the bibliography, and a link created. In L^AT_EX, the document has in general to be processed three times; the first run will print `\citation{companion2}` in the auxiliary file. This file is processed by Bib_TE_X, that generates a `bbl` file containing `\bibitem{companion2}`. On the second run, the bibliography is typeset, and the key is constructed; if it is 6, then `\bibcite{companion2}{6}` will be printed in the auxiliary file. On the last run, we know, after reading the auxiliary file, that the `\cite` command should typeset as 6.

The mechanism in `Tralics` is a bit different: there is only one run. Each `\cite` command produces a `<cit>` element, plus an entry into a biblist. At the end of the document, the bibliography is constructed, with all the necessary entries; details will be given later. This gives the equivalent of a `bbl` file, it is translated. The result of the translation is some XML element, that will be inserted somewhere in the main XML tree. Finally, a check is made to see if all references are defined. The mechanism is much simpler than in `LATEX`; this is really because, in `Tralics`, you can add an element or an attribute anywhere in the tree (at the start if you like) at any moment. In `TEX`, on the contrary, once a paragraph is typeset, you cannot modify it, and once a page is shipped out, you cannot modify the whatsits associated to it (pages numbers, in the case of `\label`, `\ref` are computed only when the page is shipped out; they are left in a `\write`, which is a special kind of whatsit).

There are some tentatives to design an XML format for bibliography data bases; none of them is really satisfactory. We give an example of an entry using the DocBook syntax:

```
<biblioentry id="abc123" type="book">
  <author>
    <surname>Flynn</surname>
    <firstname>Peter</firstname>
  </author>
  <title>Understanding SGML and XML Tools</title>
  <titleabbrev>SGML & XML Tools</titleabbrev>
  <publisher>
    <publishername>Kluwer</publishername>
    <address>Boston</address>
  </publisher>
  <isbn>0-7923-8169-6</isbn>
  <date YYYY-MM-DD="1998">1998</date>
</biblioentry>
```

This should be referenced in the text as:

```
<citation><biblioref linkend="abc123"/></citation>
```

This is the same using the TEI syntax:

```
<biblFull id="abc123" rend="book">
  <titleStmt>
    <title>Understanding SGML and XML Tools</title>
    <author>
      <persName>
        <foreName>Peter</foreName>
        <surname>Flynn</surname>
      </persName>
    </author>
    <respStmt>
      <name>http://imbolc.ucc.ie/~pflynn/books</name>
    </respStmt>
  </titleStmt>
  <extent>432</extent>
  <publicationStmt>
    <publisher>Kluwer Academic Publishers</publisher>
    <pubPlace>Boston</pubPlace>
    <idno type="isbn">0-7923-8169-6</idno>
    <date value="1998">1998</date>
```

```

    </publicationStmt>
  </biblFull>

```

This should be referenced in the text as:

```
<cit><ref target="abc123"></ref></cit>
```

These two citations were found on the Web¹. The careful reader may notice that two elements are used for the citation (in the DocBook, they are `<citation>` and `<biblioref>`, in the TEI case, they are `<cit>` and `<ref>`). Tralics uses the TEI syntax for the citation but a completely different one for the entries (the syntax is very near to Bib_TE_X). We shall explain, in the second part of this document, Chapter 6, how to convert the Tralics DTD into the TEI DTD (at least for the bibliography). A non-trivial question concerns mathematics: how can we insert math formulas like H^∞ and what about special words like: “the T_EXbook?” The main reason why Tralics does not read databases written in XML is the need of an XML parser (we have written a Bib_TE_X parser, this is more challenging).

The interaction between the main document and the bibliography is via the ‘cite key’ on the L^AT_EX level, in the XML document, this is via the Bid attribute, and for the typeset document, this is the ‘print key’. As an example, we shall consider a bbl file, created by Tralics, that contains

```
\citation{60}{footcite:thesefabien}{bid9}{foot}{phdthesis}[Sey98] ...
```

This is a temporary piece of stuff, the cite key is ‘thesefabien’, the Bid is ‘bid9’, there are two choices for the print key, ‘60’ or ‘Sey98’. The XML translation is

```
<citation from='foot' key='60' id='bid9' userid='footcite:thesefabien'
  type='phdthesis'> ...
```

As you can see, the effective print key is ‘60’. We shall explain in due time all details. Let’s start with the cite key, the only quantity that the author can choose freely. For the references from the web, this key is ‘abc123’, this is clearly a randomly chosen value, not mnemonic at all. At the start of the chapter, we have shown a reference with key ‘companion2’, this is the cite key for the second version of the L^AT_EX companion. The cite key ‘thesefabien’ is for the Ph.D. thesis of F. Seyfert. There is no constraint on the cite key for L^AT_EX: the only important thing is that the key can be printed in the auxiliary file and read by Bib_TE_X (some years ago, a colleague corrected `\cite{Christele}` to `\cite{Christèle}`, this gave an awful error; in current L^AT_EX, there seems to be no problem). On the other hand, Bib_TE_X needs an identifier. This is a character string that does not start with a digit, and contains anything but space, tabulation, double quote, percent sign, sharp sign, backslash, comma, equals sign, braces, parentheses. For XML, there are additional constraints for an ID: it has to be unique for the whole document, and some characters like a plus sign are forbidden. In a first version, we imagined to use the ‘userid’: this is formed of a prefix (of the form ‘cite:’ or ‘footcite:’, thus making it unique), followed by the cite key, where forbidden characters like the plus sign were replaced by a minus sign. However, we found an example where a rather long key differed from another one only by a forbidden character. Replacement introduced a conflict. For this reason, we added the Bid: this is automatically generated, hence is clearly unique and valid. A special feature of Bib_TE_X is that it does not create lines longer than 78 characters. It adds percent characters in a sensible position; in some cases, the choice is wrong. Here is an example:

```
\bibitem[13]{Bergamini-Champelovier-Descoubes-Garavel-Mateescu-Serwe-04-a}\RAS%
c{D.~Bergamini, D.~Champelovier, N.~Descoubes, H.~Garavel, R.~Mateescu,
  W.~Serwe},
```

As a result, you will get an error: *Undefined control sequence \RAS*. Note that there are few people who use such very long cite keys. A simple idea that works most of the time: use 4 letters for the first author, three letters for the others, two letters for the year, for instance ‘Bara-Chy-Pom02’.

¹From the web page of BiblioX: <http://www.silmaril.ie/biblioX/biblioXdoc.html>

As explained, `Tralics` cannot use an XML database; instead it will use a `bbl` file (this is some \LaTeX file, that will be translated by `Tralics`). The `bbl` can be part of the source document; in general it will be automatically constructed by `Tralics` (in the current version, `BibTeX` or any other external program can be used instead). This `bbl` file should contain, for each unsolved citation, a command that solves it (either `\citation` which is a `Tralics` command, or `\bibitem` which is a standard \LaTeX command, see section 4.2).

One question is: can the `bbl` contains other items, together with these `\bibitem` commands? In general, we have a unique `\begin{thebibliography}` at the start, a `\end{thebibliography}` at the end. The effect is to produce a chapter (or a section), in general unnumbered, whose name depends on the current language. In the case of the `Raweb`, there are more than one such environments, because of the following two facts: Each of the two database files ‘`foot`’ and ‘`refer`’ produce a set of references (the ‘`foot`’ bibliography was originally typeset as footnotes, via the `footcite` package). The third database, with name ‘`year`’, produces a sequence of sections, such as theses, books, articles, conferences, reports, etc.² Whenever `BibTeX` sees an entry with a different category than the preceding entry, it prints the `\end{thebibliography}` followed by a `\begin{thebibliography}`. Note: the modified environment takes a required argument (as usual, the longest label) and an optional argument (the name of the section title; the title itself being in the class file). As a consequence, the `bbl` files produced by the `Raweb` are incompatible with standard \LaTeX classes. Since year 2001, `BibTeX` is not used anymore for the `Raweb` and the XML result contains just a sequence of references. However, each entry has a category (this depends on the `from` and `type` attributes), entries are sorted by category. The style sheets that convert the XML to HTML or XSL/Format are assumed to add these sections (see part two of this document). A nontrivial question is then to guarantee that these two style sheets use the same splitting algorithm, and the same section titles.

The ‘`print key`’ is the value that is printed on the paper or displayed on the screen. Each `<citation>` has a `key` attribute that can be used as print key. However, an XML processor may as well ignore it, and use numbers 1, 2, 3, etc. It can even sort the entries, before assigning them a number³ (see part two of this document). In some cases, `Tralics` computes a symbolic key of the form ‘`Sey98`’. If the post-processor sorts the entries, and if the keys are not in alphabetic order, this is a bad idea.

The ‘`key`’ of an entry is a quantity defined in the database, whose purpose is to help sorting. In most cases, it is empty, (in some cases the values are junk); this value is used only in the case where no author is given (this is standard `BibTeX` practice, it means that this is rather useless). The ‘`sort key`’ of an entry is the character string used for sorting (this is `lost`⁴; `Tralics` could insert it in the resulting XML; this would allow one to merge two bibliographies). In some cases, the print key is part of the sort key. Imagine for instance a book by Samarin, Mittelbach and Goossens, written in 1993. The standard key would be `GMS93`. Assume however that the authors are taken in the given order, so that the key would be ‘`SMG93`’. Alphabetically, this is after ‘`Sey98`’, but if we sort by authors, Samarin comes before Seyfert.

4.2 Citing a document

In this paragraph, we shall explain the commands that can be put in the source document for inserting a citation, and the companion commands that solve the reference. When the `\end{document}` command is about to be translated, `Tralics` will have created a big list (maybe empty) called the ‘`biblist`’. Each item in the list has four slots: Reference, Rtype, Bid and Definition. Here Reference is the cite key, Rtype is a subtype (when merged, these two quantities give the ‘`userid`’; this subtype

²Exact ordering on page 98.

³Sorting after assigning a number is weird.

⁴Line 135 to 148 were produced with a hacked version of `Tralics` and `bst` files

is not standard L^AT_EX, you can ignore it. In some cases, two items with the same Reference and different Rtype are considered unequal, in some cases they are considered equal; thus, it is a bad idea to use the same cite key with different subtypes). The Bid is the unique id of the target, of the form ‘bid17’, and Definition is the internal number of the target of the reference (in Tralics, each XML element has an internal number). You can say: element number 25 is the target of reference ‘foo’ (syntax described later). This will solve the entry: If the entry with key foo has Bid 17, the action is to mark the entry as solved, and to add id=‘bid17’ to the element number 25. When the end of the document is sensed, the list of unsolved entries is computed, and a request is made for constructing a bbl. A warning or an error is signaled for missing items by this construction. This bbl is then translated. It is forbidden to add unresolved entries to the list. In Bib_TE_X, there is cross reference mechanism: if X has a cross reference to Y, then X must become before Y; when Y is read, its fields are used to fill missing fields in X. Unless cited explicitly, Y will not appear in the bibliography.

For the case of the Raweb, three Rtypes are defined, ‘foot’, ‘year’ and ‘refer’. There is one command, `\footcite`, to cite elements with Rtype ‘foot’ and a command, `\cite`, for anything else. We generalized this mechanism: for all commands described here, there is no difference between ‘year’, ‘refer’ and an empty Rtype.

If you say `\XMLcite{foo}`, or `\XMLcite[bar]{foo}`, this will insert a new item in the biblist, with Reference ‘foo’ and Rtype ‘bar’ (unless it already exists). It will construct a Bid for the reference. The command will call `\leavevmode`, for the case where it appears at the start of a paragraph (Remember the recommendations given above: a paragraph should start with a word, not a reference). The result of the translation is `<ref id='bid17'/>`, where ‘bid17’ should be replaced by the value of the Bid. Remember: if the optional argument is empty, omitted, ‘year’ or ‘refer’, this gives an empty Rtype. The `\cite@one` command could be defined as

```
\newcommand\cite@one[3]{%
  \XMLelement{cit}{%
    \def\temp{#1}\ifx\temp\empty\else\XMLaddatt{rend}{#1}\fi
    \XMLcite[#1]{#2}{#3}}
```

Said otherwise, `\cite@one{bar}{foo}{p25}` produces `<cit rend='bar' ><ref target='bid17'/>p25</cit>`. Some care must be taken concerning the first two arguments: they are fully expanded like `\csname`. The last argument is translated using usual rules. The code shown above is not the real code: `\XMLcite` is not called, but the result is the same. In particular `\leavevmode` is executed.

The translation ‘`\footcite {Knuth}`’ or ‘`\footcite [p.25] {Knuth}`’ is the same as ‘`\cite [foot] [] {Knuth}`’ or ‘`\cite [foot] [p.25] {Knuth}`’. The `\cite` command has two optional arguments, a type and an optional value. If only one optional argument is given, it is the value (so that ‘`\cite [p.25] {Knuth}`’ has the same meaning as in L^AT_EX). The translation of ‘`\cite [x] [y] {z}`’ is the same as ‘`\cite@one {x} {z} {y}`’ (note the order of the arguments). However, if you say ‘`\cite [p.25] {Knuth,Lamport}`’, the result is the same as ‘`\cite@one {} {Knuth} {p.25}`’, followed by ‘`\cite@one {} {Lamport} {}`’, said otherwise, the second optional argument applies only to the first citation. Between two `\cite@one` commands (that come from the same `\cite`) are inserted some `\citepunct` tokens. This is a command that can be redefined by the user. Its expansion is a comma followed by a space.

People generally say ‘`Text\footcite{blah}`’, like ‘`Text\footnote{blah}`’, without any space, because `\footcite` is assumed to produce a footnote; but this is not always the case; for this reason, the command `\footcitepre` is evaluated before insertion of the XML element associated to the citation. The default behavior is the following: if the last object on the XML tree is a normal or non-breaking space, nothing happens; otherwise, if the object is not an opening parenthesis, a space will be added. Moreover, the `\citepunct` is replaced by `\footcitesep`, a command whose translation is comma space (the idea is that you can redefine it, so that `Text\footcite{foo,bar}`

shows as ‘Text^{13,15}’, exercise left to the reader). This is a slight difference between `\footcite` and `\cite` with ‘foot’ as optional argument.

The command `\nocite` can take one optional argument (a Rtype). The effect of `\nocite{foo}` is the same as `\cite`, regarding the biblist, but it does not modify the XML tree. If you say `\nocite{*}`, this inserts a special marker, meaning: the whole database should be inserted. The Rtype is ignored in this case. Note that the correct behavior should be: Rtype is ignored only if one of ‘year’, ‘refer’ or ‘foot’.

You can say `\XMLsolvecite*[25][bar]{foo}`. The star is optional, as well as the ‘25’ and the ‘bar’. If only one optional argument is given it is the first one. This should be the identifier of an XML element (you can use `\XMLlastid`, the identifier of the last created element, or `\XMLcurrentid`, the identifier of the current element). The current element is used if the argument is missing or empty. In any case, this gives an element, say Target. The second optional argument is the Rtype. The required argument is the cite key. The result of the command is to solve the entry defined by the Reference and the Rtype. The easy case is when the reference has not yet been cited. In this case, we can use as Bid either the id of the Target, if it exists, or a new id. In this case, an attribute pair `id=‘Bid’` is added to the Target. If the entry exists in the biblist, it might be already solved, and you get an error of the form *Already solved foo*. An attribute pair `id=‘Bid’` is added to the Target, unless the Target has already an id, case where an error will be signaled, for instance *Cannot solve (element has an Id) foo* in the case

```
\cite{foo}\section{something}\XMLsolvecite{foo}
```

The problem here is the following: the section element has a Uid, this is like a Bid, it can be used as target of a `\label`. The XML norm forbids using two ids for the same element. Maybe in a future version, this will be allowed (it suffices to implement a double indirection mechanism). However, I doubt if this is a good idea: if you say `\label{foo}`, then `\ref{foo}` will produce a `<ref>` element, this is identical to the `<ref>` that comes from the `\cite`. Note that the Raweb DTD says: the target of a `<ref>` in a `<cit>` should be a `<citation>`.

If a star is given in `\XMLsolvecite`, there is a little hack. If Reference/Rtype is not found in the biblist, Tralics tries to see if there is an unsolved entry with the same Reference, Rtype arbitrary. In such a case, this entry will be solved. If there is no such entry, then a new slot is added to the reference list.

Some commands may produce strange results. Consider

```
\setbox0 =\hbox{\XMLsolvecite{foo}} \copy0 \copy0
\setbox1 =\xbox{Box}{\XMLsolvecite{bar}} \copy1 \copy1
```

This constructs two empty boxes, with an id `bid0` and `bid1`. Since the first box is unnamed, the tag will not appear in the XML tree; and no tag implies no attribute list, so that the first line is an error. On the other hand, the second box is copied twice; hence the id `bid1` appears twice in the XML tree, this is also an error (the XML is well-formed, but not valid against any DTD that says that the Bid should be an ID).

You can say `\bibitem[XX]{foo}`, the result is the same as `\par\leavevmode\XMLsolve*{foo}`. The optional argument is ignored. Note that the `\par` command terminates the current paragraph, and `\leavevmode` starts a new paragraph (in L^AT_EX, `\bibitem` calls `\item` that does more or less the same thing). The important point is that this newly created `<p>` element is the target of the reference. If you feed Tralics with the bbl of this document, you will see something like

```
<Bibliography><p id='bid0'>
David Carlisle, Michel Goossens, and Sebastian Rahtz.
De XML à PDF avec <hi rend='tt'>xmltex</hi> et Passive<TeX/>.
In <hi rend='it'>Cahiers Gutenberg</hi>, number 35-36, pages 79&ndash;114,
2000. </p>
<p id='bid1'>
```

Michel Goossens, Frank Mittelbach, and Alexander Samarin.
`<hi rend='it'>The <LaTeX/> companion</hi>.`
 Addison Wesley, 1993.</p>

4.3 Using Tralics instead of bibtex

The content of the BibT_EX database is a sequence of entries of the form

```

1 @article{example,
2   Author= "Joseph Garrigue and Didier R{\`e}my",
3   Title=  "Extending {ML} with semi-explicit higher-order polymorphism",
4   Number= "1/2",
5   Volume= 155,
6   Year=   1999,
7   Pages=  "134-169",
8   Journal= "Journal of Functional Programming",
9   Remark=  {a random example},
10  OptMonth = jan,
11  Url=     "ftp://ftp.inria.fr/INRIA/Projects/Cristal/iandc.ps.gz"}
    This is a second example.
12 @PhdThesis{thesefabien,
13   author =      {Seyfert, Fabien},
14   title =      {Problèmes extrémaux dans les espaces de Hardy,
15     Application à l'identification de filtres hyperfréquences à
16     cavités couplées},
17   school =     {Ecole de Mines de Paris},
18   year =      1998
19 }

```

After the @ character, there is a keyword, or an entry type. The recognized entry types are article, book, booklet, conference, coursnotes, inbook, incollection, manual, masterthesis, misc, phdthesis, techreport, unpublished, as well as mastersthesis, a synonym of masterthesis. These types are not part of the BibT_EX language, but are described in any good book about L^AT_EX, they are the only ones recognized by Tralics. The case is irrelevant.

There are three keywords. The first is ‘comment’. If you say `@comment{foo}`, this makes ‘foo’ a comment. Since everything outside the scope of a keyword or an entry is discarded, there is no real need for a comment keyword, or a comment character. In particular, the percent sign is not a comment character inside a BibT_EX file. If you insert a percent sign in a field, you have to remember that BibT_EX will replace newline characters by spaces, and insert newline characters in the bbl file wherever it judges adequate. Hence, the percent character will behave, in the bbl, as a comment character with a random scope. You can say `@string{Foo="bar"}`. This has as effect to define a macro, named ‘foo’ (the case is irrelevant), whose value is ‘bar’. In the example, there is a string after the equals sign, but any expression could be used, including one that uses macros. A macro must be defined before its use; it is always possible to redefine the macro. There are 12 predefined macros; there are jan, feb, mar, apr, may, jun, jul, aug, sep, oct, nov, dec, and the value of the full name of the month, for instance ‘janvier’, ‘January’ or ‘Januar’, depending on the language. Note the following hack: the macros are computed at the start of the run, using some random language (typically, english), and again, when `\begin{document}` is seen. This gives you a chance to select the correct language. The last keyword is ‘preamble’. If you say `@preamble{"foo"}`, the effect is to add the string ‘foo’ at the start of a bbl file. Instead of “foo”, a general value can be used. *Note:* Instead of braces, you can use parentheses to delimit the value of an entry, a string or the

preamble. Inside the value, you can use braces instead of double quotes. Thus `@preamble({foo})` is a valid preamble. More than one preamble can be given, they will be merged, in order.

After an entry type comes the cite key, followed by a sequence of pairs, of the form `field=value`, separated by commas. The following field names are recognized: `address`, `author`, `booktitle`, `chapter`, `crossref`, `doi`, `edition`, `editor`, `howpublished`, `institution`, `isbn`, `isrn`, `issn`, `journal`, `key`, `month`, `note`, `number`, `organization`, `pages`, `publisher`, `school`, `series`, `title`, `type`, `url`, `volume`, `year`. The case is irrelevant. If a field name is given, whose value is not in the previous list, it will be ignored. Note: some additional fields are read by `Tralics`, hence some errors can be signaled while parsing them.

The value of a field can be a number (lines 5 and 6 in the example), or a macro name (as on line 10), or a constant in braces (line 9), or a constant in double quotes (other lines). It is possible to concatenate basic fields, for instance `apr # "~1"`, via the use of the sharp operator. The way `BibTeX` handles braces, quotes and backslashes is a bit special. When `BibTeX` parses a value, there should be as many opening braces than closing braces; trying to put a backslash before a brace has no effect. If a string is delimited by double quotes, then braces are needed to hide double quotes. Special characters should be entered as `{\`e}`, never as `\`{e}`, but `Tralics` accepts `é`.

Assume that `Tralics` has seen `@article`, then an opening brace or parenthesis, followed by `example`. All fields up to the closing brace (or parenthesis) are read, but, if the entry is useless, no error is signaled in case of undefined macros, or duplicate fields. If an entry is useful, all fields are remembered; if it has a crossref to an entry `X`, then `X` becomes useful. Remember: each entry has a `Rtype`, this is in general empty; it is added as a prefix to the cite key. For instance, `thesefabien` gives `footcite:thesefabien`. In the case of a crossreference from `Y` to `X`, we use as prefix for `X` the prefix of `Y`. An entry can be useful because the user has said `\nocite{*}`. There is a special hack for the `Raweb`: we have three types of entries, `foot`, `year` and `refer`. We already mentioned that the types `year` and `refer` as the same as the empty type. The difference is that `\nocite` applies only to entries from the file `year`, never to `foot` (there is an implicit `\nocite` for `refer`).

An entry is useful because it is cited (by `\cite` or `\nocite`). Since `BibTeX` is generally case insensitive, the entry shown above is useful if you say `\cite{Example}`. However, for `LATEX`, `\cite{Foo}` and `\cite{f00}` are two different items, as a consequence, two references are needed. Thus, an entry named `foo` is ambiguous. For this reason, you should always capitalize entries in a consistent way (say, use always lowercase letters), and use the same method in the `LATEX` document.

After some manipulations, the entry is printed on the `bbl` like this (`BibTeX` version)

```

20 \citation {GR99a}{example}{article}
21 \bateurs{\bpers\RAo J.\RAB \RAB Garrigue\RAB \RAf \bpers\RAo D.\RAB \RAB
22   R{\`e}my\RAB \RAf }
23 \cititem{btitle}{Extending {ML} with semi-explicit higher-order polymorphism}
24 \cititem{bjournal}{Journal of Functional Programming}
25 \cititem{bnumber}{1/2}
26 \cititem{bvolume}{155}
27 \cititem{byear}{1999}
28 \cititem{bpages}{134--169}
29 \url{ftp://ftp.inria.fr/INRIA/Projects/cristal/iandc.ps.gz}
30 \endcitation

```

or like that (`Tralics` version)

```

31 \citation{60}{footcite:thesefabien}{bid9}{foot}{phdthesis}[Sey98]
32 \bauthors{\bpers[Fabien]{F.}{Seyfert}{}}
33 \cititem{btitle}{Problèmes extrémaux dans les espaces de Hardy, Application
34 à l'identification de filtres hyperfréquences à cavités couplées}
35 \cititem{btype}{Ph. D. Thesis}

```

```

36 \cititem{school}{Ecole de Mines de Paris}
37 \cititem{year}{1998}
38 \endcitation

```

There are some slight differences between these two entries. If you compare lines 20 and 31, you can see that the number of arguments of the `\citation` command has changed from 3 in the original version to six in the current version. The following were added: the type (here ‘foot’), the unique id (here ‘bid9’), the numerical print key (here 60). The first entry was created by BibT_EX, that cannot guess the Rtype of the reference nor the Tralics unique id. It could have computed the number 60, but we initially thought that only one of the two keys were useful (in the current version, the `\citation` command takes five arguments, plus an optional one after these.) If you compare lines 21 and 32, you can notice two differences. First, we have decided, in 2005, to add an optional argument to the `\bpers` command (it contains the full first name). This might be used for the Ra2005. The second difference is that it is impossible, in BibT_EX, to print braces inside a name. Thus we used `\RAo` for an opening brace, `\RAf` for a closing brace and `\RAb` for a pair of closing and opening braces. Omitting the first line, the fields are printed in the following order:

1. Unless the type is proceedings, the author.
2. In the case of a book or booklet, the editor.
3. The title.
4. In the case of proceedings or incollection, the editor.
5. In the case of an article, the journal, number, and volume.
6. In the case of a book or inbook, the edition, series, number, volume, publisher, address.
7. In the case of a booklet, the howpublished and address.
8. In the case of incollection, the booktitle, series, number, volume, publisher, address.
9. In the case of inproceedings or conference, the booktitle, series, number, volume, organization, publisher, editor, pages, address.
10. In the case of a manual, the organization, edition, address.
11. In the case of masterthesis, coursnotes, or phdthesis, the type, school, and address.
12. In the case of a techreport, the type, number, institution, address. (For the case of masterthesis, phdthesis and techreport, the type has a default value, that depends on the language, and is initialized together with the ‘jan’ macro).
13. In the case of misc, the howpublished, editor, booktitle, series, number, volume, publisher, address.
14. In the case of proceedings, the organization, series, number, volume, publisher, address.
15. In any case, the month, year.
16. In the case of inbook or incollection, the chapter.
17. In the case of inbook, incollection, article or proceedings, the pages.
18. In any case, the doi, url, note.

This may seem confusing (is there a standard way for formatting entries?). Note that missing fields are not printed. In some case, BibT_EX prints a message like “there’s a number but no series” or “can’t use both volume and number”. No such message is printed by Tralics.

Two keys are computed, the ‘Sey98’ or ‘GR99a’ in the example, and the sort key, which is something longer. In fact, handling the author or editor field produces four characters strings L_1 , L_2 , L_3 and L_4 . The L_4 string is the argument of the `\bauthors` or `\beditors` (see lines 21, 32). The L_1 string is ‘Sey’ or ‘GR’, the L_2 string contains the full name (it is like L_4 , without the full first name, and braces) and L_3 contains only the last name (not the first name).

We consider the author (the editor in the case of proceedings). This may give a triple L_1 , L_2 , L_3 , unless the field is missing. If it is missing, we consider the ‘key’ field. If it is not empty, then L_1 is formed of the first three characters of the field, L_2 is empty, L_3 is the field. If it is empty, we consider the editor (author, in case of proceedings). If this is empty, we consider the cite key, handle it like the ‘key’ above. Note: in the case of ‘Lo{\i}c’ the first three characters are ‘Lo{\i}’, the last two characters are ‘{\i}c’. In the case of ‘Lo\ic’, asking for the first three or last two characters gives the full string. The last two characters are the year are added to L_1 , so that we may obtain ‘GR99’. This gives the print label. The L^AT_EX companion says that you can use `year="{\SortNoop{86}}1991"`. With the rules above, the last two characters of the year are ‘91’. However, `Tralics` uses the full year, not ‘861991’ when it computes the sort key. In the case when `Tralics` processes the Raweb for, say year 2003, if a reference has type ‘year’, then its year field should not be missing, and should be ‘2003’. Otherwise an error is signaled⁵. The sort key is computed as follows: first a prefix, then the cite label, then L_2 , then the year, then the title. All characters are converted to lower case. Note: when BibT_EX converts {\E} to lower case, the result is ‘e’. Converting ‘É’ can produce strange results. Such subtleties do not exist in `Tralics` (the style sheet that converts the XML to HTML sorts all entries; how can we tell it that the author used a `\SortNoop` command?).

Because of this sort-again, we try to be clever. Said otherwise, for the Raweb, and only the Raweb, we use a prefix, formed of a letter and L_3 . The prefix 0 for an entry of Rtype ‘refer’, 1 for an entry of Rtype ‘foot’, and for entries of Rtype ‘year’, it is: 2 for book, booklet, proceedings, 3 for phdthesis, 4 for article, inbook, incollection, 5 for conference, inproceedings, 6 for manual, techreport, coursenotes, 7 for masterthesis, misc, unpublished. These numbers are indices into a table. Currently the order is 02345671. In a future version, this might be changed (however, the result should be compatible with the style sheets described in the second part of this report).

Let’s repeat: for the Raweb case, we have in the sort key a prefix that depends on the type and Rtype, followed by the author names, the print key, the full author names, the year, the title. In this case, the content of the `bbl` will be as on line 31: the first argument of `\citation` is not the print key, but the index of the reference in the table after sorting. On the other hand, for the non-Raweb case, the sort key starts with the print key, the `bbl` looks like line 20. The important point is: assume that we have two entries with the same print key, say ‘GR99’; we must change them to ‘GR99a’ and ‘GR99b’, this is easy to do when they are consecutive. The following piece of code comes from a standard `bst` file. Parsing a `bst` file is rather easy (maybe one day, `Tralics` will do it). The important point is that a postfix language is used: instead of: if a then b else c, you say: a b c if. This piece of code computes a suffix for every entry that has the same key as the previous one.

```

39 FUNCTION {forward.pass}
40 { last.sort.label sort.label =
41   { last.extra.num #1 + 'last.extra.num :=
42     last.extra.num int.to.chr$ 'extra.label :=
43   }
44   { "a" chr.to.int$ 'last.extra.num :=
45     "" 'extra.label :=
46     sort.label 'last.sort.label :=
47   }
48   if$
49 }
```

The is the companion routine, executed in reverse order. Its purpose is to add the ‘a’ suffix when the next entry has a ‘b’ suffix. There is a piece of code, not shown here, that computes the longest label. This is sometimes nonsense (consider the ‘De La Cruz’ case below).

⁵No error is signaled if you invoke `Tralics` with the switch ‘nobibyearerror’

```

50 FUNCTION {reverse.pass}
51 { next.extra "b" =
52   { "a" 'extra.label := }
53   'skip$
54   if$
55   label extra.label * 'label :=
56   extra.label 'next.extra :=
57 }

```

In summary, when Tralics is used instead of BibT_EX, the following happens. We have a big entry list, and a list of typed databases. From the entry list, we consider only unsolved ones. For each entry, a prefix is computed, for instance, ‘footcite:fabien’, by considering the Rtype, the word ‘cite:’ and the cite key. If the Rtype is anything else than ‘foot’, an empty value will be used. When an entry with cite key ‘foo’ is read from a database of type ‘bar’, the same mechanism is applied. The type of a database is currently one of ‘year’, ‘refer’ or ‘foot’ (the default being ‘year’). We plan to extend this mechanism: more than these three types can be used; ‘year’ and ‘refer’ are the same as empty, but ‘refer’ has an implicit `\nocite`.

All entries from the database files are read, and stored if useful. For each entry X that has a crossreference to Y, missing fields in X are copied from Y. After that Y is discarded (unless cited via `\cite` or `\nocite`). An error is signaled in case some references are undefined. After that, the sort label is computed, entries are sorted, the print label is computed, and everything is printed on the `bbl` file. This is `apics_.bbl` if the jobname is ‘apics’. Note the underscore in the name.

This is the XML version of the reference above, as used in the Raweb2004.

```

58 <citation from='foot' key='60' id='bid9' userid='footcite:thesefabien'
59   type='phdthesis'>
60 <bauteurs><bpers prenom='F.' part='' nom='Seyfert' junior=''</bauteurs>
61 <btitle>Problèmes extrémaux dans les espaces de Hardy,
62   Application à l'identification de filtres hyperfréquences à cavités
63   couplées</btitle>
64 <btype>Ph. D. Thesis</btype>
65 <bschool>Ecole de Mines de Paris</bschool>
66 <byear>1998</byear>
67 </citation>

```

4.4 The format of a name

We shall discuss in this section how names can be used in a BibT_EX file, and how Tralics constructs keys. We have already mentioned a procedure that gives ‘Sey’ from ‘Seyfert’. It is not satisfactory, but is used only in rare cases (when the year is strange, or a strange key has been used). The important point that, when we fetch the first three letters of `Lo\ic`, we do not obtain neither ‘Lo\’ nor ‘Lo\’’. The mechanism explained here is more subtle. The L^AT_EX companion explains that, in order to get ‘Göd’ for the key, you should use one of the first names shown here, not the others.

```

author = {A. G{\o}del and B. G{\o}del},
editor = {C. {G{\o}del} and D. {G{\o}del}}

```

The rule is that special BibT_EX characters are formed by a left brace followed by a backslash. In the case C, the brace is inside another brace. In fact, if the bibliography contains the following

```

68 @Article{GoA,
69   author = {A. G{\o}del      }, title="X"}
70 @Article{GoB,

```

```

71   author = {B. G{"o}del   }, title="X" }
72 @Article{GoC,
73   author = {C. G{"o}del   }, title="X" }
74 @Article{GoD,
75   author = {D. G{"o}del   }, title="X" }

```

Then the translation by Tralics looks like this. If you compare with lines 60 and 61, you can see that the full first name appears, empty attribute pairs part and junior are not shown.

```

76 <biblio>
77 <citation from='year' key='Ga' id='bid2' userid='cite:GoC' type='article'>
78 <bauteurs><bpers prenom='C.' nom='Gödel' prenomcomplet='C.'/></bauteurs>
79 <bttitle>X</bttitle>
80 </citation>
81 <citation from='year' key='Gb' id='bid3' userid='cite:GoD' type='article'>
82 <bauteurs><bpers prenom='D.' nom='Gödel' prenomcomplet='D.'/></bauteurs>
83 <bttitle>X</bttitle>
84 </citation>
85 <citation from='year' key='Göd' id='bid0' userid='cite:Goa' type='article'>
86 <bauteurs><bpers prenom='A.' nom='Gödel' prenomcomplet='A.'/></bauteurs>
87 <bttitle>X</bttitle>
88 </citation>
89 <citation from='year' key='Göd' id='bid1' userid='cite:GoB' type='article'>
90 <bauteurs><bpers prenom='B.' nom='Gödel' prenomcomplet='B.'/></bauteurs>
91 <bttitle>X</bttitle>
92 </citation></biblio>

```

The same file processed by Bib_T_EX gives the following keys: $\{G\{a}\}$, $\{G\{b}\}$, $G\{o\}da$ and $G\{o\}db$. The first two keys are invalid. The reason why suffixes a and b are added is that a special Bib_T_EX function removes braces and funny characters when comparing keys. Such a function is not implemented in Tralics, thus labels $G\{o\}d$ and $G\{o\}db$ are considered different, although their translation is the same. In Tralics, the best thing to do is use ‘Gödel’ as name.

Since lots of errors may be found in bibliography files, Tralics tries to be clever. First, it replaces ‘ $\{c\}$ ’ by ‘ \c ’ and ‘ $\{C\}$ ’ by ‘ \C ’. It also replaces ‘ $\{v\{c\}\}$ ’ by ‘ $\{v\ c\}$ ’. Maybe other replacements of this kind will be made in a future version. Expressions of the form $\{a\}e$ are replaced by $\{a\}e$. We also replace backslash-space by a single space.

After that, characters or group of characters are classified, this will make parsing easier. A sequence like ‘ $\{foo\}$ ’ will be considered as a single random character; something like $\{e\}$ as a single lower case letter, $\{E\}$ as a single uppercase letter. The expression $\{e\}$ will be replaced by $\{e\}$ with a warning, $\{i\}$ will be rejected (unless inside braces) because a single character is needed after backslash-accent. Commands like $\{foo\}$ are also rejected. Note that an ampersand & is an error (some people try to use this instead of ‘and’). Character categories are: space, comma, dash, and tie (this is a \sim). In a case like this,

```

93 @Article{cruz,
94   author = {Maria {\MakeUppercase{d}e La} Cruz},
95   title="X" }

```

the print key computed by Bib_T_EX is $\{d\}e\ La\}C$, this typesets as ‘De LaC’. Such a construct is not understood by Tralics, that thinks that the last name is ‘Cruz’.

If more than one author is given, in the author or editor list, you should use ‘and’ as separator. Case is irrelevant, a space is required. For instance, the following citation contains 3 authors and others. The print key is ‘AAJA+’, because the last author has a double last name.

```

96 @Article{many,
97   author = {Joe~And and And,Joe and Joe-And And others}
98   title="X" }

```

The BibT_EX transformation of this is

```

99 \bibitem[AAJA{\etalchar{+}}]{many}
100 Joe And, Joe And, Joe-And, et~al.

```

If the list is too long, you can use ‘others’ as the last name (case is important). A name has four components: von, First, Last and Junior. On line 32, you can see the value of the full first name, then the abbreviated first name, then the von part (empty) then the last name, then the junior part (empty). In Tralics, the von part is always merged with the last name. Consider somebody named Jean de la Fontaine. French rules say that the particle ‘de’ should be omitted, unless preceded by the first name or a word like ‘Monsieur’. In particular, in the dictionary, you will find him between La Follette (an American politician) and Lafontaine (a Canadian politician), not between Delacroix and Delage. More interesting is the case of Marie Joseph Gilbert Motier, marquis de La Fayette. The name of this guy is ‘Motier’, but he is known as ‘La Fayette’. Another example is William Thomson (For his work on the transatlantic cable Thomson was created Baron Kelvin of Largs in 1866. The Kelvin is the river which runs through the grounds of Glasgow University and Largs is the town on the Scottish coast where Thomson built his house.) How this guy should be cited is unclear: William Thomson or Lord Kelvin?

The simple case is when two fields are given, with a comma between. The first field is the last name, the other field is the first name. Then comes the case of three fields: last name, junior, and first name. You cannot use more than three fields, that is, you cannot give more than two commas. In the case no comma is given, we look at a ‘von’ part. This is something that starts at a lower case letter. For instance,

```

101 @Article{poussin,
102   author = {Charles Louis Xavier Joseph de la Vall{\a'e}e Poussin   },
103   title="X" }

```

This is what BibT_EX puts in the bbl file:

```

104 \bibitem[dLVP]{poussin}
105 Charles Louis Xavier~Joseph de~la Vall{\a'e}e~Poussin.
106 \newblock X.

```

The translation by Tralics is the same, but no ties are inserted (BibT_EX inserts one for the first name, the von part, the last name, see T_EXbook, page 92). The ‘De La Cruz’ example shows how you can fool BibT_EX. Tokens between names are recognized. For instance, consider:

```

107 @Article{strange,
108   author = {A-b-C and A.b.C and A~b~C and A.Bb.Cc},
109   title="X" }

```

This is how BibT_EX interprets the names. Authors number two and four have only a last name, no von part, no first name.

```

110 \bibitem[bCAbCA]{strange}
111 A~b~C, A.b.C, A~b~C, and A.Bb.Cc.
112 \newblock X.

```

This is the translation by Tralics. You can see that, for the last author, one dot has been replaced by a space: this is done in case no other way is found to split the name, but there is an upper case on each side of the dot. You can also see that BibT_EX inserts some characters (here ties) instead of dashes. Tralics keeps the dashes, whenever possible.

```

113 \citation{bCAbCB}{cite:strange}{bid3}{year}{article}
114 \bauthors{\bpers[A]{A.}{b-C}{
115           \bpers[]{}{A.b.C}}

```



```

116      \bpers[A]{A.}{b~C}{}
117      \bpers[A]{A.}{Bb.Cc}{}
118 \cititem{btitle}{X}
119 \endcitation

```

Here is another example.

```

120 @Article{strange2,
121   author = {Jean-Claude XX and J.-Ch. YY and J.-{Ch.} ZZ},
122   title="X" }

```

This is the translation by Bib_TE_X, in ‘abbrv’ mode. The format used in plain mode is `{ff~}{vv~}{ll}{, jj}`, and in abbrv mode, it is `{f.~}{vv~}{ll}{, jj}`. This is explained in any good reference about Bib_TE_X⁶

```

123 \bibitem{strange2}
124 J.-C. XX, J.-C. YY, and J.-C. ZZ.
125 \newblock X.
126 \bibitem{poussin}
127 C.~L. X.~J. de~la Vall{\a'e}e~Poussin.
128 \newblock X.

```

This is the translation by Tralics. The quantity ‘{Ch.}’ is considered as a single character. No dot is added after it, since it is terminated by a dot.⁷

```

129 \citation{XYZ}{cite:strange}{bid3}{year}{article}
130 \bauthors{\bpers[Jean-Claude]{J.-C.}{XX}{}
131          \bpers[J.-Ch.]{J.-C.}{YY}{}
132          \bpers[J.-{Ch.}]{J.-{Ch.}}{ZZ}{}}
133 \cititem{btitle}{X}
134 \endcitation

```

The print key is computed as follows: Each author gives an initial (if the name is complicated, more than one will be used, for instance Poussin gives four letters ‘dlVP’). If a single author is cited, and if it gives less than three letters, then the first three letters of its name are used (for instance, Seyfert gives ‘Sey’). If more than four authors are given, only the first three ones give an initial, there is a ‘+’ sign at the end. If ‘and others’ is given, there is also a ‘+’ sign.

We show here the sort key, as computed by Tralics, for some the entries shown above. Remember that these entries have no year field and that the title is X.

```

135 cru m. {\makeuppercase{d}e la}. cruz          x
136 g c. {g{"o}}del          x
137 g d. {g{"o}}del          x
138 g{"o}d a. g{"o}}del          x
139 g{"o}}d b. g{"o}}del          x
140 aja+ j. and j. and joe-and etal          x
141 dlvp c. l. x. j. de la vall{\a'e}e poussin          x

```

These are the keys, for the same entries, computed by Bib_TE_X, using the alpha style. You can see that Bib_TE_X uses last name and first name, whereas Tralics uses abbreviated first name then last name. The format is: `{vv{ } }{ll{ }}{ ff{ }}{ jj{ }}`

```

142 delac   dela cruz maria          x
143 god    godel a          x
144 god    godel b          x
145 g      godel c          x

```

⁶For instance *Tame the Beast*, by Nicolas Markey, available on CTAN

⁷According to N. Markey, you should say `{\relax Ch}ristopher` in Bib_TE_X

```

146 g      godel d      x
147 aaaja  and joe  and joe  joe and  et al      x
148 dlvp   de la vallee poussin charles louis xavier joseph      x

```

4.5 Commands for the bbl

The Raweb DTD explains that the following items can appear inside a bibliography entry.

- `<baddress>`, the address of the publisher.
- `<bauteurs>`, the names of the authors of the document.
- `<bbooktitle>`, the title of the document.
- `<bchapter>`, the chapter number, in case it is relevant.
- `<bedition>`, the edition number.
- `<bdoi>`, the DOI.
- `<bediteur>`, the names of the editors.
- `<bhowpublished>`, how a document is published, in unusual cases.
- `<binstitution>`, the institution, for the case of a report.
- `<bjournal>`, the journal where an article can be found.
- `<bmonth>`, month of publication.
- `<bnote>`, a note.
- `<bnumber>`, the number of a report, etc.
- `<borganization>`, organizer (of a conference ...)
- `<bpages>`, page number (or sequence of pages).
- `<bpublisher>`, the publisher of a book.
- `<bschool>`, the school for a thesis.
- `<bseries>`, the name of a series.
- `<bttitle>`, the title of the reference.
- `<btype>`, the type (of a thesis, technical report).
- `<bvolume>`, the volume.
- `<xref>`, an external link.
- `<byear>`, the year of publication.

In almost every case, if the database file contains a field ‘foo’ with value ‘bar’, the `bbl` file will contain `\cititem{bfoo}{bar}`, and this is translated into `<bfoo>bar</bfoo>`. The `\cititem` command takes two arguments. The second argument is translated as usual. The first argument is the name of the resulting element. There is a hook: in the case where `\cititem-foo` is defined, this macro is used instead of the default procedure. If the database contains a ‘url’ field, the result is a call to the `\url` command, that will produce a `<xref>` element. The `\cititem` command should be used only in a bibliography.

If the entry in the database contains a ‘author’ or ‘editor’, the `\bauthors` or `\beditors` commands will be called. These two commands must be used inside a bibliography. They take a single argument, translate it, and put the result in a `<bauteurs>` or `<beditor>` element. Note: the bibliography part of the Raweb DTD was meant to be temporary. For this reason, the names were chosen so as to replace them easily with new names (hence the prefix ‘b’); For some reason,

‘auteurs’, ‘editeur’ and attributes of ‘bpers’ have French names. Later on, we decided to modify the Tralics names, hence the ‘bauthors’ and ‘beditors’. Because ‘bauteurs’ had a final s, we added an s to both command names; not the best choice.

The `\bpers` command takes one optional argument, and 4 required arguments. The translation is an empty `<bpers>` element with following attributes: `prenomcomplet` for the optional argument, and `prenom`, `part`, `nom`, `junior` for the required arguments.

The `\citation` command constructs a `<citation>` element. It takes 5 required arguments, and an optional argument. The optional argument is ignored. Other arguments are converted to attributes. The whole text, up to `\endcitation` is translated in bibliography mode, and added to the `<citation>` element. Example:

```

149 \citation{a}{b}{c}{d}{e}
150 \cititem{foo}{bar}
151 \beditors{\bpers[a]{b}{c}{d}{e} \bpers[] {B}{C}{} \cititem{etal}{} }
152 \endcitation
    The translation is
153 <citation from='d' key='a' id='c' userid='b' type='e'>
154   <foo>bar</foo>
155   <bediteur>
156     <bpers prenom='b' part='c' nom='d' junior='e' prenomcomplet='a' />
157     <bpers prenom='B' nom='C' />
158     <etal />
159   </bediteur>
160 </citation>

```

4.6 Other commands

The `\bibliography` command takes one argument, this is a comma separated list of database files. Spaces are ignored. The command can be given more than once. This command (the last occurrence) defines the position where the bibliography should be inserted.

The command `\insertbibliohere` can be used to force the position of the bibliography. It overwrites the location specified by the previous command.

The environment ‘thebibliography’ can be used for typesetting the bibliography. There is an optional argument (ignored), a required argument (ignored), an optional argument (ignored). The result is an XML element whose name is defined by `\refname`, by default ‘Bibliography’, and whose content is formed of the translation of the environment. You can redefine this `\refname` command. An error is signaled if strange commands appear in the argument, but not for invalid characters (in particular, space cannot appear in an element name). The command can be empty. In this case, the name will not appear in the XML result.

The command `\bibliographystyle` takes one argument. Its translation is empty. The argument is remembered. This is the style to use. If the argument is ‘bibtex:’, this is an indication that BibTeX should be used instead of Tralics for the production of the `bbl`. The style can be given after the colon, or with the invocation of another command. If the argument is ‘program:foo’, this means to use foo as program. For instance `\bibliographystyle{program:cat -v}`. In this example, this will print the auxiliary file; this is not good, because the command should create the `bbl` file (its argument is `jobname.aux`, data must be written on `jobname.bbl`). A second `\bibliographystyle` command can be used for specifying the style (the default is ‘plain’). Example. Consider a file that contains these lines

```

161 \documentclass{article}
162 \begin{document}

```

```

163 \AtEndDocument{\bibitem{unused}Hey}
164 \bibliography{torture}
165 \bibliographystyle{bibtex:}
166 \cite{poussin,cruz,many,strange,unused}
167 \end{document}

```

When Tralics sees the `\end{document}` command, it evaluates it (with the hooks, etc.) After that, a `bbl` is created and translated. If there is no unsolved entry, nothing happens. If no style command indicates that BibT_EX or an external program should compute the `bbl`, then Tralics does it, as explained above. In the case of the Raweb, three database files are used: `apicsfoot_2004`, `apicsrefer_2004`, and `apics2004`. These files are typed ‘foot’, ‘refer’ and ‘year’. In the non-Raweb case, files in the list indicated by `\bibliography` are used. If a file is named ‘miaou+refer’ or ‘miaou+foot’ and does not exist, then miaou is tried instead; in this case the type will be ‘refer’ and ‘foot’ (otherwise, it is ‘year’). In the case an external program is used, a minimal auxiliary file is created. In the case of the example, it will contain

```

168 \citation{poussin}
169 \citation{cruz}
170 \citation{many}
171 \citation{strange}
172 \bibstyle{plain}
173 \bibdata{torture}

```

The database `torture.bib` contains a sequence of entries, plus the following lines. In order to understand the last line, you have to remember that character strings are always balanced against braces. Hence it is not: open brace concatenated with 1 concatenated with 1 and close brace. It is: open brace, double quote, space, sharp, etc, up to double quote, close brace.

```

174 @String{ stra= {\def}}
175 @String{ strb= "#1" }
176 @String( strc= "\mycmd " )
177 @Preamble (stra # strc # strb )
178 @Preamble( "{" #1 #1 "}")

```

After that, the external program is called, and the `bbl` file is read. In the example this gives the following. The first line is the preamble.

```

179 \def\mycmd #1{" #1 #1 "}
180 \begin{thebibliography}{1}
181
182 \bibitem{many}
183 Joe And, Joe And, Joe-And, et~al.
184 \newblock X.
185
186 \bibitem{strange}
187 A~b~Cde.
188 \newblock X.
189
190 \bibitem{poussin}
191 Charles Louis Xavier~Joseph de~la Vall{\a'e}e~Poussin.
192 \newblock X.
193
194 \bibitem{cruz}
195 Maria {\MakeUppercase{d}e La}~Cruz.
196 \newblock X.

```

197

198 `\end{thebibliography}`

After that, the bibliography is translated and inserted. The resulting XML file is shown here.

```

1 <?xml version='1.0' encoding='iso-8859-1'?>
2 <!DOCTYPE std SYSTEM 'classes.dtd'>
3 <!-- translated from latex by tralics 2.5 (pl4)-->
4 <std>
5 <bibliography>
6 <Bibliography><p id='bid2'>
7 Joe And, Joe And, Joe-And, et&nbsp;al.
8 X.</p>
9 <p id='bid3'>
10 A&nbsp;b&nbsp;b&nbsp;Cde.
11 X.</p>
12 <p id='bid0'>
13 Charles Louis Xavier&nbsp;Joseph de&nbsp;la Vallée&nbsp;Poussin.
14 X.</p>
15 <p id='bid1'>
16 Maria De La&nbsp;Cruz.
17 X.</p>
18 </Bibliography>
19 </bibliography><p><cit><ref target='bid0'/></cit>, <cit><ref target='bid1'/></cit>,
20 <cit><ref target='bid2'/></cit>, <cit><ref target='bid3'/></cit>,
21 <cit><ref target='bid4'/></cit></p>
22 <p id='bid4'>Hey</p>
23 </std>

```

Finally, we show here everything printed on the screen, including all warnings by BibTeX.

```

1 This is tralics 2.5 (pl4), a LaTeX to XML translator
2 Copyright INRIA/MIAOU/APICS 2002-2005, Jos\`e Grimm
3 Licensed under the CeCILL Free Software Licensing Agreement
4 Starting xml processing for testb.
5 Configuration file identification: standard $ Revision: 2.12 $
6 Read configuration file /user/grimm/home/cvs/tralics/.tralics_rc.
7 Bib stats: seen 5(1) entries
8 This is BibTeX, Version 0.99c (Web2C 7.3.1)
9 The top-level auxiliary file: testb.aux
10 The style file: plain.bst
11 Database file #1: torture.bib
12 Warning--empty journal in many
13 Warning--empty year in many
14 Warning--empty journal in strange
15 Warning--empty year in strange
16 Warning--empty journal in poussin
17 Warning--empty year in poussin
18 Warning--empty journal in cruz
19 Warning--empty year in cruz
20 (There were 8 warnings)
21 Used 1718 commands
22 Math stats : formulas 0, non trivial kernels 0, cells 0,
23   special 0 trivial 0, \mbox 0 large 0 small 0.
24 List stats: short 0 inc 5 alloc 1192

```

```

25 Buffer realloc 2 string 1400 size 13877; merge 4
26 Macros created 74 deleted 0
27 Save stack +14 -14
28 Attribute list search 1558(1496) found 962 in 1424 elements (1403 after boot)
29 number of ref 0, of used labels 0, of defined labels 0, of ext. ref. 0
30 Modules with 0, without 0, sections with 0, without 0
31 There was no image.
32 Output written on testb.xml (626 bytes).
33 No error found.
34 (For more information, see transcript file testb.log)

```

Some comments. Line 6 shows the name of the configuration file. If this file contains a line that starts with ‘## tralics ident rc=’ then all characters after the equals sign are printed (see line 5). Since version 2.5 (pl4), in the case where character number 30 is a dollar sign, a space will be added after it.⁸ The reason for this is that the RCS software interprets a string like ‘Revision’ in dollar signs; we do not want it to replace the 2.11 by the revision number of the L^AT_EX document. We shall explain elsewhere how to read the statistics. When Tralics says: Used 1718 commands, this is the number of entries in the hash table (the total number of commands defined). Among these, 74 were defined by `\def`, in the bootstrap phase, all others are defined in C++.

Line 7 shows the number of entries in the biblist. If some entries are solved, they are shown in parentheses. Here, we have 5-1=4 unsolved entries. If line 5 of the source file is commented out, then BibT_EX is not used, and lines 30 to 42 will be replaced by the single line ‘Seen 4 bibliographic entries’. In such a case, Tralics says: Used 1720 commands, because the two commands `\cititem-etal` and `\cititem-btitle` are dynamically created.

The standard configuration file contains a line that says that ‘article’ is an alias for ‘std’. The ‘std’ configuration defines two quantities: the name of the DTD, hence the root element, it is `<std>`, see line 4 of the XML result. It defines `xml_biblio` to be ‘bibliography’. This is the name of the element that will hold the bibliography. The default value is ‘biblio’, but it can be redefined (see line 5). Do not confuse this with the name of the element produced by the environment ‘thebibliography’, that appears line 6 in the XML result.

⁸In fact, in subsequent version, a space is added after the first dollar sign, independently of the position on the line.

Chapter 5

Other commands

5.1 Character encoding

We have to distinguish between input encoding, internal encoding and output encoding. The internal encoding of T_EX is ASCII (i.e. 65 is the internal code of the upper case letter A), at least for all characters with code between 32 and 126. The input encoding is the mechanism that converts the code of the letter A supplied by computer into the code 65. Almost all input encodings are nowadays ASCII-based, they produce the same value for the letter A; the results may be different for a character like é. The output encoding indicates for a letter, say A, which position in the font to use. We shall not discuss the output encoding here. Let's just notice that the character ‘{’ exists in the font cmtt10, but not in other text fonts of the computer modern family. If you read a version of this document that uses the original encoding (OT1), braces shown in error messages are taken from a math font, hence are upright. Some years ago, a 8bit encoding (called T1) was designed, which contains braces. You can compare Figure 1 in appendix F of the [4] (describing the font cmr10) with Table 7.32 of [6], describing ecrm1000.

The first version of T_EX was using 7bit input and output characters (but fonts and dvi files were coded on 8bits). There is an extension Ω to T_EX that accepts 16bit characters as input, using different encoding schemes. Characters that are not part of the ASCII specifications (less than 32 or greater than 126) are not guaranteed to be treated the same in all implementations. For this reason, it is wise to load the `inputenc` package, with the current encoding as argument. The effect will be that some characters, like é will become active, and expand to `\'e`. As a result: only ASCII letters are allowed in control sequence names. On the other hand, if you say `\begin{motclés}`, then L^AT_EX complains with *LaTeX Error: Environment motcl'es undefined*. Don't try to define the `motcl'es` environment: the expansion of the accent depends on the context: it is é for `\begin` and `\'e` for the macro that prints the error message. Non-ASCII characters may be printed by T_EX as `ˆˆab` (in some older version of T_EX, I had to pretend, via locale settings, that my computer did not understand English in order for it to output the guillemet as «).

A silly question concerns end-of-line markers. Some systems like Unix use LF (line feed) as line separators, some others like Macintosh use CR (carriage return) and Windows uses CR-LF. This is replaced by T_EX by a single character: the carriage return with ASCII code 13. Tralics interprets CR-LF, CR and LF alike: as an end-of-line marker. This marker will be replaced by the character whose code is in `\endlinechar`, provided that this value is in the range 0–255¹. The default value is 13, a character of category 5. The tokeniser converts this into a `\par` token, a space token or ignores it depending on the state. This space token has value 32 (but Tralics uses 10, so as to keep the same line breaks in the XML result as in the T_EX source). Note that, whenever a line is read,

¹Null characters are ignored by Tralics, so that the effective range is 1–255.

spaces at the end of the line are removed. If you want a space after a control sequence, you say something like ‘\TeX\ ’, and if this construct appears at the end of a line, the space is ignored; if the endline character has category code 5, it will be converted to a space, and everything works fine; if this character is for instance 65, you may get a strange error, like this

```
! Undefined control sequence.^^J
1.170 ...reaks in the \XML\ result as in the \TeX\^^J
                                     ^^J
? ^^J
```

We have shown here the end of line as ^^J. There are four lines: the error messages, two context lines, and the line with the prompt. The two context lines show that the space at the end of the line is removed. T_EX does not print the undefined control sequence: it assumes that it is either the last token on the first context line, or a token marked as ‘<recently read>’ or something like that; in our case, the undefined control sequence is the one obtained by replacing ^^J by the value of the endline character.

There is a way to enter special characters in T_EX, for instance ^^J is a line feed. The algorithm is the following: whenever T_EX sees two consecutive identical characters of category code 7, followed by a character whose number is x , it replaces these three characters by the character whose code is y , where $y = x - 64$ if $x \geq 64$, and $y = x + 64$ if $x < 64$. Hence ^^? yields $y = 127$ (the is the delete character). All characters with codes between 1 and 26 can be obtained using the form ^^A, ^^B, etc. The null character is ^^@, characters with code between 27 and 31 are ^^[, ^^[, ^^[, ^^[, and ^^_. Character 32 can be represented as ^^‘. All other characters are ASCII characters. This is an example of use:

```
27=\char‘\^^[, 28=\char‘\^^[, 39=\char‘\^^[, 30=\char‘\^^[, 31= \char‘\^^_
```

Because some characters in the list could be active, we have used the construction \char‘\A (with A replaced by some other character). This typesets as ‘27=ff, 28=fi, 39=fl, 30=ffi, 31= ffi’ The result depends on the font encoding (OT1 vs T1). The translation by Tralics is 27=, 28=, 39=, 30=, 31= . When T_EX switched to 8 bits, the rule changed a little bit: the previous rule applies only if $0 \leq x \leq 127$, it gives $0 \leq y \leq 127$. Another test was added: if you say ^^ab, these four characters are replaced by the single character whose code is ab (in base 16, i.e. 171 in base ten in this case). In such a case two characters are needed: a letter or a digit; only lower case letters between a and f are allowed. Thus every character in the range 0-255 has such a representation. For instance, the character 32 has can be entered as ^^20. There is one situation where the space character can be used in this way: at the end of the line, when \endlinechar is non trivial. Note that, in the case where the resulting character has category 7, it can participate in a hat-hat construct. Here is an example.

```
{1^^{^^ab2^^5e^ab3^^5e^5e^ab4\def\Abc{ok}\def\bAc{OK}\^^41bc\b^^41c}
{\catcode ‘\é=7 ééab $xé2$ %next line should produce M
éé
%$1^è=^^^AééT$ %% hat hat control-A
$1^è=^^^A$ %% hat hat control-A
}\def\msg{a message.^^J}
```

Some explanations are needed. ^^‘ is a semi colon, ^^ab is an opening French guillemet, ^^5e is a hat (recursion...), ^^41 is the uppercase letter A. The first line of the example explains that such funny characters can appear in a control sequence name. The second line shows that the hat-hat mechanism can be used with other characters than a hat. It also shows that, if the mechanism cannot be applied, a character with category 7 behaves like a superscript character, whatever its numeric value. The line that follows shows that the end-of-line character is ASCII 13, aka control-M (usually written as ^M). After that, there are two lines containing a control-A character, shown here as ^A. It is preceded by hat-hat, so that the effect should be a single A. The line that is commented out contains a control-T written as ééT (for some strange reasons, this character is

invalid in XML1.0, but valid as an entity in XML1.1, [8, 9]). The last line is just a real example of J . This character is printed by Tralics as LF, or CR-LF on Windows. This is the translation of Tralics:

```
<p>1;<2<3<4ok0K
<< <formula type='inline'>
  <math xmlns='http://www.w3.org/1998/Math/MathML'>
    <msup><mi>x</mi> <mn>2</mn> </msup></math>
  </formula> M<formula type='inline'>
    <math xmlns='http://www.w3.org/1998/Math/MathML'>
      <mrow><msup><mn>1</mn> <mi>è</mi> </msup><mo>=</mo><mi>A</mi></mrow>
    </math>
  </formula></p>
```

As said above, Ω accepts 16bit characters, using the notation abcd . This syntax works also in Tralics, but it does not produce a character that can be used in a command name. The scanner evaluates `\chardef\foo"ABCD\relax`, and produces the `\foo` token (of course, the `\foo` command is not modified). Note that this allows 16bit characters, while in Tralics, the `\char` and `\chardef` commands accept 27bit integers. Here is an example

```
\def\foo#1#2#3{#1=#2=#3=}
\foo~~~~0153~~~~0152~~~~0178
~~~~017b^^8?
```

It is translated by Tralics as `œ=Œ=Ÿ= Żx?`. The argument to `\foo` could also have been: `\oe\OE{"Y}`. The transcript file contains lines of the form:

```
[8] \foo~~~~0153~~~~0152~~~~0178
\foo #1#2#3->#1=#2=#3=
#1<-~~~~0153
#2<-~~~~0152
#3<-~~~~0178
{\char339}
{\char338}
{\char376}
```

The original version of the Tralics documentation said: Si on a un texte qui contient essentiellement des caractères 7bits, et très peu d'autres caractères, l'utilisation de caractères 16bits consomme énormément de place. This means that using a 16bit encoding consumes a lot of space if you write a French document (and even more, for an English one). The sentence has 159 ASCII characters and 6 others; these can be input using iso-8859-1 (aka latin-1) as input encoding². In $\text{T}_{\text{E}}\text{X}$, it uses 165 bytes, in Ω , it uses 330 bytes. Using a construction like `\'e` we need 177 bytes (and 7 bits per byte). Using UTF-8 requires only 171 bytes (8 bits per byte). This explains why UTF-8 is popular. We shall explain (in the second part of this document) how UTF-8 is encoded and how $\text{T}_{\text{E}}\text{X}$ may read it. In the case of Tralics, the situation is: you can (via an argument to the Tralics program) specify that the sources are encoded using UTF-8 or latin1 (this being the default). However, if the tex file contains, on the first line "utf8-encoded" UTF-8 encoding will be used, if it contains "iso-8859-1" then latin1 encoding will be used. When UTF-8 mode is enabled, Ae is read as é ; if the value of a character is greater than 256, it is read in the form abcd .

5.2 Characters and Accents

There are some commands that put an accent over a letter. You can say `a\accent 98 cde`, this works in $\text{T}_{\text{E}}\text{X}$, but not in Tralics: you will get an error, *Unimplemented command \accent*. The

²All French letters exist in iso-8859-1, except `\oe \OE, \Y`, and the Euro sign

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z |
|------------------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\backslash\hat{\ }$ | x | x | | | x | | x | x | x | x | | | | | x | | | | x | x | x | x | x | x | x | x |
| $\backslash\acute{\ }$ | x | x | | | x | | x | | x | | x | x | x | x | x | x | | | x | x | | x | x | x | x | x |
| $\backslash\grave{\ }$ | x | | | | x | | | | x | | | | | x | x | | | | | | x | x | | x | | x |
| $\backslash\text{''}$ | x | | | | x | | | x | x | | | | | x | | | | | | t | x | x | x | x | x | x |
| $\backslash\text{c}$ | | x | x | x | x | | x | x | | | x | x | | x | | | | | | x | x | x | | | | |
| $\backslash\text{u}$ | x | | | | x | | x | | x | | | | | x | | | | | | | x | | | | | |
| $\backslash\text{v}$ | x | x | x | x | x | | x | x | x | j | x | x | | x | x | | | | | x | x | x | x | | | x |
| $\backslash\tilde{\ }$ | x | | | | x | | | | x | | | | | x | x | | | | | | x | x | | | x | |
| $\backslash\text{H}$ | | | | | | | | | | | | | | | x | | | | | | x | | | | | |
| $\backslash\text{k}$ | x | | | | x | | | | x | | | | | x | | | | | | | x | | | | | |
| $\backslash\text{.}$ | x | x | x | x | x | x | | x | I | | | x | x | x | | x | | | | x | x | x | | x | x | x |
| $\backslash\text{=}$ | x | | | | x | | x | x | x | | | | | x | | | | | | | x | x | | | x | |
| $\backslash\text{r}$ | x | | | | | | | | | | | | | | | | | | | | x | w | | y | | |
| $\backslash\text{b}$ | | x | | x | | | | h | | | x | x | | x | | | | | | x | x | | | | | x |
| $\backslash\text{d}$ | x | x | | x | x | | | x | x | | x | x | x | x | x | | | | | x | x | x | x | x | x | x |
| $\backslash\text{f}$ | x | | | | x | | | | x | | | | | x | | | | | | x | x | | | | | |
| $\backslash\text{C}$ | x | | | | x | | | | x | | | | | x | | | | | | x | x | | | | | |
| $\backslash\text{T}$ | | | | | x | | | | x | | | | | | | | | | | | x | | | | | |
| $\backslash\text{V}$ | | | | x | x | | | | | | | x | x | | | | | | | | x | x | | | | |
| $\backslash\text{D}$ | x | | | | | | | | | | | | | | | | | | | | | | | | | |
| $\backslash\text{h}$ | x | | | | x | | | | x | | | | | x | | | | | | | x | | | | | x |

Table 5.1: *All possible accents.* You can put an accent on any letter, except Q. You can put accents on non-letters, for instance $\backslash\text{ae}$, see text. Some characters accept two accents. In general, you can put an accent on a lower case letter, an upper case letter. There is one exception: you cannot put a dot over a lower case I, because there is already a dot. For h, j, t w, and y, there are accents that apply only to lowercase letters.

number 98 is read, and converted to an integer. The Unicode character will be used; thus the translated result is ‘abcde’.

You can say $\backslash\text{a}'\text{e}$. This is a command introduced by \LaTeX so as to allow accents inside a tabbing. Some care must be taken. If you say $\backslash\text{a}\{\text{par}\}\{\text{b}\}$ in \LaTeX , you get an error of the form: *Paragraph ended before \@changed@cmd was complete.* The Tralics error message is: *wanted a single token as argument to \a.* If you say $\backslash\text{a}\backslash\text{foo}12$, there is a single token, and the error is: *Bad syntax of \a, argument is \foo.* In fact, the token after $\backslash\text{a}$ must be a valid accent character. After that $\backslash\text{a}'$ is handled exactly like \backslash' . You can say $\backslash\text{= U}$, the space after the command is ignored. You cannot say $\backslash\text{=}\{\text{U}\}$, the space is not removed, this is an error. In fact, the argument list of the accent command should contain exactly one token (exception: double accents will be explained later). This token should be a character, with code between 0 and 128. Hence $\backslash'\hat{\text{E}}$ is wrong, you must say $\backslash'\{\hat{\text{E}}\}$ if you want $\hat{\text{E}}$. The message is *Error in accent, command = \'; Cannot put this accent on non 7-bit character \hat{E}.* If the token $\backslash\text{i}$ is given, it will be replaced by i, so that $\backslash\text{''}\backslash\text{i}$ and $\backslash\text{''}\text{i}$ produce the same result. You can say $\backslash\text{=}\backslash\text{AE}$, $\backslash\text{=}\backslash\text{ae}$, $\backslash'\backslash\text{AE}$, $\backslash'\backslash\text{ae}$, $\backslash'\backslash\text{AA}$, $\backslash'\backslash\text{aa}$, $\backslash'\backslash\text{O}$, $\backslash'\backslash\text{o}$. The result looks like $\bar{\text{E}}\tilde{\text{e}}\acute{\text{E}}\acute{\text{e}}\acute{\text{A}}\acute{\text{a}}\acute{\text{O}}\acute{\text{o}}$.

You can put an accent on a letter only in the case where this gives a Unicode character. In the case of $\backslash\text{c}\{\text{a}\}$ and $\backslash\text{c}\{\backslash\text{a}\}$, the error message is the same: *Error in accent, command = \c; Cannot put this accent on letter a.* Table 5.1 indicates on which letters you can put an accent. See the html page <http://www-sop.inria.fr/apics/tralics/doc-chars.html> for a list of some glyphs.

Some accents are not standard. Examples:

- $\backslash\text{^}\text{a}$ gives $\hat{\text{a}}$,

Û Ä Ö Ú Í Û Ä Ö Ö Ö Ú Ø Ó É Š Ç Á Ó
 É Ò È Ì Ñ À È Ú Ó Ú Ö Ú Ö Û Ö Ú Ö Ä
 Ô Ê Æ Õ Ë Å Ö É Á Å Å Æ Ö Ö Ë Š Š

Figure 5.1: Some characters

- \’a gives á,
- \‘a gives à,
- \"a gives ä,
- \c c gives ç,
- \u a gives ů,
- \v a gives ǎ,
- \~a gives ã,
- \H o gives ó, it is redefined in the case of double accents,
- \k a gives ā,
- \. a gives â,
- \=a gives ā,
- \r a gives ȁ,
- \b b gives ȃ,
- \d a gives ȁ,
- \f a gives an inverted breve accent over a,
- \C a gives a double grave accent on a,
- \T e gives a tilde under e,
- \V d gives a circumflex below d,
- \D a gives a ring below a,
- \h a gives a hook over a.

If in the table you see ‘T’ instead of ‘x’, this means that the accent applies only on capital I. If you see h, j, t, w or y, this applies only to the lower case letter. Otherwise the accent applies to both upper case letter and lower case letter.

There is a possibility to put double accents (for Vietnamese, for instance). The following ones are recognized, for upper and lower case letters, the order of the accents is irrelevant. Inside braces, there is an accent command, optional spaces, and a character (maybe enclosed in braces).

```
\{"\=U} \{"\=A} \{"\=0} \{"\'U} \{"\'I} \{"\'U} \.{\=A} \.{\=0}
\={\~ 0} \k{\=0} \\'{\~U} \\'{\0} \\'{\=0} \\'{\=E} \\'{\.S} \c{\\' C}
\\'{\^A} \\'{\^0} \\'{\^E} \\'{\=0} \\'{\=E} \d{\=L} \d{\=R}
\’{\^ A} \’{\^ E} \H{\’U} \H{\’0} \H{\’U} \H{\’0} \H{\h U} \H{\h 0}
\H{\~U} \H{\~0} \H{\d U} \H{\d 0} \d{\^A} \d{\^0} \d{\^E} \~{\^A}
\~{\^0} \~{\^E} \h{\^A} \h{\^0} \h{\^E} \u{\’A} \u{\’A} \u{\h A}
\u{\~A} \u{\d A} \~{\’ 0} \~{\’0} \~{\’0} \u{\c E} \.{\v S} \.{\d S}
```

This is the translation.

```
&#x1E7A; &#x1DE; &#x22A; &#x1D7; &#x1E2E; &#x1DB; &#x1E0; &#x230;
&#x22C; &#x1EC; &#x1E78; &#x1FE; &#x1E52; &#x1E16; &#x1E64; &#x1E08;
&#x1EA4; &#x1ED0; &#x1EBE; &#x1E50; &#x1E14; &#x1E38; &#x1E5C;
&#x1EA6; &#x1EC0; &#x1EE8; &#x1EDA; &#x1EEA; &#x1EDC; &#x1EEC; &#x1EDE;
&#x1EEE; &#x1EE0; &#x1EF0; &#x1EE2; &#x1EAC; &#x1ED8; &#x1EC6; &#x1EAA;
&#x1ED6; &#x1EC4; &#x1EA8; &#x1ED4; &#x1EC2; &#x1EAE; &#x1EB0; &#x1EB2;
&#x1EB4; &#x1EB6; &#x1E4E; &#x1ED0; &#x1ED2; &#x1E1C; &#x1E66; &#x1E68;
```

You can see a representation on figure 5.1. We show here the \LaTeX output for these characters. $\bar{U} \bar{A} \bar{O} \ddot{U} \ddot{I} \ddot{U} \bar{A} \bar{O} \bar{O} \bar{O} \bar{U} \acute{O} \bar{O} \bar{E} \acute{S} \acute{C} \acute{A} \acute{O} \acute{E} \bar{O} \bar{E} \bar{L} \bar{R} \grave{A} \grave{E} \grave{U} \grave{O} \grave{U} \grave{O} \grave{U} \grave{O} \bar{U} \grave{O} \bar{U}$
 $\acute{O} \grave{A} \grave{O} \grave{E} \grave{A} \grave{O} \grave{E} \grave{A} \grave{O} \grave{E} \grave{A} \grave{A} \grave{A} \grave{A} \grave{A} \grave{A} \grave{O} \grave{O} \grave{O} \grave{E} \grave{S} \grave{S}$. The first character in the list is: latin capital letter u with macron and diaeresis, the second one is latin capital letter a with diaeresis and macron. The order of accents is not the same. For simplicity, in `Tralics`, this is irrelevant. You can notice that the \LaTeX output is strange. First, we have defined `\h` to be a no-op. For the figure, we used the following code:

```
\newcommand\hook@above[1]{%
\leavevmode\setbox0\hbox{#1}\dimen@ht0 \advance\dimen@.5ex
\rlap{\raise\dimen@\hbox{\kern2pt\char11}}#1}
```

This code works, provided that the font has, at position 11, something that looks like a hook (for the T1 encoding, this is a cedilla). In \LaTeX you can't put a `\`` accent on `\=U`. No error is signaled, it is just that \TeX puts the accent before the accentee in case the accentee is not a character, instead of putting it above³. You can say `\`{\`U}` because `\`U` is a character in the T1 encoding. The `\mathaccent` command has not these limitations⁴. The first character of the figure was composed via

```
\UnicodeCharacter{x1E7A}{\ensuremath{\ddot{\mbox{\=U}}}}
```

In German, the umlaut character has a special meaning. The following example shows what can be input. See the `babel` documentation for details.

```
\language=2
"a"o"u"e"i""A"O"U"i"E
"s"z"S"Z"c"C"f"F"l"L"m"M"n"N"p"P"r"R"t"t
""-""~""|""="'">"<">
```

and the `Tralics` translation.

```
äöüëïÄÖÜË
ßSSSSZckCKffFFlLLmmMMnnNNppPPrrRRttTT
--&#x0201E;&#x0201D<>
```

The previous hack does not apply if the double quote character has category code 11 (letter), is in an URL, or in a file name to be read (for instance, via `\includegraphics`).

The translation of the dash character is the following. If this character appears in an URL or while reading a file name, it is left unchanged. If its category code is 11 (letter), usually inside a verbatim environment, its translation is a dash followed by a `\textnospace`, unless you invoke `Tralics` with the `nozerowidthspace` switch, case where the translation is a single hyphen. Otherwise, a test is made for a ligature: three hyphens in a row produce `—` (mdash), and two hyphens produce `–` (ndash),

The characters `;!?<>` are handled normally if inside an URL, when translating a filename, when their category code is 11 (letter, typically, inside a verbatim), or when the current language is not French. The translation of `<` is an opening guillemet with some space after it. If the character that follows is (after expansion), a normal space, or a `~` or a `\,`, it will be discarded. The translation of `>` is a space plus the character. If the previous character is a space, it will be removed. (\TeX has a primitive `\unhskip` that can remove a space; the `Tralics` equivalent works in usual cases). The other four punctuation characters are handled like a closing guillemet. In any case, the space added by these characters is a non-breaking one.

The characters `'<>` behave in a special manner, in the same case as the guillemets. In fact, if they are doubled, French guillemets will be used instead. Thus `<<foo>>` and `'foo'` and `<foo>` behave the same, if the current language is French. Otherwise, a `\textnospace` will be added

³See the \TeX book, chapter 25

⁴See the \TeX book, chapter 26

after the character, in the same way as for a dash, namely outside an URL, file name, but if the category code is 11 (letter), and the magic switch has not been given. Example:

```
\language = 0
test ligatures: <<>>'-- et --- !?;
\language=1
test ligatures: <<>>'-- et --- !?;
test ligatures:\verb=<<>>'-- et --- !?;=
```

This is the translation

```
test ligatures: &lt;&lt;&gt;&gt;'&ndash; et &mdash; !?;
test ligatures&nbsp;;: <&nbsp;>><&nbsp;>>&ndash; et &mdash;&nbsp;;!&nbsp;sp;?
&nbsp;sp;:&nbsp;sp;;
test ligatures&nbsp;sp;:<hi rend='tt'>&lt;&#x200B;&lt;&#x200B;&gt;&#x200B;
&gt;&#x200B;'&#x200B;'&#x200B;'&#x200B;'&#x200B;'&#x200B;-&#x200B;-&#x200B;
&nbsp;sp;et&nbsp;sp;-&#x200B;-&#x200B;-&#x200B;&nbsp;sp;!?:</hi>
```

5.3 Verbatim material

We have seen a little example of verbatim code above. It shows that some `​` characters are inserted, this is so that, if the XML file is read, a double dash will not be interpreted as an en-dash. What the `\verb` command produces is a sequence of characters, whose category codes are 12, except for some, that are of category 11, namely ‘ ’-<>~&:;!<>. You can compare this with the L^AT_EX code, shown in section 2.12: the `\@noligs` command makes some characters of category code 13, the associated action is: output the character, with a zero kern in front.

You can say `\verb*+ x y+` or `\verb+ x y+`. All characters between the two plus signs are collected. Any character can be used instead of the plus sign (Try `\verb*abca` and `\verb =a= !`). In the case a star is given, spaces are replaced by `\textvisiblespace`, otherwise by `\nobreakspace`.

You can say `\DefineShortVerb\+`, after that `+foo+` is the same as `\verb+foo+`. Note that the command must be followed by something like ‘\+’ or ‘*’, i.e., a macro whose name is formed of a single character. You can say `\UndefineShortVerb\+`, this will undo the previous command. The syntax is the same. In the case where ‘+’ is a short verb character, you can say `\SaveVerb{foo}+\bar+`. This has as effect to remember in a private command all tokens that `+\bar+` gathers. When you say `\UseVerb{foo}`, these tokens are re-inserted in the input stream. Example:

```
\DefineShortVerb\+
\SaveVerb{foo}+\bar +
\UndefineShortVerb\+
\UseVerb{foo}
```

The transcript file will contain, for the `\UseVerb` command the following line.

```
\savedverb@foo ->\verbprefix {\verbatimfont \bar\nobreakspace }
```

The ‘\’ before ‘b’ is not a command delimiter, for otherwise there would have been a space after `\bar`. Note: another explanation is that the ‘b’ is not of category code 11, so that the command is `\b`; exercise: find all interpretations of this line.

In the case of a verbatim environment, all characters on the line are gathered (final spaces disappear, as usual), with category codes as explained above. As is the case of `\verb`, the `\verbatimfont` command is prepended. This is defined to be `\tt`. Moreover, `\verbatimprefix` is also added in front of the token list. In the case of the `\verb` command, there is `\verbprefix` instead. These two commands are defined as `\@empty`. You can redefine them. Each line is followed by `\par` and

`\noindent`. If the environment is followed by an empty line, or a `\par` command, this command is removed, as well as the last `\indent`. Example that shows use of the prefix commands:

```
\DefineShortVerb{\|}
\def\verbatimfont#1{{#1}}
\def\verbprefix#1{A#1A}
\def\verbatimprefix#1{B#1B}
Test: \verb+foo+ and |bar|
\UndefineShortVerb{\|}
\begin{verbatim}
line1
line2
\end{verbatim}
```

The translation is:

```
<p>Test: AfooA and AbarA</p>
<p noindent='true'>Bline1B</p>
<p noindent='true'>Bline2B</p>
<p noindent='true'></p>
```

The Verbatim environment is an extension of the verbatim environment. There is an optional argument, an association list. If you say ‘numbers=true’, then lines will be numbered, starting from 1. If say say ‘numbers=N’, where N is a number, then the number N will be used; instead of a number, you can also give a counter name. Other features defined by the fancyvrb package have not yet been implemented.

If a line number M is given, the following piece of code is inserted before the verbatim line: `{\verbatimnumberfont{M}}\space`. The funny command is `\let` equal to `\small` at the start of the run. The number is incremented for each line.

Characters after `\begin{Verbatim}`, but on the same line, are ignored. The same is true if an optional argument is given: all characters that follow the closing bracket of the optional argument are ignored. The opening bracket is only looked for on the current line (unless the end of line character is commented out).

```
\begin{Verbatim}                                [numbers=true]
TEST
\end{Verbatim}
and without
\begin{Verbatim}
[ok]TEST
\end{Verbatim}
\begin{Verbatim} %
[ok] this is handled as comment
TEST
\end{Verbatim}

\def\verbatimfont#1{{\it #1}}
\def\verbatimnumberfont{\large}
\tracingall
\count3=4
\begin{Verbatim}[counter=3]
5,one line
\end{Verbatim}
```

```

\begin{Verbatim}[counter=03]
6,one line
\end{Verbatim}
\newcounter{vbcounter}
\setcounter{vbcounter}8
\begin{Verbatim}[counter=vbcounter]
9,one line
\end{Verbatim}
\begin{Verbatim}[counter=vbcounter]
10,one line
\end{Verbatim}

```

This is the translation.

```

<p noindent='true'><hi rend='small1'>1</hi> <hi rend='tt'>TEST</hi></p>
<p noindent='true'>and without</p>
<p noindent='true'><hi rend='tt'>[ok] TEST</hi></p>
<p noindent='true'></p>
<p noindent='true'><hi rend='tt'>TEST</hi></p>
<p noindent='true'><hi rend='large1'>5</hi> <hi rend='it'>5</hi>,one line</p>
<p noindent='true'><hi rend='large1'>6</hi> <hi rend='it'>6</hi>,one line</p>
<p noindent='true'></p>
<p noindent='true'><hi rend='large1'>9</hi> <hi rend='it'>9</hi>,one line</p>
<p noindent='true'><hi rend='large1'>10</hi> <hi rend='it'>1</hi>0,one line</p>
<p noindent='true'></p>

```

You can say `\numberedverbatim` or `\unnumberedverbatim`. After that, verbatim environments will be automatically numbered or not. This does not apply to Verbatim environments. There is a command `\DefineVerbatimEnvironment` that is not yet implemented.

There is a command `\fvset` that takes an associated list as argument. If it contains ‘shows-spaces=true’ or ‘shows-spaces=false’, this changes how spaces are interpreted in a verbatim environment or command (except for `\verb*`, case where the space is always visible).

5.4 Case change

There are different commands for changing the case of letters. For instance, the translation of

```

\uppercase{Einstéin: $E=mc^2$}
\lowercase{Einstéin: $E=mc^2$}

```

is

```

<p>EINSTÉIN: <formula type='inline'>
<math xmlns='http://www.w3.org/1998/Math/MathML'>
<mrow><mi>E</mi><mo>=</mo><mi>M</mi><msup><mi>C</mi> <mn>2</mn> </msup>
</mrow></math></formula>
einstéin: <formula type='inline'>
<math xmlns='http://www.w3.org/1998/Math/MathML'>
<mrow><mi>e</mi><mo>=</mo><mi>m</mi><msup><mi>c</mi> <mn>2</mn> </msup>
</mrow></math></formula>
</p>

```

There are two tables that control these conversions: the `lc-table` and the `uc-table`. If the `lc` value of a character is non-zero, it’s the lowercase equivalent of the character; otherwise, the character

is left unchanged by `\lowercase`. The same is true for the `lc-table`. You can use `\lccode` and `\uccode` for changing these tables. They are initialized like this: for all integers x with value between ‘a’ and ‘z’, and between ‘à’ and ‘ÿ’, the `uc` value is $x - 32$, the `lc` value is x , the same holds for $x - 32$. There are four exceptions: the pair 215, 247, this is multiplication and division sign, and the pair 223, 255 this is ß and ÿ, the uppercase version of these two characters are ‘SS’ and ‘ÿ̇’, but this character is not a latin1 character, its code is 376.

You can use the two commands `\MakeUppercase` and `\MakeLowercase`. These commands have a regular syntax (in the example that follows, the `\expandafter` would be useless for `\lowercase`). They convert letters, as for `\uppercase` and `\lowercase`, plus some commands that define some characters. This example shows the list of all the recognised commands.

```
\def\foo{foo}
\def>List{{abcABCéÉ\foo
    \oe\OE\o\O\ae\AE\dh\DH\dj\DJ\l\L\ng\NG\ss\SS\th\TH}}
\expandafter\MakeUppercase>List
\expandafter\MakeLowercase>List
```

The translation is

```
ABCABCéÉfoo&#x152;&#x152;øøÆÆÐÐ&#x110;&#x110;&#x141;&#x141;
&#x14B;&#x14B;SSSSÞÞ
abcabcééfoo&#x153;&#x153;øøææðð&#x111;&#x111;&#x142;&#x142;
&#x14A;&#x14A;ßßþþ
```

5.5 Simple commands

We consider here some commands that take no arguments. They are not allowed in math mode. A new paragraph is started (via `\leavevmode`) in vertical mode.

- `\#` translates to `#`.
- `\-` is `\discretionary{-}{-}{-}` in \LaTeX , empty in \Tralics .
- `\AA` and `\aa` translate to Å and å.
- `\AE` and `\ae` translate to Æ and æ.
- `\copyright` translates to ©.
- `\dag` translates to †.
- `\ddag` translates to ‡.
- `\DH` and `\dh` translate to Ð and ð.
- `\DJ` and `\dj` translates to Đ and đ.
- `\endguillemets` translates to ».
- `\fg` translates to ».
- `\guillemets` translates to «.
- `\guillemotleft` translates to «.
- `\guillemotright` translates to ».
- `\ieme` translates to &ieme;. Something like `3\ieme` should typeset as 3^e.
- `\iemes` translates to &iemes;. Something like `3\iemes` should typeset as 3^{es}.
- `\ier` translates to &ier;. Something like `1\ier` should typeset as 1^{er}.
- `\iers` translates to &iers;. Something like `1\iers` should typeset as 1^{ers}.
- `\iere` translates to &iere;. Something like `1\iere` should typeset as 1^{re}.
- `\ieres` translates to &ieres;. Something like `1\ieres` should typeset as 3^{res}.
- `\LaTeX` translates to `<LaTeX/>`.
- `\No` and `\Numero` translate to &Numero;. This should render as N^o.

- `\no` and `\numero` translates to `&numero`; . This should render as n^o.
- `\O` and `\o` translates to \emptyset and \emptyset .
- `\og` translates to «.
- `\P` translates to ¶.
- `\S` translates to §.
- `\slash` translates to /.
- `\SS` and `\ss` translate to SS and β .
- `\TeX` translates to `<TeX/>`.
- `\TH` and `\th` translate to \mathbb{P} and \mathbb{p} .

The following commands all start with ‘text’. They are forbidden in math mode.

- `\textacutedbl` translates as ¨.
- `\textasciiacute` translates as ´.
- `\textasciicutex` translates as the Unicode character U+2032; this is known as ‘prime’. It is not the same as U+27, apostrophe, or U+B4 acute accent, or U+2B9, modifier letter prime.
- `\textasciicedilla` translates to a cedilla.
- `\textasciidieresis` translates to ¨.
- `\textasciigrave` translates to `.
- `\textasciimacron` translates as ¯.
- `\textbackslash` translates to \.
- `\textbaht` translates to ฿.
- `\textbrokenbar` translates as |.
- `\textbullet` translates to •.
- `\textcelsius` translates to °C.
- `\textcent` translates as ¢.
- `\textcircledP` translates to ©.
- `\textcopyright` translates as ©.
- `\textcurrency` translates as ¤.
- `\textdagger` translates to †.
- `\textdaggerdbl` translates to ‡.
- `\textdegree` translates as °.
- `\textdiv` translates to ÷.
- `\textdollar` translates to \$.
- `\textdong` translates to ₫.
- `\textestimated` translates to €.
- `\texteuro` translates to €.
- `\textexclamdown` translates as ¡.
- `\textflorin` translates to ₣.
- `\textfrenchfranc` translates to the symbol of the French Franc (not used any more).
- `\textgravedbl` translates to ¨.
- `\textinterrobang` translates to †.
- `\textlira` translates to ₣.
- `\textlnot` translates to ¬.
- `\textlquill` translates to {.
- `\textmho` translates to \mathbb{U} .
- `\textmu` translates to μ .
- `\textnaira` translates to ₦.

- `\textnospace` translates to U+200B (zero width space). We use this as a mark to inhibit ligatures in verbatim mode.
- `\textnumero` translates to \mathbb{N}° .
- `\texttohm` translates to Ω .
- `\textonehalf` translates to $\frac{1}{2}$.
- `\textonequarter` translates to $\frac{1}{4}$.
- `\textonesuperior` translates to 1 .
- `\textordfeminine` translates as $^{\text{a}}$.
- `\textordmasculine` translates to $^{\text{o}}$.
- `\textparagraph` translates to ¶.
- `\textperiodcentered` translates to \cdot .
- `\textpertenthousand` translates to ‰ .
- `\textperthousand` translates to ‰ .
- `\textpm` translates as \pm .
- `\textquestiondown` translates to . .
- `\textrecipe` translates to \mathbb{R} .
- `\textreferencemark` translates to \otimes .
- `\textregistered` translates as ® .
- `\texttrquill` translates to $\text{}$.
- `\textsection` translates as § .
- `\textservicemark` translates to SM .
- `\textsoftthyphen` translates as soft-hyphen, Unicode character U+00AD, (discretionary hyphen). Is this is the same as `\-?`
- `\textsterling` translates as £ .
- `\textthreequarters` translates to $\frac{3}{4}$.
- `\textthreesuperior` translates as 3 .
- `\texttimes` translates as \times .
- `\texttrademark` translates as TM .
- `\texttwosuperior` translates as 2 .
- `\textvisiblespace` translates as □ .
- `\textwon` translates as ₩ .
- `\textyen` translates as ¥ .

The following commands are accepted in text and math mode.

- `\textvisiblespace` (space, carriage return, newline, tabulation) translates as a space.
- `\!` has empty translation. This is defined by $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ as minus `\thinmuskip`, math only.
- `\,` translates as ` `, same as a tilde. This is defined in $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ as a mskip of width `\thinmuskip` in math mode, and `\thinspace` otherwise.
- `\$` translates as $\text{\$}$.
- `\%` translates as \% .
- `\&` translates as \& , in fact as `&`
- `_` translates as `__`.
- `\dots` translates as `'...'`.
- `\flat` translates as \flat .
- `\i` translates as i .
- `\j` translates as j .
- `\ldots` translates as `'...'`.
- `\natural` translates as \natural .

- `\quad` translates as ` `;
- `\qqquad` is the same as `\quad\quad`.
- `\sharp` translates as `‡`.
- `\}` translates as `}`.
- `\{` translates as `{`.
- `\nobreakspace` translates as ` `.

5.6 The fp package

Introduction

This is an implementation in C++ of the package by Michael Mehlich. It implements fixed point arithmetics in T_EX. Each number is formed by a sign, then 18 digits before the point and 18 digits after the point. Since $10^9 \leq 2^{30}$, four 32bits integers are sufficient. In the code, we shall sometimes write a number as

$$x = \sum_{i=-6}^5 b_i B^i = 10^{-18} \left(\sum_{i=0}^{11} c_i B^i \right)$$

where $B = 1000$, b_i and c_i are integers between 0 and 999. This requires 12 integers, instead of 4, but is useful for internal operations. You can say

```
\FPadd\foo{10}{3.5}
\FPmul\xbar\foo\foo
```

This will put 13.5 in `\foo` and 282.5 in `\xbar`. In verbose mode, you will see that the transcript file contains lines of the form:

```
{\FPadd}
{\FP@add}
{FPread for \FP@add=+10.}
{FPread for \FP@add=+3.5}
```

In reality, the first input line is converted into

```
\FP@add\foo10..\relax3.5..\relax
```

Most commands follow this scheme. There are some exceptions. You can use `\FPprint`. This takes one argument and prints (typesets) it. The algorithm is a bit strange: if the argument list is empty, the result is 0. If the argument is 123, ore more generally a list of tokens, where the first has category code 12 (other), then nothing happens, the arguments is translated normally. If the argument is ‘foo’, the result is ‘13.5’. More generally, if the first item is a character not of category code 12, the command behaves like `\csname`. Don’t try constructions like `\FPprint{\$x~2$}`. You can say `\FPset{gee}{foo}`. The second argument is handled as for `\FPprint`. The first argument should be a command name, or a sequence of characters that becomes a command name via `\csname`. The effect of the command is the same as `\def\gee{13.5}`.

The general mechanism for a command like `\FPadd` or `\FPsincos` is to call intermediary commands like `\FP@add` or `\FP@sincos`. These read some command names (these must be definable, no check is make, as for `\let`), then parse numerical arguments, compute results and store the results in the commands. The result is always normalized: trailing zeroes are removed as well as leading zeroes (but at least one digit is returned before the point). If the number is negative a sign is added. A special case is when the result is boolean. In this case the syntax has the form

```
\FPiflt{0.21}{0.20} Wrong\else Correct\fi
```

As a side effect, `\ifFPtest` is made equivalent to `\iftrue` or `\iffalse`. The following line is valid in Tralics, it gives an error in L^AT_EX.

```
\iffalse \FPiflt{0.21}{0.20} \bad\else \badagain\fi \fi
```

Numbers are read as follows: We assume that `\FP@add` sees a string that contains two dots and a `\relax`, see above. This means that you lose if the argument of the user command contains a `\relax`. Otherwise, we have a list A, a dot, a list B, a dot, a list C, then `\relax`. As you can see from the `\FPmul\bar{foo}\foo` example, these quantities are obtained by expanding the argument (here `\foo`) in a `\edef`. For some reason quantities `\A` and `\B` are expanded again. In a case like

```
\FPadd\foo{\noexpand\noexpand\noexpand\V}{12}
```

this gives `\V` after expansion, this is wrong. In a case like

```
\def\V{10.2}
\FPadd\foo{\noexpand\V.4}{12}
```

expansion of A is 10.2, this is equally wrong: After expansion, there should remain only digits in A and B; there can be an optional sign at the start of A: any combination of + and - characters is OK. Note that C, as well as all digits after a space in A or B are ignored. Thus, the following two lines are valid for `Tralics`, invalid in the `TEX` case.

```
\def\V{10 .2}
\FPadd\foo{\noexpand\V.4}{12}
\FPadd\foo{\V.4}{12}
```

The list of all commands

In the list that follows, `\C`, `\Ca`, `\Cb` are command names, and `\V`, `\Va`, `\Vb` are values.

- `\FPident\C\V`. This is the identity; it copies `\V` in `\C`.
- `\FPclip\C\V`. This puts in `\C` the number `\V`. In the `TEX` version, trailing zeroes are not removed by default. The `\FPclip` command has as function to removed them. In `Tralics` this is a no-op.
- `\FPprint\V`. See description above. This typesets `\V`.
- `\FPset\C\V`. See description above. This copies `\V` in `\C`.
- `\FPneg\C\V`. This copies the opposite of `\V` in `\C`.
- `\FPsgn\C\V`. This copies the sign of `\V`, that can be +1, -1 or 0, in `\C`.
- `\FPabs\C\V`. This puts in `\C` the absolute value of the number `\V`.
- `\FPadd\C\Va\Vb`. This computes the sum of `\Va` and `\Vb` and stores the result in `\C`.
- `\FPsub\C\Va\Vb`. This computes the difference of `\Va` and `\Vb` and stores the result in `\C`.
- `\FPmul\C\Va\Vb`. This computes the product of `\Va` and `\Vb` and stores the result in `\C`. The `TEX` code is beautiful: `Tralics` uses the same idea: we write each number as $\sum b_i B^i$ and $\sum c_i B^i$, then compute the sum of $b_i c_j B^{i+j}$. 144 additions and multiplications are required. One can do better, using Karatsuba's method, but there is an additional cost for finding the b_i and merging at the end. This cost is 22 multiplications and divisions. The result is a number with 72 decimal digits. An overflow is signaled in case there are more than 18 decimals before the point. Last digits are silently ignored.

```

\FP@@mul vv \relax\FP@saveshift%
\FP@@mul vu uv \relax\FP@saveshift%
\FP@@mul uu vt tv \relax\FP@saveshift%
\FP@@mul ut tu vz zv \relax\FP@saveshift%
\FP@@mul tt zu uz rv vr \relax\FP@saveshift%
\FP@@mul zt tz ur ru vq qv \relax\FP@saveshift%
\FP@@mul zz rt tr uq qu vp pv \relax\FP@saveshift%
\FP@@mul zr rz tq qt up pu vo ov \relax\FP@saveshift%
\FP@@mul rr qz zq tp pt uo ou vn nv \relax\FP@saveshift%
\FP@@mul rq qr zp pz to ot un nu vm mv \relax\FP@saveshift%
\FP@@mul qq rp pr zo oz tn nt um mu vl lv \relax\FP@saveshift%
\FP@@mul qp pq ro or zn nz tm mt ul lu kv vk \relax\FP@saveshift%
\FP@@mul pp oq qo rn nr zm mz tl lt ku uk \relax\FP@saveshift%
\FP@@mul op po nq qn rm mr zl lz tk kt \relax\FP@saveshift%
\FP@@mul oo pn np mq qm rl lr kz zk \relax\FP@saveshift%
\FP@@mul no on mp pm lq ql kr rk \relax\FP@saveshift%
\FP@@mul nn mo om pl lp qk kq \relax\FP@saveshift%
\FP@@mul mn nm lo ok pk kp \relax\FP@saveshift%
\FP@@mul mm ln nl ko ok \relax\FP@saveshift%
\FP@@mul lm ml kn nk \relax\FP@saveshift%
\FP@@mul ll km mk \relax\FP@saveshift%
\FP@@mul kl lk \relax\FP@saveshift%
\FP@@mul kk \relax\FP@saveshift%

```

Figure 5.2: A part of the T_EX code of the multiplication in fp.

- `\FPdiv\C\Va\Vb`. This computes the quotient of `\Va` and `\Vb` and stores the result in `\C`. Division is much harder than multiplication. Given x and y , if

$$X = 10^n x \quad Y = 10^m y \quad q = Q10^{m-n}$$

then

$$X = QY \iff x = qy$$

We chose n and m as large as possible, then compute the quotient Q of X and Y . The quotient q is then obtained by shifting. By construction Q is between 10 and $1/10$, shifting may underflow (this gives 0), or overflow (this is an error). When we compute the quotient of X by Y , we can multiply both numbers by 10^{18} , said otherwise, consider them as integers. Write $Q = \sum a_k 10^{-k}$. Only indices k with $k \geq 0$ are needed. In the case $a_0 = 0$, we know $a_1 \neq 0$. Write $10^k Q = Q'_k + Q''_k$, where Q'_k is an integer, and Q''_k is between 0 and 1. Write

$$X - 10^{-k} Y \cdot Q'_k = Q''_k (10^{-k} Y)$$

Let $Y'_k = 10^{-k} Y$. This quantity appears twice above. Let $X'_k = X - Y'_k Q'_k$. We have $X'_k = Q''_k Y'_k$. The algorithm is then: replace X'_k and Y'_k by their integer parts. Multiply X'_k by 10, find n the largest integer such that $nY'_k \leq 10X'_k$. This is the next digit in the quotient. Replace X'_k by $10X'_k - nY'_k$, divide Y'_k by 10. The procedure stops when Y'_k becomes zero. For instance, if we compute the inverse of the inverse of π , we find a number whose last digits are 243 instead of 238. On the other hand, if we divide 10 by $10/\pi$, we get all correct digits (this is because we lose a digit when we shift).

- `\FPmin\C\Va\Vb`. This puts in `\C` the smallest of the two values `\Va` and `\Vb`.
- `\FPmax\C\Va\Vb`. This puts in `\C` the largest of the two values `\Va` and `\Vb`.

- `\FPtruncate\C\V\I`. This truncates \V to \I digits and puts the result in \C .
- `\FPround\C\V\I`. This rounds \V to \I digits and puts the result in \C . For the two commands `\FPtruncate` and `\FPround`, the last argument has to be an integer. It is an error if this is a negative integer. Nothing happens if the integer is larger than 18. Truncating π to 4 digits gives 3.1415, rounding gives 3.1416. Rounding 1.25 to 1 digit gives 1.3. Rounding $-X$ gives the opposite of rounding X . No overflow is signaled: rounding $10^{18} - \epsilon$ may give 10^{18} .
- `\FPiflt\Va\Vb`, `\FPifgt\Va\Vb`, `\FPifeq\Va\Vb`. These three commands compare two values and set the boolean `\ifFPtest` to `\iftrue` in the case $a < b$, $a > b$ and $a = b$ respectively, to `\iffalse` otherwise. This command is then executed. It is a conditional, so that a `\fi` is required, and an `\else` is allowed.
- `\FPifneg\V`, `\FPifpos\V`, `\FPifzero\V`, `\FPifint\V`. Four command that test if the argument a satisfies $a < 0$, or $a > 0$ or $a = 0$ or a integer. The command `\ifFPtest` is set and executed as above.
- `\FPe\C`. This put the number e in \C .
- `\FPseed\V`. This is a counter that contains the seed used by the random number generator.
- `\FPrand\C`. This puts in the command a random number. The algorithm is that of Lewis, Goodman & Miller (1969). Let $q = 127773$, $m = 2147483647$. Write $m = Aq + B$, with $A = 16807$ and $B = 2836$. If s is the seed, a and b the quotient and remainder of s by q , $w = Ab - Ba$ if this is ≥ 0 , $w = Ab - Ba + m$ otherwise. Then w is the remainder of As by m . This is the new seed. The random number is w/m .
- `\FPpascal\C\I`. The last argument must be an integer, between 1 and 63. The result is one row of the Pascal triangle. For instance `\FPpascal\foo{5}` stores in `\foo` the token list `[1.,5.,10.,10.,5.,1.]`.
- `\FPexp\C\V`. This puts in \C the exponential of the number \V . The command uses the relation

$$\exp(a + b) = \exp(a) \cdot \exp(b)$$

The argument x is written as $x = a + b$, where a is an integer, and b is between 0 and 1 in magnitude (same sign as x). The quantity $\exp(a)$ is precomputed (if $a \geq 42$, we have an overflow, if $a \leq -43$, we have an underflow, and the result is zero). The other factor is computed as

$$\exp(b) = \sum b^n/n! = \sum X_k$$

Here X_{k+1} is X_k multiplied by b , divided by k . The division by k uses a special algorithm (if the digits in base 1000 of the numbers are c_i , dividing $\sum c_i B^i$ by a number k is trivial, provided $k < 1000$). We compute terms as long as X_k is no zero.

- `\FPln\C\V`. This puts in \C the logarithm of the number \V . The command uses the relation

$$\log(a \cdot b) = \log(a) + \log(b)$$

Here we use for a a power of ten, so that our number is ab , with b between 1 and 10, then a power of two so that it is between 1 and 2. The logarithms of these quantities are precomputed. Let $y = (x - 1)/(x + 1)$. If $1 \leq x \leq 2$ then $0 \leq y \leq 1/3$. Let

$$f(y) = 2 \sum y^{2k+1}/(2k+1) = 2 \sum X_k$$

We have $f'(y) = 2 \sum y^{2k} = 2/(1 - y^2)$. If we consider $f(y)$ as a function of x , say $g(x)$, its derivative is $1/x$, so that $g(x)$ is the logarithm of x . We compute X_k as $Y_k/(2k + 1)$,

division as above, $Y_{k+1} = zY_k$, where z is y^2 . This gives, for the number π with 18 digits a logarithm of 1.144729885849400161 (last digits should be 173). For the exponential, we get 23.140692632779268882 (last digits should be 8995). The error is larger, because we multiply two truncated numbers.

- `\FPpow\C\Va\Vb`. This puts in `\C` the first value raised to the power the second value.
- `\FProot\C\Va\Vb`. This puts in `\C` the first value raised to the power the inverse of the second value. This use

$$a^b = \exp(b \cdot \log a) \quad \sqrt[b]{a} = \exp(\log a/b)$$

If we compute `\FPpow\foo{4}{2}`, we get $16 - 99 \cdot 10^{-18}$ and `\FProot\foo{16}{2}` gives $4 - 51 \cdot 10^{-18}$. Other examples: For some numbers, we have computes the square and fourth root. We have then squared the result (or squared twice) and computed the difference with the initial number.

| x | 70000 | 7000 | 700 | 70 | 7 | 0.7 | 0.07 | 0.007 | 0.0007 |
|---------------------------------|---------|--------|-------|------|-----|-----|------|-------|--------|
| $10^{-18}(x - (\sqrt{x})^2)$ | 1629349 | 171653 | 17569 | 1871 | 157 | 14 | 2 | 1 | 1 |
| $10^{-18}(x - (\sqrt[4]{x})^4)$ | 2515666 | 297486 | 27306 | 2222 | 247 | 7 | 2 | 1 | 1 |

- `\FPSin\C\V`, `\FPCos\C\V`, `\FPtan\C\V`, `\FPCot\C\V`, `\FPSincos\Ca\Cb\V`, `\FPtancot\Ca\Cb\V`. These six functions compute the trigonometric functions of a value v . In some cases, both the sine and the cosine are computed. The tangent is the quotient of the sine and the cosine, the cotangent is the quotient with arguments reversed. The `\FPSincos` takes two commands: the sine will be put in the first, the cosine in the second. The command `\FPtancot` stores the tangent and the cotangent.

The first step is to reduce the argument modulo 2π . The same algorithm as for the division is used (however 2π is computed with 36 digits, so that the result is more precise than a division via FP commands). The second step is to reduce modulo $\pi/2$. This is done according to the following table, where S and C and the sine and cosine of y .

| x | y | $\sin x$ | $\cos x$ |
|-----------------------------|--------------|----------|----------|
| $0 \leq x \leq \pi/4$ | x | S | C |
| $\pi/4 \leq x \leq \pi/2$ | $\pi/2 - x$ | C | S |
| $\pi/2 \leq x \leq 3\pi/4$ | $x - \pi/2$ | C | $-S$ |
| $3\pi/4 \leq x \leq \pi$ | $\pi/2 - x$ | S | $-C$ |
| $\pi \leq x \leq 5\pi/4$ | $x - \pi$ | $-S$ | $-C$ |
| $5\pi/4 \leq x \leq 3\pi/2$ | $3\pi/2 - x$ | $-C$ | $-S$ |
| $3\pi/2 \leq x \leq 7\pi/4$ | $x - 3\pi/2$ | $-C$ | S |
| $7\pi/4 \leq x \leq 3\pi$ | $2\pi - x$ | $-S$ | C |

The quantities C and S are computed via the obvious formulas

$$\sin y = \sum (-1)^n y^{2n+1} / (2n+1)! \quad \cos y = \sum (-1)^n y^{2n} / (2n)!$$

The following piece of code can be used to test:

```
\def\T#1{
\FPSin\fooA{#1}\FPCos\fooB{#1}
\FPmul\fooA\fooA\fooA
\FPmul\fooB\fooB\fooB
\FPadd\foo\fooA\fooB\FPadd\foo\foo{-1}
\show\foo}
```

If the argument is $2x/10$, x integer, between 0 and 30, this procedure gives numbers with magnitude at most $3 \cdot 10^{-18}$.

- `\FParsin\C\V`, `\FParccos\C\V`, `\FParcsincos\Ca\Cb\V`. These commands compute the inverse sine, cosine, or both of an argument. The inverse sign is between $-\pi/2$ and $\pi/2$, the inverse cosine is between 0 and π . It is an error if the argument is not between -1 and $+1$. Consider

$$f(x) = \sum x^{2k+1} a_k / (2k+1) \quad a_k = a_{k-1} (2k-1) / 2k.$$

We have $f'(x) = \sum a_k x^{2k}$. Since $a_k = (2k-1)! / k! 2^k$, we have $f'(x) = (1-x^2)^{-1/2}$, so that f is inverse sine. If $n = 2k-1$, we have to compute $n/(n+1)$, this uses a hack. If $x \leq \sqrt{2}/2$, we compute $f(x)$, otherwise $f(y)$ with $y = \sqrt{1-x^2}$. This is the inverse cosine of x . The square root is computed exactly, it is not an approximation via logarithms and exponentials. When we compute, for x integer between 0 and 30, the inverse cosine of the cosine of x , we get an error at most $24 \cdot 10^{-18}$. The error reaches its maximum and changes sign near $\pi/4$ and $3\pi/4$.

- `\FPtan\C\V`, `\FPcot\C\V`, `\FPtancot\Ca\Cb\V`. These commands compute the inverse tangent, or cotangent or both of the argument. For the inverse tangent the result is between $-\pi/2$ and $\pi/2$. For the inverse cotangent the result is between 0 and π . Since $\tan(\pi/2-t) = 1/\tan t$, we can assume that the argument is less than one. In the case $x \geq 8/9$, we use the following:

$$y = \frac{5-4x^2}{9} \quad \frac{1}{\sqrt{1+x^2}} = \frac{2/3}{\sqrt{1-y}}.$$

If $x = \tan t$, then $1/\sqrt{1+x^2} = \cos t$. We compute this, then the inverse cosine. For $1/\sqrt{1-y}$, we use a variant of the formula shown above: We write $(2k-1)/(2k) = 1 - 1/(2k)$, and we compute $1/(2k)$ via normal division. In the case $x \leq 8/9$, we use the formula

$$\arctan x = x - x^3/3 + x^5/5 - x^7/7 + \dots$$

- `\FPlsolve\Ca\Va\Vb`. Assume that the arguments are x_1 , and A, B . This solves the equation

$$Ax + B = 0$$

and puts the result in x_1 . In `Tralics`, you will find the number of solutions in `\count0`. In the case of an equation of degree one, the result is trivial: $x = -B/A$.

- `\FPqsolve\Ca\Cb\Va\Vb\Vc`. Assume that the arguments are x_1, x_2 , and A, B, C . This solves the equation

$$Ax^2 + Bx + C = 0$$

and puts the result in x_1 , and x_2 . In `Tralics`, you will find the number of solutions in `\count0`. A special case is when $A = 0$, the equation is of degree one. Otherwise, we replace B by B/A and C by C/A . We replace B by $B/2$. Thus, the equation to be solved is $x^2 + 2Bx + C = 0$. Let $\Delta = B^2 - C$. If $\Delta < 0$, there is no solution. Otherwise, the solutions are $x_2 = -(\pm\sqrt{\Delta} + B)$, and $x_1 = C/x_2$. Here the sign of the square root is the same as B .

- `\FPcsolve\Ca\Cb\Cc\Va\Vb\Vc\Vd`. Assume that the arguments are x_1, x_2, x_3 , and A, B, C, D . This solves the equation

$$Ax^3 + Bx^2 + Cx + D = 0$$

and puts the result in x_1, x_2 , and x_3 . In `Tralics`, you will find the number of solutions in `\count0`. A special case is when $A = 0$, this is an equation of degree 2. Otherwise, let $D' = (D/A)/2$, $T = -B/(3A)$, $p = C/(3A) - T^2$ and $q = -(BC)/(6A^2) + D - T^3$. Let $\delta = p^3 - q^2$. If $\delta > 0$, there is a single solution. Let $u = \sqrt{\delta} - q$ and $v = -(\sqrt{\delta} + q)$. The solution is

$$x = \sqrt[3]{u} + \sqrt[3]{v} + T.$$

Otherwise, let $R = \pm\sqrt{|p|}$, with the same sign as q . Let $s = \arccos(q/R^3)/3$, and

$$r_1 = -\cos(s) \quad r_2 = \cos(\pi/3 - s) \quad r_3 = \cos(\pi/3 + s).$$

The solutions are $x_i = 2Rr_i + T$. The implementation of the formulas are a bit different than in the initial T_EX version, so that numerically, there can be some small differences.

- `\FPqqsolve\Ca\Cb\Cc\Cd\Va\Vb\Vc\Vd\Ve`. Assume that the arguments are x_1, x_2, x_3, x_4 and A, B, C, D, E . This solves the equation

$$Ax^4 + Bx^3 + Cx^2 + Dx + E = 0$$

and puts the result in x_1, x_2, x_3 and x_4 . In Tralics, you will find the number of solutions in `\count0`. We can assume $A \neq 0$, otherwise this is an equation of degree 3. Then we divide by A , and assume $A = 1$. A special case is $D = 0$, i.e. $Ay^2 + Cy + E = 0, x^2 = y$. We first find y . For each positive solution, we take the square root. A second special case is $A = E, B = D$. This is

$$x^4 + Bx^3 + Cx^2 + Bx + 1 = 0.$$

If $u = x + 1/x$, we get $u^2 + Bu + (C - 2) = 0$. We solve this equation. For each solution u we solve $x^2 - ux + 1 = 0$. In the general case, we consider

$$B' = B/4, T = -B', E' = E - B'D + CB'^2 - 3B'^4, D' = D - 2B'C' + 8B'^3, C' = C - 6B'^2$$

The effect is to shift the roots by T . If

$$x^4 + C'x^2 + D'x + E' = 0$$

has roots r_i , then the initial equation has roots $x_i = r_i + T$. Consider

$$P = (x^2 + s'/2 + r - D/s')(x^2 - s'/2 + r + D/s')$$

This is the original polynomial if $C = 2r - s'^2/4$ and $E = r^2 - D^2/s'^2$. Let $s = 8r - 4C$. The first equation is true if $s'^2 = s$. We must have $s > 0$. The second relation says that r must be a solution of

$$8y^3 - 4Cy^2 - 8Ey + 4CE - D^2 = 0.$$

This is called the resolvent. This, we solve this equation, compute the three roots, chose one with $s > 0$ (if this is not possible the initial equation has no real root). After that $P = 0$ becomes two equations of degree two.

Alternate syntax

The command `\FPupn` implements a postfix language that allows you to write shorter code. Here is an example

```
\FPupn\foo{7 20 2 sub 100 2000 - add +}
\testeq\foo{1925.}
\FPupn\foo{20 2 div 100 2000 / add 3 mul 2 *}
\testeq\foo{180.}
\def\mthree{-3}%there is no unary minus in this language
\FPupn\foo{ 3 abs mthree abs 3 sgn 10 * mthree sgn 100 * + + +}
\testeq\foo{-84.}
\FPupn\foo{2 3 min 400 500 max +}
\testeq\foo{502.}
\FPupn\foo{12.43745678 2 round 12.35745678 2 trunc -}
```

```

\testeq\foo{-0.09}
\FPupn\foo{e 1.2 exp + 2.3 ln + 3 4 pow + 5 6 root +}
\testeq\foo{72.302276955235951659}
\FPupn\foo{pi 0.7 - sin cos sincos - tan cot tancot +}
\testeq\foo{-2.894412996263821897}
\FPupn\foo{0.3 arcsin 0.1 * arccos 0.1 * arcsincos - arctan arccot arctancot -}
\testeq\foo{0.751779218345560029}
\FPupn\foo{3.4 seed random}
\testeq\foo{0.000023479107778276}
\FPupn\foo{1.1 2.3 3.4 pop swap copy add sub}
\testeq\foo{0.1}

```

The `\testeq` command can be used to test the code. It is an error if the two arguments are not the same. Some comments. Consider the last expression. We put 1.1, 2.3 and 3.4 on the stack. After that we pop an item. After that we swap. The stack holds 1.1 (top stack), followed by 2.3. Then we duplicate the top stack. Then we add. The topstack is now 2.2. After subtraction, we get 0.1. If you say ‘2 3 -’, the result is 1, because - and sub use arguments in a different order. The same is true for / and div. Note the order of 10 2 pow, this gives 1024. If strange words are seen, like ‘mthree’, they are replaced by `\mthree`. Note that ‘e’ and ‘pi’ are predefined.

If you don’t like postfix language, you can use `\FPeval`. Here are some examples.

```

\FPeval\foo{1000-100-10-1}
\testeq\foo{889.}
\FPeval\foo{1000+100+10+1}
\testeq\foo{1111.}
\FPeval\foo{1000-100+10+1}
\testeq\foo{911.}
\FPeval\foo{1000+100-10+1}
\testeq\foo{1091.}
\FPeval\foo{(20 - 2) + (2000-100) + 7}
\testeq\foo{1925.}
\FPeval\foo{(20/2 + 2000/100)*3*2}
\testeq\foo{180.}
\FPeval\foo{210/2/3/5}
\testeq\foo{7.}
\FPeval\foo{210*2/3/5}
\testeq\foo{28.}
\FPeval\foo{210/2*3/5}
\testeq\foo{63.}
\FPeval\foo{210/2/3*5}
\testeq\foo{175.}
\FPeval\foo{210*2*3/5}
\testeq\foo{252.}
\FPeval\foo{210*2/3*5}
\testeq\foo{700.}
\FPeval\foo{210/2*3*5}
\testeq\foo{1575.}
\FPeval\foo{abs(3) + abs(-3) + (sgn(3)* 10) + (sgn(-3) * 100)}
\testeq\foo{-84.}
\FPeval\foo{min(2:3) + max(400,500)}
\testeq\foo{502.}
\FPeval\foo{round(12.43745678,2) - trunc(12.35745678, 2)}
\testeq\foo{0.09}

```

```

\FPeval\xfoo{e + exp(1.2) + ln(2.3) + pow(3, 4) + root(5, 6)}
  \testeq\xfoo{72.302276955235951659}
\FPeval\xfooa{sin(cos(sin(0.7 - pi))) - cos(cos(sin(0.7 - pi)))}
\FPeval\xfoo{tan (cot(tan(xfooa))) + cot(cot(tan(xfooa)))}
  \testeq\xfoo{-2.894412996263821897}
\FPeval\xfooa{arcsin (arccos (arcsin(0.3)*0.1)*0.1) -
  arccos (arccos (arcsin(0.3)*0.1)*0.1)}
\FPeval\xfoo{arctan(arccot(arctan(xfooa))) - arccot(arccot(arctan(xfooa)))}
  \testeq\xfoo{0.751779218345560029}
\FPeval\foo{2+3*4+5*3^2}
  \testeq\foo{58.999999999999998665}
\FPeval\foo{3^2*5+4*3+2+1}
  \testeq\foo{59.999999999999998665}
\FPeval\foo{(+3+4)*(-5-6)}
  \testeq\foo{-77.}

```

If you wonder what happens, you can look the transcript file. You can see something like:

```

{\FPeval}
{FPpostfix 1 2 3 mul add 400 500 max sin 4 pow add}
{\FP@upn}
{FPupcmd ??}
{FPupcmd ??}
{FPupcmd ??}
{FPupcmd mul}
{FPread for \FP@upn=+3.}
{FPread for \FP@upn=+2.}
{FPupcmd add}
{FPread for \FP@upn=+6.}
{FPread for \FP@upn=+1.}
{FPupcmd ??}
{FPupcmd ??}
{FPupcmd max}
{FPread for \FP@upn=+500.}
{FPread for \FP@upn=+400.}
{FPupcmd sin}
{FPread for \FP@upn=+500.}
{FPupcmd ??}
{FPupcmd pow}
{FPread for \FP@upn=+4.}
{FPread for \FP@upn=-0.467771805322476126}
{FPupcmd add}
{FPread for \FP@upn=+0.522845423476396576}
{FPread for \FP@upn=+7.}
{FPread for \FP@upn=+7.522845423476396576}

```

The second line is the expression converted from infix to postfix. Each ‘??’ represents a string that does not start with a letter. This is generally a number.

5.7 Action before translation

Normally, translation applies only to what is between `\begin{document}` and `\end{document}`. This is a very special environment, in fact, it leaves the semantics stack pointer unchanged. There are two hooks. You can say

```
\AtBeginDocument{\foo}
\AtEndDocument{\xbar}
```

These commands remember the tokens in a special list, that is inserted in the input stream when `\begin` or `\end` is seen. After that, the meaning of the command changes: it becomes ‘evaluate now’, more precisely `\@firstofone`. The last action in the begin-document hook is to change the definition again, so that an error may be signaled, for instance *Can be used only in preamble*: `\AtBeginDocument`. On the other hand, the `\end{document}` command inserts a special marker that closes every open file, thus stopping translation at the end of the hook (the bibliography is translated after that). Before the begin of the document, you can use commands of the form

```
\documentclass[doc-opt]{doc-class}
\usepackage[pack-opt]{pack-name}
```

There are some differences with \LaTeX , in particular the lines are completely independent; in \LaTeX , options that are not used by the document class are seen by the packages. We assume moreover that pack-name is the name of a single package. If ‘doc-opt’ contains ‘useallsizes’ this is the same as if a line in the configuration file has said to use all font sizes. If it contains ‘french’ or ‘english’, this defines the default language. If it contains ‘leadingchapter’ or ‘leadingpart’ this means that the top-level section is ‘chapter’ or ‘part’, otherwise it is ‘section’. If ‘doc-class’ contains ‘book’ then the top-level section is ‘part’, it if contains ‘report’, it is ‘chapter’. If the top-level section is part, chapter, or section, the translation of `\subsection` is, respectively, a `<div3>`, `<div2>` or `<div1>` element. Moreover, an attribute pair `chapters=‘true’` or `part=‘true’` is added to the main element, so that a post-processor can decide that `<div1>` is subsection, section or chapter.

If the packages ‘calc’, ‘fp’ or ‘fancyhdr’ are loaded, then the meaning of some command changes, as explained elsewhere. If the babel package is loaded, the following languages are recognized: english, american, british, canadian, UKenglish, USenglish (these have number 0), french, francais, frenchb, acadian, canadien (these have number 1) austrian, german, germanb, naustrian, ngerman (these have number 2). The first language in the list is the default language. If a package is name ‘french’ or ‘frenchb’ or ‘german’, the default language is also set. The default language can be used in the attribute list of the main document element. Setting the default language also set the current language (value of `\language`).

The standard configuration file contains lines like these:

```
on package loaded calc CALC = "true"
on package loaded foo/bar FOO1 = "true"
on package loaded *foo/bar FOO2 = "true"
on package loaded foo/*bar FOO3 = "true"
on package loaded *foo/*bar FOO4 = "true"
```

The first line means: if the ‘calc’ package is loaded, then the CALC attribute of the main document should be set to true. The second line means: if the package ‘foo’ is loaded with the option ‘bar’, then the FOO1 attribute should be set. In the lines that follows ‘*foo’ means: anything that contains foo. You can also say

```
on_package_loaded calc CALC = "true"
on_package_option calc CALC = "true"
on_class_option article CALC = "true"
on class option */* CALC = "**"
```

In the attribute value, a star is replaced by the value of the package, or document class, and a plus sign by the options that matches (or the whole list of options). If ‘class’ is used instead of ‘package’, the the document class and its options is considered instead of the package.

5.8 Other expandable commands

Not all commands defined here can be expanded, see the list at the end of Chapter 2. Essentially, this describes the ifthen package, the calc package, the newtheorem mechanism, and some input-output commands.

- `\@whilenum`, `\@whiledim`, `\@whilesw`, `\whiledo`. Consider the following example,

```
\newcounter{ca}\newcounter{cb}
\newcommand{\printgcd}[2]{%
  \setcounter{ca}{#1}\setcounter{cb}{#2}%
  Gcd(#1,#2) =
  \whiledo{\not\(\value{ca}=\value{cb}\)}%
    {\ifthenelse{\value{ca}>\value{cb}}%
      {\addtocounter{ca}{-\value{cb}}}%
      {\addtocounter{cb}{-\value{ca}}}%
      gcd(\arabic{ca},\arabic{cb}) = }%
  \arabic{ca}.}
```

This works only if the package ‘ifthen’ is loaded. It should produce (we have inserted some `\allowbreak` commands in order to allow linebreaks in the very big formula): $Gcd(144,272) = gcd(144,128) = gcd(16,128) = gcd(16,112) = gcd(16,96) = gcd(16,80) = gcd(16,64) = gcd(16,48) = gcd(16,32) = gcd(16,16) = 16$.

The ifthen package implements `\whiledo` in terms of the L^AT_EX command `\@whilesw`. The Tralics implementation of `\whiledo` is: read arguments A and B. Evaluate A as the first argument of `\ifthenelse`. If false terminate. Otherwise, the expansion is `B\whiledo A{B}`. Using one version or the other produces the same result.

```
\makeatletter
\newif\if@whiledo
\long\def\whiledo#1#2{%
  \ifthenelse{#1}%
    {\@whiledottrue
     \@whilesw\if@whiledo\fi
     {#2}
    \ifthenelse{#1}\@whiledottrue\@whiledofalse}}%
  {}%
}
```

This is the same code, without the ifthen package.

```
\newif\iftest
\newcommand{\printgcd}[2]{%
  \setcounter{ca}{#1}\setcounter{cb}{#2}%
  \testtrue
  \@whilesw\iftest gcd(\arabic{ca},\arabic{cb}) = \fi{%
  \ifnum\value{ca}=\value{cb}\testfalse\else
  \ifnum\value{ca}>\value{cb}%
    \addtocounter{ca}{-\value{cb}}\else
    \addtocounter{cb}{-\value{ca}}\fi
```

```
\fi}%
\arabic{ca}.}
```

We propose now a method that uses `\@whilenum`. We start with two commands that put in (c, d) the numbers (a, b) , in some order, starting with the smallest. We consider also a function that takes the integers, puts them in length registers.

```
\newlength\c\newlength\cD
\newcounter{cc}\newcounter{cd}
\def\assigncounter{%
\ifnum\value{ca}<\value{cb}%
\setcounter{cc}{\value{ca}}%
\setcounter{cd}{\value{cb}}%
\else
\setcounter{cc}{\value{cb}}%
\setcounter{cd}{\value{ca}}%
\fi}
\def\assignlength{%
\ifnum\value{ca}<\value{cb}%
\setlength\cC{\value{ca}sp}%
\setlength\cD{\value{cb}sp}%
\else
\setlength\cC{\value{cb}sp}%
\setlength\cD{\value{ca}sp}%
\fi}
```

This puts $(c, c - d)$ in (a, b) .

```
\def\subtractcounter{%
\setcounter{cb}{\value{cd}}%
\setcounter{ca}{\value{cc}}%
\addtocounter{cb}{-\value{cc}}}%
\def\subtractlength{%
\setcounter{cb}{\cD}%
\setcounter{ca}{\cC}%
\addtocounter{cb}{-\cC}}%
```

Computing the gcd is now obvious: it suffices to call `assign` and `subtract`, until $c = d$ (this will happen), because the procedures leave the gcd invariant.

```
\let\printgcd\relax
\newcommand{\printgcd}[2]{%
\setcounter{ca}{#1}\setcounter{cb}{#2}%
\assigncounter
\@whilenum \value{cc}<\value{cd} Gcd(\arabic{cc},\arabic{cd}) = \do{%
\subtractcounter\assigncounter}%
\arabic{ca}.}
\printgcd{144}{272}
```

The same with dimensions

```
\let\printgcd\relax
\newcommand{\printgcd}[2]{%
\setcounter{ca}{#1}\setcounter{cb}{#2}%
\assignlength
\@whiledim \cC<\cD Xgcd(\arabic{ca},\arabic{cb}) = \do{%
\subtractlength\assignlength}%
```

```

\arabic{ca}.}
\printgcd{144}{272}

```

In verbose mode, the transcript file of Tralics will contain lines of the form shown here. It indicates the expansion of `\@whilenum` and the internal command `\@iwhilenum` associated to it.

```

\@whilenum<- \ifnum \value {cc}<\value {cd} Gcd(\arabic {cc},\arabic {cd}) =
\relax \subtractcounter \assigncounter \relax \@iwhilenum {\value {cc}<\value
{cd} Gcd(\arabic {cc},\arabic {cd}) = \relax \subtractcounter \assigncounter
\relax } \fi

```

```

\@iwhilenum<- \ifnum \value {cc}<\value {cd} Gcd(\arabic {cc},\arabic {cd}) =
\relax \subtractcounter \assigncounter \relax \expandafter \@iwhilenum \else
\expandafter \@gobble \fi {\value {cc}<\value {cd} Gcd(\arabic {cc},\arabic
{cd}) = \relax \subtractcounter \assigncounter \relax }

```

This looks like the following (in the case of `\@whilenum`, there is no `\relax` after ‘ETC’, so that the code is compatible with L^AT_EX).

```

\@whilenum<- \ifnum TEST ACTION \relax ETC \relax
\@iwhilenum {TEST ACTION \relax ETC \relax } \fi

\@iwhilenum<- \ifnum TEST ACTION \relax ETC \relax
\expandafter \@iwhilenum \else \expandafter \@gobble \fi
{TEST ACTION \relax ETC \relax }

```

We can simplify this further to the following.

```

\@whilenum<- \ifnum CODE \@iwhilenum {CODE} \fi
\@iwhilenum<- \ifnum CODE
\expandafter \@iwhilenum \else \expandafter \@gobble \fi {CODE}

```

You can see that, if the test is true, the first `\expandafter` token is evaluated; as a consequence, everything between `\else` and `\fi` is removed and we are left with `\@iwhilenum` followed by `{CODE}`; if the test is false the `\fi` is expanded first, so that `\@gobble` sees the code to gobble.

- `\loop`. This is the command described in the T_EX book. The implementation is less tricky. The following example should print 0291817161514131211101908070605040302010

```

\count0=0
\def\foo{}
\def\xbar#1#2{\xdef\foo{\foo#1#2}}
\loop \advance\count0by1 \edef\xx{\ifnum\count0<10 0\fi\the\count0 }
\expandafter\xbar\xx \ifnum\count0<20 \repeat
\show\foo

```

- `\setlength`. This command takes two argument, say `\foo` and ‘ba’. The expansion is `\foo ba\relax`, unless the `calc` package is loaded, case where the result is `\calc{\foo}{ba}`.
- `\addtolength`. This command takes two argument, say `\foo` and ‘bar’. The expansion is `\advance\foo bar\relax`, unless the `calc` package is loaded, case where the result is `\calc{\advance\foo}{bar}`. Tralics tests that the first argument is a single token; an error is signaled if this is `\skip\footins`, although this is a valid length.
- `\setcounter`. This command takes two arguments, say ‘foo’ and ‘bar’. The expansion is `\global\c@foo bar\relax`, unless the `calc` package is loaded, case where the result is `\calc{\global\c@foo}{bar}`.

- `\addtocounter`. This command takes two arguments, say ‘foo’ and ‘bar’. The expansion is `\global\advance\c@foo bar\relax`, unless the `calc` package is loaded, case where the result is `\calc{\global\advance\c@foo}{bar}`. For these last two commands, Tralics tests that the first argument is a counter. For instance, if the first argument is `\bar`, you get *Invalid token \bar found in counter name*.
- `\widthof`, `\heightof`, `\depthof`: These commands take an argument, typeset it in a box, and return the dimension of the box. Not implemented in Tralics.
- `\ratio`. This command is defined by the `calc` package to be used as `x/\ratio{u}{v}` or `x*\ratio{u}{v}`. Arguments u and v are dimensions, interpreted by the `\calc` function. The expression that comes before the operator is divided by u and multiplied by v (or the contrary). In Tralics, the quotient is computed with 6 digits; this is not exactly the same algorithm as in L^AT_EX.
- `\real`. This command is defined by the `calc` package to be used as `x*\real{u}` or `x/\real{u}`. The argument u should be a real number as in ‘1.5’.
- `\calc`. Examples of the `calc` package are given in section 2.7. The rules are the following: the `\calc` command (a private command), takes two arguments; the first argument is a token list of the form `\global\advance\count0`. Said otherwise, a variable, preceded by optional prefixes. The variable must be something that reads an integer, a dimension, or glue. The second argument is evaluated and put in the variable (or added to the variable); assignment can be global. A primitive is either a parenthesized expression, a call to `\widthof`, or the result of the command `scanint`, `scandim` or `scanglue` (whichever is used depends on the type of the variable). As a result, if a command like `\setcounter{foo}{bar}` is valid without the `calc`, it remains valid with it, and the interpretation is the same. Moreover, expressions like $a + b$, $a - b$, a/b and $a * b$ can be used, instead of `\advance`, `\multiply` or `\divide`. In the case of multiplication or division, the second argument must be an integer; as an extension, it can be `\real` or `\ratio`, as explained above, in this last case, the Tralics result may differ slightly from the L^AT_EX value.
- `\newboolean`, `\provideboolean`. These two commands take as argument a character string, for instance ‘foo’ and behave like `\newif\iffoo`. Nothing happens for `\provideboolean` if the command `\iffoo` exists, but an error is signaled by `\newboolean`.
- `\setboolean`. This command takes two arguments. The second argument should be, after translation, and modulo the case, ‘true’ or ‘false’. If the first argument is ‘foo’, then `\footrue` is executed, otherwise `\foofalse` is executed. This works well if the command has been defined by `\newboolean{foo}`.
- `\boolean`. This can be used only in an ‘ifthenelse’ test. It takes an argument, say ‘foo’. The test is true if `\iffoo` is `\iftrue`, false otherwise. This works as expected if `\setboolean` has been used to set the value of ‘foo’.
- `\equal`. This can be used only in an ‘ifthenelse’ test. It takes two arguments, puts them in commands, say `\tmpa`, and `\tmpb`, expands then via `\edef`, and compares the result via `\ifx`. The result can be true or false.
- `\isodd`. This can be used only in an ‘ifthenelse’ test. It takes an argument, say ‘foo’. The test is true if the argument is an odd number as seen by `\ifodd`. There are no such hacks as in the L^AT_EX version.
- `\isundefined`. This can be used only in an ‘ifthenelse’ test. It takes an argument, say `\foo`. The test is true if it is undefined (if the argument contains more than one token, only the first one is considered).

- `\lengthtest`. This can be used only in an ‘ifthenelse’ test. It takes an argument, say ‘1cm=2cm’. The test is true if `\ifdim` interprets this as true.
- `\not`, `\and`, `\or`, `\NOT`, `\AND`, `\OR`. These can be used only in an ‘ifthenelse’ test as boolean connectors (but could be used elsewhere with a different meaning). If you say `\not \foo \and \not \not \bar`, the condition is true if `\foo` is false and `\bar` is true. Note that if `\foo` is found true, the condition is false, whatever follows `\and`, and this is not evaluated. In the same fashion, if what precedes `\or` is true, what follows is not evaluated.
- `\(, \)`. This can be used to change the order of evaluation. For instance `\not \(\foo \and \bar\)` negates foo-and-bar.
- `\ifthenelse`. This takes three arguments. The first is evaluated using the command explained above. If true, the expansion is the second argument, otherwise the third. In a case like `\ifthenelse{1=3}{Y}{N}`, the argument is evaluated as in `\ifnum`.
- `\theorembodyfont`. This is a command that remembers its argument for use with theorems. Default value is empty.
- `\theoremstyle`. This is a command that remembers its argument for use with theorems. Default value is ‘plain’.
- `\theoremheaderfont`. This is a command that remembers its argument for use with theorems. Default value is `\bfseries`.
- `\newtheorem`. The implementation is as described in the latex companion, first edition. An example can be found on page 194. We comment here a part of the trace:

```
1 [8] \theorembodyfont{\sl}
2 [9] \theoremstyle{break}
```

This is our first theorem. It defines an environment ‘Cor’. For some reason, `\theCor` is set to `\relax`.

```
3 [10] \newtheorem{Cor}{Corollary} \relax
4 {\newtheorem}
5 {\newtheorem Cor}
6 {\let \endCor \@endtheorem}
7 {\let \theCor \relax}
8 {\def \Cor ->\@begintheorem {Corollary}\theCor {\sl }{Cor}{Cor}{break}}
```

Here you see that a counter ‘Cor’ is defined.

```
9 {\countdef \c@Cor=\count24}
10 {\def \theCor ->\arabic {Cor}}
```

Here we see that an optional argument can be used after the name. The counter ‘Exa’ depends on the counter section, and `\theExa` uses `\thesection`.

```
11 [14] \newtheorem{Exa}{Example}[section]
12 {\newtheorem}
13 {\newtheorem Exa}
14 {\let \endExa \@endtheorem}
15 {\let \theExa \relax}
16 {\def \Exa ->\@begintheorem {Example}\theExa {\sl }{Exa}{Exa}{plain}}
17 {\countdef \c@Exa=\count25}
18 {newcounter_opt}
19 {newcounter_opt->\cl@section}
20 \cl@section ->\@elt {subsection}
21 {\def \theExa ->\thesection .\arabic {Exa}}
```

Here we have an optional argument before the name. There is no counter ‘Lem’, but `\theLem` is `\theCor`.

```

22 [17] \newtheorem{Lem}[Cor]{lemma}
23 {\newtheorem}
24 {\newtheorem Lem}
25 {\let \endLem \@endtheorem}
26 {\let \theLem \relax}
27 {\def \Lem ->\@begintheorem {lemma}\theLem {\sl }{Lem}{Cor}{marginbreak}}
28 {\def \theLem ->\theCor }

```

This is now the use of one theorem. We have a command that reads 6 arguments, plus the text.

```

29 [23] \begin{Cor}
30 {\begin}
31 {\begin Cor}
32 +stack: level + 2 for environment
33 \Cor ->\@begintheorem {Corollary}\theCor {\sl }{Cor}{Cor}{break}
34 {\@begintheorem}
35 [24] This is a sentence typeset in the theorem environment \Lenv{Cor}.
36 {Push theorem 1}

```

We first increment the counter (argument 5).

```

37 \refstepcounter->\global \advance \c@Cor 1\relax {\let \@elt \@stpelt \cl@Cor }
38 {\global}
39 {\global\advance}
40 +scanint for \c@Cor->1
41 ...

```

This produces ‘Corollary 1’ from arguments 1 and 2. A part of the transcript file has been removed. but you get the idea. The `\scshape` comes from `\theoremheaderfont`.

```

42 +stack: level + 3 for brace
43 {\scshape}
44 {font change \scshape}
45 {begin-group character {}}
46 \theCor ->\arabic {Cor}
47 \arabic->\number \c@Cor
48 +scanint for \number->1
49 {Text:Corollary 1 }
50 {font restore }
51 +stack: level - 3 for brace

```

What follows is easy: the body is typeset using argument 3.

```

52 {\sl}
53 {font change \slshape}
54 Character sequence: This is a sentence typeset in the theorem environment .
55 ...

```

If the configuration file defines `xml_theorem_name`, the translation is not the same, because the meaning of `\@begintheorem` has changed.

```

56 [23] \begin{Cor}
57 {\begin}
58 {\begin Cor}
59 +stack: level + 2 for environment
60 \Cor ->\@begintheorem {Corollary}\theCor {\sl }{Cor}{Cor}{break}
61 {\@begintheorem}
62 [24] This is a sentence typeset in the theorem environment \Lenv{Cor}.
63 {Push theorem 1}

```

Arguments 1, 4 and 6 are grabbed, and put in an element or attributes. Note: arguments 5 and 6 are identical.

```

64 {Push Head 2}
65 Character sequence: Corollary.
66 {Text:Corollary}
67 {Pop 2: document_v theorem_v Head_t}
68 {Push argument 2}
69 Character sequence: Cor.
70 {Text:Cor}
71 {Pop 2: document_v theorem_v argument_t}
72 {Push argument 2}
73 Character sequence: break.
74 {Text:break}
75 {Pop 2: document_v theorem_v argument_t}

The counter is always incremented.

76 \refstepcounter->\global \advance \c@Cor 1\relax {\let \@elt \@stpelt \cl@Cor }

End is as usual.

77 {Push p 2}
78 Character sequence: This is a sentence typeset in the theorem environment .

```

- `\openin`. This command reads a small integer, an optional equals sign, and a file name, this is a sequence of characters not containing a space. It reads the content of the file. If the filename is 'foo', and no file named foo exists, then foo.tex is tried. You can read the content via `\read`. If the file is not found, empty, or completely read, it is marked as 'closed'.
- `\closein`. This command reads an integer, and closes the channel associated to the number.
- `\ifeof`. This command reads an integer, and checks if the channel is opened for reading.
- `\read`. This command reads an integer, the keyword 'to' and a command name. After that, it reads a line (more than one line can be read, because the result is always balanced according to braces), and put in the command. The line is read from the file associated to the number, unless it is closed, case where a line is read from the terminal. We give here an example:

```

{
\openin 5=tortureaux
\endlinechar=-1
\ifeof5 \badifeofatentry\fi
\read 5 to \foo\show\foo
\read 5 to \foo\show\foo
\read 5 to \foo\show\foo
\global\read 5 to \foo
\closein5\relax
\ifeof5\else\badifeofatexit\fi
}\show\foo
\ifeof3\else \badifeofnonexists\fi

```

The transcript file contains the following lines. One purpose of the group opened here is to keep the modification to the endline character local.

```

1 [9] {
2 {begin-group character {}}
3 +stack: level + 2 for brace

```

This shows the `\openin` command. You can see that the file was tortureaux.tex, it contains six lines.

```

4 [10] \openin 5=tortureaux
5 {\openin}
6 +scanint for \openin->5
7 ++ Opened file tortureaux.tex; it has 6 lines

```

The `\endlinecharacter` command specifies the value of the character that should be inserted at the end of every line read from a file. An out-of-range value means that no character should be inserted.

```

8 [11] \endlinechar=-1
9 {\endlinechar}
10 +scanint for \endlinechar->-1

```

This is a test of `\ifeof`; the test should be ‘false’ if the file has been read.

```

11 [12] \ifeof5 \badifeofatentry\fi
12 +\ifeof1
13 +scanint for \ifeof->5
14 +iftest1 false
15 +\fi1

```

The first line of the file contains ‘abc’. Thus, `\foo` should contain these letters, plus the endline character.

```

16 [13] \read 5 to \foo\show\foo
17 {\read}
18 +scanint for \read->5
19 {\show}
20 \foo=macro: ->abc.

```

The second line of the file is empty.

```

21 [14] \read 5 to \foo\show\foo
22 {\read}
23 +scanint for \read->5
24 {\show}
25 \foo=macro: ->.

```

The third line contains `\a \b {\c`, this is incomplete, so that line 4, containing `{\d} \e` must also be read.

```

26 [15] \read 5 to \foo\show\foo
27 {\read}
28 +scanint for \read->5
29 {\show}
30 \foo=macro: ->\a \b {\c {\d } \e .

```

This is to show that `\global` applies to `\read`: the command `\foo` is globally defined.

```

31 [16] \global\read 5 to \foo
32 {\global}
33 {\global\read}
34 +scanint for \read->5

```

We close the file. Since the end-of-line character is inactive, it cannot be used as a terminator for the number 5, thus the `\relax`.

```

35 [17] \closein5\relax
36 {\closein}
37 +scanint for \closein->5
38 ++ End of file tortureaux.tex
39 {\relax}

```

Here we check that the file is effectively closed. The `\relax` is automatically inserted because the `\else` is seen before the number 5 has been completely read (the `\else` is re-inserted).

```

40 [18] \ifeof5\else\badifeofatexit\fi
41 +\ifeof2
42 +\else2
43 +scanint for \ifeof->5
44 +iftest2 true
45 {\relax}
46 +\else2
47 +\fi2

```

Here you can see what happens at the end of the group: the end-line character is restored, but `\foo` is retained.

```

48 [19] }\show\foo
49 {end-group character }}
50 +stack: retaining \foo
51 +stack: restoring integer value 13 for \endlinechar
52 +stack: level - 2 for brace
53 {\show}
54 \foo=macro: ->123.

```

- `\openout`. This command reads a small integer (between 0 and 15), an optional equals sign, and a file name, this is a sequence of characters not containing a space. It opens the file for writing. In T_EX, this is an extension, and can be prefixed by `\immediate`; in Tralics, the action is always immediate.
- `\closeout`. This command reads a small integer, and closes the channel with this number. No error is signaled if the integer is not in the range 0-15. In T_EX, this is an extension, and can be prefixed by `\immediate`; in Tralics, the action is always immediate. If you write something on a file, and do not close it, it is unclear what is saved on the disk when Tralics quits.
- `\write`. This command reads an integer N , and a command and writes the argument, after full expansion, to the stream. If $N < 0$, this is the transcript file; if N is a valid channel, to which a file is associated via `\openout`, printing will go there; otherwise to the terminal and transcript file. A special case is when the $N = 18$ and the ‘shell-escape’ switch has been given: in this case, the command is evaluated by a shell.

Here is an example. Note how we inhibit expansion of commands. Note also that each `\write` command produces a single line; but newline characters can be added via `^^J`.

```

\immediate\openout 6=tout.tex
\immediate\write6{\def\noexpand\foo{\noexpand\endinput}}
\immediate\write6{\noexpand\foo^^J\noexpand\bar}
\immediate \closeout6 \relax
\input{tout}

```

This is part of the transcript file. It shows that the file has 3 lines, a first line that defines `\foo`, a line with `\foo` and `\bar`. Since `\foo` is `\endinput`, the line with `\bar` is not read.

```

1 [12] \input{tout}
2 {\input}
3 {Push argument 1}
4 Character sequence: tout.
5 {Text:tout}
6 {Pop 1: document_v argument_v}
7 {\input tout}
8 ++ Opened file tout.tex; it has 3 lines
9 [1] \def \foo {\endinput }
10 {\def}

```

```

11 {\def \foo ->\endinput }
12 [2] \foo
13 \foo ->\endinput
14 {\endinput}
15 ++ End of file tout.tex

```

- `\IfFileExists`. This command takes three arguments, say A , B , and C . The quantity A must be a character string after translation. If there is a file with name A , the test is true. Otherwise, if there is a file with name ' $A.tex$ ', the test is true. Otherwise, it is false. If the test is true, the expansion is ' B ', otherwise ' C '.
- `\InputIfFileExists`. This command takes one argument, say A . It does the same test as the previous command. If the test is false, nothing is done, otherwise, the file is input.
- `\include`. This command takes one argument, otherwise behaves like `\input`.
- `\input`. This command reads a filename as in `\openin` or `\openout`. It is redefined by \LaTeX to accept an argument. '`\input foo\input{bar}`' is valid. In the first case, tokens are read and expanded until find either a space, a non-expandable token, or `\input`.
- `\endinput`. This command has as effect to close the current file (in \Tralics , the file is read once and for all, closing the file means ignoring all unprocessed files. Here is an example. This assumes that there are no files `nohope`, `nohope.tex` neither `X.tex`. but a file `taux2.tex` that contains a call to `\mytypeout`.

```

{
  \IfFileExists{nohope}{\errmessage{bad1}}{}
  \IfFileExists{\jobname}{\errmessage{bad2}}{}
  \IfFileExists{X.tex}{\errmessage{bad3}}{}
  \def\bad{\errmessage{BAD}}\let\ybad\bad
  \def\mytypeout#1{\def\bad{\xbad}}
  \def\foo{\ifx\bad\ybad\else\let\xbad\relax\fi\let\bad\ybad}
  \let\IIFE\InputIfFileExists
  \IIFE{taux2}\IIFE{nohope}\foo\IIFE{taux2}
  \bad
}

```

Assume now that `taux` contains the 5 following lines

```

% aux file for testing tralics
% this file contains nothing useful
\mytypeout{in fileiaux1.tex}
\inputiaux2
% the file should finish with a comment

```

and that `taux` contains the 5 following lines

```

% aux file for testing tralics
% this file contains nothing useful
\mytypeout{in fileiaux2.tex}
\endinput

```

The file should finish with a `\endinput`, but not on the last line.

The following lines should work:

```

\let\mytypeout\typeout
\ifnum \tracingcommands=0 \def\mytypeout{\write -1 }\fi
\inputiaux1.tex\inputiaux2.tex
\input{iaux1.tex}\input{iaux2.tex}

```

```

\def\foo{\input taux1 \input taux2 }
\foo

```

5.9 Other non-expandable commands

- `\ignorespaces`. Tokens are full expanded, until a non-space character is seen. Spaces are ignored.
- `\xspace`. This command reads a token. It may insert a space unless a) the token is a character of category code 12, and one of `., !? : ; / ' -`, b) the token is an open brace, a close brace, or c) a space or `\`.
- `\mark`. A token list is read, the result is discarded.
- `\topmark`, `\firstmark`, `\botmark`, `\splitfirstmark`, `\splitbotmark`. The expansion is empty.
- `\penalty`. An integer is read, but nothing happens.
- `\lastkern`, `\lastpenalty`, `\lastskip`. Since Tralics not insert kerns, penalties, or glue in the boxes, it is not possible to fetch the last such quantity. In T_EX you can get an error of the form *You can't use '\lastpenalty' in vertical mode*. In Tralics the message is: *Read only variable \lastpenalty*. You can use `\the` to access the variable, but the result is zero (depending on the command, this is an integer, a dimension, or glue).
- `\unpenalty`, `\unkern`, `\unskip`. These are assumed to remove the last penalty, kern or glue. They are defined as a no-op.
- `\shipout`. This reads a box and signals an error *\shipout is undefined*.
- `\setlanguage`. This command reads an integer and does nothing.
- `\@@end`. This command can be used to stop translation. It closes all files. The end-document hooks are not applied.
- `\message`. This is like `\write-1`. The argument is expanded and printed.
- `\typeout`. This is like `\write17`. The argument is expanded and printed on the terminal and the transcript file.
- `\errmessage`. This is as above, but an error is signaled. For instance `\errmessage{This \iftrue can \else cannot\fi happen}` will give: *Error signaled at line 1: This can happen*.
- `\insert{arg}`. Unimplemented, an error is signaled.
- `\vadjust`. Unimplemented, an error is signaled.
- `\unhbox`, `\unvbox`, `\unhcopy`, `\unvcopy`. These commands read a box number, say *N*. The `\unhbox` command calls `\leavevmode` (in L^AT_EX, `\leavevmode` uses `\unhcopy`). A copy of the box register *N* is put on the main tree. In the case of `\unhbox` and `\unvbox`, the box is cleared.
- `\centering`. This command changes the current value of the internal centering number, and adds an attribute to the current paragraph (the numbers mean normal, centered, flush left, flush right, quoted).

- `\nocoming`. This sets the internal centering number to ‘normal’. The next paragraph will not have a centering attribute.
- `\href`. The command takes two arguments, say ‘foo’ and ‘bar’. The translation is an `<xref>` element, containing `\bar`, with an attribute `url`, with value `foo`.
- `\htmladdnormallink`, `\Hhref`. Like `\href`, but the order of arguments is changed.
- `\url`. This takes one argument. Normally `\url{foo}` is the same as `\href{foo}{foo}`. However, the element is translated using the value of `\urlfont`, which is empty by default. In a construction like `\href{foo}{\url{bar}}`, the `\url` command does nothing, as well as in the case `\url{\rrrt{foo}}`.
- `\rrrt`. This command takes one argument. Its translation is `\url{prefix/foo.html}`, if the argument is `foo`, where the prefix is ‘http://www.inria.fr/rrrt’. This is useful for the Raweb.
- `\caption`. This command takes an optional argument and a required argument. The optional argument is ignored. The result is a `<caption>` element. It can contain paragraphs. An implicit `\nocoming` command is executed.
- `\footnote`. This command takes one argument. The result is a `<note>` element. It can contain paragraphs. An implicit `\nocoming` command is executed. If the result contains a single `<p>` element, the `<p>` will be replaced by its content (unless in hack mode).
- `\label`. This command takes an argument, and defines it as the name of an anchor to be used by `\ref`. The anchor is an `id` attribute of the element associated to a section command, a footnote, an item in an enumeration, a formula, a figure etc. The translation is empty.
- `\eqref`. Like `\ref`, but parentheses are put around the reference.
- `\ref`. This takes an argument, that must be defined by `\label`. The translation is a `<ref>` element.
- `\part`, `\chapter`, `\section`, `\subsection`, `\subsubsection`, `\paragraph`, `\subparagraph`. To each command is assigned an integer between 0 and 6. It is decremented by 1 if the leading section is ‘chapter’, by 2 if it is ‘section’ (negative numbers are replaced by 0). If this gives N , the result is a `<divN>` element. The command takes an argument, the title, that is translated in a `<head>` element. Before the title, an optional star and an optional argument is ignored. The title is printed in the terminal if $N = 0$ (and unless the ‘silent’ switch has been given, if $N = 1$). A section can be the target of a `\label`.
- `\endsec`. This command takes an argument. Nothing happens if it is empty. If the first argument is one of the 7 section commands described above, it will be terminated; more precisely, if this section command has number N , then all divisions with number N and greater will be terminated. Example:

```

\section{A}\subsection{B}
\paragraph{C} x \label{a1}
\paragraph{D} \label{a2}x\label{a3}
\endsec{\subsection}
\paragraph{E} x\footnote{x\label{a4}}
\endsec{\part}
\paragraph{F} x
\ref{a1} \ref{a2}\ref{a3}\ref{a4}

```

This translates as

```

<Section id='uid1'><Head>A</Head>
  <Subsection id='uid2'><Head>B</Head>
    <Paragraph id='uid3'><Head>C</Head>
      <p>x</p>
    </Paragraph>
    <Paragraph id='uid4'><Head>D</Head>
      <p>x</p>
    </Paragraph>
  </Subsection>
  <Paragraph id='uid5'><Head>E</Head>
    <p>x<Footnote id='uid6' Place='as a note'><p>x</p>
      </Footnote></p>
  </Paragraph>
</Section>
<Paragraph id='uid7'><Head>F</Head>
  <p>x <ref target='uid3' />
    <ref target='uid4' /><ref target='uid4' /><ref target='uid6' />
  </p>
</Paragraph>

```

- `\aparaître`, `\toappear`. These two commands are for the Raweb. They translate as ‘à paraître’ (with the correct spelling) or ‘to appear’, depending on the current language.
- `\xmllatex`. This is a command that takes two arguments, and ignores the second. In the first, commands are replaced by their names. For instance `\xmllatex{\&\#x25;}{etc}` translates as `%`. Normally, the translation of `\&` is `&`.
- `\xmlelt`. This command takes two arguments, say `foo` and `bar`. The translation is a `<foo>` element containing the translation of `bar`. The command calls `\leavevmode`.
- `\xmleptyelt`. This command takes one argument, say `foo`. The translation is an empty `<foo>` element. The command does not call `\leavevmode`.
- `\xbox`. This command takes two arguments, say `foo` and `bar`. The translation is a `<foo>` element containing the translation of `bar`. This command behaves like `\hbox`, in that it does not read the second argument, so that category code changes are allowed. The token list `\everyxbox` is inserted in the input stream after the first argument and opening brace of the second have been read.
- `\vbox`, `\hbox`. These commands read their argument as described in the T_EX book; the result is an unnamed XML element. The token list `\everyvbox` or `\everyhbox` is inserted in the input stream after the opening brace has been read.
- `\leaders`, `\cleaders`, `\xleaders`. These commands read a box, then signal an error *leaders are not implemented*.
- `\moveleft`, `\moveright`. These commands read a dimension, and a box; but the box is not moved. Compare with `\raise` and `\lower`.
- `\vtop`. This is the same as `\vbox`.
- `\vsplit`. You can say `\vsplit0 to 3cm`. The result is the same as `\copy0`.
- `\lastbox`. This returns the last element on the stack, unless the top-stack is a character, case where the result is the empty box. Example:

```
\def\dupbox{\setbox0=\lastbox \copy0 \box0 }
\abox{foo}{ok} and \dupbox; \abox{bar}{ok}\dupbox.
```

The translation is:

```
<foo>ok</foo><p>and ; <bar>ok</bar><bar>ok</bar>.
```

- `\begin{rawxml}... \end{rawxml}`. This is like a verbatim environment, but special characters like `&` and `<` are not escaped. For instance

```
let\verbatimfont\relax
\begin{verbatim}
{\let\rm\bf \bf toto}
<!--this is a comment -->
&dollar; not &equals; EURO &excl;
\end{verbatim}
%
\begin{rawxml}
{\let\rm\bf \bf toto}
<!--this is a comment -->
&dollar; not &equals; EURO &excl;
\end{rawxml}
```

translates (with the ‘nozerowidthspace’ switch) to

```
<p Noindent='true'>{\let\rm\bf&nbsp;\bf&nbsp;toto}</p>
<p Noindent='true'>&lt;!--this&nbsp;is&nbsp;a&nbsp;comment&nbsp;--&gt;</p>
<p Noindent='true'>&amp;dollar;&nbsp;not&nbsp;&equals;
&nbsp;EURO&nbsp;&amp;excl;</p>
<p Noindent='true'>{\let\rm\bf \bf toto}
<!--this is a comment -->
&dollar; not &equals; EURO &excl;</p>
```

- `\begin{xmlelement} \end{xmlelement}`. Like `\abox`, but this is an environment. You can put a star after the name of the environment. The effect is the following. If no star is used, then horizontal mode is entered before creating the element. Otherwise, the element just created is typeset in vertical mode. If a ‘+’ sign is used instead of ‘*’, then `\leavevmode` is called, translation starts in vertical mode. For instance

```
\begin{xmlelement}{foo1} A \end{xmlelement}\par
\begin{xmlelement*}{foo2} B \end{xmlelement*}\par
\begin{xmlelement+}{foo3} C \end{xmlelement+}\par
```

translates as

```
<p> <foo1> A </foo1> </p>
<foo2><p>B </p> </foo2>
<p> <foo3><p>C </p></foo3></p>
```

- `\AddAttToCurrent`. This command takes two arguments, and constructs an attribute pair that will be added to the current element.
- `\AddAttToLast`. This command takes two arguments, and constructs an attribute pair that will be added to the last created element.
- `\addattributestodocument`. This command takes two arguments, and constructs an attribute pair that will be added to the main document element. This is an example:

```

\begin{xmlelement}{foo}
\begin{xmlelement}{subelt1}
texte1
\end{xmlelement}
\begin{xmlelement}{subelt2}
texte2
\end{xmlelement}
\AddAttToLast{sb2-att}{value1}%
\AddAttToLast{sb3-att}{}%
\AddAttToCurrent{foo-att}{att-value''}%
\end{xmlelement}

```

The translation is

```

<foo foo-att='att-value&apos;&apos;''><subelt1>texte1
</subelt1>
<subelt2 sb3-att='' sb2-att='value1'>texte2
</subelt2>
</foo>

```

The commands described above can be followed by a '*'. In that case, an existing attribute will be overwritten. Example:

```

\AddAttToCurrent{x}{1} % ignored
\AddAttToCurrent{y}{2} % ok
\AddAttToCurrent*{x}{3} % ok
\AddAttToCurrent{y}{4} % ignored

```

- `\XMLaddatt`. This command takes an optional star, an optional integer, and two arguments, and constructs an attribute pair that will be added to the element defined by the number. If no number is given, the current element is used. The following example shows usage.

```

{\everyxbox{Vest}
\setbox0=\xbox{foo}{1\xbox{bar}{2} %
  \XMLaddatt[\the\XMLlastid]{x}{1}%
  \XMLaddatt[\the\XMLcurrentid]{y}{2}%
  \XMLaddatt{y}{2}3}
\showbox0
}
{\everyxbox{West}
\setbox0=\xbox{foo}{1\xbox{bar}{2} %
  \XMLaddatt*[\the\XMLlastid]{x}{1}%
  \XMLaddatt*[\the\XMLcurrentid]{y}{2}%
  \XMLaddatt*{y}{4}3}
\showbox0
}

```

should print

```

<foo y='2'>Vest1<bar x='1'>Vest2</bar> 3</foo>
<foo y='4'>West1<bar x='1'>West2</bar> 3</foo>

```

- `\begin{itemize}` `\begin{description}` `\begin{enumerate}` `\begin{glossary}` ...
`\end{glossary}` `\end{enumerate}` `\end{description}` `\end{itemize}`.
 These commands behave as in L^AT_EX, but do not depend on quantities like 'enumiv' or `\leftmarginvi`. They can be arbitrarily nested.

- `\item`. This command takes an optional argument, that will produce a `<label>` element. What follows `\item` is translated in a `<item>` element. The end of the element is defined by a following item or the end of the list (thus, you should use this command only in a list).
- `\lhead`, `\rhead`, `\chead`, `\cfoot`, `\lfoot`, `\rfoot`. These commands are available only if the `fancyhdr` package is loaded. They take an optional argument, say A, and a required argument, say B. The default value of A is B. The result is the same as `\@IF{elh}{A}` followed by `\@IF{olh}{B}`, where ‘lh’ must be replaced by the first two characters of the command.
- `\fancyhead`, `\fancyfoot`, `\fancyhf`. These commands are available only if the `fancyhdr` package is loaded. They take two arguments, an optional one x, and a required argument y. The argument x is a sequence of letters, separated by commas. In each sequence, only letters e, o, l, c, r, h, and f are considered. If none of e or o are given, both are added; if none of l, c, r and given, all three are added; for `\fancyhead`, h is added, for `\fancyfoot`, f is added; if none of h or f are given, both are added. After that we consider all sequences XYZ, where X is one of e, o, Y is one of l, c, r, Z is one of h, f, the letters being in the list so completed (at least one sequence matches). For each match we call `\@IF{XYZ}{y}`.
- `\@IF`. This is an internal command, that takes two arguments, say x and y. The translation is `\fancyinternal {x} {\let \thepage \inert@thepage y}`.
- `\inert@thepage`. The translation is `<thepage/>`.
- `\fancyplain`. Not yet implemented. It takes two arguments, that are conditionally evaluated, depending on whether this is a “plain” page or not.
- `\fancyinternal`. The command takes two arguments x and y. It constructs a `<headings>` element, containing y, with an attribute `type` with value x. Example:

```

\newcommand\rightmark{\xmlenptyelt{rightmark}}
\newcommand\leftmark{\xmlenptyelt{leftmark}}
\newcommand\fancyplain[2]{#2}
\lhead[\fancyplain{}{\bf\thepage}]{\fancyplain{}{\sl\rightmark}}
\rhead[\fancyplain{}{\sl\leftmark}]{\fancyplain{}{\bf\thepage}}
\cfoot[\fancyplain{\bf\thepage}{}]{\fancyplain{\bf\thepage}{}}
\chead{}\lfoot{} \rfoot{}

```

The translation is

```

<headings type='olh'><hi rend='slanted'><rightmark/></hi></headings>
<headings type='elh'><hi rend='bold'><thepage/></hi></headings>
<headings type='orh'><hi rend='bold'><thepage/></hi></headings>
<headings type='erh'><hi rend='slanted'><leftmark/></hi></headings>
<headings type='ocf' />
<headings type='ecf' />
<headings type='och' />
<headings type='ech' />
<headings type='olf' />
<headings type='elf' />
<headings type='orf' />
<headings type='erf' />

```

- `\@ifundefined`. The command takes three arguments. The first argument should be a sequence of letters, for instance ‘foo’. If `\foo` is undefined, the expansion of the command is the value of the first argument, otherwise, the value of the second. Note: `\csname` is used to construct the command, so that ‘undefined’ is the same as being `\relax`. If the first argument

is `\par`, an error is signaled, of the form *Invalid token \par found while scanning command for \@ifundefined*. If the argument is `\xpar`, the error is *Undefined command \xpar*, and `\relax` gives: *Invalid token \relax found while scanning command for \@ifundefined*. Note that Tralics may complain with an extra `\endcsname`.

As an example, this should give 2.

```
\makeatletter
\def\Foo{foo}\def\Bar{bar}
\@ifundefined{Foo}{\@ifundefined{Bar}{1}{2}}{3}
```

- `\epsfbox`, `\epsffile`. These command do exactly the same thing. They take one argument, say X. It behaves as `\includegraphics[Y]{X}Z`, where Y contains ‘width=w’ and ‘height=h’, where w and h are the values of `\epsfxsize` and `\epsfysize`, provided they are not zero, and Z is the code that sets these values to zero.
- `\psfig`, `\epsfig`. The command takes an argument, which is an association list. It must contain `file=X`, or `figure=X`. If Y is the remaining of the list, the command is equivalent to `\includegraphics[Y]{X}`.
- `\includegraphics`. The command takes an optional argument and a required argument. The optional argument is an association list, with names can be clip, angle, height, width, scale. The required argument is a file name. There is no difference between ‘foo’, ‘foo.ps’ and ‘foo.eps’. Not more than one dot is allowed in the file name. See examples section 6.14.
- `\lsc`, `\fsc`. These commands are like `\textsc`.
- `\@ifstar` The command takes two arguments. After that, it looks at the next character (ignoring spaces). If the character is a star, it is read, the expansion is the first argument, otherwise the second. The following code produces B1A2.

```
\makeatletter
\def\foo{\@ifstar{A}{B}}
\foo1\foo*2
```

- `\@ifnextchar`. The command takes three arguments, a character token, say w, and two commands say A and B. After that, it looks at the next character (ignoring spaces). If the character is w, the expansion is A, else B. The character is not read. The following should produce: ‘With opt AABBBwithout opt CC’.

```
\makeatletter
\def\ifbracket{\@ifnextchar[}
\makeatother
\def\wopt{\ifbracket\xwopt\ywopt}
\def\xwopt[#1]#2{With opt #1#2}
\def\ywopt#1{Without opt #1}
\wopt[AA]{BB}\wopt{CC}
```

- `\batchmode`, `\nonstopmode`, `\scrollmode`, `\errorstopmode`. These commands do nothing in Tralics, because the only supported mode is ‘batchmode’. In T_EX, they change an interval variable, so that they can be preceded by the prefix `\global`.
- `\@gobble`, `\@gobbletwo` These commands read one or two argument. The result is empty.
- `\discretionary`. This command reads two arguments but does nothing with it. In T_EX, there is a third argument, that is typeset.
- `\HTMLset` (Raweb only command). This command reads two arguments but does nothing with them.

- `\RALabel` (Raweb only command). This command reads one argument but does nothing with it.
- `\special`, `\patterns`, `\hyphenation`. These command read one argument but do nothing with it.
- `\protect`, `\sloppy`, `\subitem`, `\htmlimage`, `\centerline`, `\immediate`, `\break`, `\nobreak`, `\allowbreak`, `\@`, `\/` `\clearpage`, `\cleardoublepage`, `\allowbreak`, These command are currently defined as doing nothing.
- `\begin{latexonly} ... \end{latexonly}`, `\begin{xmlonly} ... \end{xmlonly}`, `\begin{subequations} ... \end{subequations}`. The environment is ignored, not the content.
- `\begin{htmlonly} ... \end{htmlonly}`, `\begin{comment} ... \end{comment}`, `\begin{rawhtml} ... \end{rawhtml}`, `\begin{LaTeXonly} ... \end{LaTeXonly}`, The environment is ignored, the content also.
- `\centering`, `\raggedright`, `\raggedleft`. These commands are associated to the environments: ‘center’, ‘flushleft’, ‘flushright’. There are two other environments ‘quote’ and ‘quotation’. This gives an integer with values between 1 and 5, the normal value being 0. Whenever a new paragraph is started, the value of the integer considered. It is non-zero, an attribute is added (for instance `centerend='center'`). If you say `\begin{center}`, this will terminate the current paragraph (and the next one will be centered). If you say `\end{center}`, this will terminate the current paragraph (and the next one will not be centered). If you say `\centering`, this will add an attribute to the current paragraph.
- `\error`. This command takes an argument and signals an error.
- `\@notprerr`. After `\let\foo\@notprerr`, the command `\foo` will produce the *Can be used only in preamble: \foo error*.
- `\hrule`, `\vrule`. These commands are implemented identically in Tralics. They produce a `<rule>` element. For a `hrule`, the default height is 0.4pt, the default depth is 0pt, for a `vrule` the default width is 0.4pt. You can give the height, depth, width arguments in any order, more than once. For instance


```

\hrule \hrule width 15pt height 2pt
\vrule \vrule height 2pt depth 4pt

```

 translates as


```

<rule depth='0.0pt' height='0.4pt' />
<rule width='15.0pt' depth='0.0pt' height='2.0pt' />
<rule width='0.4pt' />
<rule width='0.4pt' depth='4.0pt' height='2.0pt' />

```
- `\ding`. You can use the `\ding` command in order to access one character of the Zapf Dingbats fonts. The command takes one argument, an integer (it is parsed by the `scanint` routine, `\ding{'\A}` is as valid as `\char'\A`). If the number is between 33 and 126, or between 161 and 239, or between 241 and 254, then you get one of the characters listed on the Tralics Web page. Otherwise, the translation is bad!
- `\mkern`. There is currently no difference between `\kern` and `\mkern`.

Chapter 6

Running Tralics

6.1 The command line arguments

If you call Tralics without arguments, you will see something like

```
This is tralics 2.5 (p15), a LaTeX to XML translator
Copyright INRIA/MIAOU/APICS 2002-2005, Jos'e Grimm
Licensed under the CeCILL Free Software Licensing Agreement
Say tralics --help to get some help
```

In any case, the first three lines are printed. Arguments are read, and interpreted as follows.

- Anything that does not start with a dash is a source file. At most one source file can be given.
- An option starts with a single or double dash; this is irrelevant.
- If the option is ‘version’, the program stops after printing the banner.
- If the option is ‘?’ or ‘help’ the program stops after printing the list of all possible switches.
- If the option is ‘verbose’, the program is rather verbose. The equivalent of the command `\tracingall` is executed.
- If the option is ‘silent’, the program is silent (less lines are printed on the terminal).
- If the option is ‘utf8’, then sources files are assumed to be UTF-8 encoded, unless the first line of a file contains ‘iso-8859-1’.
- If the option is ‘latin1’, then sources files are assume to be latin1 encoded, unless the first line of a file contains ‘utf8-encoded’ (if ‘utf8’ and ‘latin1’ are given, the last option wins, if none is given latin1 is assumed).
- If the option is ‘entnames’, then some commands as `\nobreakspace` translate as ‘ ’. Note that `\texteuro` always translates as ‘€’.
- If the option is ‘noentnames’, then some commands as `\nobreakspace` translate as ‘ ’. Note that `\alpha` always translates as ‘α’.
- If the option is ‘notrivialmath’, then even trivial math formulas such as $\$0\$$ are translated as a `<formula>` element.
- If the option is ‘nozerowidthspace’, then no special marker is inserted in a case like `\verb+---+`.

- If the option is ‘shell-escape’, then `\write18{rm \jobname.tex}` removes the file you are translating. The dash in the name is optional.
- If the option is ‘noxmlerror’, then an error will not produce an `<error>` element in the XML tree.
- If the option is ‘noundefmac’, then no error is signaled when you use an undefined command like `\foo`. Instead, the command is automatically defined to be `\foo`.
- If the option is ‘dir=XXX’, then XXX will be the “raweb dir”.
- If the option is ‘type=XXX’, then XXX will be the default type, as explained later.
- If the option is ‘noconfig’, then no configuration file will be used.
- If the option is ‘config_file=XXX’, then XXX will be used as configuration file (the underscore in the option name is optional).
- If the option is ‘find_words’, then the list of all words will be printed on the file `words`. The underscore in the name is optional.
- If the option is ‘interactivemath’, then no input file is required, expressions are read from the terminal. The translation of every math expression is printed on the terminal. The value of `\jobname` will be ‘texput’ (so that the XML result is in `texput.xml`, the transcript file is `texput.log`).
- If the option is ‘interactivebib’ is given, then a special mode is entered. It is not described here.
- If the option is ‘year=XXX’ (Raweb only), then XXX will be the default year (year 2005 starts at May, 1st, because people start writing the Raweb2004 in September 2004, and everything should be finished by March 2005).
- If the option is ‘debug’ (Raweb only), then parsing of the Raweb source continues after the first error.
- If the option is ‘check’ (Raweb only), then the document is tested for the Raweb syntax, not translated.
- If the option is ‘ps’ (Raweb only), then \LaTeX is called for converting the document into dvi then PostScript format.
- If the option is ‘xml’ (Raweb only), then Tralics is called for converting the document into XML.
- If the option is ‘xmlfo’ (Raweb only), then the XML should be translated in XSL/Format.
- If the option is ‘xmlltex’ (Raweb only), then the XML should be translated in XSL/Format, that should be converted to PostScript.
- If the option is ‘xmlhtml’ (Raweb only), then the XML should be translated in HTML.
- If the option is ‘xmllint’ (Raweb only), then an external process is called that checks the validity of the translation against the Raweb DTD.
- If the option is ‘xmlall’ or ‘all’ (Raweb only), then all options ‘xmlXXX’ are selected.
- If the option is ‘external_prog=XXX’ (Raweb only), then XXX will be used instead of ‘rahandler.pl’ as interpreter for the raweb actions defined by the xmlXXX switches. The underscore in the name is optional.

- If the option is ‘nobibyearerror’ (Raweb only), then no error is signaled for a bibliography reference with missing or wrong year.

Example. Assume that we have a file, named `xii.tex`, containing

```
\let~\catcode~'76~'A13~'F1~'j00~'P2jdefA71F~'7113jdefPALLF
PA'~FwPA;;FPAZZFLaLPA//71F71iPAHHFLPAzzFenPASSFthP;A$$FevP
A@@FfPARR717273F737271P;ADDFRgniPAWW71FPATTFvePA**FstRsamP
AGGFRruoPAqq71.72.F717271PAY7172F727171PA??Fi*LmPA&&71jfi
Fjfi71PAVVFjbigskipRPWGAUU71727374 75,76Fjpar71727375Djifx
:76jelse&U76jfiPLAKK7172F7117271PAXX71FVln0SeL71SLRyadR@oL
RrhC?yLRurtKFeLPFovPgaTLtReRomL;PABB71 72,73:Fjif.73.jelse
B73:jfiXF71PU71 72,73:PWs;AMM71F71diPAJJFRdriPAQQFRsreLPAI
I71Fo71dPA!!FRgiePBt'el@ 1TLqdrYmu.Q.,Ke;vz vzLqipip.Q.,tz;
;Lql.IrsZ.eap,qn.i. i.eLlMaesLdRcna,;!;h htLqm.MRasZ.ilK,%
s$;z zLqs'.ansZ.Ymi,/sx ;LYegseZRyal,@i;@ TLRlogdLrDsW,@;G
LcYlaDLbJsW,SWXJW ree @rzchLhzsW,;WERcesInW qt.'oL.Rtrul;e
doTsW,Wk;Rri@stW aHAHHFndZPpqr.tridgeLinZpe.LtYer.W,:jbye
```

If you call `Tralics`, with the option ‘`find_words`’, you can see that the XML file contains once drumming and drummers, twice piping and pipers, 3 times leaping and lords, 4 times dancing and ladies, 5 times milking and maids, 6 times swimming and swans, 7 times laying and geese, 8 times rings and gold, 9 times calling and birds, 10 times hens and french, 11 times doves, turtle and ‘and’, 12 times tree, pear, in, partridge, me, to, gave, love, true, my, christmas, of, day, the, on. There are 45 words with a single letter. The words: twelve, eleven, ten, nine, eight, seven, six, five, four, three, two, appear x times, where $13 - x$ is the value of the word. The words first, second, third, fourth, fifth, sixth, seventh, eighth, ninth, tenth, eleventh, twelfth appear once. Amazing isn’t it? The file was written by D. Carlisle, it is available on the CTAN. This is not really a L^AT_EX file, so that some features cannot be applied (for instance, there is no at-begin-document hook). The `\bye` command is implemented in `Tralics` for this example to compile without error.

6.2 The standard configuration file

We give here the content of the standard configuration file. Lines 2 and 3 were modified: we added the letter ‘x’ after the dollar sign, for otherwise RCS would replace the identification of the original file by the identification of the L^AT_EX file. The content of the file will be explained later. What you have to know is that some lines start with ‘Begin’, and others with ‘End’. To each Begin, there should be an associated End. Blocks can be nested. Characters after ‘End’ are ignored, so that you can say ‘BeginFoo’ followed by ‘EndBar’, although it is not recommended. All other lines should be comment lines, empty, or indented.

```
1 # This is a configuration file for tralics.
2 # $xId: tralics_rc,v 2.15 2005/08/02 09:22:56 grimm Exp $
3 ## tralics ident rc=standard $xRevision: 2.15 $
4
5 # Copyright Inria. José Grimm. Apics. 2004/2005
6 # This file is part of Tralics.
7 # (See the file COPYING in the Tralics main directory)
8 # If you modify this file, by changing the semantics of type RA,
9 # please remove the ‘standard’ on the ‘tralics ident’ line above,
10 # or replace it by ‘non-standard’.
11
12 % Some comments: comments start with % or #
```

A configuration file is split into main sections, one for each type. We start with the RA, or raweb.

```

13 ## configuration for the RA (Inria's Activity Report)
14 ## First type defined is the default
15
16 BeginType RA      % Case RA
17   DocType = raweb raweb3.dtd
18   DocAttrib = year \specialyear
19   DocAttrib = creator \tralics

```

This comment explains how to parameterize some element or attribute names that were built-in in a previous version of Tralics. We shall see later how `Language` can be used (default value is 'language'), the same for `lang_en` and `lang_fr` that have 'english' and 'french' as default value. Translation of a `\caption` produces an element `<caption>`, whose name will be changed to `<head>` by the post-processor of figures (it will be left unchanged if the caption is not in a figure or a table). The variable `xml_caption_name` can be used to change the first name, and `xml_scaption_name` can be used to change the second name. The title of a 'topics' (defined by `\declaretopics`) is in a `<t_title>` element, the name can be changed by `xml_topic_title`. A reference to a topic uses the `num` attribute; this attribute name can be changed by `att_topic_num`. The identification of an Inria Team is in `<accueil>`, this can be changed via `xml_accueil_name`. It is formed of a long name in `<projetdeveloppe>` and a short name in `<projet>`, the name of these elements can be changed via `xml_project_name` or the 'expanded' version. The section with the composition of the team is `<composition>`, its name can be changed by `xml_composition_ra_name`.

```

20   #(new)
21   # Language = "xml:lang"
22   # lang_en = "en"
23   # xml_scaption_name= "caption"
24   # xml_topic_title=""
25   # xml_project_name = "title"
26   # xml_expanded_project_name = "longtitle"
27   # xml_accueil_name = "identification"
28   # xml_composition_ra_name = "team"
29   # att_topic_num = "id"

```

Processing of the Raweb needs converting the XML output of Tralics into XSL/Format, HTML, etc., via some external commands like 'xsltproc', 'latex', etc. Details are given on the Web page. Since 2004, processing is done by an external program, so that you should look at `rahandler.pl` for explanations of how %A, %B, and the like are converted to file names.

```

30   makefo="xsltproc --catalogs -o %A.fo %B %C";
31   makehtml = "xsltproc --catalogs %B %C";
32   call_lint = "xmllint --catalogs --valid --noout %C"
33   makepdf = "pdflatex -interaction=nonstopmode %w"
34   %makedvi = "latex -interaction=nonstopmode %w"
35   % makedvi et dvips pour marie-pierre
36   %dvitops = "dvips %w.dvi -o %w.ps"
37   %makedvi = "latex -interaction=nonstopmode %w"
38   generatedvi = "latex -interaction=nonstopmode %t"
39   % old latex: "latex \nonstopmode\input{%t}"
40   generateps = "dvips %t.dvi -o %t.ps"

```

This defines the list of valid Raweb sections, themes and URs (research units). If you change these lines please: a) remove the 'standard' on line 3, or b) make sure that it matches the official

list, or c) make sure that this remains a private copy. A star after a section name says that topics are not allowed¹.

```

41 #these are new in version 2.0
42   theme_vals = "com cog num sym bio"
43   section_vals = "composition*/presentation*/fondements/domaine/logiciels/"
44   section_vals = "+resultats/contrats*/international*/diffusion*/"
45   ur_vals = "Rocquencourt//Sophia/Sophia Antipolis/Rennes//Lorraine//";
46   ur_vals = "+RhôneAlpes/Rhône-Alpes/Futurs//"

```

Due to some inertia, people continue using the obsolete environment. We make sure an error is signaled.

```

47 BeginCommands
48   \newenvironment{body}{\obsoleteEnvBody The body environment is %
49     obsolete since 2003}
50   {End of obsolete environment body}
51   \newenvironment{abstract}{\obsoleteEnvAbstract The abstract %
52     environment is obsolete since 2003}
53   {End of obsolete environment abstract}
54 EndCommands
55
56 End

```

This is an example of titlepage environment; it will be discussed later.

```

57 ## configuration for the RR (Research Report of Inria)
58 ## not yet complete
59
60 BeginType RR#      Case RR
61   DocType = rr raweb.dtd # ok ?
62 BeginTitlePage
63   \makeRR <RRstart> "" "type = 'RR'"
64   alias \makeRT "" "type = 'RT'"
65
66   <UR> -
67   \URSophia ?+<UR>
68   \URRocquencourt ?+<UR>
69   alias \URRocq
70   \Paris ?<UR> <Rocquencourt>
71   \URRhôneAlpes ?+<UR>
72   \URRennes ?+<UR>
73   \URLorraine ?+<UR>
74   \URFuturs ?+<UR>
75
76   \RRtitle <title> "pas de titre"
77   \RRetitle <etitle> "no title"
78   \RRprojet <projet> "pas de projet"
79   \RRtheme <theme> "pas de theme"
80   \motcle <motcle> "pas de motcle"
81   \keyword <keyword> "no keywords"
82   \RRresume p<resume> "pas de resume"
83   \RRabstract p<abstract> "no abstract"
84   \RRauthor + <author> <auth> "Pas d'auteurs"

```

¹This is new in version 2.5 (p17)

```

85  \RRdate <date> "\monthyearvalfr"
86  \RRNo <RRnumber> "???"
87
88  End
89  EndType
    A short definition for standard classes.
90  BeginType std#      standard latex classes
91  DocType = std classes.dtd
92  xml_biblio = "bibliography"
93  End
    Some aliases.
94  # (types Article and slides are not defined, hence this is useless)
95
96  BeginAlias
97  Article report
98  slides inriaslides foiltex
99  End
    This command has to be outside any block.
100 % this means: take the documentclass value as type name
101 Type = \documentclass
    More aliases. Note that toto matches RR (first in list) and report matches std (because 'unknown' is undefined).
105 BeginAlias
106 RR toto# ra2001
107 RA ra toto ra2001x%/etc
108 torture torture1 torture2
109 unknown report
110 std report book article minimal
111 End
    For fun.
112 ## an empty type
113 BeginType MP
114 EndType
    This is used for testing Tralics.
115 BeginType torture
116 DocAttrib = creator \tralics
117 DocType = remain raweb.dtd
118 on package loaded calc CALC = "true"
119 on package loaded foo/bar F001 = "true"
120 on package loaded *foo/bar F002 = "true"
121 on package loaded foo/*bar F003 = "true"
122 on package loaded *foo/*bar F004 = "true"
123 url_font = "\large "
124 no_footnote_hack = "false"
125 on class loaded calc CALC="true"
126
127 use_font_elt = "true"
128 xml_font_small = "font-small"
129 xml_font_small1 = "font-small1"

```

```

130 xml_font_small2 = "font-small2"
131 xml_font_small3 = "font-small3"
132 xml_font_small4 = "font-small4"
133 xml_font_large = "font-large"
134 xml_font_large1 = "font-large1"
135 xml_font_large2 = "font-large2"
136 xml_font_large3 = "font-large3"
137 xml_font_large4 = "font-large4"
138 xml_font_large5 = "font-large5"
139 xml_font_normalsize = "font-normalsize"
140 xml_font_upright = "font-upright-shape"
141 xml_font_medium = "font-medium-series"
142 xml_font_roman = "font-roman-family"
143 xml_font_it = "font-italic-shape"
144 xml_font_slanted = "font-slanted-shape"
145 xml_font_sc = "font-small-caps-shape"
146 xml_font_tt = "font-typewriter-family"
147 xml_font_sansserif = "font-sansserif-family"
148 xml_font_bold = "font-bold-series"
149
150 xml_sup_name = "font-super"
151 xml_sub_name = "font-sub"
152 xml_oldstyle_name = "font-oldstyle"
153 xml_overline_name = "font-overline"
154 xml_underline_name = "font-underline"
155
156 BeginCommands
157   % These commands are inserted verbatim in the file
158   \def\recurse{\recurse\recurse}
159 EndCommands
160 EndType
    This may be used for typesetting a bibliography, exactly like the Raweb.
161 BeginType rabib      % Case RA
162   DocType = raweb raweb3.dtd
163   DocAttrib = year \specialyear
164   DocAttrib = creator \tralics
165
166 BeginCommands
167   % These commands are inserted verbatim in the file
168   \newcommand\usebib[2]{\bibliography{#1#2,#1_foot#2+foot,#1_refer#2+refer}}
169 EndCommands
170 EndType

```

6.3 The action before translation

We have indicated above the list of all switches, some of them take an argument, and there must be a space or an equals sign between the switch and the argument (but no spaces in case of an equals sign). For instance you can say:

```
tralics -dir foo -year=2004
```

Some options are marked ‘Raweb only’; this means that there are not used, unless you are translating the Raweb. Unless you call Tralics with interactive math, or interactive bib, an input file name is required. This cannot be an absolute or relative pathname (on Unix, this means that the first character cannot be a dot or a slash). It must be the name of a T_EX file: an extension `tex` is added, so that `foo` and `foo.tex` are the same. If the input is `foo.xml`, you will see something like *Fatal error: Cannot open input file foo.xml.tex*. After the input file has been completely read, Tralics will print something like *Starting xml processing for xii*. Instead of ‘xml processing’ it could be ‘latex processing’, or ‘check’, case where you specify a raweb option that tells Tralics not to produce an XML file. Let’s ignore this special case. After that, Tralics reads the configuration file; an example is given above. If a line starts with `## tralics ident rc` (for instance, line 3), what follows the equals sign is printed. Note the hack: a space is printed after the dollar sign. You may see: *Configuration file identification: standard \$ Revision: 2.12 \$*. Note the space after the dollar sign; it is inserted by Tralics, for the same reason than we inserted a ‘x’ on line 3.

If the option ‘noconfig’ is given, then an empty configuration file is used. If the option ‘config-file=xxx’ has been used, then xxx is the name of the configuration file. In the current version of Tralics, the configuration file can be indicated in the source file, near the start, using a line of the form:

```
% Tralics configuration file 'foobar'
```

The first character on the line has to be a percent sign. Single quotes are used to delimit the file name. The first character in ‘Tralics’ can be a lower case T as well. If this fails, then the name of the configuration file is `.tralics_rc` (without the leading dot on Windows). It is looked up in the current directory first. Otherwise, in the directory specified by the ‘dir’ option; otherwise, in the home directory of the user; otherwise, a standard directory is looked at. If the shell variable `TRALICSDIR` is set, this is a good location. If `RAWEBDIR` is set, this is another good location. Otherwise, a default path is used. In all examples shown here, the current directory or the default path will be used (the default path works well for the author of the software, by definition).

Translation depends on the “type” of the document. This is computed as follows. First of all, the configuration file is looked at to see if there is a line defining ‘Type’. The standard configuration file has (line 101) `\documentclass`. This means that the value of the document class should be considered. In the case of the xii file, the following line is printed: *No \documentclass in source file*. Remember, it is not a L^AT_EX file. In the case of a very simple example like this

```
\documentclass{article}
\begin{document}
Hello, world!
\end{document}
```

the result would be *Seen \documentclass article*. Note the following hack: the `\documentclass` command can be inside a comment, provided that a single percent sign is used, not a double percent sign. The potential type is the following: if the command line option ‘type=foo’ is given, then the type is foo. If the configuration file defines a type it will be used. If no configuration file is found, or no type is found in the configuration file, or if this is `\documentclass`, then the value of the document class will be used, if any. In the case of xii, you will see *Using some default type*, in the case of the hello world, you will see *Potential type is article*.

The configuration file is analyzed, in order to see what types are defined. In the case of the default configuration file, Tralics will print: *Defined type: RA*, plus other lines with `RR`, `std`, `MP`, `torture`, and `rabib`. It happens that the potential type is none of these. Thus, the aliases are looked at. For line 97, you will see *Alias Article undefined*. This line as well as the following is useless (it’s just a test to show the alias mechanism). For line 106, you get *Alias RR does not match article*, because the line says that you can use ‘toto’ instead of ‘RR’. Six lines are printed in the transcript file. Line 110 matches, and you will see *Potential type article aliased to std*. If you call Tralics with option ‘-type=ra2004’ for this file, you will see an error message of the form: *Fatal error: Please*

add 2005 in the name of the input file. The main reason for this is that, if no type matches, and if the potential type ends with digits, these are removed, and the whole list is tried again. You will see something like *Trying type ra*, then *Potential type ra aliased to RA*, then *Using type RA*. If you use ‘ar’ instead of ‘ra’, this will match nothing, and you get: *Chosing first defined type RA*. In the case where there is no configuration file, you will see something like *Using some default type*.

If a type is found, then lines that do not concern this type are removed from the configuration file. Lines that are not in a block are evaluated. For instance, lines 116, 118, 123. An important line is 117: this defines the DocType. If it is missing, you will see some thing like *dtd is unknown from unknown.dtd (standard mode)* or, in the case of standard classes: *dtd is article from article.dtd (standard mode)*. You could also see: *dtd is raweb from raweb3.dtd (mode RAWEB2005)*. In fact, if the Type is ‘ra’ (lowercase, or uppercase) we are in ‘raweb mode’. In this case, the switches for Raweb mode apply. Moreover, the file name should be of the form ‘foo2005’ (any year can be used instead of 2005), and the document class should be ‘ra2005’ (same year). A ‘standard’ configuration file (see Copyright notice) must enter Raweb mode if the document class is ‘raxxxx’ and no ‘type=yy’ option has been given.

Assume that we call Tralics in verbose mode, with ‘type=rabib’. The transcript file will contain the following

```
dtd is raweb from raweb3.dtd (standard mode)
Ok with the config file, dealing with the TeX file...
There are 4 lines
Starting translation
[1] %% Begin bootstrap commands for latex
...
[66] %% End bootstrap commands for latex
[1] \documentclass{article}
{\documentclass}
{Push argument 1}
Character sequence: article.
{Text:article}
{Pop 1: document_v argument_v}
[183] % These commands are inserted verbatim in the file
[184] \newcommand\usebib[2]{\bibliography{#1#2,#1_foot#2+foot,#1_refer#2+refer
}}
{\newcommand}
{\newcommand \usebib}
[2] \begin{document}
...
[4] \end{document}
...
Output written on hello1.xml (218 bytes).
No error found.
(For more information, see transcript file hello1.log)
```

One interesting point here is that all lines that start with digits like ‘[123]’ are T_EX input lines that have to be translated. The document source has 4 lines. There are 66 lines (not shown here) marked ‘bootstrap commands’. They contain definitions that are written in L^AT_EX syntax, not C++. The two lines 183 and 184 come from the configuration file (with number 167 and 168 in this document). The interesting point is that they are inserted just *after* the line with the document class. As line 183 explains, the block named ‘Commands’ is copied verbatim from the configuration file to the T_EX source file (to be precise: the T_EX source file is read, once and for all, as a list of lines, before the configuration file is read, because it might contain the name of the configuration file; some lines are added to this list).

If the configuration file contains lines as 18 and 19, the effect is to add an attribute to the main element. The normal syntax is: `DocAttrib = foo "bar"`. Using a command name instead of a string means that the value of the command should be used. In effect, a line like the following is added, after the document-class line:

```
\AtBeginDocument{\addattributestodocument{Foo}{\World}}
```

On line 18, you see `\specialyear`: this is a hack; this is the value of the year as used by the Raweb (the current year, in general). On line 19, you see `\tralics`. This means that the current Tralics version should be used. Example: for this configuration file,

```
## tralics ident rc=helloconf.

BeginType article
DocType = Article classes.dtd
DocAttrib =Foo \World
DocAttrib =A \specialyear
DocAttrib =B \tralics
DocAttrib =C \today
BeginCommands
  \def\World{world}
  \def\today{\the\year/\the\month/\the\day}
End
End
```

the translation of the hello file is:

```
<?xml version='1.0' encoding='iso-8859-1'?>
<!DOCTYPE Article SYSTEM 'classes.dtd'>
<!-- translated from latex by tralics 2.5 (p17)-->
<Article Foo='world' C='2005/8/5' B='Tralics version 2.5 (p17)' A='2005'>
<p>Hello, world!
</p>
</Article>
```

6.4 Translating the Raweb

Raweb mode is entered if a configuration file is found that says that the type to use is 'RA' or 'ra'. The document class should be ra97, ra98, or, for later years, ra1999. The example has ra2003. This must match the name of the input file, which is miaou2003. The document can be translated in one of three versions: first, you may try latex, this gives miaou2003.dvi; then we have a mode in which miaou2003.tex is converted into miaou.tex, and latex can produce miaou.dvi. Finally, Tralics may produce miaou.xml, and this can be compiled into wmiaou.dvi. Historically, we had a Perl script for the conversion, this was extended to a translator, then re-written in C++; since the normal user has no more access to the source, we have added a means to parameterize the options of the 'latex' and other commands, see configuration file, lines 30 to 40. Since 2004, an external program is used, it is described on the web page of Tralics. What Tralics does is thus: convert miaou2003.tex to miaou.tex; if this file has to be processed by L^AT_EX, write it on disk, otherwise, translate it as miaou.xml. Then, relevant parameters are written to user_param.pl. Here is an example.

```
1 $:makefo='xsltproc --catalogs -o %A.fo %B %C';
2 $:makehtml='xsltproc --catalogs %B %C';
3 $:checkxml='xmllint --catalogs --valid --noout %C';
4 $:makepdf='pdflatex -interaction=nonstopmode %w';
5 $:makedvi='';
```

```

6 $::dvitops='';
7 $::generate_dvi='latex -interaction=nonstopmode %t';
8 $::generate_ps='dvips %t.dvi -o %t.ps';
9 $::raweb_dir='/user/grimm/home/cvs/raweb';
10 $::raweb_dir_src='/user/grimm/home/cvs/raweb/src/';
11 $::ra_year='2003';
12 $::no_year='miaou';
13 $::tex_file='miaou';
14 $::todo_fo=0;
15 $::todo_html=0;
16 $::todo_tex=0;
17 $::todo_lint=0;
18 $::todo_ps=0;
19 $::todo_xml=1;
20 1;

```

If some action has to be done (one of 'todoXXX' between lines 14 and 18 is 1), then 'rahandler.pl' is called. In our case, only 'xml' is set, but this action is already done. You can edit the file, and run 'rahandler' yourself. You can also use another program, via the 'externalprog' switch of Tralics. This program can ignore some of the flags, so that it is called in any case, even if there is nothing to do. Here is an example of a source file.

```

1 \documentclass{ra2003}
2 \theme{Num}
3 \isproject{YES} % \isproject{OUI} works also
4 \projet{MIAOU}{Miaou}{Mathématiques et Informatique de
5   l'Automatique et de l'Optimisation pour l'Utilisateur}
6 \def\foo{bar}
7 \UR{\URSophia\URFuturs}
8 \declaretopic{abc}{Topic abc}
9 \declaretopic{def}{Topic def}
10 \begin{document}
11 \maketitle
12 ...
13 \begin{module}{composition}{en-tete}{}
14 \begin{catperso}{Head of project team}
15 \pers{Laurent}{Baratchart}[DR INRIA]
16 \end{catperso}
17 \end{module}
18 \begin{module}{diffusion}{dif-conf}{Conferences and workshops}
19 \begin{glossaire}\glo{A}{B\par C}\glo{A1}{B1\par C1}\end{glossaire}
20 \begin{participants}
21 \pers{Laurent}{Baratchart},
22 \pers{José}{Grimm}
23 \end{participants}
24 \begin{motscle}
25 meromorphic approximation, frequency-domain identification,
26 extremal problems
27 \end{motscle}
28 \end{module}
29 \loadbiblio
30 \end{document}

```

This is what Tralics prints, for the full miaou2003 document, in verbose mode

```

This is tralics 2.5 (pl7), a LaTeX to XML translator
Copyright INRIA/MIAOU/APICS 2002-2005, Jos'e Grimm
Licensed under the CeCILL Free Software Licensing Agreement
Starting xml processing for miaou2003.
Configuration file identification: standard $ Revision: 2.14 $
Read configuration file /user/grimm/home/cvs/tralics/.tralics_rc.

```

Following lines are not printed on the terminal in non-verbose mode. You can see how the document type is found.

```

Configuration file has type \documentclass
Seen \documentclass ra2003
Potential type is ra2003
Defined type: RA
Defined type: RR
Defined type: std
Defined type: MP
Defined type: torture
Defined type: rabib
Alias Article undefined
Alias slides undefined
Alias RR does not match ra2003
Alias RA does not match ra2003
Alias torture does not match ra2003
Alias unknown undefined
Alias std does not match ra2003
Trying type ra
Alias Article undefined
Alias slides undefined
Alias RR does not match ra
Potential type ra aliased to RA
Using type RA

```

The lines that follow show the assignments from the configuration file.

```

makefo=xsltproc --catalogs -o %A.fo %B %C
makehtml=xsltproc --catalogs %B %C
makepdf=pdflatex -interaction=nonstopmode %w
generatedvi=latex -interaction=nonstopmode %t
generateps=dvips %t.dvi -o %t.ps
theme_vals=com cog num sym bio
dtd is raweb from raweb3.dtd (mode RAWEB2005)

```

Following lines are specific to the Raweb. You can see a summary of all the tests done by the program that converts miaou2003.tex to miaou.tex.

```

Ok with the config file, dealing with the TeX file...
Activity report for MIAOU (Miaou)
Mathématiques et Informatique de l'Automatique et de l'Optimisation pour l'Utilisateur
There are 138 environments
Checked 15 keyword env with 60 keywords (52 unique)
Checked 8 catperso and 31 participant(es) envs with 146 \pers
There were 2 topics
Sections (and # of modules): 1(1) 2(1) 3(2) 4(6) 5(5) 6(13) 7(4) 8(5) 9(3).

```

Whenever a section or a chapter is translated, a line is printed on the terminal. There is a complaint at the end, about a lonely module without title. A title is invented, namely '(Sans Titre)'. A non-trivial task for the post-processor is to remove it (it should not appear on the HTML pages).

```

Translating section composition
Translating section presentation
Translating section fondements

```

```

Translating section domaine
Translating section logiciels
Translating section resultats
Translating section contrats
Translating section international
Translating section diffusion
Bib stats: seen 57 entries
Seen 64 bibliographic entries
(SansTitre) Only one module seen in the section
Problem with sans titre 1
There was 1 NoTitle not handled

```

Tralics prints now statistics.

```

Used 1756 commands
Math stats : formulas 503, non trivial kernels 299, cells 10227,
  special 1 trivial 149, \mbox 5 large 0 small 118.
List stats: short 0 inc 10 alloc 43456
Buffer realloc 41 string 15750 size 610086; merge 7
Macros created 80 deleted 0
Save stack +1582 -1582
Attribute list search 7539(1494) found 3154 in 5616 elements (1401 after boot)
Number of ref 92, of used labels 36, of defined labels 73, of ext. ref. 19
Modules with 24, without 16, sections with 9, without 15
There were 6 images.
Output written on miaou.xml (250758 bytes).
No error found.
(For more information, see transcript file miaou2003.log)

```

Here you can see the call to the post-processor.

```

rahandler.pl v2.12, (C) 2004 INRIA, José Grimm, projet APICS
Postprocessor did nothing

```

Here are the commands specific to the Raweb:

- The document class should be ‘ra2003’ (the number must match the suffix of the file name). The document should start with `\maketitle` (see line 11) and end with `\loadbiblio`, see line 29. These commands are set to `\relax` by the translator.
- Some commands can be placed before `\begin{document}`. They will be evaluated in order. Then all commands outside a ‘module’ environment are executed (modules may be re-ordered). Strange results can be obtained... so that it is better not to put anything between modules.
- You use `\project`, or `\projet`, with three arguments. The first argument is the name of the Team. After converting to lower case letters, this must be the name of the file. The second argument can be empty (in this case the name of the Team is used). Any L^AT_EX command is allowed (remember: this is the official name of the Team). Then comes the long name (see example, lines 4-5).
- You use `\theme` to specify the theme. The argument must be one of those defined by the configuration file. The case is irrelevant. See example, line 2.
- You use `\isproject`, to say whether this is a “Project” or not (an argument that starts with y, Y, o or O is true). See example, line 3. Note that the Miaou team has been dissolved on 12/31/2003, and replaced by Apics, a team that had not the status of “project” until 01-01-2005.
- The command `\localisation` is obsolete.

- You use `\UR` for explaining in which UR the team is localized. Line 7 gives an example of a team in ‘Sophia’ and ‘Futurs’. The list of valid URs are defined by the configuration file (see example, lines 45, 46). The ‘UR’ prefix is not in the configuration file. The translator will see a single command, instead of all these four; in the example it is:

```
\RAstartprojet{num}{1}{Miaou}{Mathématiques et Informatique de l'Automatique
et de l'Optimisation pour l'Utilisateur}{Sophia Antipolis --- Futurs}
```

- You can use `\declaretopic`. At most ten topics can be declared. The first argument will be replaced by a unique id. For instance, the translator will see the following, instead of lines 8 and 9.

```
{\declaretopic{1}{Topic abc}\declaretopic{2}{Topic def}}
```

- You can use the module environment. Modules cannot be nested. A module is the equivalent of a `\section`, so that you cannot use the `\section` command; you can use `\subsection`. The first argument of the environment is the section name. It can be empty, meaning the same section as the previous module. The list of valid sections is defined by the configuration file. The first section cannot have more than one module. A module can have an optional argument (before the section name), a topic reference (the translator will see the numeric value). This is ignored, in the case a star follows the section name in the configuration file². Then comes the module name and the module title. The module name must be unique, and consists only in letters, digits, and dashes. The module title should not be empty (module in the ‘composition’ section are special; this is an awful hack, because the translator does not know which section is the composition section.) The translator will see

```
\begin{module}{MIAOU}{1}{en-tete}{module1}{module\textit{title}!}
... \end{module}
```

Here ‘MIAOU’ is the name of the team, ‘1’ the section number, and ‘module1’ the unique identifier associated to the module.

- You can use the ‘catperso’ environment, only in a module in section one. You can use ‘participants’ in other sections. These environments contain `\pers` commands (separated by commas in a ‘participants’).
- You can use `\pers` only as indicated above. The program checks a lot a things. Between the first name and the last name, there can be an optional argument, a particle. After the last name, there can be an optional argument.
- You should not use the ‘abstract’ and ‘body’ environments.
- You can use the ‘motscle’ environments for keywords.
- You can use a glossary (see line 19).
- You can use `\moduleref` for a reference to a module. As a consequence, each module has an implicit label; The statistics shown above say that, of the 73 labels defined, only 36 were used; in fact, 40 of them were implicit.
- The environments ‘glossary’, ‘moreinfo’, ‘participants’ and ‘keywords’ are designed for modules only. The statistics say : *Modules with 24, without 16, sections with 9, without 15*. You have to interpret this as: there are 24 modules and 9 other sectioning commands (like `\subsection`) with such an information; there is a total of 40 modules (in accordance with *Sections (and # of modules): 1(1) 2(1) 3(2) 4(6) 5(5) 6(13) 7(4) 8(5) 9(3)*).

²Feature introduced in version 2.5 (pl7)

More information is available on the Web page.

The following commands can be used in any document, but they are specific to the Raweb.

- `\RAstartprojet{a}{b}{c}{d}{e}`. The translation is a `<accueil>` element. You should not use this command, because its translation depends on parameters that are not arguments.
- `\begin{RAsection}{domaine} ... \end{RAsection}`. The translation is a `<domaine>` element; it can be the target of a `\ref`, if a `\label` is used, it can also be the target of a `\moduleref`, but this is a bad idea, because the section may disappear, if the Raweb is viewed as a sequence of topics. The first argument can be any character sequence. Note: This is automatically generated, it should not appear in a Raweb source, otherwise the result is not conforming to the Raweb DTD.
- `\begin{module}{MIAOU}{9}{dif-conf}{module28}{Some title} ... \end{module}`
The syntax is a bit different, from what is in the input source. The translation is
`<module id='uid75' html='module28'><head>Some title</head>...`
A module can be the target of a `\ref`, or a `\moduleref` via its name (in the example ‘dif-conf’). The name does not appear in the XML result. The value ‘module28’ is computed by the preprocessor. The first two arguments are ignored.
- `\begin{participants} ... \end{participants}`. The name of the environment can be ‘participants’, the final S can be omitted. The translation is an element with the same name. Locally `\par` is redefined to do nothing. No text is allowed. If you look at the example, the comma at the end of line 21 is necessary, but should not be translated.
- `\begin{catperso}{title} ... \end{catperso}`. This is like ‘participants’, but it takes an argument. No commas should be given here.
- `\declaretopic{a}{b}`. The translation is a `<topic>` element, with attribute `num` equal to `a`, containing an element `<t_titre>` containing the translation of `b`.
- `\begin{moreinfo} ... \end{moreinfo}`. The translation is a `<moreinfo>` element. Paragraphs are allowed.
- `\moduleref[y]{t}{s}{m}`. Argument `y` is optional; only empty values are allowed. Argument `t` is unused (this must be the name of the team in uppercase letters). If `m` is empty, this is reference to section `s`, otherwise to module `m`.
- `\pers{a}[b]{c}[d]`. If `b` is not empty, it will be prepended to `c`. If `c` contains a `\footnote`, the command disappears, the argument is prepended to `d`. Normally, `a` should be the first name of a person, `c` its last name, and `d` additional information. Initial and final spaces are removed from `c` and `d`. The result is then `\@pers{a}{c}{d}`.
- `\@pers{a}{c}{d}`. This command takes three arguments; if followed by a comma, the comma is read. In Raweb mode, the command is only allowed in a ‘catperso’ or ‘participants’ environment. Translation may be

```
<pers prenom='Laurent' nom='Baratchart'>DR INRIA</pers>
```

You can modify the translation of the `\pers` command, as follows:

```
\def\@pers#1#2#3{\xbox{pers}{\xbox{firstname}{#1}%
\xbox{lastname}{#2}\xbox{info}{#3}}}
\pers{Jean}[de la]{Fontaine}
\pers{Donald}{Knuth}[author of \TeX]
\pers{Leslie}{Lamport}
```

Then, the translation will be:

```
<pers><firstname>Jean</firstname><lastname>de la Fontaine</lastname>
  <info/></pers>
<pers><firstname>Donald</firstname><lastname>Knuth</lastname>
  <info>author of <TeX/></info>
</pers><pers><firstname>Leslie</firstname><lastname>Lamport</lastname>
  <info/></pers>
```

6.5 Tracing commands

In some cases, \TeX or \Tralics produce wrong results, incomprehensible error messages, and so on. In these cases, you must use specialized commands to see what happens. Since the internal structure of \TeX is not the same as \Tralics , the results in the transcript file may be different.

We have explained the command `\show`: it prints the meaning of a command (useful for a user defined command) and `\showthe` (this shows the value of a variable, counter, a token list, etc). We have also mentioned that `\showbox` prints the content of a \TeX box or XML element. There is a command `\showlists`; its effect is to indicate the global context; this is not implemented in \Tralics . The typical example is from the \TeX book. Given the test file:

```
\tracingcommands=1
\hbox{
$
\vbox{
\noindent$$
x\showlists
$$}\bye
```

This is the result of the `\showlists` command.

```
### display math mode entered at line 5
\mathord
.\fam1 x
### internal vertical mode entered at line 4
prevdepth ignored
### math mode entered at line 3
### restricted horizontal mode entered at line 2
\glue 3.33333 plus 1.66666 minus 1.11111
spacefactor 1000
### vertical mode entered at line 0
prevdepth ignored
```

This example does not compile in \Tralics : you cannot put a `\vbox` in a math formula. You cannot put a display math formula in a formula.

\TeX provides 9 commands of the form `\tracingXXXX` described earlier. Each variable defines an integer (in general, positive means verbose). There is a command `\tracingall` that turns everything on. In \Tralics , it sets `\tracingmacros`, `\tracingoutput`, `\tracingcommands` and `\tracingrestores` to 1. Only these variables are useful in \Tralics . For instance, `\tracingonline` controls whether or not anything is printed on the terminal; for \Tralics , debugging information is only printed on the transcript file. Variables like `\tracingparagraphs` and `\tracingpages` show line-break and page-break calculations, performed by \TeX but not by \Tralics . The command `\tracingoutput` shows boxes when they are shipped out (in \Tralics , the whole XML tree is printed at the end; if the command is positive, lines are printed, whenever used by the scanner), `\tracinglostchars` indicates all characters not found in the fonts (\Tralics never looks at font

properties). The command `\tracingstats` indicates that T_EX should gather all statistical information available; in Tralics, statistics are always computed; if you call it with the ‘silent’ switch, statistics are not printed on the terminal. Note that the ‘verbose’ switch calls `\tracingall`.

There are three remaining commands: `\tracingmacros` is used whenever a user command is expanded, `\tracingrestores` whenever things are popped from the save stack, and finally, `\tracingcommands` for all other commands. Let’s start with the example given on page 21. This is what you see if `\tracingoutput` is positive:

```
[4] \def\foo#1{\xbar#1}
[5] \def\xbar#1{{\itshape #1}}
[6] \foo{12}
```

It shows the input. This is what you see if `\tracingmacros` is positive:

```
\foo #1->\xbar #1
#1<-12
\xbar #1->{\itshape #1}
#1<-1
```

This is what you see if `\tracingrestores` is positive:

```
+stack: level + 2 for brace
{Push p 1}
{font restore }
+stack: level - 2 for brace
```

This is now what you see if `\tracingcommands` is positive: As you can see, some commands produce more than one line in the transcript file. For instance, a line is printed for `\def` when the command is seen, another one when the whole definition is read.

```
{\def}
{\def \foo #1->\xbar #1}
{\def}
{\def \xbar #1->{\itshape #1}}
{begin-group character {}}
{\itshape}
{font change \itshape}
Character sequence: 1.
{end-group character {}}
{Text:1}
Character sequence: 2 .
{Text:2}
}
```

This is the start of the trace on page 24:

```
[61] \begin{x}a b c \end{x}
{\begin}
{\begin x}
+stack: level + 3 for environment
```

What you can see is that `\begin` produces three lines, the second line holds the name of the environment; the last line explains that the stack pointer was changed from 2 to 3; the system remembers that the change comes from an environment, so a closing brace of a `\endgroup` command is illegal. Later on, the trace says:

```
Character sequence: ZbAY c .
{\end}
{Text:ZbAY c }
```



```
{\end x}
\endx ->by\end {y}ay
```

In TeX, instead of the first line, you would have seen:

```
{the letter Z}
{blank space }
{the letter c}
{blank space }
```

Tralics shows all characters it translates; it puts them on a single line. The character sequence is printed on the log file, because the command `\end` wants to be logged. After that, we have a line that contains ‘Text’ in braces. The text is added to the current XML element; a line is printed whenever the buffer is flushed. The buffer is flushed here because it might be used by the internal routine that scans the argument of `\end`. The trace contains also:

```
Character sequence: ay.
{\endgroup (for env)}
{Text:ay}
+stack: ending environment x; resuming document.
+stack: level - 3 for environment
Character sequence: .
```

Normally, each line of the form ‘level +3’, is followed by a line ‘level -3’, after that the current level is 2. The last line contains *Character sequence*, followed by a colon, a space, some characters, a period. Instead of ‘some characters’, you see only a space, it could be any character token with the category code of a space. In our case, it is the new line character that marks the end of the line. If the example is followed by an empty line, you will see:

```
[62]
{\par}
{Text:
}
```

What you see here is: open brace, *Text*, colon, some text, close brace. Here ‘some text’ is the space above, shown as a new line character. What the `\par` command does is: a) flush the buffer, so that the text is printed, b) remove space at end of paragraph, c) terminate the element and unwind the XML stack. The next line shows this ‘pop’. The integer 1 is the number of elements on the stack after the pop; you see the content of the stack, just before it is popped, in case the topstack is wrong. After the underscore, there is a suffix that indicates the mode (here `p_v` means that vertical mode will be entered after the pop).

```
{Pop 1: document_v p_v}
```

This is an example from page 28:

```
\E ->\expandafter
{\expandafter \E \E}
\E ->\expandafter
\E ->\expandafter
{\expandafter \expandafter \def}
{\expandafter \def \toto}
\toto ->\titi !
{\def}
{\def \titi !->7}
```

This shows that `\expandafter\foo\bar` shows all three tokens. This can be interesting if these tokens come from the expansion of other tokens. For instance, in a case like this

```
\def\y{yy}
\def\foo{\textit\y}
\expandafter\expandafter\foo
```

it is interesting to know that `\y` is expanded before `\textit`.

The next one is from page 39.

```
[346] \skip\count0=2pt plus \parindent \relax
{\skip}
+scanint for \count->0
+scanint for \skip->1
+scanint for \skip->2
+scandimen for \skip ->2.0pt
+scandimen for \skip ->3.0pt
{scanglue 2.0pt plus 3.0pt}
{\relax}
```

In T_EX, there is a big recursive function that converts characters into integers, dimensions and glue. The interesting point is the following: We have two commands `\skip` and `\relax`. The purpose of `\relax` is to stop scanning the glue, because an optional ‘minus’ term. You will not see `\parindent` nor `\count`. The T_EX output, in this case, consists of two lines. Tralics offers 6 more lines. The last line holds the glue that is effectively read and put in the register. There are three calls to the internal function ‘scanint’, the first is the number of the count register, the second is the number of the skip register, the last is the integer part of the dimension. There are two calls to the internal routine ‘scandimen’, one for each component of the glue (the shrink component is omitted, hence not read).

The next example comes from page 40.

```
1 [3506] \count0=2\ifnum\count0=\count13\fi4
2 {\count}
3 +scanint for \count->0
4 +\ifnum3532
5 +scanint for \count->0
6 +scanint for \ifnum->7
7 +\fi3532
8 +scanint for \count->13
9 +scanint for \ifnum->7
10 +iftest3532 true
11 +scanint for \count->2
12 {\relax}
13 +\fi3532
14 Character sequence: 4 .
```

We have already explained that ‘scanint’ is used to read something in case of assignment; as you can see, the procedure is also called in the case of a conditional. This example is a bit strange. Let’s explain what happens. A line of characters is read (see line L1), tokens are constructed, expanded and evaluated. The evaluator sees a first token, printed on line L2. This matches the rule: $\langle \text{simple assignment} \rangle$, in fact, the first clause, which is $\langle \text{variable assignment} \rangle$, that is defined as $\langle \text{integer variable} \rangle \langle \text{equals} \rangle \langle \text{number} \rangle$, and the first term is `\count` $\langle \text{8-bit number} \rangle$. There are two calls to ‘scanint’, the first with a range check. In order to makes things easier to understand, we given an index to each call, like S_1 , S_2 , etc.

The job of S_1 is easy: there is one digit, printed on L3. The equals character is the first unread character. It is an $\langle \text{equals} \rangle$. After the equals sign an integer is read, via S_2 . This sees the digit 2, then the conditional I_{3532} . This number was computed by Tralics, it is printed on line L4 to make debug easier. The `\ifnum` command reads two numbers, and a character between them, and compares the numbers. First number is read via S_3 . In fact, S_3 sees `\count` and calls S_4 . Procedure S_4 sees the number 0, followed by an equals sign. It prints that value on line L5. Now S_3 knows that its value is in `\count0`, this is 7, printed on line L6. After that, I_{3532} has a first

number 7, and sees the equal sign, and reads the second number via S_5 . This sees `\count` and reads a number via S_6 . This reads 13. Then comes `\fi`. The `\fi` command prints line L7 on the transcript file. This terminates I_{3532} . This is not possible: our conditional is still reading the second number. As a consequence, two tokens are pushed back, a `\fi` and a `\relax`, in this order: the `\relax` is read again first. This terminates expansion of the `\fi`.

Our procedure S_6 is programmed to fully expand tokens, and read them as long as digits are seen (even if the result overflows). It is unaware of the fate of the `\fi` token. All it knows is that the first unexpandable token after the digit 3 is `\relax`. Thus, S_6 has finished its job: the value is 13, printed on line L8. Then S_5 knows its result: the value is `\count13`, hence 7, printed on line L9. After that, I_{3532} has two numbers, they are the same, the test is true, as can be seen on line L10. Normal expansion resumes; however the condition stack has a marker that tells that the next `\fi` or `\else` matches I_{3532} . If the test had been false, the expansion of the test would have read, at high speed, all tokens up to the next `\fi` or `\else`. There are four unread tokens: a `\relax`, a `\fi`, the digit 4, and the newline character. Remember S_2 : this is a procedure that has read a digit, and wonders what follows. It does not care how complicated the task of ‘expand’ may be. It just wants a non-expandable token. In fact, the current token is `\relax`. Thus, S_2 knows it has read all digits; it prints the result of the transcript file L11. As a consequence, 2 is stored in `\count0`. The `\relax` token does nothing (let’s hope nobody has redefined it), see line L12. The conditional is terminated because of the inserted `\fi` token, line L13. After the last character on the line is translated, a new line is read, and a line of the form L1 will be printed; before a line is added to the transcript file, the internal buffer is flushed, this explains line L14. Note that the following sequence provokes an error in \TeX

```
\edef\F00{\ifnum\count0=\count13\fi}
\expandafter\def\F00{2x}
```

It is accepted by Tralics. As a consequence, the special `\relax` token inserted by \TeX always behaves like `\relax`.

6.6 Pictures and friends

We explain here the translation of some commands related to the picture environment. The syntax is unusual. In some cases, a pair of integers or a pair of real numbers are read. These numbers are multiplied by the value of the current unit of length, and the XML file contains these values, in pt, without the unit. The default value of `\unitlength` is 1pt. For instance

```
\setlength{\unitlength}{3pt}
\def\ten{10}
\put(\ten,\ten.2){x}
```

translates as `<pic-put xpos='30' ypos='30.59999'>x</pic-put>`. As the example shows, arithmetic on scaled integers is exact, but ‘10.2’ cannot be represented exactly. In some cases, arguments are converted to attributes, and errors can be signaled. In the case ‘`\put(1.2.3,4) {}`’, you will see *Missing unit (replaced by pt) {Character . of catcode 12}*, followed by three other errors. In the case of ‘`\makebox(1,2)[α]{x}`’ the error is *unexpected element formula*. Without the dollar signs, an error is signaled, the math formula is discarded, a second error is signaled with *unexpected element error*. If you invoke Tralics with ‘-noxmlerror’, the first error produces no `<error>` element, so that there is only one error.

- `\begin{picture}(A,B)(C,D) ... \end{picture}`. The first two arguments are required, the other ones are optional. Normally, there should be no text in the environment, the mode is neither horizontal nor vertical, `\par` commands are forbidden. Example:

```
\begin{picture}(1,2)
\begin{picture}(3,4)(5,6)
```

```

1=1
\end{picture}
\end{picture}

```

This translates as

```

<picture width='1' height='2'>
  <picture xpos='5' ypos='6' width='3' height='4'>
    1=1
  </picture>
</picture>

```

- `\fboxrule`, `\fboxsep`. These two dimensions are defined by L^AT_EX, unused by Tralics. Default values: 0pt and 3pt.
- `\makebox(A,B)[x]{y}`. This translates the value y , inside a group (as is the case for all ‘box’ commands that follow). The result is `<pic-framebox>` element, with attributes `width='A'`, `height='B'`, and `position='x'`.
- `\framebox(A,B)[x]{y}`. Same as `\makebox` above, but the attribute `framed` is set to true. As an example, the translation of

```

\makebox(1,2)[l]{\it x}
\makebox(1,2){\it x}
\framebox(1,2)[l]{\it x}
\framebox(1,2){\it x}

```

is

```

<pic-framebox width='1' height='2' position='l'>
  <hi rend='it'>x</hi></pic-framebox>
<pic-framebox width='1' height='2'>
  <hi rend='it'>x</hi></pic-framebox>
<pic-framebox width='1' height='2' position='l' framed='true'>
  <hi rend='it'>x</hi></pic-framebox>
<pic-framebox width='1' height='2' framed='true'>
  <hi rend='it'>x</hi></pic-framebox>

```

- `\makebox[w][x]{y}`. See alternate syntax above. The first argument must be a dimension, it is the width of the box; its value is currently ignored. The second argument must be one character of ‘lrcs’. A paragraph is started if necessary. The result is a `<mbox>` element, containing the translation of y , unless either the box contains only a `<figure>`, or if it contains only text (characters, and font changes), case where the result is y .
- `\mbox{y}`. This is exactly like `\makebox`, with a simple syntax. For instance:

```

\makebox{\it x} \mbox{\it x}
\makebox[2cm]{y\it x}
\makebox[2cm][l]{\it x}
\makebox{ $x$ } \mbox{ $x$ }
\makebox{\xbox{foo}{bar}} \mbox{\xbox{foo}{bar}}
\makebox{\includegraphics{x}} \mbox{\includegraphics{x}}

```

translates as

```

<hi rend='it'>x</hi> <hi rend='it'>x</hi>
<mbox>y<hi rend='it'>x</hi></mbox>
<mbox position='l'><hi rend='it'>x</hi></mbox>
<mbox><formula type='inline'><simplermath>x</simplermath></formula></mbox>

```

```

<mbox><formula type='inline'><simplemath>x</simplemath></formula></mbox>
<mbox><foo>bar</foo></mbox> <mbox><foo>bar</foo></mbox>
<figure rend='inline' file='x' /><figure rend='inline' file='x' />

```

- `\framebox[w][x]{y}`. This is like `\makebox`, but the result is a `<fbox>`, instead of `<mbox>`, with attribute `rend='boxed'`. In the case where the content is a `<figure>`, then no box is created but the attribute `framed` is added to the figure.
- `\fbox{y}`. Like `\framebox`, with a simple syntax. For instance

```

\framebox{\it x} \fbox{\it x}
\framebox[2cm]{y\it x}
\framebox[2pt][1]{\it x}
\framebox{$x$} \fbox{$x$}
\framebox{\xbox{foo}{bar}} \fbox{\xbox{foo}{bar}}
\framebox{\includegraphics{x}} \fbox{\includegraphics{x}}

```

translates as

```

<fbox rend='boxed'><hi rend='it'>x</hi></fbox>
<fbox rend='boxed'><hi rend='it'>x</hi></fbox>
<fbox width='56.9055pt' rend='boxed'>y<hi rend='it'>x</hi></fbox>
<fbox width='2.0pt' position='1' rend='boxed'><hi rend='it'>x</hi></fbox>
<fbox rend='boxed'><formula type='inline'><simplemath>x</simplemath>
  </formula></fbox>
<fbox rend='boxed'><formula type='inline'><simplemath>x</simplemath>
  </formula></fbox>
<fbox rend='boxed'><foo>bar</foo></fbox>
<fbox rend='boxed'><foo>bar</foo></fbox>
<figure framed='true' rend='inline' file='x' />
<figure framed='true' rend='inline' file='x' />

```

- `\scalebox{x}{y}`. The translation is a `<scalebox>` containing `y`, with a `scale` attribute whose value is `x`.
- `\rotatebox{x}{y}`. The translation is a `<pic-rotatebox>` containing `y`, with a `angle` attribute whose value is `x`.
- `\dashbox{A}(B,C)[d]{y}`. The translation is a `<pic-dashbox>` containing `y`, with a `position` attribute whose value is `d`. The three values `A`, `B` and `C` are numbers, affected by the unit length. They produce attributes `dashdim`, `width` and `height`. Example:

```

\setlength\unitlength{2pt}
\scalebox{3}{\it x}
\rotatebox{90}{\it x}
\dashbox{1.2}(2,3)[1]{\it x}
\dashbox{1.2}(2,3){\it x}

```

The translation is

```

<scalebox scale='3'><hi rend='it'>x</hi></scalebox>
<pic-rotatebox angle='90'><hi rend='it'>x</hi></pic-rotatebox>
<pic-dashbox dashdim='2.4' width='4' height='6' position='1'>
  <hi rend='it'>x</hi></pic-dashbox>
<pic-dashbox dashdim='2.4' width='4' height='6'>
  <hi rend='it'>x</hi></pic-dashbox>

```

- `\begin{minipage}[x][y][z][t]{dim} ... \end{minipage}`. Optional arguments y and t are ignored. Arguments x and z should define vertical position. The last argument is a glue. Note: the translation starts in vertical mode. The command `\nocentering` is called.
- `\parbox[x][y][z]{dim}{y}`. This should be the non-environment version of ‘minipage’, but it is much more primitive. It behaves like `\hbox` and `\xbox` for argument parsing. Example

```
\begin{center}
a
\begin{minipage}{2cm} a \end{minipage}
\begin{minipage}[c][l][b][r]{2cm plus 3pt} a \end{minipage}
\end{center}
\parbox[foo][bar][gee]{2cmPlus3mm}{some \it box content}.
```

The translation is

```
<p rend='center'>a
<minipage width='56.9055pt'><p>a </p></minipage>
<minipage inner-pos='b' pos='c' width='56.9055pt'><p>a </p></minipage></p>
some <hi rend='it'>box content</hi><p>.</p>
```

- `\bezier{N}(A,B)(C,D)(E,F)`. The first argument must be an integer, others are real numbers, affected by `\unitlength`. The result is a `<pic-bezier>` element.
- `\qbezier[N](A,B)(C,D)(E,F)`. As above, but the first argument is optional. A Bezier curve is defined by three points, you must give the coordinates. In the first implementation, you had to specify the number of intermediate points used for plotting. For instance

```
\bezier{10}(1,2)(3,4)(5,6)
\qbezier(1,2)(3,4)(5,6)
\qbezier[10](1,2)(3,4)(5,6)
```

translates as

```
<pic-bezier a1='1' a2='2' b1='3' b2='4' c1='5' c2='6' repeat='10' />
<pic-bezier a1='1' a2='2' b1='3' b2='4' c1='5' c2='6' />
<pic-bezier a1='1' a2='2' b1='3' b2='4' c1='5' c2='6' repeat='10' />
```

- `\put(A,B){x}`. The translation is a `<pic-put>` element containing the translation of x , with attributes `xpos` and `ypos` that come from A and B .
- `\xscale`, `\yscale`, `\xscaley`, `\yscalex`. These are user defined commands, that should contain real numbers. The default values are 1, 1, 0, and 0.
- `\scaleput(A,B){x}`. The translation is a `<pic-scaleput>`. It is like `\put`, but current scale values are put in the element. For instance the translation of

```
\setlength\unitlength{2pt}
\put(1,2){x}
\def\xscale{2.0}\def\yscale{3.0}\def\xscaley{4}\def\yscalex{0.5}
\scaleput(1,2){x}
```

is

```
<pic-put xpos='2' ypos='4'>x</pic-put>
<pic-scaleput xpos='2' ypos='4'
xscale='2.0' yscale='3.0' xscaley='4' yscalex='0.5'>x</pic-scaleput>
```

- `\multiput(A,B)(a,b){n}{x}`. The translation is a `<pic-multiput>` element, with all the arguments as attributes. If the command is followed by a star, then n calls of `\put` are evaluated, with the same argument x , but at positions $(A + ka, B + kb)$ instead of (A, B) .

See example of the little car on figure 6.1; the ruler was constructed with `\multiput`, a star is used for the figures.

- `\frame{x}`. This translates as `<pic-frame>` with the value of x .
- `\oval(A,B)[c]`. This translates as `<pic-oval/>` with some attributes, `xpos`, `ypos`, `spec`.
- `\line(A,B){c}`. This translates as `<pic-line/>`. The first two arguments have to be integers. Only small values are recognized by L^AT_EX. The last argument is a real number. Attributes are `xdir`, `ydir`, `width`.
- `\vector(A,B){c}`. This translates as `<pic-vector/>`. Same comments as above. Example.


```
\setlength\unitlength{2pt}
\oval(2,3) \oval(2,3)[4]
\frame{x}
\line(2,3){12.2} \vector(2,3){12.2}
```

The translation is

```
<pic-oval xpos='4' ypos='6' />
<pic-oval xpos='4' ypos='6' specs='4' />
<pic-frame>x</pic-frame>
<pic-line xdir='2' ydir='3' width='24.4' />
<pic-vector xdir='2' ydir='3' width='24.4' />
```

- `\thicklines`. The translation is `<pic-thicklines/>`.
- `\thinlines`. The translation is `<pic-thinlines/>`.
- `\linethickness{x}`. The translation is `<pic-linethickness>` with attribute `size`.
- `\arc[n](A,B){c}`. The translation is an empty `<pic-arc>` with attributes `nbsymb`, `angle`, `xpos`, `ypos` (and also `unit-length`).
- `\bigcircle[n]{c}`. The translation is an empty `<pic-bigcircle>` element with attributes `nbsymb`, `size`, (and also `unit-length`).
- `\circle{c}`. The translation is an empty `<pic-circle>`. For instance

```
\setlength\unitlength{2pt}
\arc[17](2,3){40} \arc(2,3){40}
\bigcircle[17]{40} \bigcircle{40}
\circle{40}
```

translates to

```
<pic-arc angle='40' xpos='4' ypos='6' unit-length='2' nbsymb='17' />
<pic-arc angle='40' xpos='4' ypos='6' unit-length='2' />
<pic-bigcircle size='40' unit-length='2' nbsymb='17' />
<pic-bigcircle size='40' unit-length='2' />
<pic-circle size='80' />
```

- `\curve[n](A1,B1)(A2,B2)(A3,B3), \dots, (An,Bn)`.
- `\closecurve[n](A1,B1)(A2,B2)(A3,B3), \dots, (An,Bn)`.
- `\tagcurve[n](A1,B1)(A2,B2)(A3,B3), \dots, (An,Bn)`. These produces curves.
- `\dashline[A]{B}[C](A1,B1)(A2,B2)(A3,B3), \dots, (An,Bn)`.
- `\dottedline[A]{B}[C](A1,B1)(A2,B2)(A3,B3), \dots, (An,Bn)`.
- `\drawline[A][C](A1,B1)(A2,B2)(A3,B3), \dots, (An,Bn)`. These produce dashes. Example:

```

\setlength\unitlength{2pt}
\curve[12](1,2)(3,4)(4,5)(7,8)
\tagcurve[12](1,2)(3,4)(4,5)(7,8)
\closecurve[12](1,2)(3,4)(4,5)(7,8)
\curve(1,2)(3,4)(4,5)(7,8)
\tagcurve(1,2)(3,4)(4,5)(7,8)
\closecurve(1,2)(3,4)(4,5)(7,8)
\dashline[12]{13}[14](1,2)(3,4)(4,5)(7,8)
\dottedline[12]{13}[14](1,2)(3,4)(4,5)(7,8)
\drawline[12][14](1,2)(3,4)(4,5)(7,8)

```

The translation is

```

<pic-curve unit-length='2' nbsymb='12'>1,2</pic-curve><p>(3,4)(4,5)(7,8)
<pic-tagcurve unit-length='2' nbsymb='12'>1,2</pic-tagcurve>(3,4)(4,5)(7,8)
<pic-closecurve unit-length='2' nbsymb='12'>1,2</pic-closecurve>(3,4)(4,5)(7,8)
<pic-curve unit-length='2'>1,2</pic-curve>(3,4)(4,5)(7,8)
<pic-tagcurve unit-length='2'>1,2</pic-tagcurve>(3,4)(4,5)(7,8)
<pic-closecurve unit-length='2'>1,2</pic-closecurve>(3,4)(4,5)(7,8)
<dashline arg3='14' arg2='13' arg1='12'>
  <point xpos='2' ypos='4'><point xpos='6' ypos='8'>
  <point xpos='8' ypos='10'><point xpos='14' ypos='16'></dashline>
<dottedline arg3='14' arg2='13' arg1='12'><point xpos='2' ypos='4'>
  <point xpos='6' ypos='8'><point xpos='8' ypos='10'>
  <point xpos='14' ypos='16'></dottedline>
<drawline arg3='14' arg1='12'><point xpos='2' ypos='4'>
  <point xpos='6' ypos='8'><point xpos='8' ypos='10'>
<point xpos='14' ypos='16'></drawline>

```

6.7 The title page

The configuration file explains how to gather information from a document, in order to construct a ‘title page’. This mechanism is used for producing a HTML page containing some meta-data for each Inria Research Report. The author sends to the “gescap” mailing list the beginning of the document, up to the magic command `\makeRR`. This is translated by Tralics (that does not care about missing `\end{document}`), and a post-processor extracts the `<RRstart>` element from the XML result, and converts it to HTML. Consider a sample document:

```

1 \documentclass[a4paper]{report}
2 \usepackage{RR}
3
4 \providecommand\Tralics{\xbox{Tralics}{}}
5 \def/XML{XML}
6
7 \RRtitle{Tralics, a \LaTeX\ to XML translator\Partie I}
8 \RRetitle{Tralics, a \LaTeX\ to XML translator\Part I}
9 \RRauthor{José Grimm\thanks{Email: Jose.Grimm@sophia.inria.fr}}
10
11 \RRprojet{Apics}
12 \RRtheme{\THNum}
13
14 \RRresume{
15 Dans cet article\par nous décrivons le logiciel \Tralics,\par...}
16 \RRabstract{

```



```

17 In this paper we describe \Tralics, a \LaTeX\ to \XML\ translator.}
18
19 \RRdate{Aout 2005}
20 \URSophia
21 \motcle{Latex, XML, HTML, MathML, Perl, PostScript, Pdf}
22 \keyword{Latex, XML, HTML, MathML, Perl, PostScript, Pdf}
23
24 \begin{document}
25 \makeRR
26 text

```

This is translated by Tralics as follows.

```

<?xml version='1.0' encoding='iso-8859-1'?>
<!DOCTYPE rr SYSTEM 'raweb.dtd'>
<!-- translated from latex by tralics 2.5 (p17)-->
<rr type='RR' chapters='true'>
<RRstart><UR>
<URSophia/>
</UR>
<title>Tralics, a <LaTeX/> to XML translator Partie I</title>
<etitle>Tralics, a <LaTeX/> to XML translator Part I</etitle>
<projet>Apics</projet>
<motcle>Latex, XML, HTML, MathML, Perl, PostScript, Pdf</motcle>
<keyword>Latex, XML, HTML, MathML, Perl, PostScript, Pdf</keyword>
<resume><p>
Dans cet article</p>
<p>nous décrivons le logiciel <Tralics/>,</p>
<p>...</p></resume>
<abstract><p>
In this paper we describe <Tralics/>, a <LaTeX/> to XML translator.</p></abstract>
<author><auth>José Grimm<note id='uid1'
place='foot'>Email: Jose.Grimm@sophia.inria.fr</note></auth>
</author>
<date>Aout 2005</date>
<RRnumber>????</RRnumber>
<Theme>
<THNum/>
</Theme>
<cmdp>nodefaut</cmdp>
<cmdA><error l='1' c='Undefined command '/></cmdA>
<cmdB><error l='24' c='Undefined command '/></cmdB>
<cmdC><error l='25' c='Undefined command '/></cmdC>
</RRstart>
<p>text</p>
</rr>

```

A big part of the transcript file is given later in section 6.11, we sometimes make references to it. A line number of the form S3 refers to line 3 of the source, C3 is line 3 of the configuration file, and T3 is line 3 of the transcript file.

We used the following configuration file. It starts like the standard configuration file, with some modifications near the end. In particular we have added some lines that provoke 4 errors, they will be explained later. We copy the titlepage section from the standard configuration file, with a small addition, plus some comments.

```

1 ## test file for the titlepage
2 ## tralics ident rc=Test $ Revision: 2.1 $

```

```

3
4 BeginType RR
5   DocType = rr raweb.dtd
6 BeginTitlePage
7   \makeRR <RRstart> "" "type = 'RR'" % CESS
8   alias \makeRT "" "type = 'RT'" % ACSS
9   <UR> - % E
10  \URSophia ?+<UR> % CE
11  \URRocquencourt ?+<UR> % CE
12  alias \URRocq % AC
13  \Paris ?<UR> <Rocquencourt> % CEE
14  \URRhoneAlpes ?+<UR> % CE
15  \URRennes ?+<UR> % CE
16  \URLorraine ?+<UR> % CE
17  \URFuturs ?+<UR d='true''> % CE
18  \RRtitle q<title> "pas de titre" % CES
19  \RRetitle q<etitle> "no title" % CES
20  \RRprojet <projet> "pas de projet" % CES
21  \motcle <motcle> "pas de motcle" % CES
22  \keyword <keyword> "no keywords" % CES
23  \RRresume p<resume> "pas de resume" % CES
24  \RRabstract p<abstract> "no abstract" % CES
25  \RRauthor + <author> <auth> "Pas d'auteurs" % CEES
26  \RRdate <date> "\monthyearvalfr" % CES
27  \RRNo <RRnumber> "???" % CES
28
29  \RRtheme <> +"pas de theme" % CES
30  <Theme> - % E
31  \THNum ?+<Theme> % CE
32  \THCog ?+<Theme> % CE
33  \THCom ?+<Theme> % CE
34  \THBio ?+<Theme> % CE
35  \THSym ?+<Theme> % CE
36
37  \myself \RRauthor "grimm" % CCS
38  \cmdp <cmdp> +"nodefault" % CES
39  \cmda <cmdA> A"\cmdAval" % CES
40  \cmdb <cmdB> B"\cmdBval" % CES
41  \cmdc <cmdC> C"\cmdCval" % CES
42
43 End
44 BeginCommands
45   \let\thanks\footnote
46 End
47
48 EndType

```

The syntax in the TitlePage part of a configuration file is the following: each line has some fields, that can be of type A (the word 'alias') or type C (a command, a backslash followed by some letters), or E (an element name, delimited by less-than and greater-than) or S (a string, delimited by a double quote). Before each field, you can put one or two modifiers. Only the second and the third fields can have a modifier. Comments are not necessary. The following combinations are recognized.

CESS as in `\makeRR <RRstart> "" "type = 'RR'".` This declaration has to be the first in the list. It can be given only once. No modifiers are allowed. It defines a command `\makeRR`, that can be used only once in the document, after `\begin{document}`. Each titlepage command is defined by a number, line T1 shows that this number is 0. Line T118 shows the use. On the line that follows, we see that the evaluator does not know the name associated to the command, line T119 shows that this is the command number 0, and line T120 shows the name.

The effect is to insert the `<RRstart>` element into the XML tree, after some checks (that may produce an error like on line T122). In what follows, we shall call it the TPA element. This element is formed of other elements defined by the titlepage info, the names of these elements are statically defined, their content is dynamic (i.e., the names depends on the configuration file, the content on the \TeX document). The first string is a list of attributes added to the TPA element and the second string is a list of attributes added to the document element. In our example, the first string is empty.

In the case where one of the attributes of the second string has the value ‘only title page’, then `\endinput` is evaluated just after the Titlepage command. This means that everything after the titlepage command is ignored. This is useful if you want to extract the titlepage information from a document, without converting the whole document.

AESS as in `alias \makeRT "" "type = 'RT'".` This declaration is valid only after a CESS declaration (or after another AESS declaration). It defines a command `\makeRT` that can be used instead of `\makeRR` (only one of these commands can be used). The result is the same; however it can use different attributes. (Same remark as above for special attribute values in the second string). In what follows, the `\TPA` command means one of the commands defined by this rule or the preceding one.

CEES as in `\RRauthor + <author> <auth> "Pas d’auteurs".` Note that the plus sign is required before the `<author>` element. This declaration has as side effect that the TPA element will contain a `<author>` element, formed of a number of `<auth>` elements. Initially there is only one, initialized with ‘Pas d’auteurs’.

The declaration has another effect, it defines a command `\RRauthor`, that has to be used before the `\TPA` command. It takes one argument, and creates a `<auth>` element whose content is the translation of the argument. This element is added to the end of the `<author>` element. The command can be used more than once, in the case there are multiple authors. Note that the default value is removed in case at least one value is given.

CCS as in `\myself \RRauthor "JG".` The effect is the same as `\def\myself{\RRauthor{JG}}`. However, the string argument is not translated, it is taken verbatim.

AC as in `alias \URRocq`, line C12 and T11. This makes `\URRocq` an alias for the command defined on the previous line. Aliasing is achieved via `\let`.

E as in `<UR> -`. The dash after the element is required. Another example can be `<sUR fr='unité de recherche' en='research unit'> -`. In this second example, we have an element named `<sUR>`, that has two attributes. The effect is to put, in the XML result, this element (with its attributes), and its content is a list of items declared in the configuration file (the list can be empty).

CE as in `\URSophia ?+<UR>` or `\URFuturs ?+<UR d='true'>`. This has as effect to define a command, here `\URSophia` or `\URFuturs`, that takes no argument, whose effect is to insert, to the element `<UR>` (that must be defined by a previous rule), an empty element, whose name is `<URSophia>`, and that has the attributes of `<UR>`. See lines T8 and T21, they show the element and the attribute.

CEE as in `\Paris ?<UR> <Rocquencourt>`. The effect is to define a command `\Paris`, that behaves like `\Ursophia`, but the element created is `<Rocquencourt>` instead of one named `<Paris>`.

CES as `\RRtitle <title> "pas de titre"`. This is the generic command. The element can have the modifiers `p`, `q`, `e` or `E`, and the value can have the modifiers `+ABC`. The effect is to define a command `\RRtitle` (or an environment `'RRtitle'` if the `E` modifier has been given), that can be used only before the `\TPA` command. The argument of the command (or the content of the environment) is translated, put in a `<title>` element, and added to the TPA element.

If no modifier is given for the element, paragraphs are forbidden in the argument. If you want to use paragraphs (either `\par` or `\`) you must use the `P` modifier (lower-case letter), as in lines C23, C24, T33, T35. In the same fashion, a lower case `E` means environment without paragraphs, an upper case `E` means environment with paragraphs. If the `q` modifier is given, paragraphs are forbidden, but you can use `\`, which is ignored. (In fact, the command reads an optional star, an optional argument, and the result is replaced by a space). Example, lines C18, C19, T23, T25. Note that, in this document, there is a `\` in the title, that appears as a space in the page headings. This is done by redefining `\` to `\space`, so that optional arguments are not taken into account. There is a dirty hack in Tralics.

If no modifier is given for the value, then `<title>pas de titre</title>` is added to the TPA element in case the command is never used. For instance, line C27 specifies an optional report number to be some question marks.

Near the end of the titlepage example, we define `\cmdp`, `\cmdA`, `\cmdB`, and `\cmdC` in a similar fashion, but add a modifier before the value. None of these commands is used in the T_EX file, and you can observe the following facts.

- Tralics complains with: *Error signaled at line 25 of file tptest.tex: No value given for command \cmdp.* In fact, when Tralics sees the `\makeRR` command, it notices that `\cmdp` has not been called and complains, because the plus-sign modifier means: this value is required. See transcript file, line T122, T123.
- In the case where the modifier is one of `A`, `B` or `C`, then the default value is a L^AT_EX command (in fact, a list of characters, that may be interpreted as a set of L^AT_EX commands, and must be translated). Tralics removes the double quotes, and inserts the characters in one of its buffers, or converts the string into a token list that is appended to the input token list. The non trivial point is: when is the command evaluated?
- In the case of `'A'` modifier, the command is evaluated just after the `\documentclass` command. See line T70. There may be other lines that are not in the T_EX source file that are evaluated there; for instance line C46, which is evaluated just after this one. As lines 71 to 76 show, an error is signaled, because the command is evaluated too early.
- In the case of `'B'` modifier, the command is evaluated at begin-document. More precisely `\cmdb{\cmdBval}` is tokenised, and the result is added to the token list maintained by `\AtBeginDocument`. See lines T64, T65. The second line shows the special `\endinput` marker that indicates the end of the character list. These tokens are stored somewhere, and executed after the `\begin{document}`, the result can be seen in the transcript file. Line T104 shows the document-hook, lines T106 to T112 the translation. Is the `\ignorespaces` really needed? this is the last command in the document-hook.
- In the case of `'C'` modifier, the command is evaluated when the TPA command is seen. See line T124 (the definition) and T125 (End of virtual file). The important point is that category codes in effect at this moment will be used to convert the string into a sequence of tokens.

There is special trick for the case where the name of the element associated to the command is empty. Assume that the configuration file contains `\RRtheme <> +"pas de theme"`, line C29. In the case where the user does not use `\RRtheme`, an error will be signaled, and the text will appear in the resulting XML. If the user says `\RRtheme{foo}`, then `Tralics` remembers the use and issues no complain. Moreover, it reads the argument, and pushes `foo\par` in the input stream (the reason why `\par` is executed is to make sure that `Tralics` remains in vertical mode). See lines T88-90, the `\par` is on line T91.

In fact, in the case of a theme, the input source is on line S12. The author must use commands defined in the file `RR.sty`, that contains the following code. Note how the `\TH*` commands are defined only locally, and `\@RRthmnum` has a default value that prints a warning if not overwritten.

```
\newcommand\@RRthmnum{?\@warning{*** Il manque un numéro de thème INRIA ;
  utilisez \string\RRtheme ***}}
\newcommand\RRtheme[1]{\def\RR@theme{#1}}%
\RRtheme{}
\def\eval@theme{%
\begingroup
\def\THCom{\TH@add{COM}{Systèmes communicants}}%
\def\THCog{\TH@add{COG}{Systèmes cognitifs}}%
\def\THSym{\TH@add{SYM}{Systèmes symboliques}}%
\def\THNum{\TH@add{NUM}{Systèmes numériques}}%
\def\THBio{\TH@add{BIO}{Systèmes biologiques}}%
\RR@theme\endgroup}
\newcommand\TH@add[2]{%
\ifx\@RRthmname\@empty
  \gdef\@RRthmnum{}%
  \else
    \xdef\@RRthmname{\@RRthmname\space et }%
    \xdef\@RRthmnum{\@RRthmnum\space et }%
    \xdef\@RRTHMTYPE{Thèmes}%
    \xdef\@RRthmtype{Thèmes}%
  \fi
  \xdef\@RRthmnum{\@RRthmnum#1}%
  \xdef\@RRthmname{\@RRthmname#2}%
}
```

6.8 Array and Tables

We describe here the implementation of the arrays in `Tralics`. One has to distinguish between ‘table’ which is an environment in which you can put some objects (in general tables) with a caption; like the ‘figure’ environment, this generates a floating object. On the other hand, the ‘array’ and ‘tabular’ environments can be used to create a table: the first one is designed for math only, the second for non-math material. Math tables are described in the chapter about mathematics.

Example

```
\begin{table}
\begin{tabular}{c} x \y \end{tabular}
\caption{My caption}
\label{t1}
\end{table}
\begin{tabular}{c} \ref{t1} \end{tabular}
```

The translation is as follows. As you can see, both objects have the same same. If the table contains a tabular, only one XML object is created.

```
<table id='uid1'>
  <head>My caption</head>
  <row><cell halign='center'>x</cell></row>
  <row><cell halign='center'>y</cell></row>
</table>
<p><table rend='inline'><row><cell halign='center'><ref target='uid1' /></cell>
</row></table></p>
```

The tabular environment

You can say `\begin{tabular} [pos] {cols} ... \end{tabular}` or `\begin{tabular*}{width} [pos] {cols} ... \end{tabular*}`. In both cases, the result is a `<table>` element. This element has a `vpos` attribute whose value is `t`, `b` or `c`, provided that the optional `[pos]` argument is one of `[t]`, `[b]` or `[c]`. The element has a `width` attribute with value `xx`, provided that the ‘`tabular*`’ environment has been used and the first argument evaluates to `xx` as a dimension. The resulting element consists of some `<row>` elements, each of which contains some `<cell>` elements. A more complicated example:

```
\begin{tabular*}{10pc}[b]{lrc}
\hline
a&b&c\ [2pt]
\multicolumn{1}{l}{A}&B&C\ \hline
\end{tabular*}
```

The translation here shows that the name of elements and attributes can be changed.

```
<Table VPos='b' TableWidth='120.0pt' rend='inline'>
  <Row SpaceAfter='2.0pt' TopBorder='true'>
    <Cell Align='Cleft'>a</Cell>
    <Cell Align='Cright'>b</Cell>
    <Cell Align='Ccenter'>c</Cell>
  </Row>
  <Row BottomBorder='true'>
    <Cell Align='Cleft' Cols='1'>A</Cell>
    <Cell Align='Cright'>B</Cell>
    <Cell Align='Ccenter'>C</Cell>
  </Row>
</Table>
```

Interpreting the preamble

The preamble of the array is the quantity marked ‘`{cols}`’ in the description above. This is a specification for columns. It specifies how the columns should be formatted. In standard L^AT_EX, you cannot use more columns than specified; in Tralics, this is not relevant. The T_EX primitive is called `\halign`, and L^AT_EX has to construct a preamble that matches the requirements of T_EX; it is very difficult to implement the T_EX algorithm, so that we make no attempt to implement the commands. This is an example

```
\vbox{\offinterlineskip
\hrule
\halign{&\vrule#&
\strut \quad\hfil#\quad\cr
```

| Year | Word Population |
|--------|-----------------|
| 8000BC | 5,000,000 |
| 50AD | 200,000,000 |
| 1650AD | 500,000,000 |
| 1850AD | 1,000,000,000 |
| 1945AD | 2,300,000,000 |
| 1980AD | 4,400,000,000 |

Table 6.1: A T_EX table

```

height2pt &\omit&&\omit&\cr
&Year\hfil&&Word Population&\cr
height2pt &\omit&&\omit&\cr
\noalign{\hrule}
\noalign{\vskip 2pt}
\noalign{\hrule}
height2pt &\omit&&\omit&\cr
&8000BC&&5,000,000&\cr
&50AD&&200,000,000&\cr
&1650AD&&500,000,000&\cr
&1850AD&&1,000,000,000&\cr
&1945AD&&2,300,000,000&\cr
&1980AD&&4,400,000,000&\cr
height2pt &\omit&&\omit&\cr}\hrule}

```

The table has 5 columns, of the form ABABA, because the preamble has the form `&A&B\cr`, the first `&` marks repetition. Both templates A and B are formed of a `<u>` part, then `#`, then a `<v>` part. In the table, the `#` means: “stick the text of each column entry in this place”. In the case of A, this is almost always empty. In some cases, it is ‘height2pt’, case where the values of B are `\omit`. Here `\omit` says that `<u>` and `<v>` should be omitted; the important point is that the `\strut` be omitted. This gives additional vertical space, of exactly 2pt; other rows have (at least) the vertical size of a `\strut`. Very often row are too narrow; L^AT_EX has a command `\arraystretch` that controls this. Note that A is `\vrule#` and `\vrule` is a command that accepts an optional argument. Note how horizontal rules are inserted in the table.

As the previous example shows, there are three standard column types: `c`, `l` and `r` (centered, left-aligned, right-aligned). A T_EX preamble like `\quad\hfil#\quad` corresponds to ‘`r`’ (instead of `\quad`, L^AT_EX uses some default intercolumn space that can be modified. You can also say `p{dim}`. This should typeset the column in a `\parbox[t]{dim}`. This feature is not implemented: the argument is ignored, and `p` is replaced by `c`. Note: `\parbox` currently ignores its argument in Tralics.

The `array.sty` package adds two options that take a dimension as argument: ‘`m`’ and ‘`b`’. The ‘`b`’ option is like the ‘`p`’ option, but bottom-aligned. The ‘`m`’ option should be used only in math mode (i.e. for the `array` environment, and not `tabular`). In Tralics, there is no difference between ‘`b`’, ‘`m`’ and ‘`p`’.

There is a `@{text}` option. It inserts ‘`text`’ in every row, where ‘`text`’ is processed in math mode in the ‘`array`’ environment and in LR mode in the ‘`tabular`’ and ‘`tabular*`’ environments. Such an expression suppresses the space that L^AT_EX normally inserts between columns. For instance, an `array` specification like `{l@{\hspace{1cm}}l}` says that the two columns of text should be separated by exactly one centimeter. A specification like `{@{}c@{}}` says that no additional space should be added neither of the left nor the right of the column. An `\extracolsep{wd}` command

can be used inside such an expression. It causes an extra space to appear to the left of all subsequent columns. Note that `\extracolsep` expands to `\tabskip`; this T_EX primitive is not implemented in Tralics. In fact, Tralics ignores an '@' and its argument.

You can use a | for specifying a vertical rule. However, in Tralics you cannot use double or triple rules. Sorry. There is also a !{...} options that is not implemented.

Every specification ('l', 'r', 'c', 'p', 'b', 'm') can be preceded by a `>{xx}` declaration, and followed by a `<{yy}` declaration. In case of multiple declarations, the last will be executed first. Said otherwise, `>{3}>{b}c<{a}<{z}` is the same as `>{b3}c<{za}`. The effect is to insert 'b3' before the cell in the current position, and 'za' after the cell. See the last tabular in table 6.2. This corresponds to ⟨u⟩ and ⟨v⟩ parts of a T_EX array. Note that the cell is finished when a token is sensed that indicates either a new cell, a new row or the end of the array. Technically, this means a &, a \\, or an `\end` (the end of the environment). A special marker is pushed back after the 'za'. This is a special endtemplate token in the case of a cell, and a `\cr` in the case of \\ . You should not use `\cr` or `\crcr` outside an array defined by `\halign` (this is not yet implemented). You must be careful that the 'za' (more generally, the ⟨v⟩ part) does not contain something that reads the special end marker. For instance `\def\x#1{\halign{\x&\cr 1&2\cr}}` is an error. Finally, `*{N}{text}` can be used instead of N occurrences of 'text'.

Note. At the end of Chapter 22 of the T_EXbook, Knuth gives an example of a table where the preamble is `\centerline{#}`. Such a construction cannot be done in Tralics, since a specification of the form `>{\centerline?}c<{?}` would transform into `\centerline?##` and question marks cannot be replaced by braces; you could try `>{\expandafter\centerline?}` and replace the question mark by something that expands to an open brace but contains as many open braces as closing ones, for instance `\expandafter{\iffalse}\fi`. However, it is not possible to put in the `<{?}` part something that the parser considers as a closing brace followed by some other text (otherwise, this closing brace would terminate parsing of the `<{?}` part).

Knuth says that an entry of the form `a}b{c` is legitimate, with respect to this template. This cannot be the case in Tralics, but it would be valid for a template like `>{\bgroup\bf}c<{\egroup}`. This justifies that a table has to be terminated by `\cr` or `\crcr`. In the case of Tralics, this is not needed.

New column types

You can add new column types to the list of existing one, using `\newcolumnntype`, with as argument a letter. For instance:

```
\newcolumnntype{C}{>{$}c<{$}}
\newcolumnntype{L}{>{$}l<{$}}
\newcolumnntype{R}{>{$}r<{$}}
\newcolumnntype{d}[1]{>{\rightdots{#1}}r<{\endrightdots}}
\newcolumnntype{X}{CLR}
\begin{tabular}{*{3}{|c|}d{23}X}
\end{tabular}
```

In this case, the log file will contains (line breaks added before 'r')

```
{Push tabular 2}
array preamble at start: |c||c||c|d{23}X
array preamble after X: |c||c||c|d{23}CLR
array preamble after d: |c||c||c|>{\rightdots {23}}
                        r<{\endrightdots }CLR
array preamble after C: |c||c||c|>{\rightdots {23}}
                        r<{\endrightdots }>{$}c<{$}LR
array preamble after L: |c||c||c|>{\rightdots {23}}
```



```

r<{\endrightdots }>{\$}c<{\$}>{\$}l<{\$}R
array preamble after R: |c||c||c|>{\rightdots {23}}
r<{\endrightdots }>{\$}c<{\$}>{\$}l<{\$}>{\$}r<{\$}
array preamble parse: | c | | c | | c | >>{}
r <<{} >>{} c <<{} >>{} l <<{} >>{} r <<{}

```

Whenever a tabular is seen, optional arguments are read, and then the first argument is handled. In a first pass, * is evaluated. This gives the lines marked ‘at start’. After that, the preamble contains, at toplevel (outside braces) two characters ‘d’ and ‘X’ that are defined to be new column types. These are evaluated one after the other (the order is irrelevant, here alphabetic order is used so that X is expanded first). Since the expansion was non trivial, a second try is made. Note that only a finite numbers of tries are executed. In case of recursion, strange things can happen. Note how you can use commands with arguments (here ‘d’ takes one argument, it is ‘23’).

The table is empty, on purpose, because there are two undefined macros, moreover, because, in the current version of Tralics, dollar signs have to be explicit, and not hidden in a >{}...<{} construction.

Another example

We consider here the following new column types. As you can see, one of them is the character +, another is the character _. The fact that these characters have special catcodes is irrelevant (they cannot be of catcode 1 and 2, because this would interfere with brace matching, and they cannot be of catcode 10, because space characters should be ignored in the preamble).

```

\newcolumntype{L} >{\large\bfseries 2}l <{y}|}
\newcolumntype{+} >{B}l <{D}|}
\newcolumntype_{_}{rlc<{x}}
\newcolumntype{x}{>{b}c<{a}}

```

Consider the four following tables

```

\begin{tabular*}{5cm}[b]{*{4}{_c|}}
a1&a2&a3&a4 & b1&b2&b3&b4 & c1&c2&c3&c4 & d1&d2&d3&d4\\
Wa1&Wa2&Wa3&Wa4 & Wb1&Wb2&Wb3&Wb4 & Wc1&Wc2&Wc3&Wc4 & Wd1&Wd2&Wd3&Wd4\\
\end{tabular*}

```

```

\begin{tabular}{|l|l|rr|cc|}
\hline a&b&c&d&e&f\\
aaa&bbb&ccc&ddd&eee&fff\\
\hline
A&\multicolumn{3}{+}{C}&E&F\\
\multicolumn{2}{|l|}{ab}&c&d&e&f\\
\cline{1-3}\cline{6-6}
aaa&bbb&ccc&ddd &eee&fff\\
\end{tabular}

```

```

\begin{tabular} {l >{\large 1}c <{x}| L > {\large\itshape 3}c <{z}|}
\hline A&B&C\\
\hline 10&10 &1\\
\end{tabular}

```

```

\begin{tabular} {l >{\large 1}c <{x}| L > {\large\itshape 3}x <{z}|}
\hline A&B&C\\
\hline 10&10 &1
\end{tabular}

```

| | | | | | | | | | | | | | | | |
|------|-----|------|------|-----|-----|------|-----|-----|------|------|-----|-----|-----|------|-----|
| a1a2 | a3x | a4 | b1b2 | b3x | b4 | c1c2 | c3x | c4 | d1d2 | d3x | d4 | | | | |
| Wa1 | Wa2 | Wa3x | Wa4 | Wb1 | Wb2 | Wb3x | Wb4 | Wc1 | Wc2 | Wc3x | Wc4 | Wd1 | Wd2 | Wd3x | Wd4 |

| | | | | | |
|-----|-----|-----|-----|-----|-----|
| a | b | c | d | e | f |
| aaa | bbb | ccc | ddd | eee | fff |
| A | BCD | | | E | F |
| ab | | c | d | e | f |
| aaa | bbb | ccc | ddd | eee | fff |

| | | |
|-------|-------------|-----|
| 1Ax | 2By | 3Cz |
| 1100x | 210y | 31z |

| | | |
|-------|-------------|-------|
| 1Ax | 2By | b3Cza |
| 1100x | 210y | b31za |

Table 6.2: Some L^AT_EX tables

You can see the L^AT_EX result on table 6.2. We had to change the `\tabcolsep` of the first table to 0, otherwise, it is two wide. Specifying 5cm as width gives the following warning : *Overfull \hbox (197.12044pt too wide) in alignment*. The table appears centered because it is followed by a `\kern` of 200pt. The XML translation can be found on the Web page. In order to explain what happens, we consider an example:

```
\begin{tabular*}{10pc}[b]{1>{x}r<{y}c}
\hline
a&b&c\ [2pt]
a&\omit b&c\
\multicolumn{1}{1}{A}&B&C\ \hline
\end{tabular*}
```

We explain now the translation. Line numbers refer to the transcript file given below. We do not show the start of the job (initialization). The command `\par` is redefined to do nothing. It is restored on line 107.

- A first procedure is used to start a row. See example, lines 1, 35, 63, 101. It can take an argument, that explains that some vertical spacing should be added between rows, this is the case for line 35. Tokens are read and expanded. Spaces are ignored. If the command that comes next is `\hline`, for instance lines 1 and 101, we add a bottom-border attribute to the previous row (if we are about to start the first row, we mark this as top-border for the current row). If the command is `\cline`, see syntax above, we try to add a bottom-border attribute to some cells (this is not shown on the example). This could fail, because cells can span more than one column; in this case an row of empty cells may be added. If the next token is `\end`, we do nothing (line 101). Otherwise, we create a new `\row` element (lines 3, 36, and 63) and start a cell.
- When a cell is started, the equivalent of `\begingroup` is evaluated, and a `\cell` element is started. The example table has 9 cells, started on lines 4, 12, 22, 37, 45, 53, 64, 82, 92. The ‘Push’ line indicates that a `<cell>` element is created, the line that follows show the ‘begingroup’ command. We consider the first non space token after expansion. This token may be `\omit`, as on line 47. The cell is marked omitted; otherwise the `<u>` part of the current template is added. The current token could be `\multicolumn`. This is a command that takes 3 arguments: an integer *n*, a specification *c*, a value *v*. The example on lines 66 to 69 shows how *n* is read; this might change, L^AT_EX uses the equivalent of `\setcounter` for parsing the number. The specification is handled in the same fashion as an array preamble

(lines 70-71), it should provide only one cell specification. The integer n specifies the span, it will be added as an attribute to the cell, but also used to find the position of the next cell in the row.

The cell is marked ‘omit’; in fact, we push in the current input stream the value of the $\langle u \rangle$ part, $\langle v \rangle$ part and content (last argument) of `\multicolumn`. Braces are added for some obscure reason (see lines 72 and 73).

- Translation proceeds as usual, three events may change it. Assume that we see `&`, as on lines 7, 15, 40, 77, 85. This is only allowed in a table. A final space in the XML element will be removed. A `\endtemplate` token is pushed to be read again. It will be read again on lines 9, 18, 42, 79, and 88. If the cell on the stack is not marked ‘omit’ then the $\langle v \rangle$ part of the template is inserted in the input stream (see lines 17 and 87).
- The translation of `\endtemplate` is trivial: it is an error if not in a cell. Otherwise, the current cell is finished, the equivalent of `\endgroup` is executed, and a new cell is started as above. On line 10 you can see for instance that the cell and the row are in ‘array’ mode, the tabular in horizontal mode, the paragraph and the document in vertical mode.
- Assume that we see `\`, for instance lines 25, 56, 95. We have explained the meaning of this command outside a cell. In any case, an optional argument is read (between the command and the optional argument can be a space or a newline character, so that source line 10 is read; see line 58). The behavior is like `&`, except that `\cr` is pushed instead of `\endtemplate`. (in the case of an optional argument, the code is a bit hacked, see line 30).
- The translation of `\cr` is trivial. We have two versions: one that reads a dimension, and one that does not. On lines 27-29, you can see how ‘scanglue’ reads to optional argument to `\`, on line 31, you see that ‘scanint’ reads 1703, this is the address in a table of the string that represents the glue. The command finishes the cell, finishes the row, evaluates `\endgroup` and starts a new row.
- The `\end` command, with argument ‘tabular’ or ‘tabular*’ behaves in a strange manner: If inside a cell, it behaves like `\`, except that it does not insert `\endtemplate` but itself with the argument, then a `\cr`. We have seen that `\cr` starts a new row. We have seen that no row is started if the token is `\end`. The magic is that the top-stack contains the ‘begin’, and a normal ‘end’ can be used, instead of a hacked one. In our example, the array is terminated by `\hline`, so that the end occurs when we are about to start a new row. In order to show this, we have commented out this line, and re-run the example. The result can be see on line 111, a `\cr` is inserted and evaluated on line 115, the second `\end` is evaluated on line 19.
- The commands `\cr`, `\crr`, `\span`, `\halign`, `\valign`, `\noalign` are not implemented as described in the `TEXbook`.

This is the transcript file.

```

1 [7] \hline
2 [8] a&b&c\[2pt]
3 {Push row 3}
4 {Push cell 4}
5 +stack: level + 3 for cell
6 Character sequence: a.
7 {alignment tab character &}
8 {Text:a}
9 {\endtemplate}
10 {Pop 4: document_v p_v tabular*_h row_a cell_a}
11 +stack: level - 3 for cell
12 {Push cell 4}

```

```

13 +stack: level + 3 for cell
14 Character sequence: xb.
15 {alignment tab character &}
16 {Text:xb}
17 Character sequence: y.
18 {\endtemplate}
19 {Text:y}
20 {Pop 4: document_v p_v tabular*_h row_a cell_a}
21 +stack: level - 3 for cell
22 {Push cell 4}
23 +stack: level + 3 for cell
24 Character sequence: c.
25 {\}
26 {Text:c}
27 +scanint for \->2
28 +scandimen for \->2.0pt
29 {scanglue 2.0pt}
30 {\cr withargs}
31 +scanint for \cr withargs->1703
32 {Pop 4: document_v p_v tabular*_h row_a cell_a}
33 +stack: level - 3 for cell
34 {Pop 3: document_v p_v tabular*_h row_a}
35 [9] a&\omit b&c\
36 {Push row 3}
37 {Push cell 4}
38 +stack: level + 3 for cell
39 Character sequence: a.
40 {alignment tab character &}
41 {Text:a}
42 {\endtemplate}
43 {Pop 4: document_v p_v tabular*_h row_a cell_a}
44 +stack: level - 3 for cell
45 {Push cell 4}
46 +stack: level + 3 for cell
47 Character sequence: b.
48 {alignment tab character &}
49 {Text:b}
50 {\endtemplate}
51 {Pop 4: document_v p_v tabular*_h row_a cell_a}
52 +stack: level - 3 for cell
53 {Push cell 4}
54 +stack: level + 3 for cell
55 Character sequence: c.
56 {\}
57 {Text:c}
58 [10] \multicolumn{1}{l}{A}&B&C\\\hline
59 {\cr}
60 {Pop 4: document_v p_v tabular*_h row_a cell_a}
61 +stack: level - 3 for cell
62 {Pop 3: document_v p_v tabular*_h row_a}
63 {Push row 3}
64 {Push cell 4}
65 +stack: level + 3 for cell
66 {Push argument 5}
67 Character sequence: 1.
68 {Text:1}

```

```

69 {Pop 5: document_v p_v tabular*_h row_a cell_a}
70 array preamble at start: 1
71 array preamble parse: 1
72 {begin-group character {}}
73 +stack: level + 4 for brace
74 Character sequence: A.
75 {end-group character {}}
76 +stack: level - 4 for brace
77 {alignment tab character &}
78 {Text:A}
79 {\endtemplate}
80 {Pop 4: document_v p_v tabular*_h row_a cell_a}
81 +stack: level - 3 for cell
82 {Push cell 4}
83 +stack: level + 3 for cell
84 Character sequence: xB.
85 {alignment tab character &}
86 {Text:xB}
87 Character sequence: y.
88 {\endtemplate}
89 {Text:y}
90 {Pop 4: document_v p_v tabular*_h row_a cell_a}
91 +stack: level - 3 for cell
92 {Push cell 4}
93 +stack: level + 3 for cell
94 Character sequence: C.
95 {\}
96 {Text:C}
97 {\cr}
98 {Pop 4: document_v p_v tabular*_h row_a cell_a}
99 +stack: level - 3 for cell
100 {Pop 3: document_v p_v tabular*_h row_a}
101 [11] \end{tabular*}
102 {\end}
103 {\end tabular*}
104 {\endtabular*}
105 {Pop 2: document_v p_v tabular*_h}
106 {\endgroup (for env)}
107 +stack: restoring \par=\par
108 +stack: ending environment tabular*; resuming document.
109 +stack: level - 2 for environment
110 Character sequence: .

```

Alternate version, where the final `\\hline` is commented out

```

111 [11] \end{tabular*}
112 {\end}
113 {Text:C}
114 {\end tabular*}
115 {\cr}
116 {Pop 4: document_v p_v tabular*_h row_a cell_a}
117 +stack: level - 3 for cell
118 {Pop 3: document_v p_v tabular*_h row_a}
119 {\end}
120 {\end tabular*}
121 {\endtabular*}

```

6.9 Actions declared in the configuration file

An action is defined by a name, an equals sign, and a value. Optional spaces can be used. The syntax of DocType and DocAttrib is special. In all other cases, double quotes must delimit the value. Some names are special, see for instance lines 118-122. All other names contain only letters, digits, and underscores. If the name starts with ‘xml’ this defines an element. If the name starts with ‘att’ this defines an attribute. Unless specified otherwise, if a name is given more than once, the last occurrence wins. These are all the actions:

- `makefo`, `makehtml`, `checkxml`, `makepdf`, `makedvi`, `dvitops`, `generatedvi`, `generateps`. These specify actions for the Raweb. For details see the Web page.
- `theme_vals` (Raweb only). This defines the list of all Inria Themes. See example line 42.
- `section_vals` (Raweb only). This defines the list of all valid sections in the Raweb. See examples on lines 43 and 44. If the first character is a ‘+’ sign, the remaining of the line is appended to the previous value. You must use a slash as separator.
- `ur_vals` (Raweb only). This defines the list of all valid UR (name and value) in the Raweb. See examples on lines 45 and 46. If the first character is a ‘+’ sign, the remaining of the line is appended to the previous value. You must use a slash as separator. After each name, you must give a value (if empty, the name will be used). Only letters can be used in the name.
- `Language`. The main element contains an attribute pair of the form `language='english'`. Specifying ‘Language’ in the configuration file modifies the name of the attribute.
- `lang_fr`, `lang_en`. See above. These two commands can parameterize the value of the attribute. The attribute pair is added to the main element: just after translation of the preamble in the Raweb case, by the `titlepage` command, or in the at-begin-document hook. The language chosen is the “default language”.
- `url_font`. If you specify a value ‘\foo’, then `\def\urlfont{\foo}` will be added to the document-hook. The command is empty by default. It is the font used for urls.
- `no_footnote_hack`. If the value is ‘true’, then no hack is applied to footnotes. As a consequence, the translation of

```
a\footnote{B}\footnote{C\par D}
```

is

```
<p>a<note id='uid1' place='foot'><p>B</p>
</note><note id='uid2' place='foot'><p>C</p>
<p>D</p>
</note></p>
```

The hack is that, if the footnote holds only a single `<p>` element, it will be removed, and the translation is this:

```
<p>a<note id='uid1' place='foot'>B</note>
<note id='uid2' place='foot'><p>C</p> <p>D</p>
</note></p>
```

- `xml_footnote_name`, `att_place`, `att_foot_position`. These assignments explain how to change the name of the `<note>` element, the attribute name and the attribute value in ‘place=’foot’. Thus, the translation of the example above can be:

```

<p>a<Footnote id='uid1' Place='as a note'><p>B</p>
</Footnote><Footnote id='uid2' Place='as a note'><p>C</p>
<p>D</p>
</Footnote></p>

```

- `everyjob`. If the value is `'\foo'`, then `\everyjob={\foo}` will be executed. In fact, this is the last line of the bootstrap code (it will be inserted just above line '[66]' in the example above). After these lines are translated, the value of the token list `\everyjob` is inserted in the input stream, before the first line of the input file.

- `use_font_elt`. The default translation of `{\tiny \textit{A}B}` is

```
<hi rend='small'><hi rend='it'>A</hi></hi><hi rend='small'>B</hi>
```

In this case, you cannot change the name of the element: the parameters described below will change the value of the attribute. If you say `'use_font_elt=true'`, this changes the translation as follows (the parameters described below will change the name of the element).

```
<small><it>A</it></small><small>B</small>
```

- `use_all_sizes`. By default, Tralics knows only three font sizes, normal, smaller, larger. If the switch is true, the previous example translates to

```
<small4><it>A</it></small4><small4>B</small4>
```

or

```
<hi rend='small4'><hi rend='it'>A</hi></hi><hi rend='small4'>B</hi>
```

- `xml_font_small`, `xml_font_large`, `xml_font_normalize`. These elements specify the names of small, large and normal size. The translation of `{\tiny a}{\normalsize b}{\large c}` can be changed to

```
<Small>a</Small>b<Large>c</Large>
```

Currently, no tag is inserted in the case of `\normalsize`. In some cases, this is wrong.

- `xml_font_small1`, `xml_font_small2`, `xml_font_small3`, `xml_font_small4`, `xml_font_large1`, `xml_font_large2`, `xml_font_large3`, `xml_font_large4`, `xml_font_large5`. These specify the names of all font sizes for the case where all sizes are used. For instance, the translation of

```
{\tiny a}{\scriptsize b}{\footnotesize c}{\small d}%
{\normalsize e}{\large f}{\Large h}{\LARGE h}{\huge i}{\Huge j}
```

is

```
<small4>a</small4><small3>b</small3><small2>c</small2><small1>d</small1>
e<large1>f</large1><large2>h</large2><large3>h</large3><large4>i</large4>
<large5>j</large5>
```

It can be changed to

```
<S4>a</S4><S3>b</S3><S2>c</S2><S1>d</S1>e<L1>f</L1>
<L2>h</L2><L3>h</L3><L4>i</L4><L5>j</L5>
```

- `xml_font_upright`, `xml_font_it`, `xml_font_slanted`, `xml_font_sc`. These four parameters specify the shape. The first one is currently unused. The default translation of

```
{\upshape a}{\itshape b}{\slshape c}{\scshape d}
```

is

```
a<hi rend='it'>b</hi><hi rend='slanted'>c</hi><hi rend='sc'>d</hi>
```

or (using elements):

```
a<it>b</it><slanted>c</slanted><sc>d</sc>
```

You can change it to:

```
a<hi rend='It'>b</hi><hi rend='Slanted'>c</hi><hi rend='Sc'>d</hi>
```

or

```
a<It>b</It><Slanted>c</Slanted><Sc>d</Sc>
```

- `xml_font_medium`, `xml_font_bold`. These two parameters specify the series. The first is currently unused.

- `xml_font_roman`, `xml_font_tt`, `xml_font_sansserif`. These three parameters specify the family. The first is currently unused. The translation of

```
{\bfseries a}{\ttfamily b}{\sffamily c}
```

is

```
<hi rend='bold'>a</hi><hi rend='tt'>b</hi><hi rend='sansserif'>c</hi>
```

but you can change it to

```
<bold>a</bold><tt>b</tt><sansserif>c</sansserif>
```

or

```
<B>a</B><TT>b</TT><SF>c</SF>
```

- `xml_sup_name`, `xml_sub_name`. These two parameters control the name of superscript and subscripts in text mode. The translation of

```
\textsuperscript{a}\textsubscript{b}
```

is by default

```
<hi rend='sup'>a</hi><hi rend='sub'>b</hi>
```

but can be changed to

```
<Super>a</Super><Sub>b</Sub>
```

- `xml_oldstyle_name`, `xml_overline_name`, `xml_underline_name`. These are font-like commands. The translation of

```
\overline{foo}\underline{bar}\oldstylenums{xyz}
```

is

```
<hi rend='overline'>foo</hi><hi rend='underline'>bar</hi>
<hi rend='oldstyle'>xyz</hi>
```

but it can be changed to

```
<Overline>foo</Overline><Underline>bar</Underline><Oldstyle>xyz</Oldstyle>
```

- `xml_labelitem_name`. This specifies the name of the element containing the optional argument of an `\item`. Default value is 'label'.
- `xml_item_name`. This specifies the name of the element containing an `\item`. Default value is 'item'.
- `xml_gloitem_name`. This specifies the name of the element containing the first argument of `\glo`. Default value is 'label'.
- `xml_list_name`. This is the name of the element generated by list environments like `itemize`, `enumerate` and `description`.

- `xml_glo_name`. This controls the name of the glossary.
- `att_gloss_type`. In a glossary, a `<list>` element is created with attribute `type='gloss'`. This variable can be used to change the value of the attribute.

Example:

```
\begin{description}
\item [a] etc
\begin{enumerate}
\item [b] etc
\begin{itemize}
\item [c] etc
\end{itemize}
\end{enumerate}
\end{description}
\begin{glossaire}\glo{a}{b}
\end{glossaire}
```

This is the translation

```
<list type='description'><label>a</label>
<item id='uid1'><p>etc</p>
<list type='ordered'><label>b</label>
<item id='uid2'><p>etc</p>
<list type='simple'><label>c</label>
<item id='uid3'><p>etc</p>
</item></list>
</item></list>
</item></list>
<list type='gloss'><head>Glossary</head><label>a</label><item><p>b</p>
</item></list>
```

It can be changed to:

```
<List type='description'><Label>a</Label>
<Item id='uid1'><p>etc</p>
<List type='ordered'><Label>b</Label>
<Item id='uid2'><p>etc</p>
<List type='simple'><Label>c</Label>
<Item id='uid3'><p>etc</p>
</Item></List>
</Item></List>
</Item></List>
<List type='Glossary'><Head>Glo</Head><GItem>a</GItem><Item><p>b</p>
</Item></List>
```

- `xml_head_name`. This is the name of the `<head>` element in a title.
- `xml_div0_name`, `xml_div1_name`, `xml_div2_name`, `xml_div3_name`, `xml_div4_name`, `xml_div5_name`, `xml_div6_name`. This specifies the name of a section. The default translation of

```
\part{A}\chapter{B}\section{C}\subsection{D}
\subsubsection{E}\paragraph{F}\subparagraph{G}
etc
\endsec{\chapter}ok?
```

is

```
<div0 id='uid1'><head>A</head>
<div1 id='uid2'><head>B</head>
<div2 id='uid3'><head>C</head>
<div3 id='uid4'><head>D</head>
<div4 id='uid5'><head>E</head>
<div5 id='uid6'><head>F</head>
<div6 id='uid7'><head>G</head>
<p>etc</p>
</div6></div5></div4></div3></div2></div1>
<p>ok?</p>
```

but you can use these commands to get:

```
<Part id='uid1'><Head>A</Head>
<Chapter id='uid2'><Head>B</Head>
<Section id='uid3'><Head>C</Head>
<Subsection id='uid4'><Head>D</Head>
<Subsubsection id='uid5'><Head>E</Head>
<Paragraph id='uid6'><Head>F</Head>
<Subparagraph id='uid7'><Head>G</Head>
<p>etc</p>
</Subparagraph></Paragraph></Subsubsection></Subsection></Section>
</Chapter>
<p>ok?</p>
```

- `xml_fbox_name`, `xml_scalebox_name`, `xml_box_name`. These control the names of various box commands.
- `att_boxed`. This is the name of the attributed added by `\fbox`, unless it is a `includegraphics`.
- `att_framed`. If a figure is in a `\fbox`, it will have `framed='true'`. This changes the name of the attribute.
- `att_box_width`. Name of width attribute for commands like `\framebox`.
- `att_box_scale`. This controls the name scale attribute, unless in a figure. Example

```
\fbox{foo}\scalebox{0.3}{foo}\framebox(10,10){foo}\framebox[10pt][11]{foo}
```

translates as

```
<fbox rend='boxed'>foo</fbox>
<scalebox scale='0.3'>foo</scalebox>
<pic-framebox width='51.2149' height='51.2149'
  framed='true'>foo</pic-framebox>
<fbox width='10.0pt' rend='boxed'>foo</fbox>
```

You can change it to

```
<Fbox rend='Boxed'>foo</Fbox>
<Scalebox Boxscale='0.3'>foo</Scalebox>
<Box Width='10' Height='10' Framed='true'>foo</Box>
<Fbox BoxWidth='10.0pt' rend='Boxed'>foo</Fbox>
```

- `xml_keywords_name`. This is the name of the element generated by `'motscle'` environment. The default value is `'keywords'`.

- `xml_term_name`. This is the name of the element used for each keyword. The default value is ‘term’. For instance,

```
\begin{motsclé} first, second, {(max,+)}\end{motsclé}
```

translates to

```
<keywords>
<term>first</term>
<term>second</term>
<term>(max,+)</term>
</keywords>
```

and can be changed to

```
<Keywords>
<Term>first</Term>
<Term>second</Term>
<Term>(max,+)</Term>
</Keywords>
```

- `xml_mbox_name`. This is the name of the element used by the `\mbox`, unless the result is trivial.
- `xml_rotatebox_name`. This is the name of the element generated by `\rotatebox`.
- `att_rotate_angle`. This controls the name of the `angle` attribute in the case of `\rotatebox`.
- `att_noindent`. This is the name of the `noindent` attribute.
- `xml_subfigure_name`. This is the name of the element generated by `\subfigure`. The default value is ‘subfigure’.
- `xml_text_name`. This is the name of the element generated by `\subfigure`, containing the required argument.
- `xml_subfigure_name`. This is the name of the element generated by `\subfigure`, containing the optional argument.

For instance, the translation of

```
\noindent \mbox{\xbox{a}{b}} \rotatebox{30}{x}
\subfigure[a]{b}
```

is

```
<p noindent='true'>
  <mbox><a>b</a></mbox>
  <pic-rotatebox angle='30'>x</pic-rotatebox>
  <subfigure><leg>a</leg><text>b</text></subfigure>
</p>
```

but can be changed to

```
<p Noindent='true'>
  <Mbox><a>b</a></Mbox>
  <Rotatebox RotAngle='30'>x</Rotatebox>
  <Subfigure><SFCapt>a</SFCapt><SFText>b</SFText></Subfigure>
</p>
```

- `xml_topic_name`, `xml_topic_title`, `att_topic_num`. These are used for topics (Raweb only). The default values for the elements are ‘topic’ and ‘t_titre’, for the attribute it is ‘num’.

- `xml_caption_name`, `xml_graphics_name`, `xml_figure_env_name`, `xml_table_env_name`, `xml_Table_name`. These names are used by some commands and environments. For instance, consider the following code. Normally, you put a tabular in a table, an image in a figure (as in B or D), this gives smaller XML objects.

```

\begin{figure}[htbp]
\begin{tabular}{c}a\end{tabular}
\caption{A}
\begin{table}[htbp]
\begin{tabular}{c}a\end{tabular}
\caption{B}
\end{table}
\begin{table}[htbp]
\includegraphics{a}
\caption{C}
\end{table}
\begin{figure}[htbp]
\includegraphics{a}
\caption{D}
\end{figure}

```

This is the default translation:

```

<figure rend='array' id='uid1'>
  <head>A</head>
  <p>
    <table rend='inline'>
      <row><cell halign='center'>a</cell></row></table></p>
</figure>
<table id='uid2'>
  <head>B</head><row><cell halign='center'>a</cell></row>
</table>
<table id='uid3'><head>C</head><figure rend='inline' file='a'/></table>
<figure file='a' id='uid4'><head>D</head></figure>

```

This shows how all names can be parameterized.

```

<Figureenv rend='array' id='uid1'>
  <Caption><p>A</p></Caption>
  <p>
    <Table rend='inline'>
      <Row><Cell Align='Ccenter'>a</Cell></Row></Table></p>
</Figureenv>
<Tableenv id='uid2'>
  <Caption><p>B</p></Caption>
  <Row><Cell Align='Ccenter'>a</Cell></Row></Tableenv>
<Tableenv id='uid3'>
  <Caption><p>C</p></Caption><Figure rend='inline' File='a'/></Tableenv>
<Figureenv File='a' id='uid4'>
  <Caption><p>D</p></Caption></Figureenv>

```

- `xml_xref_name`. This is the name of the element associated to `\url`.
- `xml_scaption_name`. Special caption.
- `xml_project_name`. This controls the name of the `<projet>` element.

- `xml_accueil_name`. This controls the name of the `<accueil>` element.
- `xml_composition_ra_name`. This controls the name of the `<composition>` element. In fact, in Raweb mode, when a environment ‘RAsection’ is closed, and the argument of the environment is the value of this variable, then modules inside the element are replaced by their content.
- `xml_xtheorem_name`. This controls the theorem names. This is an example of theorems

```

\theorembodyfont{\sl}
\theoremstyle{break}
\newtheorem{Cor}{Corollary}
\theoremstyle{plain}
%\newcounter{section}
\setcounter{section}{17}
\newtheorem{Exa}{Example}[section]
{\theorembodyfont{\rmfamily}\newtheorem{Rem}{Remark}}
\theoremstyle{marginbreak}
\newtheorem{Lem}[Cor]{lemma}
\theoremstyle{change}
\theorembodyfont{\small\itshape} \newtheorem{Def}[Cor]{Definition}
\theoremheaderfont{\scshape}
\def\Lenv#1{\texttt{#1}}

\begin{Cor}
  This is a sentence typeset in the theorem environment \Lenv{Cor}.
\end{Cor}
\begin{Exa}
  This is a sentence typeset in the theorem environment \Lenv{Exa}.
\end{Exa}
\begin{Rem}
  This is a sentence typeset in the theorem environment \Lenv{Rem}.
\end{Rem}
\begin{Lem}[Ben User]
  This is a sentence typeset in the theorem environment \Lenv{Lem}.
\end{Lem}
\begin{Def}[Very Impressive definition]
  This is a sentence typeset in the theorem environment \Lenv{Def}.
\end{Def}

```

This is the translation:

```

<p><hi rend='sc'>Corollary 1 </hi><hi rend='slanted'>This
is a sentence typeset in the theorem environment </hi>
<hi rend='slanted'><hi rend='tt'>Cor</hi></hi><hi rend='slanted'>.</hi></p>

<p><hi rend='sc'>Example 17.1 </hi><hi rend='slanted'>This
is a sentence typeset in the theorem environment </hi>
<hi rend='slanted'><hi rend='tt'>Exa</hi></hi><hi rend='slanted'>.</hi></p>

<p><hi rend='sc'>Remark 1 </hi>This is a sentence typeset in
the theorem environment <hi rend='tt'>Rem</hi>.</p>

<p><hi rend='sc'>lemma 2 (Ben User) </hi><hi rend='slanted'>
This is a sentence typeset in the theorem environment </hi><hi
rend='slanted'><hi rend='tt'>Lem</hi></hi><hi rend='slanted'>.</hi></p>

```

```

<p><hi rend='sc'>Definition 3 (Very Impressive definition) </hi><hi
  rend='small' /><hi rend='small'><hi rend='it'>
This is a sentence typeset in the theorem environment </hi></hi>
<hi rend='small'><hi rend='it'><hi rend='tt'>Def</hi></hi></hi>
<hi rend='small'><hi rend='it'>.</hi></hi></p>

```

This may be changed to:

```

<Xtheorem id='uid1'><p><Sc>Corollary 1 </Sc><Slanted>This
  is a sentence typeset in the theorem environment </Slanted>
  <Slanted><TT>Cor</TT></Slanted><Slanted>.</Slanted></p>
</Xtheorem>
<Xtheorem id='uid2'><p><Sc>Example 17.1 </Sc><Slanted>This
  is a sentence typeset in the theorem environment </Slanted>
  <Slanted><TT>Exa</TT></Slanted><Slanted>.</Slanted></p>
</Xtheorem>
<Xtheorem id='uid3'><p><Sc>Remark 1 </Sc>This
  is a sentence typeset in the theorem environment <TT>Rem</TT>.</p>
</Xtheorem>
<Xtheorem id='uid4'><p><Sc>lemma 2 (Ben User) </Sc><Slanted>
  This is a sentence typeset in the theorem environment </Slanted>
  <Slanted><TT>Lem</TT></Slanted><Slanted>.</Slanted></p>
</Xtheorem>
<Xtheorem id='uid5'><p><Sc>Definition 3 (Very Impressive definition) </Sc>
  <S1 /><S1><It>
  This is a sentence typeset in the theorem environment </It></S1><S1><It>
  <TT>Def</TT></It></S1><S1><It>.</It></S1></p>
</Xtheorem>

```

- `xml_theorem_head`. This controls the element that contains the optional argument of the theorem, if the alternate syntax is used (default is `alt_head`, shown as `AltHead` below).
- `xml_theorem_name`. This is used for theorems. The value of the variable is unused. Redefining the command changes the meaning of `\@begintheorem` to `\@xbegintheorem`. As a consequence the translation of the previous theorem is:

```

<Xtheorem style='break' type='Cor' id='uid1'>
  <Head>Corollary</Head>
  <p>This is a sentence typeset in the theorem environment <TT>Cor</TT>.</p>
</Xtheorem>
<Xtheorem style='plain' type='Exa' id='uid2'>
  <Head>Example</Head>
  <p>This is a sentence typeset in the theorem environment <TT>Exa</TT>.</p>
</Xtheorem>
<Xtheorem style='plain' type='Rem' id='uid3'>
  <Head>Remark</Head>
  <p>This is a sentence typeset in the theorem environment <TT>Rem</TT>.</p>
</Xtheorem>
<Xtheorem style='marginbreak' type='Lem' id='uid4'>
  <Head>lemma</Head>
  <AltHead>Ben User</AltHead>
  <p>This is a sentence typeset in the theorem environment <TT>Lem</TT>.</p>
</Xtheorem>
<Xtheorem style='change' type='Def' id='uid5'>

```

```

<Head>Definition</Head>
<AltHead>Very Impressive definition</AltHead>
<p>This is a sentence typeset in the theorem environment <TT>Def</TT>.</p>
</Xtheorem>

```

- `xml_biblio` This controls the name of the element that contains the bibliography.
- `att_box_pos`. This controls the name of the `pos` attribute in a box.
- `att_full`. If a star follows `\circle`, an attribute pair `full='true'` is added. This controls the attribute name.
- `att_table_width`. Name of `width` attribute for a table.
- `att_pos`. Used by environments like ‘minipage’, for the horizontal position.
- `att_vpos`. Used by ‘minipage’, for the vertical position.
- `att_inner_pos`. Used by ‘minipage’, for the third optional argument.
- `xml_figure_name`. This controls the name of the element produced by `\includegraphics`.
- `att_file`, `att_angle`, `att_scale`, `att_clip`, `att_width`, `att_height`. These control the attribute names set by `\includegraphics`.

The translation of

```

\fbbox{\includegraphics[clip=, angle=90, scale=3, %
width=10cm, height=3m]{A}}

```

is

```

<figure framed='true' rend='inline' height='3m'
width='10cm' scale='3' angle='90' clip='true' file='A'/>

```

You can change it to

```

<Figure Framed='true' rend='inline' Height='3m' Width='10cm' Scale='3'
Angle='90' Clip='true' File='A'/>

```

- `att_rend`, `att_fbox_rend`. These control the name of the `rend` attribute (why is ‘fbox’ a special case ?)
- `att_flush_right`, `att_flush_left`, `att_quotation`, `att_quote`, `att_centering`. Environments like ‘center’, etc., modify (locally) an integer that asks each `<p>` to get an attribute `rend` with value ‘flushed-left’, ‘flushed-right’, ‘center’ or ‘quote’. These environment execute the equivalent of `\par` just after the `\begin` (before the parameter is changed), and just after the `\end`. the `\centering` command modifies the integer and the attribute list of the current `\p`. These assignments show how to control the value of the attribute. Example:

```

\begin{center}A\end{center}
\begin{quote}A\end{quote}
\begin{quotation}A\end{quotation}
\begin{flushleft}A\end{flushleft}
\begin{flushright}A\end{flushright}

```

The default translation is

```

<p rend='center'>A</p>
<p rend='quoted'>A</p>
<p rend='quoted'>A</p>
<p rend='flushed-left'>A</p>
<p rend='flushed-right'>A</p>

```

but you can change it to

```
<p rend='Centered'>A</p>
<p rend='Quote'>A</p>
<p rend='Quotation'>A</p>
<p rend='RaggedRight'>A</p>
<p rend='RaggedLeft'>A</p>
```

- `att_cell_bottomborder`, `att_cell_leftborder`, `att_cell_rightborder`, `att_cell_topborder`. Specifying a `\hline` in a table, or vertical rules may add attributes to cells. The name is `left-border` etc. The name of the attribute can be changed via these commands.
- `att_row_spaceafter`, `att_row_spacebefore`. Inside a table, an optional argument to `\\` asks for a space before or after the current row. This is done by adding a `spacebefore` and `spaceafter` attribute. The name of the attribute can be changed via this variable.
- `att_cols`, `att_halign`, `att_cell_left`, `att_cell_right`, `att_cell_center`. The translation of a command of the form `\multicolumn{1}{c}{...}` is a cell, with attributes `halign='center'` and `cols='1'`. These parameters allow you to change the name of the attribute name or value.
- `xml_table_name`, `xml_row_name`, `xml_cell_name`. These commands allow you to change the names of the elements `<table>`, `<row>` and `<cell>`.

The default translation of

```
\begin{tabular}{|r|l|c}
\hline a&b&\multicolumn{1}{c}{x}\hline\\[2pt]
\end{tabular}
 $\begin{array}{|r|l|}
a&b&\multicolumn{1}{c}{x}\\[2pt]
\end{array}$ 
```

is

```
<table rend='inline'>
  <row spaceafter='2.0pt' top-border='true'>
    <cell left-border='true' halign='right' right-border='true'>a</cell>
    <cell halign='left' right-border='true'>b</cell>
    <cell halign='center' right-border='true' cols='1'>x</cell>
  </row>
</table>
<formula type='inline'>
  <math xmlns='http://www.w3.org/1998/Math/MathML'>
    <mtable>
      <mtr>
        <mtd columnalign='right'><mi>a</mi></mtd>
        <mtd columnalign='left'><mi>b</mi></mtd>
        <mtd colspan='1'><mi>x</mi></mtd>
      </mtr>
    </mtable>
  </math>
</formula>
```

You can parameterize the `<table>`, using all parameters shown above. Of course, it has no influence on the `<mtable>`.

```
<Table rend='inline'>
  <Row SpaceAfter='2.0pt' TopBorder='true'>
```



```

      <Cell LeftBorder='true' Align='Cright' RightBorder='true'>a</Cell>
      <Cell Align='Cleft' RightBorder='true'>b</Cell>
      <Cell Align='Ccenter' RightBorder='true' cols='1'>x</Cell>
    </Row>
  </Table>

```

- `att_prenom`, `att_full_first`, `att_nom`, `att_particule`, `att_junior`, `xml_bpers_name`. These control typesetting of `\bpers` in a bibliography. The default translation of

```
\bpers[E]{A}{B}{C}{D}
```

is

```
<bpers prenom='A' part='B' nom='C' junior='D' prenomcomplet='E' />
```

and you can change it to

```
<Pers First='A' von='B' Last='C' Junior='D' Full='E' />
```

- `xml_picture_name`. This can be used to change the name of the element created by the ‘picture’ environment (it is ‘picture’).
- `xml_bezier_name`. This specifies the name of the element produced by `\bezier`.
- `xml_put_name`. This specifies the name of the element produced by `\put`.
- `xml_arc_name`. This specifies the name of the element produced by `\arc`.
- `xml_scaleput_name`. This specifies the name of the element produced by `\scaleput`.
- `xml_multiput_name`. This specifies the name of the element produced by `\multiput`.
- `xml_line_name`. This specifies the name of the element produced by `\line`.
- `xml_vector_name`. This specifies the name of the element produced by `\vector`.
- `xml_oval_name`. This specifies the name of the element produced by `\oval`.
- `xml_dashline_name`. This specifies the name of the element produced by `\dashline`.
- `xml_dottedline_name`. This specifies the name of the element produced by `\dottedline`.
- `xml_circle_name`. This specifies the name of the element produced by `\circle`.
- `xml_bigcircle_name`. This specifies the name of the element produced by `\bigcircle`.
- `xml_curve_name`. This specifies the name of the element produced by `\curve`.
- `xml_closecurve_name`. This specifies the name of the element produced by `\closecurve`.
- `xml_tagcurve_name`. This specifies the name of the element produced by `\tagcurve`.
- `xml_thicklines_name`. This specifies the name of the element produced by `\thicklines`.
- `xml_thinlines_name`. This specifies the name of the element produced by `\thinlines`.
- `xml_linethickness_name`. This specifies the name produced by `\linethickness`.
- `att_xpos`, `att_ypos`, `att_xdir`, `att_ydir`, `att_size`, `att_dx`, `att_dy`, `att_repeat`. These attributes are used by some picture commands.

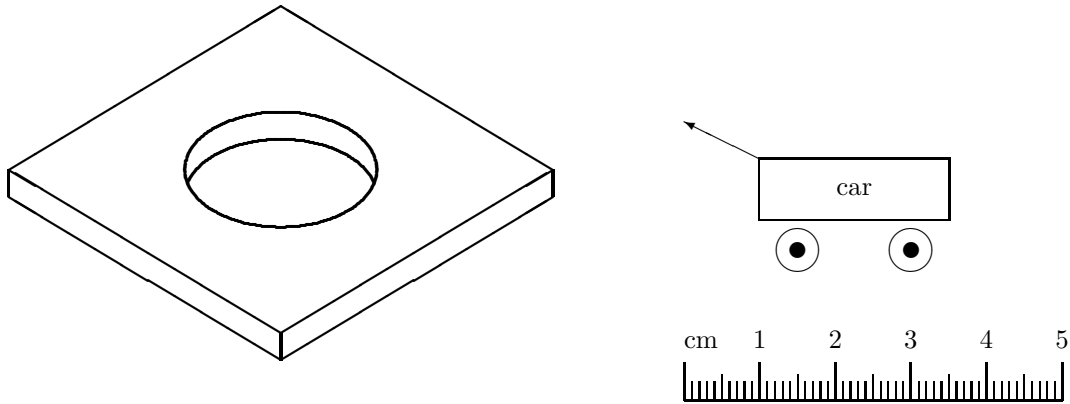


Figure 6.1: Two examples of a picture environment from [2]

- `att_xscale`, `att_xscaley`, `att_yscale`, `att_yscalex`. Commands `\xscale`, `\xscaley`, `\yscale`, and `\yscalex` read values and store them somewhere. Whenever needed, they are put as attribute on the elements. The default name of the argument is the name of the command, but this can be changed.
- `att_curve_nbpts`. Commands like `\arc`, `\bigcircle`, `\closecurve`, `\tagcurve`, `\curve` take an optional argument, the number of points. It will be stored in an attribute, named ‘nb-symb’. This parameter can change the name of the attribute.
- `att_unit_length`. Inside a picture environment, the value of `\unitlength` is stored in an attribute named ‘unit-length’. It can be changed using this command.

The example of a picture environment is

```

\setlength{\unitlength}{1.8mm}
\begin{picture}(40,30)
  \thicklines
  \multiput(20,5)(20,12){2}{\line(0,-1){2}\line(-5,3){20}}
  \multiput(20,5)(-20,12){2}{\line(5,3){20}}
  \put(20,3){\line(5,3){20}}
  \put(20,3){\line(-5,3){20}}
  \put(0,15){\line(0,1){2}}
  \linethickness{1pt}
  \put(20,5){
    \renewcommand{\xscale}{1}
    \renewcommand{\xscaley}{-1}
    \renewcommand{\yscale}{0.6}
    \renewcommand{\yscalex}{0.6}
    \scaleput(10,10){\bigcircle{10}}
    \put(0,-2){%
      \scaleput(10,10){\arc(5,0){121}}
      \scaleput(10,10){\arc(5,0){-31}}}
  }
\end{picture}

```

it is translated as

```

<picture width='204.85962' height='153.64471'>
  <pic-thicklines/>
  <pic-multiput xpos='102.42981' ypos='25.60745'

```

```

    repeat='2' dx='102.42981' dy='61.45789'>
    <pic-line xdir='0' ydir='-1' width='10.24298' />
    <pic-line xdir='-5' ydir='3' width='102.42981' />
</pic-multiput>
<pic-multiput xpos='102.42981' ypos='25.60745'
    repeat='2' dx='-102.42981' dy='61.45789'>
    <pic-line xdir='5' ydir='3' width='102.42981' />
</pic-multiput>
<pic-put xpos='102.42981' ypos='15.36447'>
    <pic-line xdir='5' ydir='3' width='102.42981' /></pic-put>
<pic-put xpos='102.42981' ypos='15.36447'>
    <pic-line xdir='-5' ydir='3' width='102.42981' /></pic-put>
<pic-put xpos='0' ypos='76.82236'>
    <pic-line xdir='0' ydir='1' width='10.24298' /></pic-put>
<pic-linethickness size='1pt' />
<pic-put xpos='102.42981' ypos='25.60745'>
    <pic-scaleput xpos='51.2149' ypos='51.2149'
        xscale='1' yscale='0.6' xscalex='-1' yscalex='0.6'>
    <pic-bigcircle size='10' unit-length='5.12149' />
</pic-scaleput>
<pic-put xpos='0' ypos='-10.24298'>
    <pic-scaleput xpos='51.2149' ypos='51.2149' xscale='1'
        yscale='0.6' xscalex='-1' yscalex='0.6'>
    <pic-arc angle='121' xpos='25.60745' ypos='0' unit-length='5.12149' />
</pic-scaleput>
<pic-scaleput xpos='51.2149' ypos='51.2149' xscale='1'
    yscale='0.6' xscalex='-1' yscalex='0.6'>
    <pic-arc angle='-31' xpos='25.60745' ypos='0' unit-length='5.12149' />
</pic-scaleput>
</pic-put>
</pic-put>
</picture>

```

You can parameterize it so as to obtain this

```

<Picture Width='204.85962' Height='153.64471'>
  <Thicklines />
  <Multiput Xpos='102.42981' Ypos='25.60745' Repeat='2'
    Dx='102.42981' Dy='61.45789'>
    <Line Xdir='0' Ydir='-1' Width='10.24298' />
    <Line Xdir='-5' Ydir='3' Width='102.42981' />
  </Multiput>
  <Multiput Xpos='102.42981' Ypos='25.60745' Repeat='2'
    Dx='-102.42981' Dy='61.45789'>
    <Line Xdir='5' Ydir='3' Width='102.42981' />
  </Multiput>
  <Put Xpos='102.42981' Ypos='15.36447'>
    <Line Xdir='5' Ydir='3' Width='102.42981' />
  </Put>
  <Put Xpos='102.42981' Ypos='15.36447'>
    <Line Xdir='-5' Ydir='3' Width='102.42981' />
  </Put>
  <Put Xpos='0' Ypos='76.82236'>
    <Line Xdir='0' Ydir='1' Width='10.24298' />
  </Put>
  <Linethickness Size='1pt' />
  <Put Xpos='102.42981' Ypos='25.60745'>
    <Scaleput Xpos='51.2149' Ypos='51.2149' Xscale='1' Yscale='0.6'

```

```

XscaleY='-1' YscaleX='0.6'>
<Bigcircle Size='10' Ulength='5.12149'/>
</Scaleput>
<Put Xpos='0' Ypos='-10.24298'>
  <Scaleput Xpos='51.2149' Ypos='51.2149' Xscale='1'
    Yscale='0.6' XscaleY='-1' YscaleX='0.6'>
    <Arc Angle='121' Xpos='25.60745' Ypos='0' Ulength='5.12149'/>
  </Scaleput>
  <Scaleput Xpos='51.2149' Ypos='51.2149' Xscale='1'
    Yscale='0.6' XscaleY='-1' YscaleX='0.6'>
    <Arc Angle='-31' Xpos='25.60745' Ypos='0' Ulength='5.12149'/>
  </Scaleput>
</Put>
</Put>
</Picture>

```

A second example:

```

\newcounter{cms}
\setlength{\unitlength}{1mm}
\begin{picture}(50,39)
\put(0,7){\makebox(0,0)[bl]{cm}}
\multiput*(10,7)(10,0){5}{% remove the * for latex
  \addtocounter{cms}{1}\makebox(0,0)[b]{\arabic{cms}}}
\put(15,20){\circle{6}}
\put(30,20){\circle{6}}
\put(15,20){\circle*{2}}
\put(30,20){\circle*{2}}
\put(10,24){\framebox(25,8){car}}
\put(10,32){\vector(-2,1){10}}
\multiput(1,0)(1,0){49}{\line(0,1){2.5}}
\multiput(5,0)(10,0){5}{\line(0,1){3.5}}
\thicklines
\put(0,0){\line(1,0){50}}
\multiput(0,0)(10,0){6}{\line(0,1){5}}
\end{picture}

```

The translation is

```

<Picture Width='142.26303' Height='110.96516'>
  <Put Xpos='0' Ypos='19.91682'>
    <Box Width='0' Height='0' BoxPos='bl'>cm</Box>
  </Put>
  <Put Xpos='28.4526' Ypos='19.91682'>
    <Box Width='0' Height='0' BoxPos='b'>1</Box></Put>
  <Put Xpos='56.90521' Ypos='19.91682'>
    <Box Width='0' Height='0' BoxPos='b'>2</Box></Put>
  <Put Xpos='85.35782' Ypos='19.91682'>
    <Box Width='0' Height='0' BoxPos='b'>3</Box></Put>
  <Put Xpos='113.81042' Ypos='19.91682'>
    <Box Width='0' Height='0' BoxPos='b'>4</Box></Put>
  <Put Xpos='142.26303' Ypos='19.91682'>
    <Box Width='0' Height='0' BoxPos='b'>5</Box></Put>
  <Put Xpos='42.67891' Ypos='56.90521'><Circle Size='17.07156'/></Put>
  <Put Xpos='85.35782' Ypos='56.90521'><Circle Size='17.07156'/></Put>
  <Put Xpos='42.67891' Ypos='56.90521'>
    <Circle Full='true' Size='5.69052'/></Put>
  <Put Xpos='85.35782' Ypos='56.90521'>

```

```

    <Circle Full='true' Size='5.69052' /></Put>
  <Put Xpos='28.4526' Ypos='68.28625'>
    <Box Width='71.13152' Height='22.76208' Framed='true'>car</Box></Put>
  <Put Xpos='28.4526' Ypos='91.04834'>
    <Vector Xdir='-2' Ydir='1' Width='28.4526' /></Put>
  <Multiptut Xpos='2.84526' Ypos='0' Repeat='49' Dx='2.84526' Dy='0'>
    <Line Xdir='0' Ydir='1' Width='7.11314' />
  </Multiptut>
  <Multiptut Xpos='14.2263' Ypos='0' Repeat='5' Dx='28.4526' Dy='0'>
    <Line Xdir='0' Ydir='1' Width='9.9584' /></Multiptut>
  <Thicklines/>
  <Put Xpos='0' Ypos='0'><Line Xdir='1' Ydir='0' Width='142.26303' /></Put>
  <Multiptut Xpos='0' Ypos='0' Repeat='6' Dx='28.4526' Dy='0'>
    <Line Xdir='0' Ydir='1' Width='14.2263' />
  </Multiptut>
</Picture>

```

6.10 A sample configuration file

This is the file we used in the section above.

```

## tralics ident rc=helloconf.
BeginType article
DocType = Article classes.dtd
#DocAttrib =Foo \World
DocAttrib =A \specialyear
DocAttrib =B \tralics
no_footnote_hack = "true"
xml_footnote_name = "Footnote"
att_place = "Place"
att_foot_position = "as a note"
everyjob="\message{OK}"
url_font= "\it"
att_flush_right = "RaggedLeft"
att_flush_left = "RaggedRight"
att_quote = "Quote"
att_quotation = "Quotation"
att_centering = "Centered"
att_cell_bottomborder = "BottomBorder"
att_cell_topborder = "TopBorder"
att_cell_leftborder = "LeftBorder"
att_cell_rightborder = "RightBorder"
att_row_spaceafter = "SpaceAfter"
att_row_spacebefore = "SpaceBefore"
att_halign="Align"
att_cell_left = "Cleft"
att_cell_right = "Cright"
att_cell_center = "Ccenter"
xml_table_name = "Table"
xml_cell_name = "Cell"
xml_row_name = "Row"
xml_fbox_name = "Fbox"
xml_scalebox_name = "Scalebox"
xml_box_name = "Box"
att_framed="Framed"
xml_figure_name = "Figure"

```

```
att_file = "File"
att_angle = "Angle"
att_rotate_angle = "RotAngle"
att_scale = "Scale"
att_clip = "Clip"
att_width = "Width"
att_height = "Height"

xml_bezier_name = "Bezier"
xml_put_name = "Put"
xml_scaleput_name = "Scaleput"
xml_multiput_name = "Multiput"
xml_arc_name = "Arc"
xml_line_name = "Line"
xml_vector_name = "Vector"
xml_oval_name = "Oval"
xml_circle_name = "Circle"
xml_bigcircle_name = "Bigcircle"
xml_dashline_name = "Dashline"
xml_drawline_name = "Drawline"
xml_dottedline_name = "Dottedline"
xml_curve_name = "Curve"
xml_closecurve_name = "CloseCurve"
xml_tagcurve_name = "Tagcurve"
xml_thicklines_name = "Thicklines"
xml_thinlines_name = "Thinlines"
xml_linethickness_name = "Linethickness"
xml_subfigure_name = "Subfigure"
xml_keywords_name = "Keywords"
xml_term_name = "Term"
xml_mbox_name = "Mbox"
xml_topic_name = "Topic"
xml_topic_title = "TopicTitle"
att_inner_pos = "InnerPos"
att_vpos = "VPos"
att_noindent = "Noindent"
att_boxed = "Boxed"
att_box_width = "BoxWidth"
att_table_width = "TableWidth"
att_pos = "Pos"
att_full = "Full"
att_xpos = "Xpos"
att_ypos = "Ypos"
att_xdir = "Xdir"
att_ydir = "Ydir"
att_rotate_angle = "RotAngle"
att_topic_num = "Num"

xml_font_small= "Small"
xml_font_large= "Large"
xml_font_normal= "Normal"
use_all_sizes="true"
xml_font_small1 = "S1"
xml_font_small2 = "S2"
xml_font_small3 = "S3"
xml_font_small4 = "S4"
```

```
xml_font_large1 = "L1"
xml_font_large2 = "L2"
xml_font_large3 = "L3"
xml_font_large4 = "L4"
xml_font_large5 = "L5"
xml_font_upright = "Upright"
xml_font_sc = "Sc"
xml_font_slanted = "Slanted"
xml_font_it = "It"
xml_font_sansserif = "SF"
xml_font_bold = "B"
xml_font_tt = "TT"
xml_sup_name = "Super"
xml_sub_name = "Sub"
xml_oldstyle_name = "Oldstyle"
xml_overline_name = "Overline"
xml_underline_name = "Underline"
xml_div0_name= "Part"
xml_div1_name= "Chapter"
xml_div2_name= "Section"
xml_div3_name= "Subsection"
xml_div4_name= "Subsubsection"
xml_div5_name= "Paragraph"
xml_div6_name= "Subparagraph"
att_full_first ="Full"
att_prenom="First"
att_nom="Last"
att_particule="von"
att_junior ="Junior"
xml_bpers_name = "Pers"
xml_caption_name = "Caption"
xml_graphics_name = "Graphics"
xml_figure_env_name = "Figureenv"
xml_table_env_name = "Tableenv"
xml_Table_name = "TTable"
xml_xref_name = "Xref"
xml_rotatebox_name = "Rotatebox"
xml_list_name = "List"
xml_head_name = "Head"
xml_scaption_name = "Scaption"
xml_glo_name= "Glo"
xml_project_name = "Projet"
xml_accueil_name = "Accueil"
xml_composition_ra_name
xml_theorem_name = "XXtheorem"
xml_theorem_head = "AltHead"
xml_xtheorem_name = "Xtheorem"
xml_biblio "Biblio"
att_box_pos = "BoxPos"
att_box_scale="Boxscale"
xml_picture_name = "Picture"
att_unit_length = "Ulength"
att_curve_nbpts = "Nbpts"
att_xscale = "Xscale"
att_xscaley = "XscaleY"
att_yscale = "Yscale"
```

```

att_yscalex = "YscaleX"
att_repeat = "Repeat"
att_dx = "Dx"
att_dy = "Dy"
att_size = "Size"
xml_labelitem_name = "Label"
xml_item_name = "Item"
xml_gloitem_name = "GItem"
xml_texte_name = "SFTText"
xml_leg_name = "SFCapt"
att_gloss_type = "Glossary"
use_font_elt ="true"
att_cols = "Cols"

End
End

```

6.11 Trace of titlepage

This is a part of the transcript file for the titlepage command.

```

1 Defining \makeRR as \TitlePageCmd 0
2   main <RRstart -- type = 'RR' />
3 Defining \makeRT as \TitlePageCmd 1
4   main <RRstart -- type = 'RT' />
5 Defining \UR as \TitlePageCmd 2
6   ur_list <UR />
7 Defining \URSophia as \TitlePageCmd 3
8   ur <URSophia />
9 Defining \URRocquencourt as \TitlePageCmd 4
10  ur <URRocquencourt />
11 Defining \URRocq as alias to \URRocquencourt
12 Defining \Paris as \TitlePageCmd 6
13  ur <Rocquencourt />
14 Defining \URRhôneAlpes as \TitlePageCmd 7
15  ur <URRhôneAlpes />
16 Defining \URRennes as \TitlePageCmd 8
17  ur <URRennes />
18 Defining \URLorraine as \TitlePageCmd 9
19  ur <URLorraine />
20 Defining \URFuturs as \TitlePageCmd 10
21  ur <URFuturs d='true' />
22 Defining \RRtitle as \TitlePageCmd 11
23  usual <title /> (flags -par)
24 Defining \RRetitle as \TitlePageCmd 12
25  usual <etitle /> (flags -par)
26 Defining \RRprojet as \TitlePageCmd 13
27  usual <projet />
28 Defining \motcle as \TitlePageCmd 14
29  usual <motcle />
30 Defining \keyword as \TitlePageCmd 15
31  usual <keyword />
32 Defining \RRresume as \TitlePageCmd 16
33  usual <resume /> (flags +par)
34 Defining \RRabstract as \TitlePageCmd 17

```

RT n° 309


```

35     usual <abstract/> (flags +par)
36 Defining \RRauthor as \TitlePageCmd 18
37     list <author/> and <auth/>
38 Defining \RRdate as \TitlePageCmd 19
39     usual <date/>
40 Defining \RRNo as \TitlePageCmd 20
41     usual <RRnumber/>
42 Defining \RRtheme as \TitlePageCmd 21
43     usual </> (flags +list)
44 Defining \Theme as \TitlePageCmd 22
45     ur_list <Theme/>
46 Defining \THNum as \TitlePageCmd 23
47     ur <THNum/>
48 Defining \THCog as \TitlePageCmd 24
49     ur <THCog/>
50 Defining \THCom as \TitlePageCmd 25
51     ur <THCom/>
52 Defining \THBio as \TitlePageCmd 26
53     ur <THBio/>
54 Defining \THSym as \TitlePageCmd 27
55     ur <THSym/>
56 Defining \myself as \TitlePageCmd 28
57     list? <grimm/> and <auth/>
58 Defining \cmdp as \TitlePageCmd 29
59     usual <cmdp/> (flags +list)
60 Defining \cmda as \TitlePageCmd 30
61     usual <cmdA/> (flags +A)
62 Defining \cmdb as \TitlePageCmd 31
63     usual <cmdB/> (flags +B)
64 [1] \cmdbf{\cmdBval}
65 ++ End of virtual file.
66 Defining \cmdc as \TitlePageCmd 32
67     usual <cmdC/> (flags +C)
68 [1] \documentclass[a4paper]{report}
69 ...
70 [1] \cmda{\cmdAval}
71 {(Unknown)}
72 {\titlepage 30}
73 {\titlepage 30=\cmda}
74 {Push cmdA 1}
75 Error signaled at line 1 of file tptest.tex:
76 Undefined command \cmdAval.
77 ...
78 [8] \RRetitle{Tralics, a \LaTeX\ to XML translator\\Part I}
79 {(Unknown)}
80 {\titlepage 12}
81 {\titlepage 12=\RRetitle}
82 {Push etitle 1}
83 ...
84 [12] \RRtheme{\THNum}
85 {(Unknown)}
86 {\titlepage 21}
87 {\titlepage 21=\RRtheme}
88 {(Unknown)}
89 {\titlepage 23}
90 {\titlepage 23=\THNum}

```

```

91 {\par}
92 [13]
93 ...
94 [24] \begin{document}
95 {\begin}
96 {\begin document}
97 +stack: level + 2 for environment
98 {\document}
99 +stack: ending environment document; resuming document.
100 +stack: level - 2 for environment
101 +stack: level set to 1
102 [1] \let\do\noexpand\ignorespaces
103 ++ End of virtual file.
104 atbegindocumenthook= \cldb {\cmdBval }\let \AtBeginDocument \@notprerr \let \do
105 \noexpand \ignorespaces
106 {(Unknown)}
107 {\titlepage 31}
108 {\titlepage 31=\cldb}
109 {Push cmdB 1}
110 Error signaled at line 24 of file tptest.tex:
111 Undefined command \cmdBval.
112 {Pop 1: document_v cmdB_t}
113 {\let}
114 {\let \AtBeginDocument \@notprerr}
115 {\let}
116 {\let \do \noexpand}
117 {\ignorespaces}
118 [25] \makeRR
119 {(Unknown)}
120 {\titlepage 0}
121 {\titlepage 0=\makeRR}
122 Error signaled at line 25 of file tptest.tex:
123 No value given for command \cmdp.
124 [1] \cmdc{\cmdCval}
125 ++ End of virtual file.
126 {(Unknown)}
127 {\titlepage 32}
128 {\titlepage 32=\cmdc}
129 {Push cmdC 1}
130 Error signaled at line 25 of file tptest.tex:
131 Undefined command \cmdCval.
132 {Pop 1: document_v cmdC_t}
133 {Push p 1}
134 [26] text
135 ...
136 Output written on tptest.xml (1059 bytes).
137 There were 4 errors.
138 (For more information, see transcript file tptest.log)

```

6.12 Bootstrap code

At the start of the run, a given number of commands are defined using the following L^AT_EX code:

```

[1] %% Begin bootstrap commands for latex
[2] \def\lq{'}\def\rq{'}\def\lbrack{[]}\def\rbrack{[]}%
[3] \countdef\count@=255 \dimendef\dimen@=0

```

```

[4] \newdimen\z@ \z@=0pt \chardef\@ne=1 \chardef\tw@=2
[5] \chardef\active=13 \def\@empty{}%
[6] \countdef\m@ne=10 \m@ne=-1 %
[7] \dimendef\epsfxsize11 \dimendef\epsfysize12\relax
[8] \def\sloppypar{\par\sloppy}\def\endsloppypar{\par}%
[9] {\catcode'\_ =13\global\let\_}%
[10] {\catcode'\# =13\global\let#\#}%
[11] {\catcode'\& =13\global\def&\char'&}%
[12] \catcode'\~ =13\def~{\nobreakspace}%
[13] \newdimen\evensidemargin \newdimen\oddsidemargin
[14] \def\null{\hbox{}}%
[15] \let\dots\cdots \let\dotsb\cdots \let\dotsi\cdots \let\dotsm\cdots
[16] \let\mathbbm\mathbb
[17] \let\Href\htladdnormallink\def\urlfont{}%
[18] \newenvironment{cases}{\left\{\begin{array}{l}}{\end{array}\right.}%
[19] \def\binom{\genfrac{}{}{z@}}{\def\dbinom{\genfrac{}{}{Opt}0}
\def\tbinom{\genfrac{}{}{Opt}1}}%
[20] \def\space{ }
[21] \def\thinspace{ } \def\@firstofone#1{#1}%
[22] \let\verbatimfont\tt\let\verbatimnumberfont\small
[23] \newskip\itemsep\newdimen\itemindent
[24] \newskip\labelsep\newdimen\labelwidth
[25] \newdimen\leftmargin\newdimen\rightmargin
[26] \newdimen\leftmargini\newdimen\leftmarginii
[27] \newdimen\leftmarginiii\newdimen\leftmarginiv
[28] \newdimen\leftmarginv\newdimen\leftmarginvi
[29] \newskip\parsep \let\pounds\textsterling
[30] \long\def \@gobble #1{\long\def \@gobbletwo #1#2{
[31] \def\markboth#1#2{\gdef\@themark{#1}\mark{#1}{#2}}
[32] \def\markright#1{\expandafter\markboth\@themark{#1}}
[33] \let\chaptermark\@gobble\let\sectionmark\@gobble
[34] \let\subsectionmark\@gobble\let\subsubsectionmark\@gobble
[35] \let\paragraphmark\@gobble\let\subparagraphmark\@gobble
[36] \newdimen\fbboxsep\fbboxsep = 3pt \newdimen\fbboxrule
[37] \def\xscale{1.0}\def\yscale{1.0}\def\xscaley{0.0}\def\yscalex{0.0}
[38] \def\@height{height} \def\@depth{depth} \def\@width{width}
[39] \def\@minus{minus}\def\@plus{plus}\def\hb@xt@{\hbox to}
[40] \def\stretch#1{\z@ \@plus #1fill\relax}
[41] \newskip\fill \fill = Opt plus 1fill
[42] \theoremstyle{plain}\theoremheaderfont{\bfseries}
[43] \def\@namedef#1{\expandafter\def\csname #1\endcsname}
[44] \def\@nameuse#1{\csname #1\endcsname}
[45] \def\I{I}\def\J{J}
[46] \countdef\c@page=0 \c@page=1\def\c1@page{}
[47] \def\thepage{\@arabic\c@page}
[48] \def\@roman#1{\romannumeral#1}\def\@Roman#1{\Romannumeral#1}
[49] \pretolerance=100 \tolerance=200
[50] \def\mod#1{\~\@mod\;#1}
[51] \def\bmod#1{\;\@mod\;#1}
[52] \def\pmod#1{\pod{\@mod\; #1}}
[53] \def\pod#1{\~(#1)}
[54] \newcount\c@bottomnumber\newcommand\bottomfraction{.3}
[55] \newcount\c@topnumber\newcommand\topfraction{.7}
[56] \def\date#1{\xmlelt{date}{#1}}
[57] \newcounter{enumi}\newcounter{enumii}\newcounter{enumiii}
\newcounter{enumiv}

```

```

[58] \newcommand\labelitemi{\textbullet}
[59] \newcommand\labelitemii{\normalfont\bfseries \textendash}
[60] \newcommand\labelitemiii{\textasteriskcentered}
[61] \newcommand\labelitemiv{\textperiodcentered}
[62] \newcounter {part}\newcounter {chapter}\newcounter {section}[chapter]
      \newcounter {subsection}[section]\newcounter {subsubsection}[subsection]
      \newcounter {paragraph}[subsubsection]
      \newcounter {subparagraph}[paragraph]
[63] \def\refname{Bibliography}\def\footcitesep{, }\def\citepunct{, }
[64] \newcommand\newblock{\hskip .11em plus.33em minus.07em}
[65] %% End bootstrap commands for latex

```

The transcript file of Tralics contains a line for each use of `\dimendef` or friends. Here is the list of all standard definitions.

```

{\countdef \count@=\count255}
{\dimendef \dimen@=\dimen0}
{\dimendef \z@=\dimen13}
{\chardef \@ne=\char1}
{\chardef \tw@=\char2}
{\chardef \active=\char13}
{\countdef \m@ne=\count10}
{\dimendef \epsfxsize=\dimen11}
{\dimendef \epsfysize=\dimen12}
{\dimendef \evensidemargin=\dimen14}
{\dimendef \oddsidemargin=\dimen15}
{\skipdef \itemsep=\skip11}
{\dimendef \itemindent=\dimen16}
{\skipdef \labelsep=\skip12}
{\dimendef \labelwidth=\dimen17}
{\dimendef \leftmargin=\dimen18}
{\dimendef \rightmargin=\dimen19}
{\dimendef \leftmargini=\dimen20}
{\dimendef \leftmarginii=\dimen21}
{\dimendef \leftmarginiii=\dimen22}
{\dimendef \leftmarginiv=\dimen23}
{\dimendef \leftmarginv=\dimen24}
{\dimendef \leftmarginvi=\dimen25}
{\skipdef \parsep=\skip13}
{\dimendef \fboxsep=\dimen26}
{\dimendef \fboxrule=\dimen27}
{\skipdef \fill=\skip14}
{\countdef \c@page=\count0}
{\countdef \c@bottomnumber=\count11}
{\countdef \c@topnumber=\count12}
{\countdef \c@enumi=\count13}
{\countdef \c@enumii=\count14}
{\countdef \c@enumiii=\count15}
{\countdef \c@enumiv=\count16}
{\countdef \c@part=\count17}
{\countdef \c@chapter=\count18}
{\countdef \c@section=\count19}
{\countdef \c@subsection=\count20}
{\countdef \c@subsubsection=\count21}
{\countdef \c@paragraph=\count22}
{\countdef \c@subparagraph=\count23}

```

6.13 The puzzle

The only requirements for the xii file is that `~` is an active character, `\\` has category code 0, `%` is a comment character, the end of line character is as usual. The file modifies the category code of 7, F, j and P, in such a way that ‘jdefjx71F71P’ is the same as ‘`\def\x#1{#1}`’. This is one way of making the code incomprehensible, the other is to use commands so that ‘six’, ‘geese’ and ‘laying’ are replaced by ‘/sx’, ‘Yegse’ and ‘RyalD’. The programs make the letter H active and defines it via ‘AHHFLP’. For those who want to write puzzles like this one: is it possible to avoid doubling the H? Without using all these strange commands, the file could be written as

```
\let~\catcode ~'A13 \defA#1{~'#113\def}
AZZ{ }APP{\par}AXX#1{\bigskip On the #1 day of Christmas my true love gave to me}
ABB{PZAZZ{and }a partridge in a pear tree.}
ACC{Ptwo turtle doves}
ADD{Pthree french hens}
AEE{Pfour calling birds}
AFF{Pfive gold rings}
AGG{Psix geese a laying}
AHH{Pseven swans a swimming}
AII{Peight maids a milking}
AJJ{Pnine ladies dancing}
AKK{Pten lords a leaping}
ALL{Peleven pipers piping}
AMM{Ptwelve drummers drumming}
\def\F#1#2#3{ }AVV#1\fi{\fi#1}
ATT#1 #2,#3:{\if.#3.\elseT#3:\fiX{#1}\U#1 #2,#3:}
\def\U#1#2#3#4 #5,#6{\F#1#2#3#5\if:#6\elseV\U#6\fi}
Ttwelfth M,eleventh L,tenth K,ninth J,eighth I,seventh H,sixth G,fifth
F,fourth E,third D,second C,first B,:\bye
```

The size of the file is 698 characters (compare to the 767 of the xii file). Note how the double loop is constructed. The xii file is made obscure by replacing B, C, D etc., by expression that have the same expansion, using instead of `\F` a command that uses some characters (because ‘nine’ and ‘ninth’ start with the same letters, etc.) An interesting point is that we can write a smaller file, with all loops unrolled, replacing the last 5 lines by the following (this makes a total of 642 characters):

```
X{first}BX{second}CBX{third}DCBX{fourth}EDCBX{fifth}FEDCBX{sixth}GFEDCB
X{seventh}HGFEDCBX{eighth}IHGFEDCBX{ninth}JIHGFEDCBX{tenth}KJIHGFEDCB
X{eleventh}LKJIHGFEDCBX{twelfth}MLKJIHGFEDCB\bye
```

6.14 Images

We give here some examples of the `\includegraphics` command. We consider a file with the following content. Notice that the clip attribute is set to true if ‘clip’ appears in the list, whether or not a value has been given. A colon and an underscore in a file name is never interpreted. The extension is always removed:

```
\let\IC=\includegraphics
\def\FILE{Logo-INRIA-couleur}
\IC[angle=0,width=3cm,clip=]{\FILE}
\IC[angle=20,width=.5\textwidth,height=.3\textheight]{\FILE.ducon}
\IC[angle=0,width=\columnwidth,height=\textheight]{\FILE.foo_bar}
{\language=1 a:c
\IC[angle=0, =foo,,width=3cm,scale=1,scale=2,clip]{.././a_b:c}
```

```

}
\framebox{\includegraphics{x.ps}}
We continue with an example of \epsfbox.
{
  \setlength\epsfxsize{50pt}
  \setlength\epsfysize{60pt}
  \epsfbox{x.ps}
  \setlength\epsfysize{70pt}
  \epsfbox{x.eps}
  \epsfbox{x.epsf}
}

```

Tralics pretends that there are 4 different images. The translation is:

```

<figure rend='inline' clip='true' width='3cm' file='Logo-INRIA-couleur' />
<figure rend='inline' height='6.cm' width='7.5cm'
  angle='20' file='Logo-INRIA-couleur' />
<figure rend='inline' height='20.cm' width='15.cm' file='Logo-INRIA-couleur' />
a&nbsp;b;c
<figure rend='inline' clip='true' scale='2' width='3cm' file='../a_b:c' />
<figure framed='true' rend='inline' file='x_' /></p>
<figure height='60.0pt' width='50.0pt' rend='inline' file='x' />
<figure height='70.0pt' rend='inline' file='x' />
<figure rend='inline' file='x' />

```

The file ‘\jobname.img’ contains the following. The last number is the number of times the image was included. The second number explains in which format the file has been found. Whether or not the image file is found is irrelevant. The information given in the file is for information only.

```

# images info, 1=ps, 2=eps, 4=epsi, 8=epsf, 16=pdf, 32=png, 64=gif
see_image("Logo-INRIA-couleur",1+16,3);
see_image("../a_b:c",0,1);
see_image("x_",0,1);
see_image("x",0,3);

```


Index

`\!`, 48, 75, 120
`\"`, 113
`(`, 75
`\(`, 73
`)`, 75
`\)`, 73
`\,`, 75, 120
`\-`, 118
`..`, 75
`\.`, 113
`\/`, 148
`\:`, 48, 75
`\;`, 48, 75
`<`, 75
`\=`, 113
`>`, 75
`\>`, 48, 75
`\@`, 148
`\@@end`, 141
`\@Alph`, 42, 69
`\@Roman`, 42
`\@alph`, 42, 69
`\@arabic`, 42, 68
`\@fnsymbol`, 42, 69
`\@gobble`, 147
`\@gobbletwo`, 147
`\@ifnextchar`, 147
`\@ifstar`, 147
`\@ifundefined`, 146
`\@mathcal`, 78
`\@mod`, 76
`\@ne`, 52
`\@pers`, 163
`\@roman`, 42
`\@stpelt`, 68
`\@whiledim`, 69, 131
`\@whilenum`, 69, 131
`\@whilesw`, 69, 131
`[`, 75
`\[`, 73
`\#`, 118
`#`, 19, 51
`$`, 7, 72
`\%`, 75, 120
`%`, 52, 66
`\&`, 75, 120
`\``, 113
`\|`, 47
`\^`, 112
`ˆ`, 21, 110
`_`, 75, 120
`\``, 113
`\~`, 113
`˜`, 48
`\$`, 75, 120
`|`, 75
`\|`, 77
`\{`, 75, 121
`\}`, 75, 121
`]`, 75
`\]`, 73
`\a`, 68, 112
`\AA`, 118
`\aa`, 118
`\above`, 82
`\abovedisplayshortskip`, 34
`\abovedisplayskip`, 34
`\abovewithdelims`, 82
`\accent`, 9, 111
`\active`, 6
`\acute`, 77
`\addattributestodocument`, 144
`\AddAttToCurrent`, 32, 144
`\AddAttToLast`, 32, 144
`address`, 96
`\addtocounter`, 42, 68, 134
`\addtolength`, 42, 68, 133
`\adjdemerits`, 36
`\advance`, 34, 42, 54
`\AE`, 118
`\ae`, 118
`\afterassignment`, 32
`\aftergroup`, 67
`\aleph`, 76
`align`, 73

`\allowbreak`, 148
`\Alph`, 42, 69
`\alph`, 42, 69
`\alpha`, 75
`\amalg`, 76
`\AND`, 135
`\and`, 135
`\angle`, 75
`\aparaître`, 143
`\approx`, 76
`apr`, 95
`\arabic`, 42
`\arc`, 172, 198, 199
`\arccos`, 76
`\arcsin`, 76
`\arctan`, 76
`\arg`, 76
`array`, 74, 178
`article`, 95
`\ast`, 76
`\asymp`, 76
`\AtBeginDocument`, 130
`\AtEndDocument`, 25, 130
`\atop`, 82
`\atopwithdelims`, 82
`att-angle`, 196
`att-box-pos`, 196
`att-box-scale`, 191
`att-box-width`, 191
`att-boxed`, 191
`att-cell-bottomborder`, 197
`att-cell-center`, 197
`att-cell-left`, 197
`att-cell-leftborder`, 197
`att-cell-right`, 197
`att-cell-rightborder`, 197
`att-cell-topborder`, 197
`att-centering`, 196
`att-clip`, 196
`att-cols`, 197
`att-curve-nbpts`, 199
`att-dx`, 198
`att-dy`, 198
`att-fbox-rend`, 196
`att-file`, 196
`att-flush-left`, 196
`att-flush-right`, 196
`att-foot-position`, 187
`att-framed`, 191
`att-full`, 196
`att-full-first`, 198
`att-gloss-type`, 190
`att-halign`, 197
`att-height`, 196
`att-inner-pos`, 196
`att-junior`, 198
`att-noindent`, 192
`att-nom`, 198
`att-particule`, 198
`att-place`, 187
`att-pos`, 196
`att-prenom`, 198
`att-quotation`, 196
`att-quote`, 196
`att-rend`, 196
`att-repeat`, 198
`att-rotate-angle`, 192
`att-row-space-after`, 197
`att-row-space-before`, 197
`att-scale`, 196
`att-size`, 198
`att-table-width`, 196
`att-topic-num`, 152, 192
`att-unit-length`, 199
`att-vpos`, 196
`att-width`, 196
`att-xdir`, 198
`att-xpos`, 198
`att-xscale`, 199
`att-xscaley`, 199
`att-ydir`, 198
`att-ypos`, 198
`att-yscale`, 199
`att-yscalex`, 199
`aug`, 95
`author`, 96

`\b`, 68, 113
`\backslash`, 76
`\badness`, 37
`\bar`, 77
`\baselineskip`, 34
`\batchmode`, 147
`\bauthors`, 97, 103
`\beditors`, 103
`\begin`, 23
`\begingroup`, 23
`\belowdisplayshortskip`, 34
`\belowdisplayskip`, 34
`\beta`, 75
`\bezier`, 171, 198
`\bf`, 5, 46, 78
`\bfseries`, 46, 189
`\bgroup`, 26, 66

| | |
|---------------------------------------|--|
| <code>\bibitem</code> , 94 | <code>\C</code> , 68, 113 |
| <code>\bibliography</code> , 104 | <code>\c</code> , 68, 113 |
| <code>\bibliographystyle</code> , 104 | <code>\cal</code> , 78 |
| <code>\Big</code> , 84 | calc, 43 |
| <code>\big</code> , 14, 84 | <code>\calc</code> , 134 |
| <code>\bigbreak</code> , 50 | <code>\cap</code> , 76 |
| <code>\bigcap</code> , 76 | <code>\caption</code> , 142 |
| <code>\bigcirc</code> , 76 | cases, 74 |
| <code>\bigcirc</code> , 172, 198, 199 | <code>\catcode</code> , 6, 38 |
| <code>\bigcup</code> , 76 | catperso, 162, 163 |
| <code>\Bigg</code> , 84 | cc, 39 |
| <code>\bigg</code> , 84 | <code>\cdot</code> , 76 |
| <code>\Biggl</code> , 84 | <code>\cdots</code> , 77 |
| <code>\biggl</code> , 84 | center, 148, 196 |
| <code>\Bigl</code> , 84 | <code>\centering</code> , 141, 148, 196 |
| <code>\bigl</code> , 84 | <code>\centerline</code> , 148 |
| <code>\bigodot</code> , 76 | <code>\cfoot</code> , 146 |
| <code>\bigoplus</code> , 76 | chapter, 96 |
| <code>\bigotimes</code> , 76 | <code>\chapter</code> , 142, 190 |
| <code>\bigskip</code> , 49 | <code>\char</code> , 9, 10, 110 |
| <code>\bigsqcup</code> , 76 | <code>\chardef</code> , 26 |
| <code>\bigtriangledown</code> , 76 | <code>\chead</code> , 146 |
| <code>\bigtriangleup</code> , 76 | check, 150 |
| <code>\biguplus</code> , 76 | <code>\check</code> , 77 |
| <code>\bigvee</code> , 76 | checkxml, 187 |
| <code>\bigwedge</code> , 76 | <code>\chi</code> , 75 |
| <code>\bindnasrepma</code> , 76 | <code>\circ</code> , 76 |
| <code>\binoppenalty</code> , 35 | <code>\circle</code> , 172, 196, 198 |
| <code>\bm</code> , 11 | <code>\citation</code> , 97, 104 |
| <code>Bmatrix</code> , 74 | <code>\cite</code> , 93 |
| <code>bmatrix</code> , 74 | <code>\cite@one</code> , 93 |
| <code>\boldmath</code> , 11 | <code>\citepunct</code> , 93 |
| book, 95 | <code>\cititem</code> , 103 |
| booklet, 95 | <code>\cleaders</code> , 143 |
| booktitle, 96 | <code>\cleardoublepage</code> , 148 |
| <code>\boolean</code> , 134 | <code>\clearpage</code> , 148 |
| <code>\bot</code> , 75 | <code>\cline</code> , 183 |
| <code>\botmark</code> , 69, 141 | <code>\closecurve</code> , 172, 198, 199 |
| <code>\bottom</code> , 75 | <code>\closein</code> , 137 |
| <code>\bowtie</code> , 76 | <code>\closeout</code> , 139 |
| <code>\Box</code> , 76 | <code>\clubpenalty</code> , 35 |
| <code>\box</code> , 38 | <code>\clubsuit</code> , 76 |
| <code>\boxmaxdepth</code> , 37 | cm, 39 |
| <code>bp</code> , 39 | <code>\columnwidth</code> , 38 |
| <code>bpers</code> , 97 | comment, 95, 148 |
| <code>\bpers</code> , 104, 198 | conference, 95 |
| <code>\break</code> , 148 | configfile, 150 |
| <code>\breve</code> , 77 | <code>\cong</code> , 76 |
| <code>\brokenpenalty</code> , 35 | <code>\coprod</code> , 76 |
| <code>\bullet</code> , 76 | <code>\copy</code> , 26 |
| <code>\bye</code> , 151 | <code>\copyright</code> , 118 |
| | <code>\cos</code> , 76 |

| | |
|---|---|
| <code>\cosh</code> , 76 | <code>\diamond</code> , 76 |
| <code>\cot</code> , 76 | <code>\diamondsuit</code> , 76 |
| <code>\coth</code> , 76 | <code>\dim</code> , 76 |
| <code>\count@</code> , 26 | <code>\dimendef</code> , 26 |
| <code>\countdef</code> , 26 | <code>\ding</code> , 148 |
| <code>coursenotes</code> , 95 | <code>dir</code> , 150 |
| <code>\cr</code> , 181, 184 | <code>\discretionary</code> , 147 |
| <code>\crrcr</code> , 181, 184 | <code>\displayindent</code> , 37 |
| <code>crossref</code> , 96 | <code>displaymath</code> , 73 |
| <code>\csc</code> , 76 | <code>\displaystyle</code> , 14, 80, 83 |
| <code>\csname</code> , 22, 28, 51, 68 | <code>\displaywidowpenalty</code> , 35 |
| <code>\cup</code> , 76 | <code>\displaywidth</code> , 37 |
| <code>\curve</code> , 172, 198, 199 | <code>\div</code> , 76 |
| | <code>\divide</code> , 34, 42 |
| <code>\D</code> , 68, 113 | <code>\DJ</code> , 118 |
| <code>\d</code> , 68, 113 | <code>\dj</code> , 118 |
| <code>\dag</code> , 118 | <code>document</code> , 5, 24, 95 |
| <code>\dagger</code> , 76 | <code>\documentclass</code> , 5, 130 |
| <code>\dashbox</code> , 170 | <code>doi</code> , 96 |
| <code>\dashline</code> , 172, 198 | <code>\dot</code> , 77 |
| <code>\dashv</code> , 76 | <code>\doteq</code> , 76 |
| <code>\day</code> , 36 | <code>\dots</code> , 75, 120 |
| <code>dd</code> , 39 | <code>\dottedline</code> , 172, 198 |
| <code>\ddag</code> , 118 | <code>\doublehyphendemerits</code> , 36 |
| <code>\ddagger</code> , 76 | <code>\Downarrow</code> , 75, 77 |
| <code>\ddddot</code> , 77 | <code>\downarrow</code> , 75, 77 |
| <code>\ddot</code> , 77 | <code>\dp</code> , 38 |
| <code>\dot</code> , 77 | <code>\drawline</code> , 172 |
| <code>\dots</code> , 77 | <code>\dump</code> , 7 |
| <code>debug</code> , 150 | <code>dvitops</code> , 187 |
| <code>dec</code> , 95 | |
| <code>\DeclareRobustCommand</code> , 23 | <code>\edef</code> , 29, 30, 59 |
| <code>\declaretopic</code> , 162, 163 | <code>edition</code> , 96 |
| <code>\def</code> , 7 | <code>editor</code> , 96 |
| <code>\defaultthyphenchar</code> , 36, 45 | <code>\egroup</code> , 26, 67 |
| <code>\defaultskewchar</code> , 36, 45 | <code>\eject</code> , 50 |
| <code>\DefineShortVerb</code> , 115 | <code>\ell</code> , 75 |
| <code>\DefineVerbatimEnvironment</code> , 117 | <code>\else</code> , 54, 56, 69 |
| <code>\deg</code> , 76 | <code>em</code> , 39 |
| <code>\delcode</code> , 38 | <code>\em</code> , 47 |
| <code>\delimiter</code> , 88 | <code>\emergencystretch</code> , 37 |
| <code>\delimiterfactor</code> , 36 | <code>\emph</code> , 6, 47 |
| <code>\delimitershortfall</code> , 37 | <code>\empty</code> , 63 |
| <code>\Delta</code> , 75 | <code>\emptyset</code> , 75 |
| <code>\delta</code> , 75 | <code>\end</code> , 23 |
| <code>\depthof</code> , 134 | <code>\endcitation</code> , 104 |
| <code>description</code> , 145, 190 | <code>\endgroup</code> , 23, 24 |
| <code>\det</code> , 76 | <code>\endguillemets</code> , 118 |
| <code>\dfrac</code> , 77 | <code>\endinput</code> , 69, 139, 140 |
| <code>\DH</code> , 118 | <code>\endlinechar</code> , 36, 47, 59, 109 |
| <code>\dh</code> , 118 | <code>\endsec</code> , 142, 190 |
| <code>\Diamond</code> , 76 | <code>\ensuremath</code> , 79, 84 |

entnames, 149
enumerate, 145, 190
enumi, 37
enumii, 37
enumiii, 37
enumiv, 37
\epsfbox, 147, 211
\epsffile, 147
\epsfig, 147
\epsfxsize, 38
\epsfysize, 38
\epsilon, 75
eqnarray, 73
\eqno, 59, 71
\eqref, 142
\equal, 60, 134
equation, 73
\equiv, 76
\eres, 47
\errhelp, 35
\errmessage, 141
\error, 148
\errorcontextlines, 37
\errorstopmode, 147
\escapechar, 28, 36
\eta, 75
\evensidemargin, 38
\everycr, 35
\everydisplay, 35, 74
\everyhbox, 32, 35, 143
everyjob, 188
\everyjob, 35, 188
\everymath, 35, 74
\everypar, 35
\everyvbox, 32, 35, 143
\everyxbox, 32, 35, 143
ex, 39
\exhyphenpenalty, 35
\exists, 76
\exp, 76
\expandafter, 22, 27, 52, 54, 55, 68
externalprog, 150

\f, 68, 113
\fam, 36, 46, 88
\fancyfoot, 146
\fancyhead, 146
\fancyhf, 146
\fancyinternal, 146
\fancyplain, 146
\fbbox, 170, 191
\fbboxrule, 38, 169

\fbboxsep, 38, 169
feb, 95
\fg, 118
\fi, 29, 54, 56, 69
fill, 39
\fill, 35
\fillbreak, 50
\finalhyphendemerits, 36
findwords, 150
\firstmark, 69, 141
\flat, 75, 120
\floatingpenalty, 36
flushleft, 148, 196
flushright, 148, 196
\fnsymbol, 42, 69
\font, 9
\fontdimen, 45
\fontencoding, 45
\fontfamily, 45
\fontname, 68
\fontseries, 45
\fontshape, 45
\fontsize, 45
foot, 93, 96
\footcite, 93
\footcitepre, 93
\footcitesep, 93
\footnote, 142, 163, 187
\footnotesize, 47, 188
\forall, 76
\FPabs, 122
\FPadd, 122
\FParccos, 126
\FParcsin, 126
\FParcsincos, 126
\FPclip, 122
\FPcos, 125
\FPcot, 125, 126
\FPcsolve, 126
\FPdiv, 123
\FPe, 124
\FPeval, 128
\FPexp, 124
\FPifeq, 124
\FPifgt, 124
\FPifint, 124
\FPiflt, 124
\FPifneg, 124
\FPifpos, 124
\FPifzero, 124
\FPindent, 122
\FPln, 124

| | |
|--|--|
| <code>\FPsolve</code> , 126 | <code>\guillemets</code> , 118 |
| <code>\FPmax</code> , 123 | <code>\guillemotleft</code> , 17, 118 |
| <code>\FPmin</code> , 123 | <code>\guillemotright</code> , 17, 118 |
| <code>\FPmul</code> , 122 | |
| <code>\FPneg</code> , 122 | <code>\H</code> , 68, 113 |
| <code>\FPpascal</code> , 124 | <code>\h</code> , 68, 113 |
| <code>\FPpow</code> , 125 | <code>\halign</code> , 184 |
| <code>\FPprint</code> , 121, 122 | <code>\hangafter</code> , 36 |
| <code>\FPqqsolve</code> , 127 | <code>\hangindent</code> , 37 |
| <code>\FPqsolve</code> , 126 | <code>\hat</code> , 77 |
| <code>\FPrand</code> , 124 | <code>\hbadness</code> , 36 |
| <code>\FProot</code> , 125 | <code>\hbar</code> , 75 |
| <code>\FPround</code> , 124 | <code>\hbox</code> , 23, 80, 83, 94, 143 |
| <code>\FPseed</code> , 37, 124 | <code>\hdots</code> , 77 |
| <code>\FPset</code> , 121, 122 | <code>\heartsuit</code> , 76 |
| <code>\FPsgn</code> , 122 | <code>\heightof</code> , 134 |
| <code>\FPsin</code> , 125 | <code>help</code> , 149 |
| <code>\FPsincos</code> , 125 | <code>\hfil</code> , 48 |
| <code>\FPsub</code> , 122 | <code>\hfill</code> , 48 |
| <code>\FPtan</code> , 125, 126 | <code>\hfilneg</code> , 48 |
| <code>\FPtancot</code> , 125, 126 | <code>\hfuzz</code> , 37 |
| <code>\FPtruncate</code> , 124 | <code>\hline</code> , 183 |
| <code>\FPupn</code> , 127 | <code>\hoffset</code> , 37 |
| <code>\frac</code> , 77 | <code>\holdinginserts</code> , 37 |
| <code>\frame</code> , 172 | <code>\hom</code> , 76 |
| <code>\framebox</code> , 169, 170, 191 | <code>\hookleftarrow</code> , 76 |
| <code>\frown</code> , 76 | <code>\hookrightarrow</code> , 76 |
| <code>\fsc</code> , 147 | <code>howpublished</code> , 96 |
| <code>\futurelet</code> , 27, 62 | <code>\Href</code> , 142 |
| <code>\fvset</code> , 117 | <code>\href</code> , 142 |
| | <code>\hrule</code> , 148 |
| <code>\Gamma</code> , 75 | <code>\hsize</code> , 37 |
| <code>\gamma</code> , 75 | <code>\hskip</code> , 48, 49, 80 |
| <code>\gcd</code> , 76 | <code>\hspace</code> , 48, 49, 83 |
| <code>\gdef</code> , 23, 68 | <code>\hss</code> , 48 |
| <code>\ge</code> , 76 | <code>\ht</code> , 38 |
| <code>generatedvi</code> , 187 | <code>\htmladdnormallink</code> , 142 |
| <code>generateps</code> , 187 | <code>\htmlimage</code> , 148 |
| <code>\genfrac</code> , 77, 79 | <code>htmlonly</code> , 148 |
| <code>\geq</code> , 76 | <code>\HTMLset</code> , 147 |
| <code>\geqslant</code> , 76 | <code>\Huge</code> , 47, 188 |
| <code>\gets</code> , 76 | <code>\huge</code> , 47, 188 |
| <code>\gg</code> , 76 | <code>\hyphenation</code> , 148 |
| <code>\glo</code> , 190 | <code>\hyphenpenalty</code> , 35 |
| <code>\global</code> , 19, 42, 67, 138 | |
| <code>\globaldefs</code> , 36 | <code>\i</code> , 75, 112, 120 |
| <code>glossaire</code> , 190 | <code>\ieme</code> , 47, 118 |
| <code>glossary</code> , 145 | <code>\iemes</code> , 47, 118 |
| <code>\glue</code> , 32 | <code>\ier</code> , 47, 118 |
| <code>\goodbreak</code> , 50 | <code>\iere</code> , 47, 118 |
| <code>\grave</code> , 77 | <code>\ieres</code> , 118 |
| <code>\gtrless</code> , 76 | <code>\iers</code> , 47, 118 |

| | |
|---|-------------------------------------|
| <code>\if</code> , 57, 69 | <code>\isundefined</code> , 134 |
| <code>\ifcase</code> , 56, 69 | <code>\it</code> , 8, 46, 78 |
| <code>\ifcat</code> , 57, 69 | <code>\Item</code> , 190 |
| <code>\ifdim</code> , 58, 69 | <code>\item</code> , 7, 20, 146 |
| <code>\ifeof</code> , 59, 69, 137 | <code>\itemindent</code> , 38 |
| <code>\iff</code> , 76 | itemize, 145, 190 |
| <code>\iffalse</code> , 58, 69 | <code>\itemsep</code> , 35 |
| <code>\IfFileExists</code> , 140 | <code>\itshape</code> , 8, 46, 188 |
| <code>\ifFPtest</code> , 121 | |
| <code>\ifhbox</code> , 59, 69 | <code>\j</code> , 120 |
| <code>\ifhmode</code> , 59, 69 | jan, 95 |
| <code>\ifinner</code> , 59, 69 | <code>\jmath</code> , 76 |
| <code>\ifmmode</code> , 59, 69 | <code>\jobname</code> , 68 |
| <code>\ifnextchar</code> , 61 | <code>\Join</code> , 76 |
| <code>\ifnum</code> , 54, 58, 69 | journal, 96 |
| <code>\ifodd</code> , 69 | jul, 95 |
| <code>\ifthenelse</code> , 56, 135 | jun, 95 |
| <code>\iftrue</code> , 58, 69 | |
| <code>\ifvbox</code> , 59, 69 | <code>\k</code> , 68, 113 |
| <code>\ifvmode</code> , 59, 69 | <code>\kappa</code> , 75 |
| <code>\ifvoid</code> , 59, 69 | <code>\ker</code> , 76 |
| <code>\ifx</code> , 27, 59, 69 | <code>\kern</code> , 32, 80 |
| <code>\ignorespaces</code> , 47, 141 | key, 96, 98 |
| <code>\Im</code> , 75 | (field in an entry), 92 |
| <code>\imath</code> , 76 | cite, 91 |
| <code>\immediate</code> , 148 | print, 91, 92 |
| in, 39 | sort, 92, 97 |
| <code>\in</code> , 76 | |
| inbook, 95 | <code>\L</code> , 118 |
| <code>\include</code> , 69, 140 | <code>\l</code> , 118 |
| <code>\includegraphics</code> , 147, 196, 210 | <code>\label</code> , 79, 142 |
| incollection, 95 | <code>\labelsep</code> , 35 |
| <code>\indent</code> , 49 | <code>\labelwidth</code> , 38 |
| <code>\inf</code> , 76 | <code>\Lambda</code> , 75 |
| <code>\infty</code> , 75 | <code>\lambda</code> , 75 |
| <code>\injl</code> , 76 | <code>\land</code> , 76 |
| <code>\input</code> , 69, 140 | lang-en, 152, 187 |
| <code>\InputIfFileExists</code> , 140 | lang-fr, 152, 187 |
| <code>\inputlineno</code> , 37 | <code>\langle</code> , 75, 77 |
| <code>\insert</code> , 141 | Language, 152, 187 |
| <code>\insertbibliohere</code> , 104 | <code>\language</code> , 36, 130 |
| institution, 96 | <code>\LARGE</code> , 47, 188 |
| <code>\int</code> , 13, 76 | <code>\Large</code> , 47, 188 |
| interactivebib, 150 | <code>\large</code> , 47, 188 |
| interactivemath, 150 | <code>\lastbox</code> , 143 |
| <code>\interlinepenalty</code> , 36 | <code>\lastkern</code> , 141 |
| <code>\iota</code> , 75 | <code>\lastpenalty</code> , 37, 141 |
| isbn, 96 | <code>\lastskip</code> , 141 |
| <code>\isodd</code> , 134 | <code>\LaTeX</code> , 118 |
| <code>\isproject</code> , 161 | LaTeXonly, 148 |
| isrn, 96 | latexonly, 148 |
| issn, 96 | latin1, 149 |
| | <code>\lbrace</code> , 75, 77 |

| | |
|--|---|
| <code>\lccode</code> , 38 | <code>\Longleftarrow</code> , 76 |
| <code>\lceil</code> , 75, 77 | <code>\longleftarrow</code> , 76 |
| <code>\ldots</code> , 75, 120 | <code>\longmapsto</code> , 76 |
| <code>\le</code> , 76 | <code>\Longrightarrow</code> , 76 |
| <code>\leaders</code> , 143 | <code>\longrightarrow</code> , 76 |
| <code>\leavevmode</code> , 49, 93 | <code>\loop</code> , 54, 69, 133 |
| <code>\left</code> , 6, 79 | <code>\looseness</code> , 36 |
| <code>\Leftarrow</code> , 76 | <code>\lor</code> , 76 |
| <code>\leftarrow</code> , 76 | <code>\lower</code> , 88 |
| <code>\leftharpoondown</code> , 76 | <code>\lowercase</code> , 67, 117 |
| <code>\leftharpoonup</code> , 76 | <code>\lsc</code> , 147 |
| <code>\lefthyphenmin</code> , 37 | <code>\mag</code> , 36 |
| <code>\leftmargin</code> , 38 | <code>\makeatletter</code> , 25 |
| <code>\leftmargini</code> , 38 | <code>\makeatother</code> , 25 |
| <code>\leftmarginii</code> , 38 | <code>\makebox</code> , 80, 169 |
| <code>\leftmarginiii</code> , 38 | <code>makedvi</code> , 187 |
| <code>\leftmarginiv</code> , 38 | <code>makefo</code> , 187 |
| <code>\leftmarginv</code> , 38 | <code>makehtml</code> , 187 |
| <code>\leftmarginvi</code> , 38 | <code>\MakeLowercase</code> , 118 |
| <code>\Leftrightarrow</code> , 76 | <code>makepdf</code> , 187 |
| <code>\leftrightharrow</code> , 76 | <code>\MakeUppercase</code> , 118 |
| <code>\leftskip</code> , 34 | <code>manual</code> , 95 |
| <code>\lengthtest</code> , 135 | <code>\mapsto</code> , 76 |
| <code>\leq</code> , 76 | <code>mar</code> , 95 |
| <code>\leqno</code> , 71 | <code>\mark</code> , 141 |
| <code>\leqslant</code> , 76 | <code>mastersthesis</code> , 95 |
| <code>\let</code> , 26 | <code>masterthesis</code> , 95 |
| <code>\lfloor</code> , 75, 77 | <code>math</code> , 73 |
| <code>\lfoot</code> , 146 | <code>\mathaccent</code> , 88, 114 |
| <code>\lg</code> , 76 | <code>\mathattribute</code> , 87 |
| <code>\lgroup</code> , 75, 77 | <code>\mathbb</code> , 12, 78 |
| <code>\lhead</code> , 146 | <code>\mathbbm</code> , 78 |
| <code>\lim</code> , 76 | <code>\mathbf</code> , 11, 78 |
| <code>\liminf</code> , 76 | <code>\mathbin</code> , 13, 84 |
| <code>\limits</code> , 84 | <code>\mathcal</code> , 12, 78 |
| <code>\limsup</code> , 76 | <code>\mathchar</code> , 87 |
| <code>\line</code> , 172, 198 | <code>\mathchardef</code> , 26, 88 |
| <code>\linebreak</code> , 50 | <code>\mathchoice</code> , 14, 77, 79, 84 |
| <code>\linepenalty</code> , 35 | <code>\mathclose</code> , 13, 84 |
| <code>\lineskip</code> , 34 | <code>\mathcode</code> , 38, 88 |
| <code>\lineskiplimit</code> , 37 | <code>\mathinner</code> , 13, 84 |
| <code>\linethickness</code> , 172, 198 | <code>\mathit</code> , 78 |
| <code>\linewidth</code> , 38 | <code>\mathmi</code> , 87 |
| <code>\ll</code> , 76 | <code>\mathmn</code> , 87 |
| <code>\lmoustache</code> , 75, 77 | <code>\mathmo</code> , 87 |
| <code>\ln</code> , 76 | <code>\mathnormal</code> , 78 |
| <code>\lnot</code> , 76 | <code>\mathop</code> , 13, 83, 84 |
| <code>\localisation</code> , 161 | <code>\mathopen</code> , 13, 84 |
| <code>\log</code> , 76 | <code>\mathord</code> , 13, 84 |
| <code>\long</code> , 19, 22, 52 | <code>\mathpunct</code> , 13, 84 |
| <code>\Longleftarrow</code> , 76 | <code>\mathrel</code> , 13, 84 |
| <code>\longleftarrow</code> , 76 | |

`\mathring`, 77
`\mathrm`, 78
`\mathsf`, 78
`\mathsurround`, 37
`\mathtt`, 78
`\mathversion`, 11
matrix, 74
`\max`, 76
`\maxdeadcycles`, 36
`\maxdepth`, 37
may, 95
`\mbox`, 80, 83, 169
`\mdseries`, 46
`\meaning`, 68
`\medbreak`, 50
`\medmuskip`, 35, 48
`\medskip`, 49
`\message`, 141
`\mho`, 76
`\mid`, 76
`\min`, 76
minipage, 171
misc, 95
`\mkern`, 148
mm, 39
`\models`, 76
module, 162
`\moduleref`, 163
month, 96
`\month`, 36
`\motscle`, 8
`\moveleft`, 143
`\moveright`, 143
`\mp`, 76
`\mu`, 75
`\multicolumn`, 81, 183
`\multimap`, 76
`\multiply`, 34, 42
`\multipt`, 171, 198
`\muskip`, 38
`\muskipdef`, 26

`\nabla`, 75
`\natural`, 75, 120
`\ne`, 76
`\narrow`, 76
`\neg`, 76
`\neq`, 76
`\newboolean`, 57, 134
`\newbox`, 29
`\newcolumnntype`, 181
`\newcommand`, 22

`\newcount`, 29
`\newcounter`, 29, 42
`\newdimen`, 29
`\newenvironment`, 24
`\newif`, 57
`\newlength`, 42, 58
`\newline`, 50
`\newlinechar`, 36
`\newskip`, 29
`\newtheorem`, 8
`\newtoks`, 29, 41
`\NG`, 118
`\ng`, 118
`\ni`, 76
`\No`, 118
`\no`, 119
no-footnote-hack, 187
`\noalign`, 184
nobibyearerror, 151
`\nobreak`, 148
`\nobreakspace`, 48, 78, 115, 121
`\nocentering`, 35, 142, 171
`\nocite`, 94, 96
noconfig, 150
noentnames, 149
`\noexpand`, 29, 30, 57, 68
`\noindent`, 49, 73
`\nolimits`, 84
`\nolinebreak`, 50, 84
`\nonscript`, 84
`\nonstopmode`, 147
`\nonumber`, 84
`\nopagebreak`, 50
`\normalfont`, 8
`\normalsize`, 47, 188
`\NOT`, 135
`\not`, 76, 135
note, 96
`\notin`, 76
notrivialmath, 149
noundefmac, 150
nov, 95
noxmlerror, 150
nozerowidthspace, 149
`\nu`, 75
`\nulldelimiterspace`, 37
number, 96
`\number`, 29, 42, 52, 64, 68
`\numberedverbatim`, 117
`\Numero`, 118
`\numero`, 119
`\nwarrow`, 76

$\backslash 0$, 118
 $\backslash o$, 118, 119
 oct, 95
 $\backslash oddsidemargin$, 38
 $\backslash odot$, 76
 $\backslash Oe$, 118
 $\backslash oe$, 118
 $\backslash of$, 79
 $\backslash og$, 119
 $\backslash oint$, 76
 $\backslash oldstylenums$, 189
 $\backslash Omega$, 75
 $\backslash omega$, 75
 $\backslash ominus$, 76
 $\backslash omit$, 183
 $\backslash openin$, 137
 $\backslash openout$, 139
 $\backslash oplus$, 76
 $\backslash OR$, 135
 $\backslash or$, 69, 135
 organization, 96
 $\backslash oslash$, 76
 $\backslash otimes$, 76
 $\backslash outer$, 19
 $\backslash output$, 35
 $\backslash outputpenalty$, 36
 $\backslash oval$, 172, 198
 $\backslash over$, 14, 21, 82
 $\backslash overbrace$, 77
 $\backslash overfullrule$, 37
 $\backslash overleftarrow$, 77
 $\backslash overline$, 77, 189
 $\backslash overrightarrow$, 77
 $\backslash overset$, 77
 $\backslash overwithdelims$, 82
 $\backslash owns$, 76

 $\backslash P$, 119
 page, 37
 $\backslash pagebreak$, 50
 pages, 96
 $\backslash par$, 49, 66
 paragraph, 37
 $\backslash paragraph$, 142, 190
 $\backslash parallel$, 76
 $\backslash parbox$, 171, 180
 $\backslash parfillskip$, 35
 $\backslash parindent$, 37, 49
 $\backslash parsep$, 35
 $\backslash parshape$, 35
 $\backslash parskip$, 34
 part, 37

 $\backslash part$, 142, 190
 $\backslash partial$, 75
 participant, 163
 participante, 163
 participantes, 163
 participants, 163
 $\backslash patterns$, 148
 $\backslash pausing$, 36
 pc, 39
 $\backslash penalty$, 32, 141
 $\backslash perp$, 76
 $\backslash pers$, 163
 phdthesis, 95
 $\backslash Phi$, 75
 $\backslash phi$, 75
 $\backslash Pi$, 75
 $\backslash pi$, 75
 picture, 168
 $\backslash pm$, 76
 pmatrix, 74
 $\backslash postdisplaypenalty$, 36
 $\backslash Pr$, 76
 preamble, 25, 95
 $\backslash prec$, 76
 $\backslash preceq$, 76
 $\backslash predisdisplaypenalty$, 36
 $\backslash predisplaysize$, 37
 $\backslash pretolerance$, 35
 $\backslash prime$, 14
 $\backslash proclaim$, 8
 $\backslash prod$, 76
 $\backslash project$, 161
 $\backslash projet$, 161
 $\backslash projlim$, 76
 $\backslash propto$, 76
 $\backslash protect$, 30, 148
 $\backslash provideboolean$, 134
 $\backslash providecommand$, 23
 ps, 150
 $\backslash psfig$, 147
 $\backslash Psi$, 75
 $\backslash psi$, 75
 pt, 39
 publisher, 96
 $\backslash put$, 171, 198

 $\backslash qbezier$, 171
 $\backslash quad$, 48, 75, 121
 $\backslash quad$, 48, 75, 121
 quotation, 148, 196
 quote, 148, 196

 $\backslash r$, 68, 113

`\radical`, 88
`\raggedleft`, 148
`\raggedright`, 148
`\raise`, 88
`\RALabel`, 148
`\rangle`, 75, 77
`\RAstartprojet`, 163
`\ratio`, 43, 134
`rawhtml`, 148
`\rbrace`, 75, 77
`\rceil`, 75, 77
`\Re`, 75
`\read`, 137
`\real`, 43, 134
`\ref`, 142
`refer`, 93, 96
`\refname`, 104
`\refstepcounter`, 43, 68
`\relax`, 30, 40, 42, 54, 79, 138
`\relpenalty`, 35
`\renewcommand`, 22
`\renewenvironment`, 22
`\rfloor`, 75, 77
`\rfoot`, 146
`\rgroup`, 75, 77
`\rhead`, 146
`\rho`, 75
`\right`, 6, 79
`\Rightarrow`, 76
`\rightarrow`, 76
`\rightharpoondown`, 76
`\rightharpoonup`, 76
`\righthyphenmin`, 37
`\rightmargin`, 38
`\rightskip`, 35
`\rm`, 46, 78, 83
`\rmfamily`, 46
`\rmoustache`, 75, 77
`robust`, 22, 30
`\Roman`, 42, 69
`\roman`, 42, 69
`\Romannumeral`, 42, 68
`\romannumeral`, 42, 64, 68
`\root`, 77
`\rotatebox`, 170, 192
`\rrrt`, 142

`\S`, 119
`\SaveVerb`, 68, 115
`\sb`, 26
`\sbox`, 64
`\sc`, 47

`\scalebox`, 170, 191
`\scaleput`, 171, 198
`scanint`, 55
`school`, 96
`\scriptscriptstyle`, 14, 80, 83
`\scriptscritpfont`, 46
`\scriptsize`, 47, 188
`\scriptspace`, 37
`\scriptstyle`, 14, 80, 83
`\scritpfont`, 46
`\scrollmode`, 147
`\scshape`, 46, 188
`\searrow`, 76
`\sec`, 76
`\section`, 5, 142, 190
`section-vals`, 187
`\selectfont`, 8, 45
`sep`, 95
`series`, 96
`\setboolean`, 58, 134
`\setbox`, 26
`setcounter`, 42
`\setcounter`, 68, 133
`\setlanguage`, 141
`\setlength`, 42, 68, 133
`\setminus`, 76
`\sf`, 46, 78
`\sfcode`, 38
`\sffamily`, 46, 189
`\sharp`, 75, 121
`shell-escape`, 150
`\shipout`, 141
`\show`, 33, 164
`\showbox`, 32, 164
`\showboxbreadth`, 32, 36
`\showboxdept`, 32, 36
`\showlists`, 164
`\showthe`, 33, 34, 164
`\Sigma`, 75
`\sigma`, 75
`silent`, 149
`\sim`, 76
`\simeq`, 76
`\sin`, 13, 76
`\sinh`, 76
`\skip`, 38, 58
`\skipdef`, 26
`\sl`, 46, 78
`\slash`, 119
`\sloppy`, 148
`\slshape`, 46, 188
`\small`, 47, 188

`\smallbreak`, 50
`\smallint`, 76
`\smallskip`, 49
`\smile`, 76
`sp`, 39
`\sp`, 26
`space`, 5, 6, 20, 23, 51, 62, 66
`\space`, 47, 48, 63
`\spaceskip`, 35
`\spadesuit`, 76
`\span`, 184
`\special`, 32, 148
`split`, 73
`\splitbotmark`, 69, 141
`\splitfirstmark`, 69, 141
`\splitmaxdepth`, 37
`\splittopskip`, 35
`\sqcap`, 76
`\sqcup`, 76
`\sqrt`, 77
`\sqsubset`, 76
`\sqsubseteq`, 76
`\sqsupset`, 76
`\sqsupseteq`, 76
`\square`, 76
`\SS`, 118, 119
`\ss`, 118, 119
`\stackrel`, 77
`\star`, 76
`\stepcounter`, 43
`string`, 95
`\string`, 22, 28, 29, 30, 68
`subequations`, 148
`\subfigure`, 192
`\subitem`, 148
`subparagraph`, 37
`\subparagraph`, 142, 190
`subsection`, 37
`\subsection`, 5, 142, 190
`\subset`, 76
`\subseteq`, 76
`subsubsection`, 37
`\subsubsection`, 142, 190
`\succ`, 76
`\succeq`, 76
`\sum`, 76
`\sup`, 76
`\supset`, 76
`\supseteq`, 76
`\surd`, 75
`\swarrow`, 76
`\T`, 68, 113
`table`, 178
`\tabskip`, 35
`tabular`, 178
`\tagcurve`, 172, 198, 199
`\tan`, 76
`\tanh`, 76
`\tau`, 75
`techreport`, 95
`\TeX`, 119
`\texorpdfstring`, 5
`\text`, 80, 83
`\textacutedbl`, 119
`\textasciicircum`, 119
`\textasciicute`, 119
`\textasciicutes`, 119
`\textasciiccedilla`, 119
`\textasciidieresis`, 119
`\textasciigrave`, 119
`\textasciimacron`, 119
`\textbackslash`, 119
`\textbaht`, 119
`\textbf`, 46
`\textbrokenbar`, 119
`\textbullet`, 119
`\textcelsius`, 119
`\textcent`, 119
`\textcircledP`, 119
`\textcopyright`, 119
`\textcurrency`, 119
`\textdagger`, 119
`\textdaggerdbl`, 119
`\textdegree`, 119
`\textdiv`, 119
`\textdollar`, 119
`\textdong`, 119
`\textestimated`, 119
`\texteuro`, 16, 119
`\textexclamdown`, 119
`\textflorin`, 119
`\textfont`, 46
`\textfrenchfranc`, 119
`\textgravedbl`, 119
`\textheight`, 38
`\textinterrobang`, 119
`\textit`, 46, 188
`\textlira`, 119
`\textlnot`, 119
`\textlquill`, 119
`\textmd`, 46
`\textmho`, 119
`\textmu`, 119
`\textnaira`, 119

| | |
|---|--|
| <code>\textnospace</code> , 120 | <code>\theta</code> , 75 |
| <code>\textnumero</code> , 120 | <code>\thicklines</code> , 172, 198 |
| <code>\textohm</code> , 120 | <code>\thickmuskip</code> , 35, 48 |
| <code>\textonehalf</code> , 120 | <code>\thinlines</code> , 172, 198 |
| <code>\textonequarter</code> , 120 | <code>\thinmuskip</code> , 35, 48 |
| <code>\textonesuperior</code> , 120 | <code>\tilde</code> , 77 |
| <code>\textordfeminine</code> , 120 | <code>\time</code> , 36 |
| <code>\textordmasculine</code> , 120 | <code>\times</code> , 76 |
| <code>\textparagraph</code> , 120 | <code>\tiny</code> , 47, 188 |
| <code>\textperiodcentered</code> , 120 | title, 96 |
| <code>\textpertenthousand</code> , 120 | <code>\to</code> , 76 |
| <code>\textperthousand</code> , 120 | <code>\toappear</code> , 143 |
| <code>\textpm</code> , 120 | <code>\toks</code> , 38 |
| <code>\textquestiondown</code> , 120 | <code>\toksdef</code> , 26 |
| <code>\textrecipe</code> , 120 | <code>\tolerance</code> , 35 |
| <code>\textreferencemark</code> , 120 | <code>\top</code> , 75 |
| <code>\textregistered</code> , 120 | <code>\topmark</code> , 69, 141 |
| <code>\textrm</code> , 46 | <code>\topskip</code> , 35 |
| <code>\textrquill</code> , 120 | <code>\tracingall</code> , 164 |
| <code>\textsc</code> , 46 | <code>\tracingcommands</code> , 36, 164, 165 |
| <code>\textsection</code> , 120 | <code>\tracinglostchars</code> , 36, 164 |
| <code>\textservicemark</code> , 120 | <code>\tracingmacros</code> , 36, 164, 165 |
| <code>\textsf</code> , 46 | <code>\tracingonline</code> , 36, 164 |
| <code>\textsl</code> , 46 | <code>\tracingoutput</code> , 36, 164 |
| <code>\textsofthyphen</code> , 120 | <code>\tracingpages</code> , 36, 164 |
| <code>\textsterling</code> , 120 | <code>\tracingparagraphs</code> , 36, 164 |
| <code>\textstyle</code> , 14, 83 | <code>\tracingrestores</code> , 36, 164, 165 |
| <code>\textsubscript</code> , 189 | <code>\tracingstats</code> , 36, 165 |
| <code>\textsuperscript</code> , 47, 189 | <code>\tralicsversion</code> , 68 |
| <code>\textthreequarters</code> , 120 | <code>\triangle</code> , 75 |
| <code>\textthreesuperior</code> , 120 | <code>\triangleleft</code> , 76 |
| <code>\texttimes</code> , 120 | <code>\triangleright</code> , 76 |
| <code>\texttrademark</code> , 120 | true, 39 |
| <code>\texttt</code> , 46 | <code>\tt</code> , 5, 46, 78 |
| <code>\texttwosuperior</code> , 120 | <code>\ttfamily</code> , 46, 189 |
| <code>\textup</code> , 46 | <code>\tw@</code> , 52 |
| <code>\textvisiblespace</code> , 115, 120 | type, 96, 150 |
| <code>\textwidth</code> , 38 | <code>\typeout</code> , 141 |
| <code>\textwon</code> , 120 | |
| <code>\textyen</code> , 120 | <code>\u</code> , 68, 113 |
| <code>\textystyle</code> , 80 | <code>\uccode</code> , 38 |
| <code>\tfrac</code> , 77 | <code>\uchyph</code> , 36 |
| <code>\TH</code> , 118, 119 | <code>\UndefineShortVerb</code> , 115 |
| <code>\th</code> , 118, 119 | <code>\underbrace</code> , 77 |
| <code>\the</code> , 6, 29, 33, 34, 69 | <code>\underleftarrow</code> , 77 |
| thebibliography, 104 | <code>\underline</code> , 77, 189 |
| <code>\theme</code> , 161 | <code>\underrightarrow</code> , 77 |
| theme-vals, 187 | <code>\underset</code> , 77 |
| <code>\theorembodyfont</code> , 135 | <code>\unhbox</code> , 141 |
| <code>\theoremheaderfont</code> , 135 | <code>\unhcopy</code> , 141 |
| <code>\theoremstyle</code> , 135 | <code>\unitlength</code> , 38, 168 |
| <code>\Theta</code> , 75 | <code>\unkern</code> , 141 |

| | |
|--|--|
| <code>\unnumberedverbatim</code> , 117 | <code>\verbatimnumberfont</code> , 116 |
| <code>\unpenalty</code> , 141 | <code>\verbatimprefix</code> , 115 |
| unpublished, 95 | verbose, 149 |
| <code>\unskip</code> , 47, 141 | <code>\verbprefix</code> , 115 |
| <code>\unvbox</code> , 141 | version, 149 |
| <code>\unvcopy</code> , 141 | <code>\Vert</code> , 75, 77 |
| <code>\Uparrow</code> , 75, 77 | <code>\vert</code> , 75, 77 |
| <code>\uparrow</code> , 75, 77 | <code>\vfil</code> , 49 |
| <code>\Updownarrow</code> , 75, 77 | <code>\vfill</code> , 49 |
| <code>\updownarrow</code> , 75, 77 | <code>\vfilneg</code> , 49 |
| <code>\uplus</code> , 76 | <code>\vfuzz</code> , 37 |
| <code>\uppercase</code> , 117 | <code>Vmatrix</code> , 74 |
| <code>\upshape</code> , 46, 188 | <code>vmatrix</code> , 74 |
| <code>\Upsilon</code> , 75 | <code>\voffset</code> , 37 |
| <code>\upsilon</code> , 75 | volume, 96 |
| <code>\UR</code> , 162 | <code>\vrule</code> , 148 |
| ur-vals, 187 | <code>\vsize</code> , 37 |
| url, 96 | <code>\vskip</code> , 49, 80 |
| <code>\url</code> , 142, 187, 193 | <code>\vspace</code> , 49 |
| url-font, 187 | <code>\vsplit</code> , 143 |
| <code>\urlfont</code> , 187 | <code>\vss</code> , 49 |
| use-allsizes, 188 | <code>\vtop</code> , 143 |
| use-font-elt, 188 | <code>\wd</code> , 38 |
| <code>\usepackage</code> , 130 | <code>\wedge</code> , 76 |
| <code>\UseVerb</code> , 68, 115 | <code>\whiledo</code> , 131 |
| utf8, 149 | <code>\widehat</code> , 77 |
| <code>\V</code> , 68, 113 | <code>\widetilde</code> , 77 |
| <code>\v</code> , 68, 113 | <code>\widowpenalty</code> , 35 |
| <code>\vadjust</code> , 141 | <code>\widthof</code> , 134 |
| <code>\valign</code> , 184 | <code>\wp</code> , 75 |
| <code>\value</code> , 42, 68 | <code>\wr</code> , 76 |
| <code>\varepsilon</code> , 75 | <code>\write</code> , 30, 32, 139 |
| <code>\varkappa</code> , 75 | <code>\xbox</code> , 94, 143 |
| <code>\varphi</code> , 75 | <code>\xdef</code> , 28, 29 |
| <code>\varpi</code> , 75 | <code>\Xi</code> , 75 |
| <code>\varrho</code> , 75 | <code>\xi</code> , 75 |
| <code>\varsigma</code> , 75 | <code>\xleaders</code> , 143 |
| <code>\vartheta</code> , 75 | xml, 150 |
| <code>\vbadness</code> , 36 | xml-accueil-name, 194 |
| <code>\vbox</code> , 143 | xml-arc-name, 198 |
| <code>\vcenter</code> , 88 | xml-bezier-name, 198 |
| <code>\Vdash</code> , 76 | xml-biblio, 196 |
| <code>\vdash</code> , 76 | xml-bigcircle-name, 198 |
| <code>\vdots</code> , 77 | xml-box-name, 191 |
| <code>\vec</code> , 77 | xml-bpers-name, 198 |
| <code>\vector</code> , 172, 198 | xml-caption-name, 152, 193 |
| <code>\vee</code> , 76 | xml-cell-name, 197 |
| <code>\verb</code> , 25 | xml-circle-name, 198 |
| <code>\Verbatim</code> , 116 | xml-closecurve-name, 198 |
| verbatim, 116 | xml-composition-ra-name, 152, 194 |
| <code>\verbatimfont</code> , 115 | |

xml-curve-name, 198
 xml-dashline-name, 198
 xml-div0-name, 190
 xml-div1-name, 190
 xml-div2-name, 190
 xml-div3-name, 190
 xml-div4-name, 190
 xml-div5-name, 190
 xml-div6-name, 190
 xml-dottedline-name, 198
 xml-expanded-project-name, 152
 xml-fbox-name, 191
 xml-figure-env-name, 193
 xml-figure-name, 196
 xml-font-bold, 189
 xml-font-it, 188
 xml-font-large1, 188
 xml-font-large2, 188
 xml-font-large3, 188
 xml-font-large4, 188
 xml-font-large5, 188
 xml-font-medium, 189
 xml-font-normalsize, 188
 xml-font-roman, 189
 xml-font-sansserif, 189
 xml-font-sc, 188
 xml-font-slanted, 188
 xml-font-small, 188
 xml-font-small1, 188
 xml-font-small2, 188
 xml-font-small3, 188
 xml-font-small4, 188
 xml-font-tt, 189
 xml-font-upright, 188
 xml-fontr-large, 188
 xml-footnote-name, 187
 xml-glo-name, 190
 xml-gloitem-name, 189
 xml-graphic-name, 193
 xml-head-name, 190
 xml-item-name, 189
 xml-keywords-name, 191
 xml-labelitem-name, 189
 xml-line-name, 198
 xml-linethickness-name, 198
 xml-list-name, 189
 xml-mbox-name, 192
 xml-multiput-name, 198
 xml-oldstyle-name, 189
 xml-oval-name, 198
 xml-overline-name, 189
 xml-picture-name, 198
 xml-project-name, 152, 193
 xml-put-name, 198
 xml-rotatebox-name, 192
 xml-row-name, 197
 xml-scalebox-name, 191
 xml-scaleput-name, 198
 xml-scaption-name, 152, 193
 xml-sub-name, 189
 xml-subfigure-name, 192
 xml-sup-name, 189
 xml-table-env-name, 193
 xml-Table-name, 193
 xml-table-name, 197
 xml-tagcurve-name, 198
 xml-term-name, 192
 xml-texte-name, 192
 xml-theorem-head, 195
 xml-theorem-name, 195
 xml-thicklines-name, 198
 xml-thinlines-name, 198
 xml-topic-name, 192
 xml-topic-title, 152
 xml-underline-name, 189
 xml-vector-name, 198
 xml-xref-name, 193
 xml-xtheorem-name, 194
 \XMLaddatt, 33, 145
 xmlall, 150
 \XMLcite, 93
 \XMLcurrentid, 33, 37, 94
 xmlelement, 144
 \xmlelt, 143
 \xmleptyelt, 143
 xmlfo, 150
 xmlhtml, 150
 \XMLlastid, 33, 37, 94
 \xmllatex, 143
 xmllint, 150
 xmlonly, 148
 \XMLsolvecite, 94
 xmltex, 150
 \xscale, 171, 199
 \xscaley, 171, 199
 \xspace, 69, 141
 \xspaceskip, 35
 year, 93, 96, 150
 \year, 36
 \yscale, 171, 199
 \yscalex, 171, 199
 \z@, 37
 \zetaeta, 75

Bibliography

- [1] David Carlisle, Michel Goossens, and Sebastian Rahtz. De XML à PDF avec `xmltex` et PassiveT_EX. In *Cahiers Gutenberg*, number 35-36, pages 79–114, 2000.
- [2] Michel Goossens, Frank Mittelbach, and Alexander Samarin. *The L^AT_EX companion*. Addison Wesley, 1993.
- [3] José Grimm. Outils pour la manipulation du rapport d’activité. Technical Report RT-0265, Inria, 2002.
- [4] Donald E. Knuth. *The T_EXbook*. Addison Wesley, 1984.
- [5] Philippe Louarn. Une expérience d’utilisation de LaTeX : le Rapport d’activité de l’INRIA. *Cahiers Gutenberg*, (0):17–24, apr 1988.
- [6] Frank Mittelbach, Michel Goossens, Johannes Braams, David Carlisle, and Chris Rowley. *The L^AT_EX companion, second edition*. Addison Wesley, 2004.
- [7] The Unicode Consortium. *The Unicode Standard, version 4.0*. Addison Wesley, 2003.
- [8] W3C. Extensible Markup Language (XML) 1.0 (Third Edition). <http://www.w3.org/TR/REC-xml/>, 1998. Third edition published in 2004.
- [9] W3C. Extensible Markup Language (XML) 1.1. <http://www.w3.org/TR/xml11/>, 2004.

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 3 |
| 1.1 | A short history of the Raweb | 3 |
| 1.2 | Birth of Tralics | 4 |
| 1.3 | Main objectives | 5 |
| 1.4 | An example | 6 |
| 1.5 | Some remarks on the Translation | 7 |
| 1.6 | Category codes and characters | 9 |
| 1.7 | Considerations about mathematics | 11 |
| 1.8 | Some subtleties of T _E X | 15 |
| 1.9 | Language options | 16 |
| 2 | Expansion | 19 |
| 2.1 | Defining new commands | 19 |
| 2.2 | Defining commands in L ^A T _E X | 22 |
| 2.3 | Some small examples | 25 |
| 2.4 | Variables in T _E X | 31 |
| 2.5 | All the variables | 34 |
| 2.6 | Using the variables | 39 |
| 2.7 | Counters | 42 |
| 2.8 | Fonts | 45 |
| 2.9 | Spaces | 47 |
| 2.10 | Conditional expansion | 50 |
| 2.11 | Conditionals in T _E X | 55 |
| 2.12 | A nontrivial command <code>\verb</code> | 66 |
| 2.13 | Expandable tokens | 68 |
| 3 | Mathematics | 71 |
| 3.1 | Introduction | 71 |
| 3.2 | The basic objects | 73 |
| 3.3 | Parsing a math formula | 77 |
| 3.4 | Translation of arrays | 81 |
| 3.5 | Trivial math | 81 |

| | | |
|----------|--|------------|
| 3.6 | Conversion to XML | 82 |
| 3.7 | Final math mode hacks | 86 |
| 3.8 | Extensions | 87 |
| 4 | Translating a bibliography | 89 |
| 4.1 | Introduction | 89 |
| 4.2 | Citing a document | 92 |
| 4.3 | Using Tralics instead of bibtex | 95 |
| 4.4 | The format of a name | 99 |
| 4.5 | Commands for the bbl | 103 |
| 4.6 | Other commands | 104 |
| 5 | Other commands | 109 |
| 5.1 | Character encoding | 109 |
| 5.2 | Characters and Accents | 111 |
| 5.3 | Verbatim material | 115 |
| 5.4 | Case change | 117 |
| 5.5 | Simple commands | 118 |
| 5.6 | The fp package | 121 |
| 5.7 | Action before translation | 130 |
| 5.8 | Other expandable commands | 131 |
| 5.9 | Other non-expandable commands | 141 |
| 6 | Running Tralics | 149 |
| 6.1 | The command line arguments | 149 |
| 6.2 | The standard configuration file | 151 |
| 6.3 | The action before translation | 155 |
| 6.4 | Translating the Raweb | 158 |
| 6.5 | Tracing commands | 164 |
| 6.6 | Pictures and friends | 168 |
| 6.7 | The title page | 173 |
| 6.8 | Array and Tables | 178 |
| 6.9 | Actions declared in the configuration file | 187 |
| 6.10 | A sample configuration file | 202 |
| 6.11 | Trace of titlepage | 205 |
| 6.12 | Bootstrap code | 207 |
| 6.13 | The puzzle | 210 |
| 6.14 | Images | 210 |



Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)

Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)

Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
<http://www.inria.fr>
ISSN 0249-0803