MECHANIZED MATHEMATICS AND ITS APPLICATIONS, VOL. 5, NO. 2, DECEMBER 2006, 19:31

# Automatic Translation in Formalized Mathematics

*Grzegorz Bancerek*

Białystok Technical University
bancerek@mizar.org

**Abstract –** This paper describes the automatic translation process used in producing the *Formalized Mathematics (a computer assisted approach)* journal. We report the current state of the implementation (bara version) as well as the improvements and future directions planned for the development of the process. In the description, we concentrate on the part of translation which is dependent on the author of the Mizar article. By indicating how symbols and Mizar formulae are translated, the readability of the output can be improved.

**Keywords –** formalized mathematics, automatic translation, text generation

## 1. Introduction

*Formalized Mathematics (a computer assisted approach)* (FM) publishes the contents of Mizar articles (see [Bon]) submitted to the Mizar Mathematical Library (MML). The first step is to translate the article contents into English and this translation is done automatically. The material published concerns the surface part of Mizar articles. Proofs forming the inner parts are not yet translated but other elements - theorems, definitions, schemes, and reservations and global sets, if necessary - are.

The process of translation is mechanized (see [B&C 1991,B&C 1992,Try]) but some portions of the final result depend on the author/editor. This dependence appears on several levels:

- Preparation of a summary and division of the article into titled sections. This is not an element of the translation process, but dividing the translated text into sections may have some impact on the homogeneity and transparency of the final result.
- Variables and reservations. Single letter variables in an article are preserved and others are abbreviated into single letters with indices. Reservations are translated in the place where they appear. So, the author of a Mizar article must be careful about the distance between a reservation of a variable and its use because this can impact readability.
- Formulae formulation. The process of translation works through several levels of generalization and it always uses the pattern from the most specific level fitting the formula translated. The formulation of theorems, definientia, and schemes in an appropriate manner may make use of those (more) specific patterns which were prepared to improve the translation.
- Existing formats. When an existing format is used to denote a new concept, it is translated according to the existing translation pattern for the format. In some situations, this may differ from the author's intentions.
- New symbols and formats. When a new format (with a new symbol or an old one) is used to denote a new concept, the translation pattern for this format is generated. The generation

Manuscript received February 27, 2006; revised November 22, 2006.

depends on the symbol itself and the format (see Section 8. for more details). The author of a Mizar article or the editor may change generated patterns according to their needs, but these changes are practically impossible after publication. The description and impact of translation patterns is given in Section 4. and the sections following it.

With regard to the above items, the author of a Mizar article is advised to pay attention to the following checkpoints when previewing[1] the article: definitions of new concepts, variable representations, parentheses in expressions with new functors, types and registrations with new attributes, and formulae with new predicates or attributes.

## 2. Historical Notes

The first attempt to increase the readibility of Mizar articles/abstracts was done by Piotr Rudnicki and Andrzej Trybulec in 1989 (see [R&T]).

The journal of *Formalized Mathematics* was established by the Association of Mizar Users (SUM, an acronym taken from the Polish name: *Stowarzyszenie Użytkowników Mizara*) in 1990. The first four volumes (1990–93) were published by Université Catholique de Louvain under the agreement of SUM and the Fondation Philippe le Hodey (copyrights), ISSN 0777-4028. From the fifth volume, *Formalized Mathematics* was published by Warsaw University, Białystok Branch, ISSN 1426-2630. After the transformation of the Białystok Branch into the University of Białystok in 1997, the latter became the publisher (with the same ISSN number).

The first issue of the journal was experimental and the contents did not differ much from the Mizar abstracts (see [EX2]). The second issue was produced to be much more similar to mathematical papers written in English (see [Mat]). The sentence structure of Mizar formulae (quantifiers, conjunctions, and some atomic formulae) and types were translated into English. The remaining elements were expressed with mathematical symbolisms available in TEX.

A new implementation of the Mizar system and new features in the Mizar language concerning attributes as well as a need for improvement in the quality of translation enforced redesign and re-implementation of the translation system in object oriented style. This was done and is still continuously developed by the author.

In 1996, an electronic extension of *Formalized Mathematics* was established[2]. In contrast to the paper edition, the electronic extension was dynamic and could follow changes in the Mizar system and in the MML as well as reflect improvements to the translation process. As a result, there were translations of Mizar articles updated according to their current form in the MML.

The journal contents (current and archival issues) are available from the site: `fm.mizar.org`. The previewing and proof-checking can also be done from the Internet.

[1] A previewing service is available under URL: `http://megrez.mizar.org/fm/preview.html`

[2] URL address: `http://mizar.org/JFM/`

## 3. The Process - Technical Details

The process of translation may be viewed as a rewriting system (see [Ban]). A Mizar article is reduced to some abstract form and then the analyzing rules are applied. After that, translation patterns for formulae and formats are used and finally the filling text, summary, and section titles are added.

These steps and the management of translation patterns are realized by the following programs.

| Program | param. | input files | output files |
|---|---|---|---|
| accom | $1 | $1.miz, PREL | ACCOM($1) |
| fmparse | $1 | $1.miz, ACCOM($1) | $1.fma, $1.nfr |
| newfmfrm | $1 -l$2 | $2.frd, ACCOM($1), PREL($1) | $1.fmn, $1.fmd |
| addfmfrm | $1 -l$2 | $2.frd, $1.fmn | $2.$-$ |
| fmfrm | $1 -l$2 | $2.frd, $1.nfr, $1.fmn, ACCOM($1) | $1.fmf, $1.fmz |
| resvar | $1 | $1.miz, ACCOM($1) | $1.ire |
| fmnotats | $1 | PREL($1) | $1.fms |
| fmanalyz | $1 | $1.fma, $1.fmf, $1.ire, $1.fmv, ACCOM($1) | $1.?, $1.fmv |
| jformath | $3 [/d\|/f] | [formath.set,] $3.lar, ∗.bnt, ∗.fms, ∗.? | $3.tex |
| latex | $3 | $3.tex, ∗.?, formath.cls, fom10.clo, mizarfrm.tex [, ∗.bbl] | $3.dvi, $3.aux |

In the table above, $1 stands for the name of a Mizar article, $2 stands for the name of a database file “$2.frd” with translation patterns, and $3 stands for the name of a list of Mizar articles (usually a code for the journal issue) in the file “$3.lar”. The question marks in “$1.?” and “∗.?” stand for the section number of the Mizar article and ∗ indicates the names of Mizar articles from the list in file $3.lar. ACCOM($1), PREL($1), and PREL stand for Mizar internal format files which are created by the accom program or are available from the Mizar directories. The contents of the files are as follows:

**$1.fma** - abstract description of the surface part of article $1.miz (reservations, definitions, theorems, schemes, and global sets)

**$1.nfr** - new formats introduced in article $1.miz

**$1.fmd** - generated translation patterns (with identification) of old formats existing in $2.frd which are used in definitions in article $1.miz

**$1.fmn** - generated (or re-edited by the editor or the author of an article) translation patterns (with identification) of new formats from article $1.miz (not yet introduced into $2.frd)

**$2.frd** - database file of translation patterns (with identification)

**$1.fmf** - translation patterns (without identification) of formats ordered according to $1.frm from ACCOM($1) and $1.nfr

**$1.fmz** - format identifications ordered according to $1.frm from ACCOM($1) and $1.nfr

**$1.ire** - information for reserved variables if they are used in elements translated

**$1.fms** - (signature) list of articles introducing notation (constructors) used in translated elements from article $1.miz

**$1.fmv** - representation of variables

**$1.?, $*$.?** - final LaTeX input including a translation of the section number ? from article $1.miz (or $*$.miz)

**$3.lar** - list of article names

**$1.bnt** - bibliographic notes of the article, summary, and section titles

**formath.set** - basic information of the publication issue

**formath.cls, fom10.clo, mizarfrm.tex** - LaTeX style files

## 4. Translation Patterns

Translations on the level of atomic formulae, terms, and types are based on *translation patterns* for formats. A format is a vocabulary symbol with an arity. For example,

| | |
|---|---|
| `- 1 1` | for subtraction ($x-y$) of reals, complex numbers, vectors, etc. |
| `- 0 1` | for opposite ($-x$) real, complex number, vector, etc. |
| `set 0` | for the type *set*, |
| `Function 2` | for a function from a set (first argument) into a set (second argument), |
| `in 1 1` | for the membership relation ($x \in y$), |
| `[ ] 2` | for an ordered pair ($\langle x, y\rangle$). |

There may be more than one format with the same symbol taking different numbers of arguments. In format descriptions, we take into consideration only visible arguments and do not consider their types or forms. The subtraction of vectors has the same format as the subtraction of reals although it has one extra (invisible) argument - the vector space.

Translation patterns for formats are grouped in the $1.fmn, $1.fmd, and $2.frd files, according to the vocabularies from which the format symbol comes. Each pattern is described by 3 or 4 lines. The first two (three, in the case of a bracket format) include the identification of the format. Specifically, the first line describes the format as: the format kind (one of the letters: `O K G J U R M L V`), the symbol number, and the (left, right) argument numbers. For a bracket functor (format kind `K`), the vocabulary and the number of the closing bracket are also given. The second line includes the symbol kind and the symbol itself. For a bracket functor, the symbol of the closing bracket is given in the third line.

```
#HIDDEN
K1 2 L1 vHIDDEN                  M1 0          R1 1 1
K[                               Mset          R=
L]
```

The third line (fourth, in the case of a bracket format) includes the control header and the proper pattern. For a selector functor (kind `U`), the fourth line includes the complemented pattern of translation of the associated field. For predicates (kind `R`), this line will include the complemented pattern of translation of the negative form. For types (`M` and `L`), this will be a description of the complemented pattern of translation of the plural form. The control header should not be repeated with the complemented patterns.

```
                                   ha set #0     m #1 = #2
mc#1#2; \langle #1, #2\rangle      sets #0       #1 \neq #2
```

The $1.fmf files include translation patterns without identification, i.e., the control headers with the proper patterns and the complemented patterns, if necessary. The `#1`, `#2`, ... in patterns indicate the places of arguments. The `#0` in a translation pattern for a type indicates the place of qualified variables. For example, the pattern

```
function #0 from #1 into #2
```

for the formula

```
for f being Function of A,B holds ...
```

renders

for every function $f$ from $A$ into $B$ ...

in most contexts.

A control header is a sequence of letters, digits, and other characters followed by a space. The formal grammar of the control header is given at `http://megrez.mizar.org/fm/control-header.html`. Terms for the grammar given in this text are written in `typewriter` font.

According to the information carried by the control headers we may divide all patterns into 4 categories:

1. Patterns for **functors**: `O` - proper functor, `K` - bracket functor, `G` - aggregate functor, `J` - forgetful functor (uses a symbol of kind `G`), and `U` - selector functor and field name.
2. Patterns for **predicates** (letter `R`).
3. Patterns for **types**: `M` - proper type, `L` - structure type (uses a symbol of kind `G`).
4. Patterns for **attributes** (letter `V`).

The control headers determine the behavior of the translator when a given format is translated. For example, patterns appear in verbal or symbolic mode and the control options (`TeX-mode` for functors and types and `Predicate-kind` for predicates) determine the destination of the pattern. Namely, `Math-mode` and `Symbolic` send patterns into symbolic mode (i.e., they are put between single $'s in TEX) and all others are put into verbal mode (i.e., no $'s are used). Adjectives are always verbal. In more complicated cases, when verbal patterns appear as arguments of symbolic patterns, a special technique of closing and opening the TEX math-mode inside a formula is used. The goal of these techniques is to preserve hyphenation in verbal mode and produce well balanced spacing in symbolic mode.

When a functor with a verbal pattern is translated, the word 'the' may be written before the pattern. This will happen for patterns with a small 'h' designating `Horizontal-mode`. A capital 'H' will cause translation without 'the'. For examples, see [EX6, Def. 15] and [EX3, Def. 9] which use the following patterns.

```
#ALG_1
O5 0 2
ONat_Hom
hol; natural homomorphism of #1 w.r.t. #2
#TSP_2
O1 0 2
OStone-retraction
Hosl(2)#1#2; Stone-retraction of #1 onto #2
```

Articles (*a*, *an*) in verbal phrases are determined by the `Article` or `Adjective-kind` control option from the pattern which is put first in the phrase. In cases of `HAS-`, `SAT-`, and `Noun-adjective`, articles should be written in the proper pattern as necessary.

## 5. Parentheses

Parentheses around terms are put in the translation process independently from those given in a Mizar article. The parentheses in a Mizar article are used to get correct identification according to the state of priorities which may differ from mathematical practice. Moreover, they can vary from author to author.

The translation process to calculate the parentheses uses the following header options (from patterns for functors):

`Bracket-ability` The expression built by the pattern may be taken into brackets if the pattern is `Open`. Otherwise, it does not need extra parentheses (e.g., the pattern already includes parentheses).

`Operation-kind` This option determines the operation kind of the pattern and argument contexts. The context is the position of an argument with respect to the operation symbol.

| | | | |
|---|---|---|---|
| `Prefix` | $\mathrm{oper}\,x$ | `Postfix` | $x\,\mathrm{oper}$ |
| `Upper-prefix` | ${}^{\mathrm{oper}}x$ | `Upper-postfix` | $x^{\mathrm{oper}}$ |
| `Lower-prefix` | ${}_{\mathrm{oper}}x$ | `Lower-postfix` | $x_{\mathrm{oper}}$ |
| `Overfix` | $\overline{x}$ | `Underfix` | $\underline{x}$ |
| `Infix` | $x_1 \circ x_2$ | `Non-associative-infix` | $x_1 \star x_2$ |
| `Right-associative-infix` | $x_1 \Rightarrow x_2$ | `Circumfix` | $\sqrt{x}$ |
| `Other-operation` | $\frac{x_1}{x_2}$ | | |

`Forcing` If the pattern should be put into parentheses in contexts other than those resulting from the operation kind, then all of the necessary operation kinds may be listed in this option. For example, if $\sqrt{x}$ is an argument of `Lower-postfix` (`w`), brackets should be put around it: $(\sqrt{x})_*$. For this, the control header and the proper pattern for `#1-root #2` are as follows:

```
    moc{w}@9#1#2; \sqrt[#1]{#2}
```

The standard forcing will force parentheses in the following situations: ${}^{\mathrm{oper}_1}x^{\mathrm{oper}_2}$, ${}^{\mathrm{oper}_1}x_{\mathrm{oper}_2}$, ${}_{\mathrm{oper}_1}x^{\mathrm{oper}_2}$, ${}_{\mathrm{oper}_1}x_{\mathrm{oper}_2}$, $\overline{x}_{\mathrm{oper}}$, ${}_{\mathrm{oper}}\overline{x}$, $\underline{x}^{\mathrm{oper}}$, ${}^{\mathrm{oper}}\underline{x}$. In the first situation, we get $({}^{\mathrm{oper}_1}x)^{\mathrm{oper}_2}$ or ${}^{\mathrm{oper}_1}(x^{\mathrm{oper}_2})$ depending on which of the two, $\mathrm{oper}_1$ or $\mathrm{oper}_2$, is the main operation and which is the suboperation.

`Priority` This option classifies operations into 4 groups: weak (values 0 and 1), additive (2–4), multiplicative (5–7), and strong (8, 9). If a weaker operation is an argument of a stronger one, it indicates a possible argument to be put into parentheses. If this option is omitted, the standard priority is assigned.

`Argument-context` If arguments should be treated differently from the standard given by `Operation-kind` or `Priority`, then the correct treatment may be given by this header option. For example, the translation pattern for `Funcs(#1, #2)` (the set of all functions from #1 into #2, see [EX2]) is as follows

```
moq(2)/1m/2q#1; #2^{#1}
```

The impact of the header can be observed in [EX1, (1)–(5)] where

```
Funcs(X, Funcs(Y, Z))
```

was translated to $(Z^Y)^X$. An example of another pattern is

```
moi(1,3)@0/1w@9#2; #1_{#2}{:=} #3
```

which is used in [EX10, Def. 20, (51)].

`List-of-bracket-disabled-arguments` The arguments specified in the list may not be taken into parentheses.

The main rule is to use as few brackets as possible without loss of meaning of an expression. For example, in the expression "`(A \/ B) \/ C`" the brackets are removed to produce: $A \cup B \cup C$ unless the exception below (associativity rule) holds. A symmetric convention for `Right-associative-infix` is applied. In situations "$\mathrm{oper}_1\, \mathrm{oper}_2\, x$" and "$x\, \mathrm{oper}_2\, \mathrm{oper}_1$", the parentheses are forced if the priority in the context of the argument of $\mathrm{oper}_1$ is much stronger than the priority of the suboperation $\mathrm{oper}_2$. However, brackets are always put in expressions like $(x^{\mathrm{oper}})^{\mathrm{oper}}$, $(x_{\mathrm{oper}})_{\mathrm{oper}}$, ${}^{\mathrm{oper}}({}^{\mathrm{oper}}x)$, and ${}_{\mathrm{oper}}({}_{\mathrm{oper}}x)$. For cases which do not appear in forcing lists, the priorities from the context and suboperation indicate whether or not parentheses should be applied.

There are two extra rules: symmetry and associativity. The symmetry rule is used when both arguments in an infix operation are of the same kind. In this situation, both are put into brackets if one of them should be. Examples may be found in [EX8, Def. 2 and Def. 3]. In the first example, the expression

$$(\text{the carrier of } S_1) \cup \text{the carrier of } S_2$$

was changed to

$$(\text{the carrier of } S_1) \cup (\text{the carrier of } S_2)$$

The associativity rule is used when variants of the associativity law are presented, as for example in

$$(x+y)+z = x+(y+z).$$

Then all brackets are introduced although the grouping to the left (or to the right in case of `Right-associative-infix`) is assumed in the standard way. For examples, see [EX5, (61), p. 315], [EX8, (10) and (27)], or [EX7, (26) and (34)].

## 6. Multipredicates

The direct translation of a Mizar formula is preceded by an analyzing procedure which may change the structure of the formula. These changes concern only the order of conjuncts and adjectives. They do not change the meaning of the formula but serve to improve readability.

To explain the analyzing process we must first introduce some terminology. Mizar atomic formulae fall into 3 groups: predicate formulae, qualifying formulae (e.g., $x$ is a function), and adjective formulae (e.g., $x$ is non empty). The last two always have verbal patterns

“`#1 is #2`” or “`#1 has #2`” or “`#1 satisfies #2`”

with a distinct subject. Predicate formulae may have symbolic or verbal patterns. Some of these verbal patterns: `IS-`, `HAS-`, `SAT-phrase`, `PRES-phrase`, and `INH-phrase`, have distinct subjects. Atomic formulae with distinct subjects are called *subjected phrases*. Subjected phrases fall into 5 categories

**IS** predicates with the `IS-phrases` control option, all qualifying formulae, and adjective formulae with attributes with `A-`, `AN-`, `Empty-`, or `Noun-adjective` control option

**HAS, SAT, PRES, INH** predicates with the `HAS`(`SAT`, `PRES`, `INH`, respectively)`-phrases` control option and adjective formulae with attributes with the `HAS`(`SAT`, `PRES`, `INH`, respectively)`-adjective` control option

The first step in analyzing a conjunction is the grouping of subjected phrases with the same subject into multipredicates. A multipredicate includes 3 collections - one for each category of subjected phrases. The translation pattern for a multipredicate depends on the quantities of those collections and on whether or not it is negated. If all collections are non empty and there is no negation, then it will look like

```
#1 is <IS-collection>, has <HAS-collection>,
     and satisfies <SAT-collection>
```

If a multipredicate is quantified by a variable which is the subject of the multipredicate, the ellipsis is used. In such situations, the following patterns are applied.

```
[not] for x holds M(x)          => [not] every $x$ <M>
[not] for x being T holds M(x)  => [not] every <T> <M>
[not] ex x st M(x)              => there exists [no] $x$ which <M>
[not] ex x being T st M(x)      => there exists [no] <a T> which <M>
```

The letter `M` stands for a multipredicate and `T` for a type. However, `<M>`,`<T>`, and `<a T>` stand for the translation of the multipredicate, the type, and the type with an article. For example, see [EX5, Def. 18] where

```
for I being Instruction of S holds I is IC-good;
```

was translated into

Every instruction of $S$ is IC-good.

Also, in theorem [EX1, (41)] the formula

```
for X being non empty set, S,T being non empty Poset
ex F being map of UPS(S, T|^X), UPS(S, T)|^X
 st F is commuting isomorphic
```

was translated into

> For every non empty set $X$ and all non empty posets $S$, $T$ holds there exists a map from $\mathrm{UPS}(S,T^X)$ into $\mathrm{UPS}(S,T)^X$ which is commuting and isomorphic.

Similar processing is used in the translation of types with adjectives and in the translation of registrations. The adjectives are grouped in three collections: `HAS`, `SAT`, and others. The translation patterns for types with adjectives also depend on quantities of the collections. If all collections are non empty, the pattern looks like

```
[a|an] <others-adjectives> <type> with <HAS-adjectives>
     and satisfying <SAT-adjectives>
```

This pattern was used in [EX4, (37)] for

```
for N being Hausdorff topological_semilattice
            with_open_semilattices Lawson (complete TopLattice)
 holds N is with_compact_semilattices
```

and should be rendered[3] as

> Every Hausdorff Lawson complete top-lattice with open semilattices and satisfying conditions of topological semilattice has compact semillatices.

This special processing causes a problem when a type with adjectives appears in a multipredicate in a qualifying formula.

## 7. Definition Style and Validity

The control options `Definition-style`, `Predicate-validity`, and `Type-validity` are used when translating definitions introducing appropriate concepts. `Definition-style` concerns functor definitions

```
func F -> specification (means|equals) ...
```

and is applied as follows:

[3] In the published version of this example, the previous implementation did not include the word *and* before *satisfying*.

| header | pattern |
|---|---|
| | `A-` or `AN-singular-functor` |
| hos, hon | `The <F> is <a specification> and is defined ...` |
| mos, mon | `The functor <F> is <a specification> and is defined ...` |
| | `Plural-functor` |
| hop | `The <F> constitute <a specification> defined ...` |
| | `Normal-functor` |
| ho | `The <F> yields <a specification> and is defined ...` |
| mo | `The functor <F> yields <a specification> and is defined ...` |

For example in [EX9, Def. 17], it is "The open neighborhoods of $p$ constitute a relational structure defined by . . . ".

When a Mizar article introduces an antonym which is formed in standard way (e.g., `#1 does not overlap #2` as an antonym for `#1 overlaps #2`) the definition of it does not need to be given in the translation. The value `Predicate-repeated` of the control option `Predicate-validity` is designed for such and similar situations. Similarly, if in a Mizar article there is a synonym of a type which gives the plural form of it, it does not need to be mentioned in the translation. The value `Type-repeated` in `Type-validity` is designed for such situations.

## 8. Conventions in Pattern Generation

Generated patterns for aggregates, brackets, and proper functors have `Math-mode` option by default. Other generated patterns are assumed to be verbal phrases. This classification may be changed for specific patterns unless they are for types, attributes, and field names which should always be verbal phrases. To generate a new translation pattern for verbal phrases the following operations on the symbol are applied:

- replacement of all substrings of the form "xX" (a letter followed by a capital letter) by "x X",
- replacement of underscore characters by spaces,
- conversion of all letters into lower case.

For example, the "`is_FreeGen_of`" symbol from the MML is converted into the phrase "is free gen of".

If the generation concerns an attribute format, then the following rules are applied.

| symbol (converted) | pattern generated |
|---|---|
| being <noun> | `b <noun>` |
| having <noun> | `h <noun>` |
| with <noun> | `x <noun>` |
| satisfying <noun> | `s <noun>` |
| <adjective> | `a <adjective>` or `n <adjective>` |

If the generation concerns a predicate format, then the following rules are applied. The first column indicates the number of left arguments.

| | symbol (converted) | pattern generated |
|---|---|---|
| 1 | is <phrase> | i <phrase> |
| 1 | has <phrase> | j <phrase> |
| 1 | satisfies <phrase> | s <phrase> |
| 2 | are <phrase> | h #1 and #2 are <phrase><br>#1 and #2 are not <phrase> |
| 2 | are not <phrase> | h #1 and #2 are not <phrase><br>#1 and #2 are <phrase> |
| | <phrase> | h <phrase><br>not <phrase> |

Symbolic patterns for predicates as well as verbal patterns for proper functors are not recognized so the author of a Mizar article should be sure to give the correct control header in such cases. The generation rules of patterns for functors are given in the table below.

| format | pattern generated |
|---|---|
| #ORDERS_1, G1 2, GRelStr | mc#1#2; \langle #1,#2 \rangle |
| #ORDERS_1, J1 1, GRelStr | hol#1; rel str of #1 |
| #ORDERS_1, U1 1, UInternalRel | hosl#1; internal rel of #1 |
| #HIDDEN, K1 2 L1 vHIDDEN, K[, L] | mc#1#2; [#1, #2] |
| #CARD_4, O1 0 1, Ocf<br>#FUNCT_2, O1 0 2, OFuncs | mol; \mathop{\rm cf} #1<br>mol#1#2; \mathop{\rm Funcs}(#1, #2) |
| #AMI_3, O2 1 1, O:=<br>#AMI_3, O2 2 1, O:= | moi; #1 \mathop{\rm :=} #2<br>moi#1#2; (#1, #2) \mathop{\rm :=} #2 |
| #OPPCAT_1, O1 1 0, Oopp | mor; #1 \mathop{\rm opp} |

There is no special processing for type patterns.

## 9. Further Work

The most recent plans concern the improvement of the quality of articles published. They include the following items:

– automatic variable representation, reservations, and sets
– new categories for subjected phrases and more detailed analysis of the order of adjectives in each category
– a richer database of translation patterns for formulae and variations of translation patterns for formats
– upgrading of the system to XML/XSLT technology which should result in a more flexible and more easily modifiable tool

Plans for the long term include the translation of proofs. This will be done in several steps. First, the references from proofs will be extracted and after some processing they will be translated. The next step will consist of a shallow translation of the proof (with the extraction applied for subproofs). Finally, some pruning of the proofs will be done before the shallow translation with extraction. The pruning will apply statistical analysis and an automated subject classification procedure for the MML. The translation of all proofs in whole is not planned.

The automatic generation of preliminaries for an article and each section outlined in [B&C 1992] are still not realized. This work is also dependent on the statistical analysis of the notation and terminology used and the theorems referred to in articles and should take into consideration the subject classification automatically developed for the MML.

On the other hand, the translation system cannot work well without direct contact with the authors of Mizar articles. One aspect that needs to be cleared is the possibility of involving authors in the previewing of translations before submission. Another advantageous aspect of such contact with authors concerns the impact authors will have on the development of a rich collection of translation patterns. The previewing process already allows authors to test interactively some possibilities of the translation system, but this area must be extended. A second role of the previewing step is to speed up the publication process. If the process of submission to the MML and the proof-reading process for publication in FM are separated it causes some additional problems: the Mizar article may change (revisions), there may not be necessary feedback on the Mizar article, and the author may lose interest.

In the plans connected with those aspects there is development of the web site: `fm.mizar.org`. It, among the FM-related things, should give direct access to the translation system as well as enable the download of the system. The previewer and proof-reader are already available at `http://megrez.mizar.org/fm/`.

As a by-product the web site includes aid tools for mathematical publications. It means that every text written in Mizar (even those still under development which may include checker errors) may be previewed and the results can be used as preliminary versions of mathematical papers (with partial mechanical verification).

**Acknowledgments**: The author would like to thank Roman Matuszewski, the Editor of FM, for his long lasting cooperation in the maintenance of the journal. The author would also like to thank Adam Grabowski who maintains the electronic extension of FM. The author also expresses his gratitude to Pauline N. Kawamoto and Piotr Rudnicki for their help during preparation of this paper.

## References

[Ban] G. Bancerek, *Reduction Relations*, Formalized Mathematics, Vol. 5, No. 4, 1996, pp.469-478.

[B&C 1991] G. Bancerek and P. L. Carlson, *Semi automated translation for mathematics*, Sprache-Kommunikation-Informatik, Akten des 26. Linguistischen Kolloquiums, Niemeyer, Poznan, 1991, pp.131-136.

[B&C 1992] G. Bancerek and P. L. Carlson, *Mizar and machine translation for mathematics documents*, `http://www.mizar.org/project/banc_carl93.ps`.

[Bon] E. Bonarska, *An Introduction to PC Mizar*, Fondation Philippe le Hodey, 1990.

[Mat] R. Matuszewski, *Preface*, Formalized Mathematics, Vol. 1, No. 2, 1990, p.255.

[R&T] P. Rudnicki and A. Trybulec, *A Collection of TeXed Mizar Abstracts*, Technical Report TR 89-18, University of Alberta, 1989, `http://mizar.org/project/TR-89-18.ps`.

[Try] A. Trybulec, *Introduction*, Formalized Mathematics, Vol. 1, No. 1, 1990, pp.7-8.

[EX1] G. Bancerek and A. Naumowicz, *Function spaces in the category of directed suprema preserving maps*, Formalized Mathematics, Vol. 9, No. 1, 2001, pp.171-177.

[EX2] C. Byliński, *Functions from a set to a set*, Formalized Mathematics, Vol. 1, No. 1, 1990, pp.153-164.

[EX3] Z. Karno, *Maximal Kolmogorov subspaces of a topological space as Stone retracts of the ambient space*, Formalized Mathematics, Vol. 5, No. 1, 1996, pp.125-130.

[EX4] A. Korniłowicz, *Meet continuous lattices revisited*, Formalized Mathematics, Vol. 9, No. 2, 2001, pp.249-254.

[EX5] A. Korniłowicz, *On the composition of macro instructions of standard computers*, Formalized Mathematics, Vol. 9, No. 2, 2001, pp.303-316.

[EX6] M. Korolkiewicz, *Homomorphisms of algebras. Quotient universal algebra*, Formalized Mathematics, Vol. 4, No. 1, 1993, pp.109-113.

[EX7] R. Milewski, *The ring of polynoms*, Formalized Mathematics, Vol. 9, No. 2, 2001, pp.339-346.

[EX8] Y. Nakamura and G. Bancerek, *Combining of circuits*, Formalized Mathematics, Vol. 5, No. 2, 1996, pp.283-295.

[EX9] A. Trybulec, *Moore–Smith convergence*, Formalized Mathematics, Vol. 6, No. 2, 1997, pp.213-225.

[EX10] A. Trybulec, Y. Nakamura, and P. Rudnicki, *The* $\mathbf{SCM}_{\mathrm{FSA}}$ *computer*, Formalized Mathematics, Vol. 5, No. 4, 1996, pp.519-528.