`omtext`: Semantic Markup for Mathematical Text Fragments in LaTeX*

Michael Kohlhase
Jacobs University, Bremen
`http://kwarc.info/kohlhase`

November 19, 2015

Abstract

The `omtext` package is part of the sTeX collection, a version of TeX/LaTeX that allows to markup TeX/LaTeX documents semantically without leaving the document format, essentially turning TeX/LaTeX into a document format for mathematical knowledge management (MKM).

This package supplies an infrastructure for writing OMDoc text fragments in LaTeX.

*Version v1.1 (last revised 2015/11/04)

Contents

1 Introduction

The `omtext` package supplies macros and environment that allow to mark up mathematical texts in SiTeX, a version of TeX/LaTeX that allows to markup TeX/LaTeX documents semantically without leaving the document format, essentially turning TeX/LaTeX into a document format for mathematical knowledge management (MKM). The package supports direct translation to the OMDoc format [**Kohlhase:OMDoc1.2**]

2 The User Interface

2.1 Package Options

`showmeta` The `omtext` package takes a single option: `showmeta`. If this is set, then the metadata keys are shown (see [**Kohlhase:metakeys:ctan**] for details and customization options).

2.2 Mathematical Text

`omtext` The `omtext` environment is used for any text fragment that has a contribution to a text that needs to be marked up. It can have a title, which can be specified via the `title=` `title` key. Often it is also helpful to annotate the `type` key. The standard relations `type=` from rhetorical structure theory `abstract`, `introduction`, `conclusion`, `thesis`, `comment`, `antithesis`, `elaboration`, `motivation`, `evidence`, `transition`, `note`, `annote` are recommended as values. Note that some of them are unary relations `for=` like `introduction`, which calls for a target. In this case, a target using the `for` key should be specified. The `transition` relation is special in that it is binary (a "transition between two statements"), so additionally, a source should be specified `from=` using the `from` key.

Note that the values of the `title` and `type` keys are often displayed in the text. `display=` This can be turned off by setting the `display` key to the value `flow`. Sometimes we want to specify that a text is a continuation of another, this can be done by `continues=` giving the identifier of this in the `continues` key.

Finally, there is a set of keys that pertain to the mathematical formulae in `functions=` the text. The `functions` key allows to specify a list of identifiers that are to be `theory=` interpreted as functions in the generate content markup. The `theory` specifies a EdN:1 module (see [**KohAmb:smmssl:svn**]) that is to be pre-loaded in this one[1] Finally, `verbalizes=` `verbalizes` specifies a (more) formal statement (see [**Kohlhase:smms:svn**]) that EdN:2 this text verbalizes or paraphrases.[2]

2.3 Phrase-Level Markup

`\phrase` The `phrase` macro allows to mark up phrases with semantic information. It takes `verbalizes=` an optional `KeyVal` argument with the keys `verbalizes` and `type` as above and `type=`

[1] EdNote: this is not implemented yet.
[2] EdNote: MK:specify the form of the reference.

style
class
index
\nlex
\nlcex
\coreft
\corefs

`style`, `class`, `index` that are disregarded in the LaTeX, but copied into the generated content markup.

We use the `\nlex{`⟨*phrase*⟩`}` for marking up phrases that serve as natural language examples and `\nlcex{`⟨*phrase*⟩`}` for counter-examples (utterances that are not acceptable for some reason). In natural language examples, we sometimes use "co-reference markers" to specify the resolution of anaphora and the like. We use the `\coreft{`⟨*phrase*⟩`}{`⟨*mark*⟩`}` to mark up the "target" of a co-reference and analogously `\corefs` for coreference source – e.g. for an anaphoric reference. The usage is the following:

```
\nlex{If \coreft{a farmer}1 owns \coreft{a donkey}2,
        \corefs{he}2 beats \corefs{it}2.}
```

is formatted to

If a farmer[1] owns a donkey[2], he$_2$ beats it$_2$.

\sinlinequote

The `sinlinequote` macro allows to mark up quotes inline and attribute them. The quote itself is given as the argument, possibly preceded by the a specification of the source in a an optional argument. For instance, we would quote Hamlet with

```
\sinlinequote[Hamlet, \cite{Shak:1603:Hamlet}]{To be or not to be}
```

which would appear as "*To be or not to be*" Hamlet, (Shakespeare 1603) in the text. The style in which inline quotations appear in the text can be adapted by specializing the macros `\@sinlinequote` — for quotations without source and `\@@sinlinequote` — for quotations with source.

\@sinlinequote
\@@sinlinequote

2.4 Block-Level Markup

sblockquote

The `sblockquote` environment is the big brother of the `\sinlinequote` macro. It also takes an optional argument to specify the source. Here the four internal macros `\begin@sblockquote` to `\end@@sblockquote` are used for styling and can be adapted by package integrators. Here a quote of Hamlet would marked up as

\begin@sblockquote
\end@@sblockquote

```
\begin{sblockquote}[Hamlet, \cite{Shak:1603:Hamlet}]\obeylines
  To be, or not to be: that is the question:
  Whether 'tis nobler in the mind to suffer
\end{sblockquote}
```

and would render as

> *To be, or not to be: that is the question:*
> *Whether 'tis nobler in the mind to suffer*
>
> Hamlet, (Shakespeare 1603)

\lec The \lec macro takes one argument and sets it as a comment at the end of the line, making sure that if the content is too long it is pushed into a new line. We use it internally for placing the of source of the sblockquote environment above. The \@@lec actual appearance of the line end comment is determined by the \@@lec macro, which can be customized in the document class.

2.5 Index Markup

The omtext package provides some extensions for the well-known indexing macros of LaTeX. The main reason for introducing these macros is that index markup in OMDoc wraps the indexed terms rather than just marking the spot for cross-referencing. Furthermore the index commands only indexes words unless noindex thenoindex option is set in the \usepackage. The omtext package and class EdN:3 make the usual \index macro undefined[3].

\indextoo The \indextoo macro renders a word and marks it for the index. Sometimes, we want to index a slightly different form of the word, e.g. for non-standard plurals: while \indextoo{word}s works fine, we cannot use this for the word "datum", which has the plural "data". For this we have the macro \indexalt \indexalt, which takes another argument for the displayed text, allowing us to use \indexalt{data}{datum}, which prints "data" but puts "datum" into the index.

The second set of macros adds an infrastructure for two-word compounds. Take for instance the compound "OMDoc document", which we usually want to add \twintoo into the index under "OMDoc" and "document". \twintoo{OMDoc}{document} is a variant of \indextoo that will do just this. Again, we have a version that prints a variant: This is useful for situations like this the one in Figure 1:

```
We call group \twinalt{Abelian}{Abelian}{group}, iff \ldots
```

will result in the following

> We call group Abelian, iff ...

and put "Abelian Group" into the index.

Example 1: Index markup

The third set of macros does the same for two-word compounds with adjectives, \atwintoo e.g. "wonderful OMDoc document". \atwin{wonderful}{OMdoc}{document} will make the necessary index entries under "wonderful" and "document". Again, \atwinalt we have a variant \atwinalt whose first argument is the alternative text.

All index macros take an optional first argument that is used for ordering the respective entries in the index.

[3] EdNote: implement this and issue the respective error message

3 Limitations

In this section we document known limitations. If you want to help alleviate them, please feel free to contact the package author. Some of them are currently discussed in the sTeX GitHub repository [**sTeX:github:on**].

1. none reported yet

4 Implementation

The `omtext` package generates two files: the LaTeX package (all the code between ⟨*package⟩ and ⟨/package⟩) and the LaTeXML bindings (between ⟨*ltxml⟩ and ⟨/ltxml⟩). We keep the corresponding code fragments together, since the documentation applies to both of them and to prevent them from getting out of sync.

The initial setup for LaTeXML:

```
⟨*ltxml⟩
package LaTeXML::Package::Pool;
use strict;
use LaTeXML::Package;
use LaTeXML::Util::Pathname;
⟨/ltxml⟩
```

4.1 Package Options

We declare some switches which will modify the behavior according to the package options. Generally, an option `xxx` will just set the appropriate switches to true (otherwise they stay false).[4]

EdN:4

```
⟨*package⟩
\newif\if@omtext@mh@\@omtext@mh@false
\DeclareOption{mh}{\@omtext@mh@true}
\newif\ifindex\indextrue
\DeclareOption{noindex}{\indexfalse}
\DeclareOption*{\PassOptionsToPackage{\CurrentOption}{modules}}
\ProcessOptions
\ifindex\makeindex\fi
⟨/package⟩
⟨*ltxml⟩
DeclareOption('mh', sub { AssignValue ('@omtext' => 1);
PassOptions('modules','sty',ToString(Digest(T_CS('\CurrentOption')))); });
DeclareOption('noindex','');
DeclareOption(undef,sub {PassOptions('modules','sty',ToString(Digest(T_CS('\CurrentOption'))));
ProcessOptions();
⟨/ltxml⟩
```

```
⟨*package⟩
\if@omtext@mh@\RequirePackage{omtext-mh}\fi
\RequirePackage{xspace}
\RequirePackage{modules}
\RequirePackage{comment}
\RequirePackage{mdframed}
\RequirePackage{latexsym}
⟨/package⟩
⟨*ltxml⟩
if(LookupValue('@omtext')) {RequirePackage('omtext-mh');}
RequirePackage('xspace');
```

[4] EdNote: need an implementation for LaTeXML

```
RequirePackage('modules');
RequirePackage('lxRDFa');
RequirePackage('latexsym');
⟨/ltxml⟩
```

4.2 Metadata

All the OMDoc elements allow to specify metadata in them, which is modeled by the `omdoc:metadata` element. Since the content of this element is precisely controlled by the Schema, we can afford to auto-open and auto-close it. Thus metadata elements from various sources will just be included into one `omdoc:metadata` element, even if they are supplied by different STEX bindings. Also we add numbering and location facilities.

```
⟨*ltxml⟩
Tag('omdoc:metadata',afterOpen=>\&numberIt,afterClose=>\&locateIt,autoClose=>1,autoOpen=>1);
⟨/ltxml⟩
```

the `itemize`, `description`, and `enumerate` environments generate `omdoc:li`, `omdoc:di` with `autoclose` inside a CMP. This behavior will be overwritten later, so we remember that we are in a `CMP` by assigning `_LastSeenCMP`.

```
⟨*ltxml⟩
Tag('omdoc:CMP', afterOpen => sub {AssignValue('_LastSeenCMP', $_[1], 'global');return;});#$
⟨/ltxml⟩
```

the `itemize`, `description`, and `enumerate` environments originally introduced in the `omtext` package do double duty in OMDoc, outside a `CMP` they are transformed into a `<omgroup layout='itemize`description`enumerate'>`, where the text after the macros `\item` come to be the children. If that is only text, then it is enclosed in an `<omtext><CMP>`, otherwise it is left as it is. The optional argument of the `\item` is transformed into the `<metadata><dc:title>` of the generated `\item` element.

```
⟨*ltxml⟩
DefParameterType('IfBeginFollows', sub {
   my ($gullet) = @_;
   $gullet->skipSpaces;
                    my $next = $gullet->readToken;
                    $gullet->unread($next);
                    $next = ToString($next);
                    #Hm, falling back to regexp handling, the $gullet->ifNext approach didn't wo
                    return 1 unless ($next=~/^\\begin/);
                    return;
                  },
 reversion=>'', optional=>1);
⟨/ltxml⟩
```

4.3 Mathematical Text

We define the actions that are undertaken, when the keys are encountered. The first set just records metadata; this is very simple via the `\addmetakey` infrastruc-

ture [**Kohlhase:metakeys:ctan**]. Note that we allow math in the `title` field, so we do not declare it to be `Semiverbatim` (indeed not at all, which allows it by default).

```
⟨*package⟩
\srefaddidkey{omtext}
\addmetakey[]{omtext}{functions}
\addmetakey*{omtext}{display}
\addmetakey{omtext}{for}
\addmetakey{omtext}{from}
\addmetakey{omtext}{type}
\addmetakey*{omtext}{title}
\addmetakey*{omtext}{start}
\addmetakey{omtext}{theory}
\addmetakey{omtext}{continues}
\addmetakey{omtext}{verbalizes}
\addmetakey{omtext}{subject}
⟨/package⟩
⟨*ltxml⟩
DefKeyVal('omtext','functions','CommaList');
DefKeyVal('omtext','display','Semiverbatim');
DefKeyVal('omtext','for','Semiverbatim');
DefKeyVal('omtext','from','Semiverbatim');
DefKeyVal('omtext','type','Semiverbatim');
DefKeyVal('omtext','title','Plain'); #Math mode in titles.
DefKeyVal('omtext','start','Plain'); #Math mode in start phrases
DefKeyVal('omtext','theory','Semiverbatim');
DefKeyVal('omtext','continues','Semiverbatim');
DefKeyVal('omtext','verbalizes','Semiverbatim');
⟨/ltxml⟩
```

The next keys handle module loading (see [**KohAmb:smmssl:ctan**]).

```
% \ednote{MK: need to implement these in LaTeXML, I wonder whether there is a general
% mechanism like numberit.}\ednote{MK: this needs to be rethought in the light of
% |\usemodule|. It is probably obsolete. Is this used? Is this documented?}
⟨*package⟩
\define@key{omtext}{require}{\requiremodules{#1}{sms}}
\define@key{omtext}{module}{\message{module: #1}\importmodule{#1}\def\omtext@theory{#1}}
```

`\st@flow` We define this macro, so that we can test whether the `display` key has the value `flow`

```
\def\st@flow{flow}
```

We define a switch that allows us to see whether we are inside an `omtext` environment or a statement. It will be used to give better error messages for inline statements.

```
\newif\if@in@omtext\@in@omtextfalse
```

`omtext` The `omtext` environment is different, it does not have a keyword that marks it. Instead, it can have a title, which is used in a similar way. We redefine the `\lec`

macro so the trailing \par does not get into the way.

```
\def\omtext@pre@skip{\smallskip}
\def\omtext@post@skip{}
\providecommand{\stDMemph}[1]{\textbf{#1}}
\newenvironment{omtext}[1][]{\@in@omtexttrue%
\bgroup\metasetkeys{omtext}{#1}\sref@label@id{this paragraph}%
\def\lec##1{\@lec{##1}}%
\ifx\omtext@display\st@flow\else\omtext@pre@skip\par\noindent%
\ifx\omtext@title\@empty%
\ifx\omtext@start\@empty\else\stDMemph{\omtext@start}\xspace\fi%
\else\stDMemph{\omtext@title}:\xspace%
\ifx\omtext@start\@empty\else\omtext@start\xspace\fi%
\fi% omtext@title empty
\fi% omtext@display=flow
\ignorespaces}
{\egroup\omtext@post@skip\@in@omtextfalse}
⟨/package⟩
⟨*ltxml⟩
DefEnvironment('{omtext} OptionalKeyVals:omtext',
  "<omdoc:omtext "
     . "?&GetKeyVal(#1,'id')(xml:id='&GetKeyVal(#1,'id')')() "
     . "?&GetKeyVal(#1,'type')(type='&GetKeyVal(#1,'type')')() "
     . "?&GetKeyVal(#1,'for')(for='&GetKeyVal(#1,'for')')() "
            . "?&GetKeyVal(#1,'from')(from='&GetKeyVal(#1,'from')')()>"
  . "?&GetKeyVal(#1,'title')(<dc:title>&GetKeyVal(#1,'title')</dc:title>)()"
  .     "?&GetKeyVal(#1,'start')(<ltx:text class='startemph'>&GetKeyVal(#1,'start')</ltx:text>)
                  .     "#body"
                ."</omdoc:omtext>");
⟨/ltxml⟩
```

4.4 Phrase-level Markup

\phrase For the moment, we do disregard the most of the keys

```
⟨*package⟩
\srefaddidkey{phrase}
\addmetakey{phrase}{style}
\addmetakey{phrase}{class}
\addmetakey{phrase}{index}
\addmetakey{phrase}{verbalizes}
\addmetakey{phrase}{type}
\addmetakey{phrase}{only}
\newcommand\phrase[2][]{\metasetkeys{phrase}{#1}%
\ifx\prhase@only\@empty\only<\phrase@only>{#2}\else #2\fi}
⟨/package⟩
⟨*ltxml⟩
DefKeyVal('phrase','id','Semiverbatim');
DefKeyVal('phrase','style','Semiverbatim');
DefKeyVal('phrase','class','Semiverbatim');
DefKeyVal('phrase','index','Semiverbatim');
```

```
DefKeyVal('phrase','verbalizes','Semiverbatim');
DefKeyVal('phrase','type','Semiverbatim');
DefKeyVal('phrase','only','Semiverbatim');
DefConstructor('\phrase OptionalKeyVals:phrase {}',
       "<ltx:text %&GetKeyVals(#1) ?&GetKeyVal(#1,'only')(rel='beamer:only' content='&GetKeyVal
⟨/ltxml⟩
```

`\coref*`

```
⟨*package⟩
\providecommand\textsubscript[1]{\ensuremath{_{#1}}}
\newcommand\corefs[2]{#1\textsubscript{#2}}
\newcommand\coreft[2]{#1\textsuperscript{#2}}
⟨/package⟩
⟨*ltxml⟩
DefConstructor('\corefs{}{}',
  "<ltx:text class='coref-source' stex:index='#2'>#1</ltx:text>");
DefConstructor('\coreft{}{}',
  "<ltx:text class='coref-target' stex:index='#2'>#1</ltx:text>");
⟨/ltxml⟩
```

`\n*lex`

```
⟨*package⟩
\newcommand\nlex[1]{\green{\sl{#1}}}
\newcommand\nlcex[1]{*\green{\sl{#1}}}
⟨/package⟩
⟨*ltxml⟩
DefConstructor('\nlex{}',"<ltx:text class='nlex'>#1</ltx:text>");
DefConstructor('\nlcex{}',"<ltx:text class='nlcex'>#1</ltx:text>");
⟨/ltxml⟩
```

`sinlinequote`

```
⟨*package⟩
\def\@sinlinequote#1{``{\sl{#1}}''}
\def\@@sinlinequote#1#2{\@sinlinequote{#2}~#1}
\newcommand\sinlinequote[2][]
{\def\@opt{#1}\ifx\@opt\@empty\@sinlinequote{#2}\else\@@sinlinequote\@opt{#2}\fi}
⟨/package⟩
⟨*ltxml⟩
DefConstructor('\sinlinequote [] {}',
              "<ltx:quote type='inlinequote'>"
               . "?#1(<dc:source>#1</dc:source>\n)()"
               . "#2"
            . "</ltx:quote>");
⟨/ltxml⟩
```

4.5 Block-Level Markup

`sblockquote`

```
⟨*package⟩
```

```
\def\begin@sblockquote{\begin{quote}\sl}
\def\end@sblockquote{\end{quote}}
\def\begin@@sblockquote#1{\begin@sblockquote}
\def\end@@sblockquote#1{\def\@@lec##1{{\rm ##1}}\@lec{#1}\end@sblockquote}
\newenvironment{sblockquote}[1][]
  {\def\@opt{#1}\ifx\@opt\@empty\begin@sblockquote\else\begin@@sblockquote\@opt\fi}
  {\ifx\@opt\@empty\end@sblockquote\else\end@@sblockquote\@opt\fi}
⟨/package⟩
⟨*ltxml⟩
DefEnvironment('{sblockquote} []',
  "<ltx:quote>?#1(<ltx:note role='source'>#1</ltx:note>)()#body</ltx:quote>");
⟨/ltxml⟩
```

sboxquote

```
⟨*package⟩
\newenvironment{sboxquote}[1][]
{\def\@@src{#1}\begin{mdframed}[leftmargin=.5cm,rightmargin=.5cm]}
{\@lec{\rm\@@src}\end{mdframed}}
⟨/package⟩
⟨*ltxml⟩
DefEnvironment('{sboxquote} []',
  "<ltx:quote class='boxed'>?#1(<ltx:note role='source'>#1</ltx:note>)()#body</ltx:quote>");
⟨/ltxml⟩
```

The line end comment macro makes sure that it will not be forced on the next line unless necessary.

\lec The actual appearance of the line end comment is determined by the \@@lec macro, which can be customized in the document class. The basic one here is provided so that it is not missing.

```
⟨*package⟩
\providecommand{\@@lec}[1]{(#1)}
\def\@lec#1{\strut\hfil\strut\null\nobreak\hfill\@@lec{#1}}
\def\lec#1{\@lec{#1}\par}
⟨/package⟩
⟨*ltxml⟩
DefConstructor('\lec{}',
   "\n<omdoc:note type='line-end-comment'>#1</omdoc:note>");
⟨/ltxml⟩
```

\my*graphics We set up a special treatment for including graphics to respect the intended OMDoc document structure. The main work is done in the transformation stylesheet though.

```
⟨ltxml⟩RawTeX('
⟨*ltxml | package⟩
\newcommand\mygraphics[2][]{\includegraphics[#1]{#2}}
\newcommand\mycgraphics[2][]{\begin{center}\mygraphics[#1]{#2}\end{center}}
\newcommand\mybgraphics[2][]{\fbox{\mygraphics[#1]{#2}}}
\newcommand\mycbgraphics[2][]{\begin{center}\fbox{\mygraphics[#1]{#2}}\end{center}}
```

```
⟨/ltxml | package⟩
⟨ltxml⟩');
```

4.6 Index Markup

\omdoc@index this is the main internal indexing command. It makes sure that the modules necessary for interpreting the math in the index entries are loaded. If the `loadmodules` key is given, we import the module we are in otherwise all the currently imported modules. We do not have to require the module files, since the index is a the end of the document. If the `at` key is given, then we use that for sorting in the index.

```
⟨*package⟩
\addmetakey{omdoc@index}{at}
\addmetakey[false]{omdoc@index}{loadmodules}[true]
\newcommand\omdoc@index[2][]{\ifindex%
\metasetkeys{omdoc@index}{#1}%
\@bsphack\begingroup\@sanitize%
\ifx\omdoc@index@loadmodules\@true%
\protected@write\@indexfile{}{\string\indexentry%
{\ifx\omdoc@index@at\@empty\else\omdoc@index@at @\fi%
{\string\importmodules{\@ifundefined{mod@id}\imported@modules\mod@id}%
#2}}{\thepage}}%
\else%
\protected@write\@indexfile{}{\string\indexentry%
{\ifx\omdoc@index@at\@empty\else\omdoc@index@at @\fi#2}{\thepage}}%
\fi% loadmodules
\endgroup\@esphack\fi}%ifindex
```

Now, we make two interface macros that make use of this:

\indexalt

```
\newcommand\indexalt[3][]{{#2}\omdoc@index[#1]{#3}}              % word in text and index
⟨/package⟩
⟨*ltxml⟩
DefConstructor('\indexalt[]{}{}',
       "<omdoc:idx>"
      .  "<omdoc:idt>#2</omdoc:idt>"
      .  "<omdoc:ide ?#1(sort-by='#1')()>"
      .    "<omdoc:idp>#3</omdoc:idp>"
      .  "</omdoc:ide>"
      ."</omdoc:idx>");
⟨/ltxml⟩
```

\indextoo

```
⟨*package⟩
\newcommand\indextoo[2][]{{#2}\omdoc@index[#1]{#2}}              % word in text and index
⟨/package⟩
⟨*ltxml⟩
DefConstructor('\indextoo[]{}',
       "<omdoc:idx>"
```

```
      .  "<omdoc:idt>#2</omdoc:idt>"
      .  "<omdoc:ide ?#1(sort-by='#1')()>"
      .    "<omdoc:idp>#2</omdoc:idp>"
      .  "</omdoc:ide>"
      ."</omdoc:idx>");
⟨/ltxml⟩
```

`\@twin` this puts two-compound words into the index in various permutations

```
⟨*package⟩
\newcommand\@twin[3][]{\omdoc@index[#1]{#2!#3}\omdoc@index[#1]{#3!#2}}
```

And again we have two interface macros building on this

`\twinalt`

```
\newcommand\twinalt[4][]{#2\@twin[#1]{#3}{#4}}
⟨/package⟩
⟨*ltxml⟩
DefConstructor('\twinalt[]{}{}{}',
       "<omdoc:idx>"
      .  "<omdoc:idt>#2</omdoc:idt>"
      .  "<omdoc:ide ?#1(sort-by='#1')()>"
      .    "<omdoc:idp>#2</omdoc:idp>"
      .    "<omdoc:idp>#3</omdoc:idp>"
      .  "</omdoc:ide>"
      ."</omdoc:idx>");
⟨/ltxml⟩
```

`\twinalt`

```
⟨*package⟩
\newcommand\twintoo[3][]{{#2 #3}\@twin[#1]{#2}{#3}} % and use the word compound too
⟨/package⟩
⟨*ltxml⟩
DefConstructor('\twintoo[]{}{}',
       "<omdoc:idx>"
      .  "<omdoc:idt>#2 #3</omdoc:idt>"
      .  "<omdoc:ide ?#1(sort-by='#1')()>"
      .    "<omdoc:idp>#2</omdoc:idp>"
      .    "<omdoc:idp>#3</omdoc:idp>"
      .  "</omdoc:ide>"
      ."</omdoc:idx>");
⟨/ltxml⟩
```

EdN:5 `\@atwin` this puts adjectivized two-compound words into the index in various permutations[5]

```
⟨*package⟩
\newcommand\@atwin[4][]{\omdoc@index[#1]{#2!#3!#4}\omdoc@index[#1]{#3!#2 (#4)}}
```

and the two interface macros for this case:

[5] EdNote: what to do with the optional argument here and below?

\@atwinalt

```
\newcommand\atwinalt[5][]{#2\@atwin[#1]{#3}{#4}{#4}}
⟨/package⟩
⟨*ltxml⟩
DefConstructor('\atwinalt[]{}{}{}{}',
       "<omdoc:idx>"
      .  "<omdoc:idt>#2</omdoc:idt>"
      .  "<omdoc:ide ?#1(sort-by='#1')()>"
      .    "<omdoc:idp>#2</omdoc:idp>"
      .    "<omdoc:idp>#3</omdoc:idp>"
      .    "<omdoc:idp>#4</omdoc:idp>"
      .  "</omdoc:ide>"
      ."</omdoc:idx>");
⟨/ltxml⟩
```

\atwintoo

```
⟨*package⟩
\newcommand\atwintoo[4][]{{#2 #3 #4}\@atwin[#1]{#2}{#3}{#4}}         % and use it too
⟨/package⟩
⟨*ltxml⟩
DefConstructor('\atwintoo[]{}{}{}',
       "<omdoc:idx>"
      .  "<omdoc:idt>#2 #3</omdoc:idt>"
      .  "<omdoc:ide ?#1(sort-by='#1')()>"
      .    "<omdoc:idp>#2</omdoc:idp>"
      .    "<omdoc:idp>#3</omdoc:idp>"
      .    "<omdoc:idp>#4</omdoc:idp>"
      .  "</omdoc:ide>"
      ."</omdoc:idx>");
⟨/ltxml⟩
```

4.7 LaTeX Commands we interpret differently

The first think we have to take care of are the paragraphs, we want to generate OMDoc that uses the `ltx:p` element for paragraphs inside `CMP`s. For that we have modified the DTD only to allowed `ltx:p` elements in `omdoc:CMP` (in particular no text). Then we instruct the `\par` macro to close a `ltx:p` element if possible. The next `ltx:p` element is then opened automatically, since we make `ltx:p` and `omdoc:CMP` autoclose and autoopen.

```
⟨*ltxml⟩
Tag('omdoc:CMP', autoClose=>1, autoOpen=>1);
Tag('omdoc:omtext', autoClose=>1, autoOpen=>1);
Tag('ltx:p', autoClose=>1, autoOpen=>1);
⟨/ltxml⟩
```

the rest of the reinterpretations is quite simple, we either disregard presentational markup or we re-interpret it in terms of OMDoc.[6]

EdN:6

[6] EdNote: MK: we should probably let LaTeXML deal with these and allow more text in the omdoc+ltxml.xsl

```
⟨*package⟩
\def\omspace#1{\hspace*{#1}}
⟨/package⟩
⟨*ltxml⟩
DefConstructor('\footnote[]{}',
       "<omdoc:note type='foot' ?#1(mark='#1')>#2</omdoc:note>");
DefConstructor('\footnotemark[]',"");
DefConstructor('\footnotetext[]{}',
       "<omdoc:note class='foot' ?#1(mark='#1')>#2</omdoc:note>");
⟨/ltxml⟩
```

4.8 Providing IDs for OMDoc Elements

To provide default identifiers, we tag all OMDoc elements that allow `xml:id` attributes by executing the `numberIt` procedure below. Furthermore, we use the `locateIt` procedure to give source links.

```
⟨*ltxml⟩
Tag('omdoc:omtext',afterOpen=>\&numberIt,afterClose=>\&locateIt);
Tag('omdoc:omgroup',afterOpen=>\&numberIt,afterClose=>\&locateIt);
Tag('omdoc:CMP',afterOpen=>\&numberIt,afterClose=>\&locateIt);
Tag('omdoc:idx',afterOpen=>\&numberIt,afterClose=>\&locateIt);
Tag('omdoc:ide',afterOpen=>\&numberIt,afterClose=>\&locateIt);
Tag('omdoc:idt',afterOpen=>\&numberIt,afterClose=>\&locateIt);
Tag('omdoc:note',afterOpen=>\&numberIt,afterClose=>\&locateIt);
Tag('omdoc:metadata',afterOpen=>\&numberIt,afterClose=>\&locateIt);
Tag('omdoc:meta',afterOpen=>\&numberIt,afterClose=>\&locateIt);
Tag('omdoc:resource',afterOpen=>\&numberIt,afterClose=>\&locateIt);
Tag('omdoc:recurse',afterOpen=>\&numberIt,afterClose=>\&locateIt);
Tag('omdoc:imports',afterOpen=>\&numberIt,afterClose=>\&locateIt);
Tag('omdoc:theory',afterOpen=>\&numberIt,afterClose=>\&locateIt);
Tag('omdoc:ignore',afterOpen=>\&numberIt,afterClose=>\&locateIt);
Tag('omdoc:ref',afterOpen=>\&numberIt,afterClose=>\&locateIt);
⟨/ltxml⟩
```

We also have to number some LaTeXML tags, so that we do not get into trouble with the OMDoctags inside them.

```
⟨*ltxml⟩
Tag('ltx:p',afterOpen=>\&numberIt,afterClose=>\&locateIt);
Tag('ltx:tabular',afterOpen=>\&numberIt,afterClose=>\&locateIt);
Tag('ltx:thead',afterOpen=>\&numberIt,afterClose=>\&locateIt);
Tag('ltx:td',afterOpen=>\&numberIt,afterClose=>\&locateIt);
Tag('ltx:tr',afterOpen=>\&numberIt,afterClose=>\&locateIt);
Tag('ltx:caption',afterOpen=>\&numberIt,afterClose=>\&locateIt);
Tag('ltx:Math',afterOpen=>\&numberIt,afterClose=>\&locateIt);
⟨/ltxml⟩
```

The `numberIt` procedure gets the prefix from first parent with an `xml:id` attribute and then extends it with a label that reflects the number of preceding siblings, provided that there is not already an identifier. Additionally, it estimates an XPointer

position in the original document of the command sequence which produced the tag. The `locateIt` subroutine is a sibling of `numberIt` as it is required as an `afterClose` handle for tags produced by LaTeX environments, as opposed to commands. `locateIt` estimates an XPointer end position of the LaTeX environment, allowing to meaningfully locate the entire environment at the source.

```
⟨*ltxml⟩
sub numberIt {
  my($document,$node,$whatsit)=@_;
  my(@parents)=$document->findnodes('ancestor::*[@xml:id]',$node);
  my $prefix= (@parents ? $parents[$#parents]->getAttribute('xml:id')."." : '');
  my(@siblings)=$document->findnodes('preceding-sibling::*[@xml:id]',$node);
  my $n = scalar(@siblings)+1;
  my $id = ($node -> getAttribute('xml:id'));
  my $localname = $node->localname;
  $node->setAttribute('xml:id'=>$prefix."$localname$n") unless $id;
  my $about = $node -> getAttribute('about');
  $node->setAttribute('about'=>'#'.$node->getAttribute('xml:id')) unless $about;
  #Also, provide locators:
  my $locator = $whatsit && $whatsit->getProperty('locator');
  #Need to inherit locators if missing:
  $locator = (@parents ? $parents[$#parents]->getAttribute('stex:srcref') : '') unless $locator
  if ($locator) {
    # There is a BUG with namespace declarations (or am I using the API wrongly??) which
    # does not recognize the stex namespace. Hence, I need to redeclare it...
    my $parent=$document->getNode;
    if(! defined $parent->lookupNamespacePrefix("http://kwarc.info/ns/sTeX"))
      { # namespace not already declared?
        $document->getDocument->documentElement->setNamespace("http://kwarc.info/ns/sTeX","stex
      }
    $node->setAttribute('stex:srcref'=>$locator);
  }return;}

sub locateIt {
  my($document,$node,$whatsit)=@_;
  #Estimate trailer and locator:
  my $locator = $node->getAttribute('stex:srcref');
  return unless $locator; # Nothing to do here...
  my $trailer = $whatsit && $whatsit->getProperty('trailer');
  $trailer = $trailer->getLocator if $trailer;
  $trailer = $locator unless $trailer; # bootstrap
  # TODO: Both should be local, or both remote, any mixture or undefinedness will produce garba
  my $file_path = LookupValue('SOURCEFILE');
  my $baselocal = LookupValue('BASELOCAL');
  # Hmm, we only care about relative paths, so let's just do a URL->pathname map
  $file_path=~s/^\w+\:\/// if $file_path;
  $baselocal=~s/^\w+\:\/// if $baselocal;
  if ($file_path && $baselocal && ($locator =~ s/^([^\#]+)\#/\#/)) {
    my $relative_path = pathname_relative($file_path,$baselocal);
    $locator = $relative_path.$locator;
```

```
  }
  if ($locator =~ /^(.+from=\d+;\d+)/) {
    my $from = $1;
    if ($trailer =~ /(,to=\d+;\d+.+)$/) {
      my $to = $1;
      $locator = $from.$to;
    } else { Error("stex","locator",undef, "Trailer is garbled, expect nonsense in stex:srcref
  } else { Error("stex","locator",undef, "Locator \"$locator\" is garbled, expect nonsense in s
  my $parent = $document->getNode;
  if(! defined $parent->lookupNamespacePrefix("http://kwarc.info/ns/sTeX"))
    { # namespace not already declared?
      $document->getDocument->documentElement->setNamespace("http://kwarc.info/ns/sTeX","stex",
    }
  $node->setAttribute('stex:srcref' => $locator);
  return;
}
⟨/ltxml⟩#$
```

4.9 Miscellaneous

Some shortcuts that use math symbols but are not mathematical at all; in particular, they should not be translated by LaTeXML.

```
⟨*package⟩
\newcommand\hateq{\ensuremath{\hat=}\xspace}
\newcommand\hatequiv{\ensuremath{\hat\equiv}\xspace}
\@ifundefined{ergo}%
{\newcommand\ergo{\ensuremath{\leadsto}\xspace}}%
{\renewcommand\ergo{\ensuremath{\leadsto}\xspace}}%
\newcommand{\reflect@squig}[2]{\reflectbox{$\m@th#1\rightsquigarrow$}}%
\newcommand\ogre{\ensuremath{\mathrel{\mathpalette\reflect@squig\relax}}\xspace}%
⟨/package⟩
⟨*ltxml⟩
DefMacro('\hateq','\@hateq\xspace');
DefConstructor('\@hateq',"\x{2259}");
DefMacro('\hatequiv','\@hatequiv\xspace');
DefConstructor('\@hatequiv',"\x{2A6F}");
DefMacro('\ergo','\@ergo\xspace');
DefConstructor('\@ergo',"\x{219D}");
DefMacro('\ogre','\@ogre\xspace');
DefConstructor('\@ogre',"\x{2B3F}");
⟨/ltxml⟩
```

4.10 Finale

We need to terminate the file with a success mark for perl.

```
⟨ltxml⟩1;
```