ProTem

Eric Hehner

ProTem is a programming system that serves as both programming language and operating system, and includes a theorem prover to check each step of program composition. This document is an informal specification of ProTem. Formal specifications of the data types and program semantics can be found in the book *a Practical Theory of Programming* (with syntactic differences). “ProTem” also means “for now”, short for the Latin words “pro tempore”.

Programming languages and operating system languages have a lot of functionality in common, but differ greatly in syntax and terminology. These differences are historical, accidental, and unnecessary. They complicate a programmer's life with no benefit. For example, a file is just a variable; file update and storage are just assignment. By unifying the programming language and the operating system commands, both gain in functionality. Communication channels and file piping are as useful in programming as they are in operating systems. Directories and permissions are useful in large-scale multi-programmer programs. Conditional execution (if) and indexed loops (for) are useful operating system commands.

ProTem is also designed for easy proof of correctness, including functionality, time requirements, and space requirements. To that end, loops can be constructed by labeling any block of code with a specification, and then using the label within the block of code. For example,

```
« n≥0 ⇒ n' = 0 » [if n>0 [n:= n-1. « n≥0 ⇒ n' = 0 »]]
```

The proof methods are the subject of the book *a Practical Theory of Programming* and paper **Specified Blocks**. They do not require preconditions, postconditions, or invariants. If proof is not wanted, then an ordinary identifier can be used as label. For example,

```
loop [if n>0 [n:= n-1. loop]]
```

A primary design criterion is to make ProTem a small, easy-to-learn, easy-to-use language. The size of a language can be measured by the number of symbols and by the complexity of grammar structure, which can be measured by the number of nonterminals. ProTem has 8 keywords. (C has 28, Python has 35, Pascal has 36, Haskell has 37, Ada has 62, MS Basic has 205.) ProTem is presented by a Presentation Grammar, which has just the structure that a programmer needs to know, not all the structure that a parser needs for parsing. It has 2 nonterminals (program and data) plus some informally defined kinds of names. (There is also an LL(1) grammar with 23 nonterminals and an LR(0) grammar with 13 nonterminals at the end of this document. For comparison, the Haskell grammar has 68 nonterminals, and the Python grammar has 87 nonterminals.) The design ethos demands an extremely good reason for adding a new feature to ProTem that requires a new keyword or syntax. That same design ethos will not tolerate any addition to the 2 nonterminals in the Presentation Grammar.

To judge ease of use, you need to use the language, but you may get a sense of the ease of use (and of the beauty of the language, if that's of interest) from reading example programs. For that purpose, there are example programs near the end of this document.

The design of ProTem is complete except for the following. We need to describe and compose picture and sound elements. We need to define touchpad and touchscreen gestures. We may need to define regions of documents and regions of the screen to be clickable links. An implementation is partly written.
## Contents

### Introduction
- if-Program
- case-Program
- for-Program
- Named Program
- Plan
- Dictionary Definition
- Synonym Definition

### Symbols
- Numbers
- Characters
- Binary Values
- Identities
- Bunches
- Sets
- Strings
- Lists
- Conditional Data
- Functions
- result-Expressions

### Presentation Grammar
- Dictionary Definition
- Intentionally Omitted Features
- Implementation Philosophy
- Predefined Names
- Example Programs
- Portation Simulation
- Quote Notation Lengths
- Huffman Codes
- Read Password

### Data
- Format
- Commands
  - Edit
  - Abort
  - Session
  - Permit
  - Identity
  - Undo
  - Names
  - Memo
  - Display
  - Context
  - Verify

### Type Transfer
- Miscellaneous
  - Intentionally Omitted Features
  - Implementation Philosophy
  - Predefined Names

### Scope
- Programs
- Variable Definition
- Assignment
- Constant Definition
- Data Definition
- Data Recursion
- Constant v Data Definition
- Program Definition
- Measuring Unit Definition
- Forward Definition
- Name Removal
- Output and Input
- Sequential Composition
- Concurrent Composition
- Channel Definition

### Programs
- Type Transfer
Symbols

ProTem has 8 keywords, plus 5 kinds of lexeme, and 68 other symbols; altogether they are:

```plaintext
case  else  for  if  new  old  plan  result
```

case  else  for  if  new  old  plan  result

number  text  name  comment  identity

```
“  ” “  ” «  » `  '  ,  ,..  ;  ;;  ;..  .  :  ::  :=  =
≠  <  >
≤
≥
!  !!  \?
(  )  { }  [ ]  \[  \]
〈  〉  ⟦  ⟧
_  |  ||
∞
&  %  +  –
×  /  \→
↔
⊤
⊥
∧
∨
^  ^^  @  *  ~
cè
#
☐
⊲
⊳
⊨
⫤

Some of the ProTem symbols may not be on your keyboard. Here are the substitutes.

for “  ” use "  ”
for ` use '  
for = use <=
for ≠ use /
for 〈 use <
for › use >
for ≥ use >=
for ≤ use <=
for ∨ use /
for ∧ use /
for \ use \/
for < use <
for > use >

The names `infinity`, `true`, and `false` are predefined, and redefinable.

A number is formed as one or more decimal digits, with an optional decimal point between digits. A decimal point must have at least one digit on each side of it. Here are four examples.

```
0     275     27.5     0.2
```

A text begins with a left-double-quote, continues with any number of any characters (but a double-quote (left or right) within a text must be underlined), and concludes with a right-double-quote. Characters within a text are not limited to any alphabet. Here are five examples.

```
"    "abc"    "don't"    "Just say "no"."    "♠♣♥♦"
```

A name is either simple or compound. A simple name is either plain or fancy. A plain simple name begins with a letter (from an alphabet), and continues with any number of letters and digits, except that keywords cannot be names. A fancy simple name begins with «  », and continues with any number of any characters (not limited to any alphabet) except «  », and ends with »  ». Within a fancy simple name, blank spaces, tabs, and new lines are not significant. A compound name is composed of two or more simple names joined with underscore characters. For examples:

- plain simple names: x  Al  george  refStack
- fancy simple names: «William & Mary»  «x' ≥ x»
- compound names: ProTem_grammars_LL1  DCS_«grad recruiting»_«2016-9-8»

A comment begins with `  ` and ends at the end of the line. Characters within a comment are not limited to any alphabet. For example: `  I      ProTem

An identity identifies a person. Ideally it is biometric, such as a fingerprint or iris scan. Or it could be an encoded (encrypted) password. It is used to grant access to programs and data.

Presentation Grammar

At each point in a program, a name is one of

- `newname`: a simple name that is not defined in the current scope,
- `oldname`: a compound name that is not defined in its dictionary
- `newname`: a simple name that is defined in the current scope,
- `oldname`: a compound name that is defined in its dictionary
An oldname is defined as one of: variablename, constantname, dataname, programname, channelname, unitname, or dictionaryname.

There are 32 ways of forming a program. Each way will be explained later. Some examples and pronunciations are shown on the right side.

\[
\begin{align*}
\text{new } & \text{newname : data := data} & \text{create variablename : type := initial value} \\
\text{new } & \text{newname := data} & \text{create constantname and evaluate data} \\
\text{new } & \text{newname = data} & \text{create dataname but don't evaluate data} \\
\text{new } & \text{newname [ program ]} & \text{create programname but don't execute program} \\
\text{new } & \text{newname ? data ! data} & \text{create channelname ? type ! initial value} \\
\text{new } & \text{newname #1} & \text{create measuring unitname} \\
\text{new } & \text{newname _} & \text{create dictionaryname} \\
\text{new } & \text{newname oldname} & \text{create synonym} \\
\text{new } & \text{newname} & \text{forward definition} \\
\text{old } & \text{oldname} & \text{remove or hide} \\
\text{variablename := data} & \text{assign variable to value} \\
\text{channelname ! data} & \text{to channel send output} \\
\text{channelname ? data} & \text{from channel receive input in this pattern} \\
\text{channelname ? data ! channelname} & \text{from channel receive input in this pattern and echo} \\
\text{channelname ? ! channelname} & \text{from channel receive corrected input and echo} \\
\text{simplename [ program ]} & \text{create programname and execute program} \\
\text{plan } & \text{simplename : data [ program ]} & \text{execute (call) named program} \\
\text{plan } & \text{simplename :: data [ program ]} & \text{plan, parameter is constantname} \\
\text{plan } & \text{simplename ! data [ program ]} & \text{plan, parameter is variablename} \\
\text{plan } & \text{simplename ? data [ program ]} & \text{plan, parameter is output channelname} \\
\text{plan } & \text{simplename ? data [ program ]} & \text{plan, parameter is input channelname} \\
\text{plan } & \text{data} & \text{plan, data argument} \\
\text{plan } & \text{variable} & \text{plan, variable argument} \\
\text{plan } & \text{channel} & \text{plan, channel argument} \\
\text{plan } & \text{program} & \text{sequential composition} \\
\text{plan } & \text{parallel program} & \text{concurrent composition} \\
\text{if } & \text{data [ program ]} & \text{if-program} \\
\text{if } & \text{data [ program ] else [ program ]} & \text{if-else-program} \\
\text{case } & \text{data [ program ]} & \text{case-program} \\
\text{case } & \text{data [ program ] else [ program ]} & \text{case-else-program} \\
\text{for } & \text{simplename := data [ program ]} & \text{for-program, create constantname} \\
\text{[ program ]} & \text{program parentheses}
\end{align*}
\]

There are 56 ways of expressing data. Each way will be explained later. Some examples and pronunciations are shown on the right side.

\[
\begin{align*}
\text{number} & \text{number} \\
\infty & \text{infinity, the infinite number} \\
\text{data & data} & \text{complex number, data } +i\times\text{ data} \\
\text{data \%} & \text{percentage, divide by 100} \\
+ \text{ data} & \text{plus, identity} \\
- \text{ data} & \text{minus, negation, not} \\
\text{data + data} & \text{plus, addition} \\
\text{data – data} & \text{minus, subtraction}
\end{align*}
\]
data × data
data / data
data ^ data
data ^^ data

\top
\bot
data ∧ data
data ∨ data
data = data
data ≠ data
data < data
data > data
data ≤ data
data ≥ data

data , data
data ... data
data ’ data
data : data

 ∈ data

{ data }

\sim data

\not data
text
data ; data
data ;.. data
data \setminus data
data ↔ data
data * data

* data
[
data ]

# data
data data
data @ data

\langle \text{ simplename} : \text{ data} \rightarrow \text{ data} \rangle

data \rightarrow data

\square data
data \mid data

\text{ variablename}

\text{ constantname}

\text{ dataname}

\text{ channelname} ??

\text{ channelname} !!

\text{ unitname}

data ⊨ data ⊨ data

\textbf{result} \text{ simplename} : \text{ data} := \text{ data} [\text{ program }]

(book)

\text{ data parenthese}

Here is the precedence (order of evaluation) of the forms of data.

<table>
<thead>
<tr>
<th>precedence level</th>
<th>forms of data</th>
<th>operators</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>number text name</td>
<td>$\top$ $\perp$ $\infty$ $!$ $!$ ( ) [ ] { } $\langle$ $\rangle$</td>
</tr>
<tr>
<td>1</td>
<td>data data postfix % ? !</td>
<td>$!$ $!$</td>
</tr>
<tr>
<td></td>
<td>prefix + - $&amp;$ $&amp;$ $|$ $|$</td>
<td>$\times$ $\div$ $#$ $\sim$ $\square$ $\ast$ $\vdash$ $\leftarrow$ $\rightarrow$ $\wedge$ $\vee$ $\top$ $\bot$ $\infty$</td>
</tr>
<tr>
<td>2</td>
<td>infix $\times$ $\div$ $|$ $|$</td>
<td>$\leftarrow$ $\rightarrow$ $\wedge$ $\vee$</td>
</tr>
</tbody>
</table>
|                  | infix $+$ $-$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;$ $;
Characters

A character is a text of length 1. We leave it to each implementation to list the characters, and to state their order. In addition to the character symbols such as “a” (small a) and “ ” (space), there are six predefined character names: delete (backspace), tab, nl (new line, next line, return, enter), click, doubleclick, and end (the end-of-file character). Predefined functions suc and pre give the successor and predecessor in the character order. Predefined functions charnat and natchar map between characters and their (possibly extended ASCII or unicode) numeric encodings. Character combinations, for example, shift-option-a, also have numeric encodings.

Binary Values

The two binary values are ⊤ and ⊥. Negation is –, conjunction is ∧, disjunction is ∨.

The infix 2-operand operators = and ≠ apply to all data in ProTem with a binary result; the two operands may even be of different types. The order operators < > ≤ ≥ apply to real numbers (including rationals, integers, and naturals), to characters, to binary values, to sets (subset, superset), to strings of ordered items, and to lists of ordered items, with a binary result. In the number order, –∞ is smallest and ∞ is largest. In the character order, “ ” comes first and end comes last. In the binary order, ⊥ is below ⊤, so ≤ is implication. The postfix operator !! applies to channels, and has a binary result saying whether there is written but unread data on the channel.

Identities

Identity values are obtained from biometric scanners, or from password encoding (encrypting). The textid function maps a text to an identity. Identities can be named, used in data structures, assigned, and communicated. They are used in permits (see Permit) and to obtain access to a persistent scope (see Session). They are used as signatures in documents.

Bunches

There are several predefined bunch names:

<table>
<thead>
<tr>
<th>bunch</th>
<th>description</th>
<th>examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>null</td>
<td>empty</td>
<td></td>
</tr>
<tr>
<td>nat</td>
<td>all natural numbers</td>
<td>0, 1, 2</td>
</tr>
<tr>
<td>int</td>
<td>all integer numbers</td>
<td>–2, –1, 0, 1, 2</td>
</tr>
<tr>
<td>rat</td>
<td>all rational numbers</td>
<td>1/2</td>
</tr>
<tr>
<td>real</td>
<td>all real numbers</td>
<td>2^(1/2)</td>
</tr>
<tr>
<td>com</td>
<td>all complex numbers</td>
<td>(–1)^(1/2), 3&amp;4</td>
</tr>
<tr>
<td>char</td>
<td>all characters</td>
<td>“a”</td>
</tr>
<tr>
<td>bin</td>
<td>both binary values</td>
<td>⊤, ⊥</td>
</tr>
<tr>
<td>text</td>
<td>all texts (character strings)</td>
<td>“abc”</td>
</tr>
<tr>
<td>everyone</td>
<td>all identities</td>
<td></td>
</tr>
<tr>
<td>all</td>
<td>all ProTem items</td>
<td></td>
</tr>
</tbody>
</table>

Any number, character, binary value, identity, set, string of elements, and list of elements is an elementary bunch, or synonymously, an element. For example, the number 2 is an elementary bunch, or element. Every expression is a bunch expression, though not all are elementary.

Bunch union is denoted by a comma:

A , B

A union B

For example,
2, 3, 5, 7
is a bunch of four integers. There is also the notation
\[x..y\]
x to y
where \(x\) and \(y\) are integers or \(\infty\) or \(-\infty\) or characters that satisfy \(x \leq y\). Note that \(x\) is included and \(y\) is excluded. For example, \(0..10\) is a bunch consisting of the first ten natural numbers, and \(5..5\) is the empty bunch \(null\).

For any \(A\) and \(B\),
\[A : B\]
\(A\) is included in \(B\)
is binary. The size (or cardinality) of \(A\) is \(\varepsilon A\). For examples, \(\varepsilon(0, 1) = 2\) and \(\varepsilon{null} = 0\) and \(\varepsilon(a..b) = b – a\).

Bunches are equal if and only if they consist of the same elements, ignoring order and multiplicity.

In ProTem, all operators whose precedence is before that of bunch union, except \(\varepsilon\) and \(\forall\), distribute over bunch union. Infix \(*\) distributes in its left operand only. For examples,
\[-(3, 5) = -3, -5\]
\[(2, 3)+(4, 5) = 6, 7, 8\]
This makes it easy to express the plural naturals (\(nat+2\)), the even naturals (\(nat \times 2\)), the square naturals (\(nat^2\)), the natural powers of two (\(2^n\)), and many other things.

Bunches serve as a type structure in ProTem, as the contents of sets, and other uses.

Sets

A set is formed by enclosing a bunch in set braces. For examples, \({0, 2, 5}\), \({0..100}\), \({null}\), \({nat}\). The inverse of set formation is the content operator ~. For example, \(~\{0, 1\} = 0, 1\). The size (or cardinality) of a set, traditionally written \(|S|\), is therefore \(\varepsilon\sim S\) in ProTem. For examples, \(\varepsilon\sim\{0, 1\} = 2\) and \(\varepsilon\sim{null} = 0\). The element relation, traditionally written \(x\in S\), is therefore \(x\sim S\) in ProTem. The union operator, traditionally \(\cup\), is \(\lor\) in ProTem. The intersection operator, traditionally \(\cap\), is \(\land\). Subset, traditionally \(\subseteq\), is \(\leq\); strict subset is \(<\); superset is \(\supseteq\); strict superset is \(>\). The power operator \(\forall\) takes a bunch as operand and produces all sets that contain only elements of the bunch. For example, \(\forall(0, 1) = \{null\}, \{0\}, \{1\}, \{0, 1\}\).

Strings

There is a predefined string name:
\[nil\]
the empty string
Any number, character, binary value, identity, list, and function is a one-item string, or synonymously, an item. For example, the number 2 is a one-item string, or item.

String join is denoted by a semi-colon:
\[S; T\]
\(S\) join \(T\)
For example,
\[2; 3; 5; 7\]
is a string of four integers. There is also the notation
\[x..y\]
x to y (same pronunciation as \(x..y\))
where \(x\) and \(y\) are integers or characters that satisfy \(x \leq y\). Again, \(x\) is included and \(y\) is excluded. For examples, \(0..10\) is a string consisting of the first ten natural numbers, and \(5..5 = null\).
The length of a string is obtained by the $\leftrightarrow$ operator. For examples, $\leftrightarrow(2; 3; 5; 7) = 4$, and $\leftrightarrow(x; y) = y-x$.

A string is indexed by the \ operator. Indexing is from 0. For example, $(2; 3; 5; 7)\langle 2 \rangle = 5$. A string can be indexed by a string. For example, $(3; 5; 7; 9)\langle 2; 1; 2 \rangle = 7;5;7$.

If $S$ is a string and $n$ is an index of $S$ and $i$ is any item, then $S\langle n\rangle_i$ is a string like $S$ except that item $n$ is $i$. For example, $(3; 5; 9)\langle 2 \rangle 8 = 3; 5; 8$. This operator associates from left to right, so $(3; 5; 9)\langle 2 \rangle 8\langle 1 \rangle 7 = ((3; 5; 9)\langle 2 \rangle 8)\langle 1 \rangle 7 = (3; 5; 8)\langle 1 \rangle 7 = 3; 7; 8$. And $(3; 5; 9)\langle 2 \rangle 8\langle 2 \rangle 7 = ((3; 5; 9)\langle 2 \rangle 8)\langle 2 \rangle 7 = (3; 5; 8)\langle 2 \rangle 7 = 3; 5; 7$.

A text is a more convenient notation for a string of characters. 
“abc” = “a”; “b”; “c”
“He said “Hi”.” = “H”; “e”; “ ”; “s”; “a”; “i”; “d”; “ ”; “”; “H”; “i”; “”; “”
“abcdefg”$(3;..6) = “def”

Strings are equal if and only if they have the same length, and corresponding items are equal. They are ordered lexicographically. For examples, 
3; 5 < 3; 5; 2 < 3; 6

A bunch of items is an item. Since string join precedes bunch union on the precedence table, we have 
$(3, 4); (5, 6) = 3; 5, 3; 6, 4; 5, 4; 6$ 
A string is an element (elementary bunch) if and only if all its items are elements.

If $S$ is a string and $n$ is a natural number, then 
$n*S$ $n$ copies of $S$ or $n$ S’s
* $S$ strings of $S$ or any number of S’s
is a string, and
is a bunch of strings. For examples, 
$3*5 = 5;5;5$
$3*(4, 5) = 4;4;4, 4;4;5, 4;5;4, 4;5;5, 5;4;4, 5;4;5, 5;5;4, 5;5;5$
*$5 = nil, 5, 5;5, 5;5;5, 5;5;5;5, and so on

The * operator distributes over bunch union, but in its left operand only. 
null*5 = null
$(2,3)*5 = 2*5, 3*5 = 5;5, 5;5;5$

Using this semi-distributivity, we have
*a = nat*a

Lists

A list is a packaged string. It can be written as a string enclosed in square brackets. For example,
[0; 1; 2]

The list operators are domain, content, indexing, pointer indexing, join, composition, selective union, and comparisons. Let $L$ and $M$ be lists, let $n$ be a natural number, and let $p$ be a string of natural numbers. The list operators are:
$\Box L$ domain of $L$
$\sim L$ content of $L$
#$L$ length of $L$
\[ L \, n \] \quad L \text{ at } n \, , \, L \text{ at index } n
\]
\[ L \, @ \, p \] \quad L \text{ at } p \, , \, L \text{ at pointer } p
\]
\[ L \, ; \, M \] \quad L \text{ join } M
\]
\[ L \, | \, M \] \quad L \text{ otherwise } M \, , \, \text{the selective union of } L \text{ and } M
\]
\[ i \rightarrow x \, | \, L \] \quad \text{index } i \text{ is item } x \text{ and otherwise } L
\]

plus the comparisons \[ L = M \, , \, L \neq M \, , \, L > M \, , \, L \leq M \, , \, L \geq M \, . \]

Here are some examples.

- \[ \square[10; \, 11; \, 12] = 0, 1, 2 \] \quad \text{the domain of a list}
- \[ \sim[10; \, 11; \, 12] = 10; 11; 12 \] \quad \text{the content of a list}
- \[ #[10; \, 11; \, 12] = 3 \] \quad \text{the length, or number of items, in a list}
- \[ [10;..20]_{5} = 15 \] \quad \text{indexing starts at zero}
- \[ [\, 2; \, 3; \, 4; \, 5; \, 6; \, 7\, ] \, @ \, (2; \, 1; \, 0) = 6 \]
- \[ [0;..10];[10;..20] = [0;..20] \] \quad \text{joining lists}
- \[ [10;..20]_{3; \, 6; \, 5} = [13; \, 16; \, 15] \] \quad \text{composition } (LM)n = L(Mn)

By using the \, @ \, operator, a string acts as a pointer to select an item from within an irregular structure. If the list \[ L \, | \, M \] \text{ is indexed with } n \, , \, \text{the result is either } Ln \text{ or } Mn \text{ depending on whether } n \text{ is in the domain } (0..\#L) \text{ of } L \, . \text{ If it is, the result is } Ln \, , \text{otherwise the result is } Mn \, .

\[ [10; \, 11] \, | \, [0;..10] = [10; \, 11; \, 2;..10] \]
\[ 1\rightarrow 21 \, | \, [10; \, 11; \, 12] = [10; \, 21; \, 12] \]

The index can be a string, as in

\[ (0;1) \rightarrow 6 \, | \, [[0; \, 1; \, 2]] \]
\[ [3; \, 4; \, 5] = [[0; \, 6; \, 2];
\[ [3; \, 4; \, 5]] \]

When a string or list is indexed by a structure, the result has the same structure as the index. For examples

\[ (10;..20) \setminus [2; \, (3, \, 4); \, [5; \, 6; \, 7]] = [12; \, (13, \, 14); \, [15; \, 16; \, 17]] \]
\[ [10;..20] \setminus [2; \, (3, \, 4); \, [5; \, 6; \, 7]] = [12; \, (13, \, 14); \, [15; \, 16; \, 17]] \]

Let \[ S = 10; \, 11; \, 12 \] . Then
\[ S\{0, \, \{1, \, [2; \, 1]; \, 0\}\} \]
\[ S0, \, \{S1, \, \{S2, \, S1\}; \, S0\} \]
\[ = 10, \, \{11, \, [12; \, 11]; \, 10\} \]

Let \[ L = [10; \, 11; \, 12] \] . Then
\[ L\{0, \, \{1, \, [2; \, 1]; \, 0\}\} \]
\[ = L0, \, \{L1, \, \{L2, \, L1\}; \, L0\} \]
\[ = 10, \, \{11, \, [12; \, 11]; \, 10\} \]

An operator that does not apply to a list can be composed with a list. For example,
\[ - \, [3; \, 5; \, 2] = [-3; \, -5; \, -2] \]

Lists are equal if and only if they are the same length and corresponding items are equal. They are ordered lexicographically.
\[ [3; \, 5] < [3; \, 5; \, 2] < [3; \, 6] \]

The list brackets \, [ \, ] \, distribute over bunch union. For example,
\[ [0, \, 1] = [0], \, [1] \]

Thus \[ [10*\text{nat}] \] \, is all lists of length 10 whose items are natural, and \[ [4*[6*\text{real}]] \] \, is all 4 by 6 arrays of reals.
Conditional Data

The 3-operand expression \( x \vdash y = z \), pronounced “if \( x \) then \( y \) else \( z \)”, has binary operand \( x \), but \( y \) and \( z \) are of arbitrary type. For example,
\[
y \neq 0 \vdash x/y = \text{“nan”}
\]
If \( y \neq 0 \) has value \( \top \), then this data expression has number value \( x/y \). If \( y \neq 0 \) has value \( \bot \), then this data expression has text value “nan”. This operator associates from right to left so that it can be evaluated from left to right. For example,
\[
(a \vdash b = c \vdash d = e) = (a \vdash b = (c \vdash d = e))
\]
If \( a \) has value \( \top \), then this expression has value \( b \), with no need to evaluate further.

Functions

A function defines a parameter; that is its only job. Let \( p \) (parameter) be any simple name, let \( D \) (domain) be any expression (but not using \( p \) ), and let \( B \) (body) be any expression (possibly using \( p \) as a constant name for an element of \( D \) ). Then \( \langle p: D \rightarrow B \rangle \) is a function with parameter \( p \), domain \( D \), and body \( B \). For example,
\[
\langle n: \text{nat} \rightarrow n+1 \rangle \quad \text{map } n \text{ in } \text{nat} \text{ to } n+1
\]
is the successor function on the natural numbers. The parameter name begins its scope at \( \langle \) and ends its scope at \( \rangle \) (see Scope). The \( \square \) operator gives the domain of a function. For example,
\[
\square \langle n: \text{nat} \rightarrow n+1 \rangle = \text{nat}
\]
The \# operator gives the size of the function, which is the size of its domain. For example,
\[
\# \langle n: (0..10) \rightarrow n+1 \rangle = 10
\]
A function of \( n+1 \) parameters is a function of 1 parameter whose body is a function of \( n \) parameters. For example, the maximum function
\[
\langle a: \text{real} \rightarrow \langle b: \text{real} \rightarrow a \geq b \Rightarrow a = b \rangle \rangle
\]
has two parameters. The notation for applying a function to an argument is the same as that for indexing a list: adjacency. If \( f \) is a function of two parameters, then \( f x y \) applies \( f \) to \( x \) and \( y \). Caution: in some languages, applying \( f \) to \( x \) and \( y \) is \( f(x,y) \). In ProTem, comma is bunch union, and function application distributes over bunch union. So in ProTem, \( f(x,y) = f x, f y \).

The predefined function \( \text{form} \) has four parameters. The first three parameters say how to format a number, and the last is the number to be formatted. For example, \( \text{form} 4 \ 1 \ 10 \ pi = \text{“3.1416^^0”} \).
We can define a new function
\[
\text{new} \ myform = \text{form} 4 \ 1 \ 10
\]
by supplying just three parameters, and then apply it to a number to be formatted:
\[
myform \ pi = \text{“3.1416^^0”}
\]
When the body of a function does not use its parameter, there is a syntax that omits the angle brackets \( \langle \rangle \) and unused name. For example,
\[
2 \rightarrow 3
\]
means \( \langle n: 2 \rightarrow 3 \rangle \) or choose any other parameter name.

Allowing the body of a function to be a bunch generalizes the function to a relation. For example, \( \text{nat} \rightarrow \text{bin} \) can be viewed in either of the following two ways: it is a function (with unused and therefore omitted parameter) that maps each natural to \( \text{bin} \); it is all functions with domain at least \( \text{nat} \) and range at most \( \text{bin} \). As an example of the latter view, we have
\[
\langle i: \text{int} \rightarrow \text{mod} \ i \ 2 \ 0 \rangle : \text{nat} \rightarrow \text{bin}
\]
The function \( \langle f: (\text{int} \rightarrow \text{int}) \rightarrow f \ 2 \rangle \) is “higher order”, which means it has a function-valued parameter.
Let \( f \) and \( g \) be functions such that \( f \) is not in the domain of \( g \) (\( \neg f \in \text{dom } g \)). Then \( g \circ f \) is the composition of \( g \) and \( f \).

\[
(g \circ f)(x) = g(f(x))
\]

If \( f \) and \( g \) are functions, then

\[
\begin{align*}
\n & \text{f otherwise } g \\
\n & \text{“the selective union of } f \text{ and } g \\
\n\end{align*}
\]

is a function that behaves like \( f \) when applied to an argument in the domain of \( f \), and otherwise behaves like \( g \).

\[
\begin{align*}
\square(f \lor g) &= \square f \lor \square g \\
(f \lor g)(x) &= (x: \square f = f x \lor g x)
\end{align*}
\]

Composition and selective union can have any mixture of operators, lists, and functions as operands. For examples,

\[
\begin{align*}
& -[3; 5; 2] = [-3; -5; -2] \\
& \text{even } [3; 5; 2] = [\bot; \bot; \top] \\
& (\text{even}) x = -(\text{even } x)
\end{align*}
\]

Argumentation comes before bunch union in precedence, and so it distributes over bunch union.

\[
(f, g)(x, y) = f x, f y, g x, g y
\]

If you want to apply a function to a bunch without distributing over the elements of the bunch, you must package the bunch as a set.

result-Expressions

A result-expression allows us to use a program to compute data. It has the form

\[
\text{result } \text{simplename} : \text{data} := \text{data} \llbracket \text{program} \rrbracket
\]

A local variable is defined with a type and initial value. Then the program is executed. The result is the final value of the newly defined local variable. We have not yet presented programs, but the following example, which approximates the base of the natural logarithms \( e \), should give the idea.

\[
\begin{align*}
\text{result } \text{sum: rat} := 1 \\
& \llbracket \text{new } \text{term: rat} := 1. \\
& \quad \text{for } i := 1;..15 \llbracket \text{term} := \text{term} / i. \text{sum} := \text{sum} + \text{term} \rrbracket \rrbracket
\end{align*}
\]

There are no side effects. Nonlocal variables become constants within the program; their values may be used, but assigning them is not permitted. Input from and output to nonlocal channels are not permitted.

All the ways of expressing data can be combined arbitrarily, without restriction. Here is a function whose body is a result-expression. It expresses the number of times \( 2 \) is a factor of \( n \).

\[
\begin{align*}
& (n: \text{nat}+1) \rightarrow \text{result } \text{f: 0..n} := 0 \\
& \quad \llbracket \text{new } \text{m: 1..n+1} := n. \\
& \quad \quad \text{loop } \llbracket \text{if even } m \llbracket \text{f} := f+1. \text{m} := m/2. \text{loop} \rrbracket \rrbracket \rrbracket
\end{align*}
\]

A result-variable begins its scope after \( \llbracket \) and ends its scope at the corresponding \( \rrbracket \) (see Scope). Consequently, the result-variable can be any simple name, even one that has already been defined in the scope that encloses the result-expression. The type and initial value of the result-variable cannot use the result-variable.

Type Transfer

There are seven predefined type transfer functions: \textit{bintext}, \textit{textbin}, \textit{numtext}, \textit{textnum}, \textit{timetext}, \textit{texttime}, and \textit{textid}. Each converts between text and another type. For examples,

\[
\begin{align*}
\text{numtext } 123 &= \text{“}123\text{”}, \text{textnum } \text{“}123\text{”} = 123, \text{numtext } (2 \times 3) &= \text{“}6\text{”}, \text{and textnum } \text{“}2 \times 3\text{”} = 6.
\end{align*}
\]
Type transfer functions are applied automatically whenever a binary, number, or time is used in a context that requires a text, or a text is used in a context that requires a binary, number, time, or identity. There is no function to convert an identity to text. For example, “123” + 1 = textnum “123” + 1 = 123+1 = 124

Output to the screen, denoted ‘!', requires a text. So ! 123 places a number where a text should be. So numtext is applied automatically ( ! numtext 123 ) resulting in a text ( ! “123”) as required for output. If numtext has been redefined (see Scope), the predefined numtext is used, and similarly for the other type transfer functions. For fine control over the format of the resulting text, use the predefined function form.

The function charnat encodes a character as a number (possibly extended ASCII or unicode), and natchar decodes a number as a character. These functions are never applied automatically.

Scope

A simple name is defined in these six ways: by the keyword new, as a named program, as a function parameter, as a plan parameter, as a for-index, or as a result-variable. We shall come to each of these shortly. The scope of a simple name is the part of a program in which the name is defined. We shall also come to the ways of composing larger programs from smaller programs using program brackets ⟦⟧. Scopes are limited by ⟦⟧, and by ⟨ ⟩. The opening bracket ⟦ or ⟨ opens a scope, and the corresponding closing bracket ⟧ or ⟩ closes the scope.

A simple name defined using the keyword new must be new, not already defined since the most recent scope opener ⟦. Its scope extends from its definition through all following sequentially composed programs to the scope closer ⟧ corresponding to the most recent scope opener ⟦. But it may be covered by a redefinition in an inner scope. Using new x=2 and new x=3 as example definitions, and letting A, B, C, D, and E stand for arbitrary program forms (but not new or old), in

⟦A. new x=2. B. [ C. new x=3. D]. E⟧

the definition of x as the number 2 is not yet in effect in A, but it is in effect in B, C, and E. The definition that makes x the number 3 is in effect in D. None of A, B, C, D, or E can contain a redefinition of x unless it is within further scope limiters ⟦⟧ or ⟨ ⟩.

A name defined by new can become undefined by the keyword old, ending its scope early. So in

new x=2. A. old x. B

the definition of x is in effect in A but not in B. Within B, the name x has the same meaning (if any) that it had before the definition new x=2. After old x, the name x is again new and available for definition. However,

new x=2. [old x. A]

is not allowed; a scope cannot be ended by old within a subscope.

A scope can be nested inside another scope, which can be nested inside another, and so on. Outside all scope limiters is the persistent scope. A name defined by new in the persistent scope is called a persistent definition. Its scope ends only with old. Its scope does not end with the end of a computing session, not even by switching off the power. Persistent variables serve as “files”. Part of the persistent scope is private to each programmer (possibly but not necessarily stored locally), and part may be shared among a group of programmers (not stored locally) (see Permit).
Outside the persistent scope is the predefined scope where the predefined names are defined. They are usable in all your scopes unless you cover them by redefining the names (and even then; see Predefined Names). Adding a new predefined name or ending the scope of a predefined name requires the right identity (see Session).

**Programs**

Some program constructs are concerned with names: creating a name (new), deleting a name (old). Other program constructs are variable assignment, input, output, and a variety of ways of combining programs to form larger programs. All programs, including those that create and delete names, are executed in their turn, just like variable assignments and input and output.

**Variable Definition**

Variable definition has the form

```
new newname : data := data
```

The newname becomes a variablename. Here is an example variable definition.

```
new x: nat := 5
```

This defines $x$ to be a variable assignable to any element in $\text{nat}$, and initially assigned to 5. There is no such thing as an “uninitialized variable” nor the “undefined value” in ProTem. In a variable definition, the data after : is called the “type” of the variable, and the data after := is called the “initial value”. The type can be anything except the empty bunch, and the initial value must be an element of the type. The type and initial value can depend on previously defined names, including variables. For example,

```
new y: 0..2x:= x
```

defines $y$ as a variable whose value can be any natural number from (including) 0 up to (excluding) twice the current value of $x$ (the value of $x$ at the time this definition is executed), with initial value equal to the current value of $x$. But the type and initial value cannot make use of the name being defined.

Here are three more examples.

```
new s: [10*int] := [10*0]
new t: text := ""
new u: (0..20)*char := "abc"
```

In the first example, $s$ is defined as a variable that can be assigned to any list of ten integers, and is initially assigned to the list of ten zeroes. In the middle example, $text$ is a predefined bunch equal to $\text{*char}$, so $t$ can be assigned to any text, and is initially assigned to the empty text. In the last example, $u$ is defined as a variable that can be assigned to any text of length less than 20, and is initially assigned to the text “abc”.

**Assignment**

A variable can be reassigned by the assignment program. It has the form

```
variable := data
```

Here are two examples using the definitions of the previous subsection.

```
x := x+1
s := 3 → 5 | s
```

The data on the right of := must be an element in the type of the variable on the left of :=. As in the examples, the data on the right of := can make use of the variable on the left of :=.
Constant Definition

Constant definition has the form

\[ \textbf{new} \ \text{newname} := \text{data} \]

The newname becomes a constantname. The data on the right of := cannot make use of the name on the left of :=. The newly created constantname cannot be reassigned. Here are three constant definitions.

\begin{align*}
\text{new} \ size & := 10 \\
\text{new} \ piBy2 & := \pi / 2 \\
\text{new} \ range & := 0..size
\end{align*}

where \( \pi \) is a predefined constant name.

A constant may use variables to express its value. For example

\[ \text{new} \ xplus1 := x + 1 \]

The current value of variable \( x \) is used to evaluate \( x + 1 \), and \( xplus1 \) expresses that value. Variable \( x \) may later be reassigned to another value, but that does not affect the value of \( xplus1 \). For example,

\begin{align*}
\text{new} \ x & := 3. \quad \text{`x has value 3} \\
\text{new} \ xplus1 & := x + 1. \quad \text{`x has value 3 and xplus1 has value 4} \\
x & := 5 \quad \text{`x has value 5 and xplus1 has value 4}
\end{align*}

Data Definition

Data definition has the form

\[ \textbf{new} \ \text{newname} = \text{data} \]

The newname becomes a dataname (note = rather than := as in constant definitions).

\[ \text{new} \ xplus2 = x + 2 \]

makes the value of \( xplus2 \) depend on the value of variable \( x \). As \( x \) changes value, \( xplus2 \) changes value so that \( xplus2 = x + 2 \) is always \( \top \). In the constant definition of \( xplus1 \) earlier, \( x + 1 \) is evaluated once, at definition time. By contrast, in the data definition of \( xplus2 \), \( x + 2 \) is not evaluated at definition time; it is evaluated every time \( xplus2 \) is used in a context that requires its value. For example,

\begin{align*}
\text{new} \ x & := 3. \\
\text{new} \ xplus2 & := x + 2. \\
! xplus2 & \quad \text{`prints 5} \\
x & := 6. \\
! xplus2 & \quad \text{`prints 8}
\end{align*}

A data definition can depend indirectly on a variable. For example,

\[ \text{new} \ twoxplus4 = 2 \times xplus2 \]

makes \( twoxplus4 \) depend indirectly on the value of variable \( x \). In this context, the value of \( xplus2 \) is not required, so it is not evaluated. If \( x \) currently has value 3, then \( twoxplus4 \) prints 10. If \( x \) currently has value 6, then \( twoxplus4 \) prints 16.

Data Recursion

In a variable definition, the type and initial value cannot depend on the variable being defined. For example,

\[ \text{new} \ bad; 0..2 \times bad := \text{bad} \quad \text{`illegal} \]

is not allowed due to the two occurrences of \( \text{bad} \) to the right of the colon. Likewise a constant
definition cannot be recursive.

Data definition does allow recursion. The next two examples define \texttt{fact} and \texttt{div} to be the factorial function and integer division function for natural numbers.

\begin{verbatim}
new fact = 0 \rightarrow 1 \langle n : (\text{nat}+1) \rightarrow n \times \text{fact} (n-1) \rangle

new div = \langle a : \text{nat} \rightarrow \langle d : (\text{nat}+1) \rightarrow a < d \models 0 \models a \models 2 \times \text{div} (a/2) \ d = 1 + \text{div} (a-d \ d) \rangle
\end{verbatim}

Here is a bunch of texts (a grammar). This bunch includes the text “a+b+a–a”, and many more.

\begin{verbatim}
new term = “a”, “b”, term; “+”; term , term; “–”; term
\end{verbatim}

This recursive definition is equivalent to the nonrecursive definition

\begin{verbatim}
new term = (\langle a, “b” \rangle, *\langle (“+”, “–”) \rangle, \langle (“a”, “b”) \rangle)
\end{verbatim}

Here is a function that eats arguments until it is fed argument 0 .

\begin{verbatim}
new eat = \langle n : \text{nat} \rightarrow n=0 \models 0 \models \text{eat} \rangle
\end{verbatim}

So \texttt{eat 5 2 0} = 0 , and \texttt{eat 4 7 3 8 0} = 0 , and \texttt{eat 1} 2 = \texttt{eat} .

The next example defines all binary trees with integer nodes.

\begin{verbatim}
new tree = [\text{nil}], [\text{tree}; \text{int}; \text{tree}]
\end{verbatim}

The final example is a pure, baseless recursion.

\begin{verbatim}
new rec = rec
\end{verbatim}

Whenever \texttt{rec} is used, its evaluation is nonterminating.

\textbf{Constant v Data Definition}

A constant definition evaluates its data once, at definition time, whereas a data definition evaluates its data each time its value is required. If the data is fully evaluated, there is no difference. For example, there is no difference between these two definitions:

\begin{verbatim}
new five:= 5
new five = 5
\end{verbatim}

When there are no variables used to express the value (neither directly nor indirectly), there is no semantic difference between data definition and constant definition, but there may be an efficiency difference. Compare these two definitions.

\begin{verbatim}
new six:= 5+1
new six = 5+1
\end{verbatim}

If the value of \texttt{six} is never required, the data definition ( = ) is more efficient. If the value of \texttt{six} is required once, they are equally efficient. If the value of \texttt{six} is required two or more times, the constant definition ( := ) is more efficient. Here is a more interesting comparison.

\begin{verbatim}
new double:= \langle n : (0,..10) \rightarrow 2\times n \rangle
new double = \langle n : (0,..10) \rightarrow 2\times n \rangle
\end{verbatim}

The constant definition causes the function to be evaluated by applying it to all its arguments and storing the results. In effect, the function is evaluated to the list

\[ [0; 2; 4; 6; 8; 10; 12; 14; 16; 18] \]

Then, when the value of \texttt{double} applied to an argument is required, that argument indexes the list. The data definition does not evaluate the function. Each time the value of \texttt{double} applied to an argument is required, the body of the function is evaluated. Which one is more efficient depends on the size of the domain, the complexity of the result, and the number of times the definition is used.
Program Definition

Program definition has the form

\textbf{new} \ \textit{newname} \ [ \ \textit{program} \ ]

The \textit{newname} becomes a programname. Program definition gives a program a name, but does not execute the program. For example,

\texttt{new switchends \ [s:= 0 \rightarrow s \ 9 \ 19 \rightarrow s \ 0 \ 1 \ s]}

Execution of this definition creates the program name \textit{switchends}, but does not execute program \texttt{[s:= 0 \rightarrow s \ 9 \ 19 \rightarrow s \ 0 \ 1 \ s]}. After execution of this definition, the name \textit{switchends} can be used to call (cause execution of) the program it names. Program definitions can be recursive. Predefined program names include \textit{await}, \textit{exec}, \textit{ok}, \textit{stop}, \textit{wait}.

A fancy name can be used as a specification. For example,

\texttt{new \ « \ x' > x \ » \ [x:= x+1]}

The specification \texttt{« \ x' > x \ »} is implemented (refined, implied) by the program \texttt{[x:= x+1]}. A prover is invoked by the \texttt{esc v} command (see Verify). If the specification is written within the language that the prover understands, the prover attempts to prove that the specification is implemented (refined, implied) by the program. If the program makes use of a specification, the inner specification is used in the outer proof. For example,

\texttt{new \ « \ x' = 0 \ » \ [if \ x\neq 0 \ [x:= x-1. \ « \ x' = 0 \ »]]}

In the program, the specification \texttt{« \ x' = 0 \ »} means exactly what it says, rather than the program that it names. Thus the use of specifications makes complicated fixed-point semantics unnecessary. If the prover fails to understand the specification, or fails to prove the refinement, it informs the programmer, and treats the specification as just a name. (See the paper Specified Blocks.)

Measuring Unit Definition

There are three predefined units of measurement. They are \textit{g}, representing mass in grams, \textit{m}, representing distance in meters, and \textit{s}, representing time in seconds. A unit of measurement has all the properties of an unknown positive finite real number constant. So, for example, we write \texttt{10\times m/s} for the speed 10 meters per second. And we can define

\texttt{new \ km:= 1000\times m}

to make \textit{km} be a kilometer, and

\texttt{new \ h:= 3600\times s}

to make \textit{h} be an hour. So \texttt{1\times m/s = 3.6\times km/h} evaluates to \texttt{T}. To assign a variable to a quantity with units attached, the variable's type must have compatible units attached. For example,

\texttt{new \ speed: \ real\times m/s:= 3.6\times km/h}

assigns \textit{speed} to \texttt{1\times m/s}.

You can create a new unit of measurement, unrelated to the existing units. Measuring unit definition has the form

\texttt{new \ \textit{newname} \ #1}

The \textit{newname} becomes a unitname. For example,

\texttt{new \ sheet \ #1}

Creates a new unit of measurement called the \textit{sheet}. Now you can define the related units

\texttt{new \ quire:= 25\times sheet}
\texttt{new \ ream:= 20\times quire}

Now you can define a variable using the new units.

\texttt{new \ order: \ nat\times sheet:= 3\times ream}

This assigns \textit{order} to \texttt{1500\times sheet}.
When the value $5 \times m/s$ is converted to text by `numtext`, the result is “5 m/s” without the $\times$ sign and without evaluating the unknown real value $m/s$. And `textnum” 5 m/s” = $5 \times m/s$. Similarly for all units of measurement. One more example: `numtext (2x3xkm/h) = “1.6667 m/s”.

**Forward Definition**

Forward definition has the form

```
new newname
```

The newname becomes either a dataname or a programname. For example

```
new abc
```

is a notice that a definition will follow later in the same scope. In a data definition or program definition, the scope of the name being defined starts immediately. A forward definition allows mutual recursion by starting the scope of a data name or program name even before its definition. For example, using ··· to stand for uninteresting things, in

```
new f:= 3. [new f. new g = ···f···g···. new f = ···f···g···. ···]
```

the inner $f$ and $g$ are each defined in terms of both of them. Without the forward definition of $f$ (following []), $g$ would be defined in terms of the earlier constant definition `new f:= 3`.

**Name Removal**

Names defined with the keyword `new` can be undefined with the keyword `old`. Name removal has the form

```
old oldname
```

Ironically, by saying `old x`, the name $x$ becomes available for reuse as a new name. Even though a name becomes undefined, what it named will remain as long as there is an indirect way to refer to it. For example,

```
new s: *all:= nil.
new push [[plan x: all [s:= s;x]].
new pop [s:= s(0;..<>s–1)].
new top = s(<>s–1).
new empty = s=nil.
old s
```

The names `push`, `pop`, `top`, and `empty` are now defined and ready for use. The name $s$ was defined for the purpose of defining the other names, and then removed, leaving the other names dependent upon an anonymous variable.

The predefined names include `randNat`, `randNatInit`, and `randNatNext`. They might have been defined as:

```
new big:= 2^31.
new rv: 0..big:= 123456789.
new randNat = (from: nat -> (to: nat -> floor (from + (to–from)xrv/big)))).
new randNatInit [[plan seed: 0..big [rv:= seed]].
new randNatNext [rv:= mod (rv x 5^13) big].
old big. old rv
```

Constant `big` and variable `rv` are now hidden; their names are removed, but `randNat`, `randNatInit`, and `randNatNext` still use them. We can use these definitions as follows:

```
randNatInit 5555555555.
randNatNext.
! randNat 0 10
```
The following sequence swaps the data names \( i \) and \( j \).
\[
\text{new } t = i. \text{ old } i. \text{ new } i = j. \text{ old } j. \text{ new } j = t. \text{ old } t
\]

**Output and Input**

Each channel is defined to transmit a specific type of value. The output channels *screen* and *printer*, and the input channel *keys* are predefined to transmit text. The input channel *microphone* and the output channel *speaker* are predefined to transmit sound. Input channel *time* transmits times. We can define local channels to transmit any type of value (see Channel Definition).

Output has the form
\[
\text{channelname }! \text{ data}
\]
Channel *screen* accepts text, which is displayed on the screen. The program
\[
\text{screen! “Hi there.”}
\]
sends the text “Hi there.” to the screen. Output is buffered so it will be available when *screen* is ready to receive it. Texts can be joined and sent together.

\[
\text{screen! “Answer = ”; numtext x; nl}
\]
where *numtext* is a predefined function that converts from a number to a text, and *nl* is the newline character, or next line character, or return character. Function *numtext* can be omitted (see Type Transfer). When *screen* or *printer* receives the delete (backspace) character, the last previously sent character is deleted; when the *tab* character is received, some number of spaces are substituted; when the *nl* character is received, further characters start on a new line.

The keyboard is a program that runs concurrently with other programs; you don't need to initiate it; it is already running. It monitors what key combinations are pressed, and for what duration, and creates a string of characters. The shift-A combination is a single character “A”. Likewise the control-Q combination is a single character. The click button is just a key like any other; *click* is a character, and *doubleclick* is a character.

Input has two forms: without echo, and with echo. The first form, without echo, is
\[
\text{channelname }? \text{ data}
\]
Text from the keyboard (including the click button) can be received from channel *keys*. Five characters of input are received from channel *keys* by saying
\[
\text{keys? 5*char}
\]
What follows ? is called the pattern (or grammar). If input is not yet available, it is awaited. The input read is the earliest input on that channel that has not yet been read. The *tab*, *delete*, and *nl* characters may be part of the input; no corrections are made. The input is not echoed on the screen. The shortest input that fits the pattern is read. The program
\[
\text{keys? text; nl}
\]
reads text up to and including the first *nl* character, but
\[
\text{keys? text}
\]
just inputs the empty text.

Patterns *digit* and *numpat* are predefined as follows.
\[
\text{new digit := “0”, “1”, “2”, “3”, “4”, “5”, “6”, “7”, “8”, “9”}
\]
\[
\text{new numpat := (“+”, “–”, “”); digit; *digit; (“.”; digit; *digit;), “”}
\]
To receive a text that can be interpreted as a number, possibly preceded or followed by spaces, possibly preceded by a sign, ending in a new line character, input
\[
\text{keys? “” ; numpat; “” ; nl}
\]
Without *nl*, leading spaces and an optional sign and the first digit are read.
When input is received, it is referred to by the channel name followed by `??`. After the previous example input, we might have the assignment

\[ x := \text{textnum}(\text{keys}??) \]

where `textnum` is a predefined function that converts from a text to a number. Function `textnum` can be omitted (see Type Transfer).

If `c` is the name of an input channel, then `c!!` is a binary expression with value `⊤` if there is written but unread data on the channel, and `⊥` if there is not. For example,

```
if keys!![keys? char. screen! keys??] else [screen! “Are you still there?”]
```

Input on a channel that does not currently have any written-but-not-yet-read data waits until data is written to the channel by a concurrent program.

If channel `c` is defined to input text, the program

```
c? “y”, “n”
```

inputs one character, either “y” or “n”, from channel `c`. If the first available character on channel `c` is “a”, or more generally, if the input on the channel does not fit the pattern, what happens is undefined. Here are three options.

- The program cannot be executed, so execution ends.
- An error message is sent to channel `screen` to say that the input is unacceptable, and execution ends.
- An error message is sent to channel `screen` to say that the input is unacceptable, and the sender is given another opportunity to send an input that fits the pattern.

What happens depends on the implementation and on the channel. Perhaps the last option is appropriate for channel `keys`, and the first is appropriate for a secure channel.

An input program consisting of an input channel name, a question mark, and a pattern, does not echo the input on the screen; the input is invisible. This is useful, for example, when reading a password (see Read Password). To input and echo together, character by character, add an exclamation mark and the output channel name for the echo.

```
channelname ? data ! channelname
```

For example,

```
keys? text; nl ! screen
```

This program inputs, from channel `keys`, text to and including the first `nl` character, and outputs the same on channel `screen`. Each character is echoed as it is input. A space \(\rightarrow\), tab \(\rightarrow\), new line \(\rightarrow\), or delete character \(\rightarrow\) is displayed visibly as a graphic symbol.

In an input with echo, the pattern between the input question mark and the echo exclamation mark can be omitted. This results in a special pattern called the correcting pattern. For example,

```
keys?!.screen
```

The correcting pattern reads a line of text to and including the first `nl` character, but this text is corrected according to `delete` (backspace) characters. The `nl` character is consumed, but not included in the value read. After this input, the value of `keys??` is the corrected text, not including the final `nl`. If, during correction, there are more `delete` characters than other characters, the extra `delete` characters are ignored. The echo displays space, tab, and new line characters as spaces, tabs, and new lines, and displays delete characters as deletions of previous characters.

An output channel name can be omitted, in which case the output channel is `screen`. For example,

```
! “Hello World”
```

prints `Hello World` on the screen. An input channel name can be omitted, in which case the input channel is `keys`. For example,
reads one character from \texttt{keys}, with no echo. The most recent text read on channel \texttt{keys} can be referred to as just \texttt{??}. And \texttt{!!} is a binary expression saying whether there is written but unread data on channel \texttt{keys}. If the input channel is omitted, and the name \texttt{keys} has been redefined, the input channel is the predefined channel \texttt{keys}. If the output channel is omitted, and the name \texttt{screen} has been redefined, the output channel is the predefined channel \texttt{screen}.

In a context requiring a number, the expression \texttt{textnum (keys??)} can be written \texttt{textnum (??)} or \texttt{keys??} or \texttt{??}. But it cannot be written \texttt{textnum keys??} because that is parsed as (\texttt{textnum keys})?? . And it cannot be written \texttt{textnum ??} because the compiler will complain that \texttt{textnum} is not a channel. Similarly for \texttt{bintext (keys!!)}.

The most common form of input
\texttt{?!
}
reads one line from \texttt{keys}, correcting it according to \texttt{delete} characters, up to the first \texttt{nl}, which is not included in the value of \texttt{??}, and echoes character by character to \texttt{screen}.

In summary, output and input are, respectively,
\begin{verbatim}
  channelname ! data
  channelname ? pattern ! channelname
\end{verbatim}
After reading input, the input most recently read is referred to as \texttt{channelname}.

The binary expression
\begin{verbatim}
  channelname !
\end{verbatim}
says whether there is written but unread data on the channel. If the channelname is \texttt{keys} or \texttt{screen} it can be omitted. If input echoing is not wanted, omit both \texttt{!} and the echo channelname. In an input program with echo, omitting the pattern results in the correcting pattern.

**Sequential Composition**

Sequential composition is denoted by a period (point, dot). According to the grammar, it is an infix connective; in other words, the period comes between and joins two programs.
\begin{verbatim}
  program . program
\end{verbatim}
In the persistent scope, each program is executed as soon as it is keyed in. The end of the sequence of keystrokes comprising a program to be executed is recognized by the period that will join it to the sequentially next program, after execution of the just completed program. So, in the persistent scope, the period feels more like a program terminator than a program joiner.

ProTem can be used as a calculator. In the persistent scope, the program
\begin{verbatim}
  ! 2+2.
\end{verbatim}
with a following period and new line character, immediately prints \texttt{4}. In full, it is
\begin{verbatim}
  screen! numtext (2+2).
\end{verbatim}
The program
\begin{verbatim}
  ! 2+2
\end{verbatim}
followed by a new line character but without a period, is not executed until a period and another new line character are entered. The program and period and new line
\begin{verbatim}
  new temp:= 2+2.
\end{verbatim}
saves the result of the calculation under the name \texttt{temp}, perhaps for use in further calculation. The name \texttt{temp} persists from session to session until it is ended by \texttt{old temp}. The program and period and new line
new myfiles_.
immediately creates a dictionary within which programs and data can be stored and edited and used.
The program and following period and comment and new line character
new myfiles_prog $\![\text{"2+2="}.$
$\!\!2+2]\] "this is a comment
immediately saves a new program named prog in existing dictionary myfiles. The saved program
is not immediately executed. The first period comes between two output programs, joining them.
There is no period following the last of these two output programs. The new line following the first
period does not indicate the completion of the program definition. It does indicate that the part of
the program definition that came before the new line character is no longer correctable by delete
(backspace) characters. The period at the end indicates the completion of the program definition, but
the program definition remains correctable by delete characters until a new line character following
the comment is typed. Further corrections can be made using the editor command $\text{esc e}$(see Edit).

Concurrent Composition

Concurrent composition has the form

\[ \text{program} \parallel \text{program} \]

The concurrent composition of programs \( P \), \( Q \), and \( R \) is \( P \parallel Q \parallel R \). A variable defined before the
concurrent composition remains a variable in at most one of the programs in the concurrent
composition; in all the other programs of the concurrent composition, it becomes a constant. For example,

\[
\text{new } a : \text{nat} := 1 \parallel \text{new } b : \text{nat} := 2. \text{new } c = a+b. \! \!a := 4. \text{ A} \parallel \! \!b := 8. \text{ B}. \text{ C}
\]

In the second concurrent composition, variable \( a \) can be reassigned in one of the concurrent
programs, but not in both; it is reassigned in the left program. Likewise variable \( b \) can be
reassigned in one of the concurrent programs, but not in both; it is reassigned in the right program.
At the start of \( A \), variable \( a \) has value 4 , constant \( b \) has value 2 , and data \( c \) has value 6 . At the
start of \( B \), constant \( a \) has value 1 , variable \( b \) has value 8 , and data \( c \) has value 9 . If \( A \)
does not reassign \( a \), and \( B \) does not reassign \( b \), then at the start of \( C \), variable \( a \) has value 4 ,
variable \( b \) has value 8 , and data \( c \) has value 12 . Concurrent programs cannot affect each other
through assignments of variables. For co-operation, programs can communicate with each other on
channels defined for the purpose (see Channel Definition).

Channel Definition

Channel definition has the form

\[
\text{new } \text{newname} \ ? \ \text{data} ! \ \text{data}
\]
The newname becomes a channelname. The definition

\[
\text{new } c ? \text{nat}! 0
\]
defines \( c \) to be a new local channel that transmits values of type \text{nat} , with 0 as initial output and
input. It can be used for output and input. Now \( c?? \) refers to the most recent input on the channel, and
\( c!! \) is a binary expression saying whether there is written but unread data on channel \( c \). Before there has been any output or input, \( c?? \) refers to the initial output and input supplied in the
channel definition, and \( c!! \) is \( \perp \). The type of the channel cannot use the name of the channel being
defined. Only one of the programs that are concurrent with each other can use a channel for output.
More than one of the concurrent programs can use the same channel for input only if the concurrent
composition is not sequentially followed by a program that uses that channel for input. When
concurrent programs read from the same channel, they read the same inputs independently.
**if-Program**

An if-program has the form

```
if data [ program ]
```

The if-program

```
if b [P]
```

is executed as follows: binary expression \( b \) is evaluated; if its value is \( \top \), then program \([P]\) is executed; if its value is \( \bot \), then program \([P]\) is not executed. An if-else-program has the form

```
if data [ program ] else [ program ]
```

The if-else-program

```
if b [P] else [Q]
```

is executed as follows: binary expression \( b \) is evaluated; if its value is \( \top \), then program \([P]\) is executed and program \([Q]\) is not executed; if its value is \( \bot \), then program \([P]\) is not executed and program \([Q]\) is executed.

**case-Program**

A case-program has the form

```
case data [ program ]
```

For example, the case-program with 3 cases,

```
case n [P. Q. R]
```

is executed as follows: natural expression \( n \) is evaluated; then one of the sequence of programs within the \([ ]\) brackets is executed. (The periods between \( P \) and \( Q \), and between \( Q \) and \( R \), do not indicate sequential execution.) These programs, called cases, are numbered in order 0, 1, 2, and so on, and each is in a new scope. In the example, case 0 is \([P]\), case 1 is \([Q]\), and case 2 is \([R]\).

In the example, if \( n \) has value 1, then just \([Q]\) is executed. If \( n \) is equal to or greater than the number of cases, an error message is printed and execution stops. The example case-program is equivalent to

```
if n=0 [P] else [if n=1 [Q] else [if n=2 [R] else ![“error: case index too large”. stop]]]
```

It is allowed, but senseless, for any of the cases to be just a simple name definition. For example, if \( n \) is 1, and \( Q \) is new \( x=2 \), then \([\text{new } x=2]\) is executed, defining \( x \) and then immediately ending the scope of \( x \). Typically, the structure is something like

```
case ...[[ ... ]]. [[[ ... ]]. [[[ ... ]]]]
```

so that each case can include useful definitions and a sequence of programs.

A case-else-program has the form

```
case data [ program ] else [ program ]
```

For example, the case-else-program with 3 cases,

```
case n [P. Q. R] else [S]
```

is the same as the case-program, but if \( n \) is equal to or greater than the number of programs before else, then the program \([S]\) after else is executed. The example case-else-program is equivalent to

```
if n=0 [P] else [if n=1 [Q] else [if n=2 [R] else [S]]])
```
for-Program

A for-program has the form
  for simplename := data [ program ]

The simplename becomes a constantname. Here is a nest of for-programs, or for-loops, that computes the transitive closure of $A: [n*[n*bin]]$.

  for $j:= 0;..n$
  [ for $i:= 0;..n$
    [ for $k:= 0;..n$
      [ if $A i j \land A j k$
        \[
        A:= (i;k) \rightarrow \top \mid A
        \]
      ]
    ]
  ]

The if-program

  if $A i j \land A j k [A:= (i;k) \rightarrow \top \mid A]$ can be restated as
  
  $A:= (i;k) \rightarrow (A i k \lor (A i j \land A j k)) \mid A$

if you prefer. The name being defined by for is known only within the loop body, and it is known there as a constant, and so it is not assignable. It is called a for-index. In the example, each of the for-indexes takes values $0, 1, 2$, and so on up to but not including $n$.

For a second example, here is the sieve of Eratosthenes.

  new $n:= 1000$.
  new prime: $n*bin:= 2*\bot; (n–2)*\top$.
  for $i:= 2;..\text{ceil} (\sqrt n)$ [ if prime\forall \[ for $j:= i;..\text{ceil} (n/i)$ \[ prime:= prime \land (i+j= \bot) \]\]\]}

A for-index is “by initial value”, so

  for $i:= x; x \[ x:= i+1 \]$ increases $x$ by 1, not 2.

This next example prints the natural numbers forever.

  for $n:= 0;..\infty$ [ ! $n; \ "\"$]

After the := we can have any string expression; the index stands for each item in the string, in sequence. We can also have any bunch expression; the index stands for each element of the bunch, concurrently. As an example (note the use of .. rather than ;.. as earlier),

  for $i:= 0;..\#L$ [ $L:= i \rightarrow 0 \mid L$]

makes the items of $L$ be 0, concurrently. We could also write either of these:

  for $i:= \Box L$ [$L:= i \rightarrow 0 \mid L$]
  $L:= [\#L*0]$

The domain of the for-index can also be a bunch of strings, or a string of bunches, and so on, so that sequential and concurrent execution can be nested within each other. (Note: distribution and factoring laws are not applied; the structure of the expression is the structure of execution.)

A for-index begins its scope after [ and ends its scope at the corresponding ] . Consequently, the for-index can be any simple name, even one that has already been defined in the scope that encloses the for-loop. The domain of the for-index cannot use the for-index.

Named Program

A named program has the form
  
  simplename [ program ]

The simplename becomes a programname within the program that it names. It begins its scope after [ and ends its scope at the corresponding ] (see Scope). Consequently, the name can be any simple
The prover fails to understand the specification names. Thus the use of specification names in the program is attached to the program (like a program definition), and the program is executed (unlike a program definition). One purpose of this naming is to make loops. Here is a two-dimensional search for \( x \) in an \( n \times m \) array \( A \) of integers (that is, \( A: [n*[m*int]] \)).

\[
\text{new } i: \text{nat} := 0.
\text{tryThisI} \begin{cases} \text{if } i=n \mid ![x; " does not occur."] \\
\text{else } \text{[new } j: \text{nat} := 0. \\
\text{tryThisI} \begin{cases} \text{if } j=m \mid ![i:= i+1. \text{tryThisI}] \\
\text{else } \text{if } A \ i \ j = x \mid ![x; " occurs at "; i; " "; j] \\
\text{else } ![j:= j+1. \text{tryThisJ}]]]] \\
\end{cases}
\end{cases}
\]

The next example is a fast remainder program, assigning natural variable \( r \) to the remainder when natural \( a \) is divided by positive natural \( d \), using only addition and subtraction.

\[
r:= a.
\text{outerloop} \begin{cases} \text{if } r \geq d \mid [\text{new } dd: \text{nat} := d. \\
\text{innerloop} \begin{cases} \text{if } r < dd \mid \text{outerloop \ else [innerloop]} \\
\end{cases}
\end{cases}
\]

The use of a program name is semantically a call; it means the same as replacing it with the program it names (including the \([ ]\) brackets). The fast remainder example means the same as

\[
r:= a.
\text{outerloop} \begin{cases} \text{if } r \geq d \mid [\text{new } dd: \text{nat} := d. \\
\text{innerloop} \begin{cases} \text{if } r < dd \mid [\text{if } r \geq d \mid [\text{new } dd: \text{nat} := d. \\
\text{innerloop} \begin{cases} \text{if } r < dd \mid \text{outerloop} \\
\text{else [innerloop]} \\
\end{cases}
\end{cases}
\end{cases}
\end{cases}
\]

The uses \text{outerloop} and \text{innerloop} were replaced by the programs they name. They reappear, and again they mean the programs they name. Although semantically they are calls, in the previous two examples they are last actions (tail recursions), so they are implemented as branches (jumps, go to's).

The next example illustrates that named programs provide general recursion, not just tail recursion. It computes the Fibonacci numbers \( x:= \text{fib } n \) and \( y:= \text{fib } (n+1) \) in \( \log n \) time.

\[
\text{Fib} \begin{cases} \text{if } n=0 \mid ![x:= 0. \ y:= 1] \\
\text{else } \text{if odd } n \mid ![n:= (n-1)/2. \ \text{Fib. } n:= x. \ x:= x^2 + y^2. \ y:= 2*n*x*y + y^2]\ \\
\text{else } ![n:= n/2 - 1. \ \text{Fib. } n:= x. \ x:= 2*x*xy + y^2. \ y:= n^2 + y^2 + x]]
\end{cases}
\]

As in a program definition, a fancy name can be used as a specification. For example, \( \langle x' > x \rangle \mid [x:= x+1] \)

The specification \( \langle x' > x \rangle \) is implemented (refined, implied) by the program \([x:= x+1] \). A prover is invoked by the \text{esc v} command (see Verify). If the specification is written within the language that the prover understands, the prover attempts to prove that the specification is implemented (refined, implied) by the program. If the program makes use of a specification, the inner specification is used in the outer proof. For example,

\( \langle x' = 0 > x \rangle \mid [[x \neq 0 \mid ![x:= x-1. \ \langle x' = 0 > ]] \]

In the program, the specification \( \langle x' = 0 > \) means exactly what it says, rather than the program that it names. Thus the use of specifications makes complicated fixed-point semantics unnecessary. If the prover fails to understand the specification, or fails to prove the refinement, it informs the
programmer, and treats the specification as just a name.  (For more on proving, see the paper
Specified Blocks.)

Suppose a name is defined within a loop. For example, the name \( a \) in

\[
\text{infiniteLoop } \begin{array}{c}
\text{[new } a:= \text{“a”}. \text{!a. } \text{infiniteLoop}] \\
\end{array}
\]

Executing this loop prints an infinite sequence of the letter “a”. Replacing the call with the called
program, it is equivalent to

\[
\text{infiniteLoop } \begin{array}{c}
\text{[new } a:= \text{“a”}. \text{[new } a:= \text{“a”}. \text{!a. } \text{infiniteLoop}] \\
\end{array}
\]

In a general recursion, each call opens a new scope, and each new definition hides but does not
destroy the previous definition. But when the recursive call is the last action performed in the named
program (a tail recursion), as in this example, the old scope and its definitions cannot be used again,
so the new scope replaces the old one; the scopes and variables do not pile up.

Let \( name \) be a new name (not defined in the local scope), and let \( program \) be a program, possibly
using the name \( name \). Then the following three lines are equivalent to each other.

\[
\begin{array}{c}
\text{name } \begin{array}{c}
\text{[program]} \\
\end{array} \\
\text{[new } name \text{[program]. } name] \\
\text{new } name \text{[program]. } name. \text{ old } name
\end{array}
\]

Plan

A plan is a program with a parameter. There are four forms of plan. The first is

\[
\text{plan } \begin{array}{c}
\text{simplename} : \text{data } \begin{array}{c}
\text{[program]} \\
\end{array}
\end{array}
\]

The `simplename` is being introduced as a parameter within the program. It can be any simple name,
even one that has already been defined in the current scope. Its type (after the `:`) cannot make use
of the parameter. The scope of the parameter is from `[[` to `]]`. For example,

\[
\text{plan } y: \text{real } \begin{array}{c}
\text{[x:= x\times y]} \\
\end{array}
\]

A plan can be argumented in the same way that lists are indexed and functions are argumented. The
argument provides a value for the parameter. For example,

\[
\text{plan } y: \text{real } \begin{array}{c}
\text{[x:= x\times y]} \\
3
\end{array}
\]

is the same as

\[
x:= x\times 3
\]

A program with \( n+1 \) parameters is a program with \( 1 \) parameter whose body is a program with \( n \)
parameters. For example, here is a program with two parameters, followed by two arguments.

\[
\text{plan } x: \text{int } \begin{array}{c}
\text{[plan } y: \text{int } \begin{array}{c}
\text{[z:= x+y]} \\
3
4
\end{array}
\end{array}
\]
\]

equivalent to \( z:= 3+4 \). Here is a program to find the maximum value in nonempty list \( L \) in
\( \log (#L) \) time. ( \( L \) is a variable, and its initial value is destroyed in the process.) We define \( \text{findmax } i j \) to find the maximum in the segment of \( L \) from index \( i \) to (but not including) index \( j \),
reporting the result as \( L i \).

\[
\text{new } \text{findmax } \begin{array}{c}
\text{[plan } i: \square L \text{[plan } j: \square L+1 \\
\text{[if } j\geq i+2 \text{[findmax } i \text{[div (i+j) 2] } \| \text{findmax } \text{[div (i+j) 2] } j. \\
L:= i \rightarrow (L i \lor L \text{[div (i+j) 2] } L)]]] \\
\end{array}
\]

After execution of \( \text{findmax } 0 \text{(#L)} \), the maximum value in the initial list is \( L 0 \).

In the previous paragraphs, the parameter is a constant (note the single colon); it is not assignable.
It is “by initial value”, so

\[
\text{plan } i: \text{int } \begin{array}{c}
\text{x:= i. y:= i]} \\
(x+1)
\end{array}
\]

assigns both \( x \) and \( y \) to a value one greater than \( x \)’s initial value.
The second form of plan

```
plan simpasename :: data [ program ]
```

(note the double colon) creates a variable parameter. For example,

```
plan x:: int [ x:= 3 ]
```

A plan with a variable parameter applies to a variable argument. But it cannot be applied to a variable appearing in the plan. This restriction is required for reasoning about the plan. This example plan can be applied to any input channel that delivers text. But the plan

```
plan x:: int [ x:= 3. y:= 4 ]
```

cannot be applied to variable \( y \). Here is a plan named \( \text{norm} \) to reduce rational \( \text{num/denom} \) to lowest terms. If \( a=8 \) and \( b=12 \), then after \( \text{norm a b} \), we have \( a=2 \) and \( b=3 \).

```
new norm [plan num:: nat+1 [plan denom:: nat+1 ` normalize num/denom

    [new gcd = \langle a: (nat+1) \rightarrow \langle b: (nat+1) \rightarrow ` greatest common divisor of \( a \) and \( b \n
    a=b \equiv a \leq b \equiv \gcd a (b-a) \equiv gcd (a-b) b\rangle\)]

    new g:= gcd num denom. num:= num/g. denom:= denom/g]]]
```

The main use for variable parameters is probably to affect many files in the same way; for example, a plan to sort files.

The next form of plan

```
plan simpasename ! data [ program ]
```

creates a plan with an output channel parameter. For example,

```
plan c! text [ c! “abc” ]
```

This plan can be applied to any channel that receives text. A plan with a channel parameter cannot be applied to a channel appearing in the plan. This example plan can be applied to any output channel, even one named \( c \), because that \( c \) is nonlocal, and is not the local channel \( c \) appearing in the plan. But

```
plan c! text [ c! “abc”. d! “def” ]
```

cannot be applied to channel \( d \).

The final form of plan

```
plan simpasename ? data [ program ]
```

creates a plan with an input channel parameter. For example,

```
plan c? text [ c? 3*char. screen! c?? ]
```

This plan can be applied to any input channel that delivers text. But

```
plan c? text [ c? text. d? text ]
```

cannot be applied to channel \( d \). The channel names \( \text{keys} \) and \( \text{screen} \) cannot be omitted when they are used as an argument for a channel parameter.

The following program \( \text{pps} \) has three channel parameters. On the first, \( a \), it reads the coefficients of a rational power series; on the second, \( b \), it reads the coefficients of another rational power series; on the last, \( c \), it writes the coefficients of the product power series.

```
new pps [plan a? rat [plan b? rat [plan c! rat

    [ a? rat || b? rat. c! a??×b??.

    new a0:= a?? \|| new b0:= b?? || new d? rat! 0.

    pps a b d || [a? rat || b? rat. c! a0×b??+a??×b0.

    loop [a? rat || b? rat || d? rat. c! a0×b??+d??+a??×b0. loop]]]]]]
```
A plan can be named; a plan can be argumented; and a plan can be the body of a plan. A plan cannot be composed sequentially or concurrently; a plan cannot be used in an if-program, nor in a case-program, nor in a for-program, nor in a result-expression. A plan with several parameters that has been partially argumented is still a plan; a plan that has been fully argumented is no longer a plan, and can be used wherever any program can be used. For examples,

\begin{verbatim}
plan y: real [x:= xxy] 3. z:= 2 `this is allowed
plan y: real [x:= xxy]. z:= 2 `this is not allowed
\end{verbatim}

**Dictionary Definition**

Dictionaries are the way you organize your programs and data. You can create as many dictionaries as you want. Dictionary definition has the form

```
new newname_
```

The newname becomes a dictionaryname. To create a new dictionary named `abc`, write

```
new abc_
```

(It does not matter whether there are spaces between the name and the underscore.) Now you can define names within this dictionary. A name being defined in a dictionary must not already be defined in that dictionary. Each name in a dictionary is defined, using the keyword `new` and a compound name, to be one of the following: a variable name, a constant name, a data name, a program name, a channel name, a unit name, or a dictionary name. For example,

```
new abc_x:= 2
```

defines `x` in dictionary `abc` to be the constant `2`. (It does not matter whether there are spaces before or after the underscore.) This constant can then be used as `abc_x`. To define new dictionary `def` within dictionary `abc` write

```
new abc_def_
```

When a name in a dictionary is defined to be a dictionary, this dictionary also can contain names, some of which can be defined as dictionaries, and so on. So a dictionary can be a tree structure. Suppose there is a dictionary named `ProTem` within which there is a dictionary named `grammars` within which there is a text named `LL1`. Its name is `ProTem_grammars_LL1`.

A dictionary that is not within another dictionary obeys the scope rules. In other words, if you define a dictionary within scope brackets `[ ]`, the dictionary becomes undefined at the end of the scope, just like any other simple name definition. And its scope can be ended early by `old`. For example,

```
old abc
```

And, like any other simple name, its scope cannot be ended by `old` within a subscope. When a dictionary becomes undefined, so do all the names within it. When a name becomes undefined, what it named remains in existence, anonymously, as long as something refers to it.

Names within a dictionary do not obey the normal scope rules. Instead, they obey the scope rules of the dictionary they are within. For example, if we define dictionary `abc` outside a local scope, and constant `x` in dictionary `abc` within the local scope, the definition of `x` within `abc` remains in effect past the end of the local scope because the definition of `abc` remains in effect. The name `abc_x` will no longer be defined when `abc` is no longer defined. The name `abc_x` can become undefined earlier by using `old`, even within a subscope. For example,

```
new abc_. [new abc_x:= 2]. screen! abc_x. [old abc_x]
```

The name `abc_x` is defined after the first `[ ]` scope, but not after the second `[ ]` scope.

There is one predefined dictionary named `predefined` containing all predefined names (see Predefined Names).
Synonym Definition

Synonym definition has the form

new newname oldname

The newname becomes a synonym for the oldname. One use is to shorten all names that are deep within several dictionaries. For example, if dictionary \( a \) contains dictionary \( b \), which contains dictionary \( c \), which contains dictionary \( d \), which contains variable \( x \), then

\[
\text{new } x \ a_{b_{c_{d_x}}}
\]

shortens the name \( a_{b_{c_{d_x}}} \) to just \( x \). The definition

\[
\text{new } d \ a_{b_{c_{d}}}
\]

shortens all names within \( a_{b_{c_{d}}} \), for example, from \( a_{b_{c_{d_x}}} \) to \( d_{x} \).

Format

Although it is not part of the ProTem language, here are some suggested formatting (indentation) rules. The choice of alternative depends on the length of component data and programs.

\[
\begin{align*}
\text{A. } & B \\
\text{or} \\
& A. \\
& B \\
\hline
& A \ || \ B \\
\text{or} \\
& A \\
& \ | | \\
& B \\
\hline
& \text{if } A \ [B] \ \text{else } [C] \\
\text{or} \\
& \text{if } A \ [B] \\
& \text{else } [C] \\
\text{or} \\
& \text{if } A \\
& [B] \\
& \text{else } [C] \\
\hline
\text{plan } x: A \ [B] \\
\text{or} \\
& \text{plan } x: A \\
& [B] \\
\end{align*}
\]

More indentation would show the structure better, but it would crowd programs onto the right side of the page.
Commands

There are 11 commands in ProTem. They are not presented in the grammar, and they cannot be part of a stored program. They can be used only by a human at a keyboard. A command may be given at any time; it does not have to respect the grammatical structure of a program; it interrupts execution. Each command is the escape character combined with a letter. The commands are:

- `esc e` enter or exit editor for saved program or data
- `esc a` abort execution of program
- `esc s` quit current session and start new session
- `esc p` change dictionary permits
- `esc i` change identity
- `esc u` undo current session
- `esc n` print names defined in current scope or in dictionary
- `esc m` attach or modify or retrieve memo to defined name
- `esc d` display source or object code for saved program or data
- `esc c` generate context comments
- `esc v` verify program according to a specification

Edit

The edit command `esc e` is used to modify an existing persistent program or data definition. It invokes a dialogue using keys and screen to determine which definition, and then invokes an editor. In the editor, `esc e` exits the editor, and asks if you want to throw away the old definition, and save and compile the new definition. If the new definition has an error, you receive an error message, an error comment is inserted into the saved source, and the compiled object code, when executed, prints “unable to execute [definition name]”. If you want to create a definition using the editor, first create the definition, for example, `new p [ok]`, and then invoke the editor to modify it. If you want to delete a definition, use `old`.

Abort

It is essential to be able to abort the execution of a program, especially if you suspect that its execution will take forever. Use `esc a` to abort execution.

Session

Sessions are defined for security and error recovery. When the computer is turned on, a session begins. When some idle time passes (how much time is a parameter of the system and may be set to infinity), a session ends and a new one begins. When the computer is turned off, a session ends. The `esc s` command causes the current session to end and a new session to begin.

Sessions do not define the lifetime of definitions. A definition in the persistent scope, outside all `[]` pairs, lasts from the execution of the definition (new) to the execution of the corresponding name removal (old). This may be less than a session, or more than a session. Turning off the computer should not cut the power instantly, but should first cause the values of any variables in the persistent scope to be saved in nonvolatile memory.

At the start of each session, channels are reinitialized, and the programmer’s identity is requested. If a programmer does not yet have an identity, they are invited to create one, and their personal persistent scope is initially empty. If they do have an identity, it is used to connect the programmer
to their saved persistent scope. It is also used to determine a programmer's right to read or write in each dictionary (see Permit).

The predefined name session is a text consisting of all keystrokes since the start of the current session. (This is quite practical: an hour's hard work produces only 10kbytes of keystrokes.) This text can be saved as a record of work done, or for error recovery (see Undo).

Permit

Each dictionary has two permit-lists of identities of people:
- read-permit-list: those who are permitted to read what is in the dictionary
- write-permit-list: those who are permitted to add new contents, change the current contents, delete old contents, and change these two permit-lists

When a dictionary is created, its two permit-lists are just the dictionary's creator. A permit-list is changed by means of the [esc p] command. The command starts a dialogue using keys and screen to ask which dictionary, to check if the person issuing the command is on the write-permit-list for that dictionary, and then to determine what change is wanted and make the change.

Permit-lists are not actually lists, but bunches of identities. The name everyone is predefined as the bunch of all identities. Dictionary predefined has read-permit-list everyone and write-permit-list only its creator.

If you have just been sent an identity on channel id, you can save it under the name Josh by the constant definition

new Josh:= id??

You might start a changeable group of identities named condoBoard by the variable definition

new condoBoard: everyone:= Josh, Claire, Martin, Amanda

This is how you can maintain permit-lists to grant limited access to dictionaries you create.

Identity

The [esc i] command is used to change identity, retaining one's persistent scope and dictionary permits. This command is used if one's identity is forgotten or compromised.

Undo

The command [esc u] undoes a session (except for inputs and outputs and session). Implementing it requires capturing the state at the start of a session. On many computers, returning to the prior state may be cheap; nonvolatile memory (that does not require power) contains the state as it was at the start of the current session, and volatile memory (that requires power) contains the current state.

After undo, you can capture the current value of session, let's call it recovery, then reassign (or edit) recovery, and then execute the result by writing exec recovery. This gives us perfectly flexible error recovery for the modest cost of a keystroke file.
Names

The command \texttt{esc n} begins a dialogue using \textit{keys} and \textit{screen} to determine whether you want the names defined in the current scope, or the names defined in a dictionary; if the latter, it determines which dictionary and checks whether you are on the read-permit-list. It does not print the names in subdictionaries of the selected dictionary.

Memo

Each definition can optionally have a memo attached to it. The memo might explain the purpose or use of the definition. It is there to be read by a human, not for execution. A memo is similar to a comment that you would make at the point of definition, but differs in that you can retrieve it anytime. The command \texttt{esc m} starts a dialogue using \textit{keys} and \textit{screen} to determine which name (simple or compound), whether you want to attach a new memo, modify an existing memo, or retrieve an existing memo, and checks whether you are on the appropriate permit-list. For example, you may say that you want to attach the memo

\begin{quote}
This variable accumulates the sum of the products.
\end{quote}
to name \texttt{x}. Asking for the memo attached to predefined name \texttt{e} prints

\begin{quote}
\texttt{e := 2.718281828459045} (approximately) \texttt{constant} The base of the natural logarithms.
\end{quote}

Display

The command \texttt{esc d} starts a dialogue using \textit{keys} and \textit{screen} to determine the name (simple or compound) of the program or data whose source or object code you want to view, and checks whether you are on the read-permit-list.

Context

The command \texttt{esc c} starts a dialogue using \textit{keys} and \textit{screen} to determine the program, bracketed by \texttt{[ ]}, for which context comments are wanted. The comments are then generated. These comments say which nonlocal names are used, and in what way they are used. Here is the format.

\begin{quote}
`input: on these channels
`output: on these channels
`use: the values of these variables, constants, data names, units, and function names
`assign: these variables
`call: these program names and plan names
`refer: to these dictionaries
\end{quote}

If there already are comments in this format, they are replaced. (For examples of context comments, see \textbf{Example Programs}.) Additionally, a programmer may want to include comments like

\begin{quote}
`spec: specification
`pre: precondition
`post: postcondition
`inv: invariant
\end{quote}

but these are not generated by \texttt{esc c}.

Verify

The command \texttt{esc v} starts a dialogue using \textit{keys} and \textit{screen} to determine the program, bracketed by \texttt{[ ]} and named by a fancy name, for which verification is wanted. The verification is then attempted. (See \textbf{Program Definition} and \textbf{Named Program}.)
As a character within a text, the left- and right-double-quote characters must be underlined. For example, “Just say “no”.”. As a character within a text, an underlined left- and right-double-quote character must be underlined again. And so on. Thus every character can occur within a text. But we cannot write a self-reproducing expression with this convention. For that purpose, we need another convention, such as repeating the left- and right-double-quote characters within a text. For example, “Just say ““no””.”. Using this convention, here is a self-reproducing expression (perform the indexing to see what you get).

```
"""
(0;0;(0;..32);31;31;(1;..31))
```

Sounds and pictures are data structures. This part of ProTem is not yet designed. Perhaps a picture is an element of \([x*y*0..z])\] where \(x\) is the number of pixels in the horizontal direction, \(y\) is the number of pixels in the vertical direction, and \(z\) is the number of pixel values. A picture could therefore be expressed in the same way as any other two-dimensional array, and one could refer to the pixel in column 3 and row 4 of picture \(p\) as \(p\ 3 \ 4\). Perhaps a movie is a string of pictures. The operations on movies would be those of strings, such as substring and join. To help in the creation of movies, one of the pixel values should be transparent, and one of the operations on pictures should be overlaying one picture on another.

Predefined **silence** is a sound, and predefined **sound** is all sounds. Sounds are input on channel **microphone**; pictures are input on channel **camera**. A constant can be defined as a sound or picture. A variable can be assigned to a sound or picture. Sounds and pictures can be included in a data structure, and manipulated using the operators on that data structure. Sounds can be output on channel **speaker**; pictures can be output on some not-yet-determined channel.

The ProTem equivalent of enumerated type is shown here.

```protem
new color:= "red", "green", "blue".
new brush: color:= "red"
```

The ProTem equivalent of the record type (structure type) is as follows.

```protem
new person:= "name" → text | "age" → nat.
new p: person:= "name" → "Josh" | "age" → 16
```

The fields of \(p\) can be selected by an argument, for example

```protem
! p "name"
```

prints the text “Josh”. The value of \(p\) can be changed using a function arrow and selective union.

```protem
p:= "age" → 17 | p.
p:= "name" → "Amanda" | "age" → 2
```

We can even have a whole file (string) of records

```protem
new file: *person:= nil
```
and join new records onto its end.

```protem
file:= file; p
```

The efficiency of pointers is obtained through the use of the predefined function **point**. When applied to a list argument, it yields the deep domain of the list. For example,

```protem
point [10; [11; 12]; 13] = 0 , 1;(0, 1) , 2 = 0 , 1;0 , 1;1 , 2
```

The use of **point** is a signal to the implementation that its strings of natural numbers will be used only as indexes into the list (and the implementation will check that this is so). For example, we can define a linked list \(G\) as follows.
new G: [*("name" → text | "next" → point G)]:= ["name" → end | “next” → 0].

new first: point G:= 0.

We can add a constant to first or subtract a constant from it, for example

first:= first+1

and similarly for the “next” field of each record of G. But we can ultimately use them only as indexes into G, for example

first:= G@first “next”

G:= first → (“name” → “Aaron” | “next” → first) | G

With this limited use, the implementation of these indexes can be memory addresses. This way we obtain all the performance benefits of pointers without destroying the logic of our language.

The previous example, with linked list G, does not show the full generality of point. Here is a tree-structured example.

new tree = [nil], [tree; all; tree].
new t: tree:= [nil].
new p: point t:= nil

To move p down to the left in the tree we reassign it this way:

p:= p; 0

To move it down to the right, reassign it this way:

p:= p; 2

Thus p is a string of indexes indicating a subtree t@p of t. We can replace this subtree with tree s using the assignment

t:= p → s | t

We can express the information at the node indicated by p as

t@p 1  or  t@(p; 1)

and we can replace the information at this node with the integer 6 using the assignment

t:= (p; 1) → 6 | t

To move up in the tree, we just remove the final item of p, and to make that easy, the predefined

new back = 〈p: (*nat) → p\(0;..↔p–1)〉

allows us to move p up to its parent by writing

p:= back p

The “procedure”, “method”, or “function” of some other programming languages is a combination of naming, scope, and parameter(s). For example,

new transform [plan magnification: real [plan translation: real

[x:= magnification×x + translation]]]

Here is a definition of a plan with one parameter

new translate [transform 1]

formed by providing one argument to a two-parameter plan. To provide an argument for just the second parameter is a little more awkward, but not too bad.

new magnify [plan magnification: real [transform magnification 0]]

We can now obtain a three-times magnification of x in either of these ways.

magnify 3
transform 3 0

In some other programming languages, the “function” is a combination of naming, scope, parameter(s), and result-expression. For example,

new factorial = 〈n: nat → result f: nat:= 1 [for i:= 1;..n+1 [f:= f×i]]〉

Exception handling is provided by | or if or ⊨ ⫤. For example,
new divide = ⟨dividend: com → (divisor: com → 
   divisor = 0 ⊨ “zero divide” = dividend / divisor)⟩

Then
divide: com → com → (com, “zero divide”)
The selective union operator applies its left side to an argument if that argument is in the stated domain of its left side; otherwise it applies its right side. Let us define
new weekday = ⟨d: (0,..7) → 1 ≤ d ≤ 5⟩

Then in the expression
(weekday | all→“domain error”) i
if i fails to be an integer in the range 0,..7 , the left side of | “catches” the exception and “throws” it to the right side, where it is “handled”.

Input choice, as in CSP, can be obtained as follows.

inputchoice ⌜if c!! ⌜c? numpat; nl. P] else ⌜if d!! ⌜d? numpat; nl. Q] else [inputchoice]⌝⌝

In the persistent scope, ProTem functions as an operating system, where programs are executed as soon as they are entered. Unix directories are dictionaries. Unix files are variables. Unix cd is ProTem's synonym definition. Unix rm is ProTem's old. Unix mv is a synonym definition followed by old. The commands esc n and esc m are the Unix ls and man commands. ProTem's esc p replaces Unix's chmod. The effect of Unix pipes is obtained by channel parameters. For example, suppose trim is a plan to trim off leading and following blanks and tabs from lines of text, and sort is a plan to sort texts. (Please excuse the informal body since it's not the point.)

new trim ⌜plan in? text ⌜plan out! text ⌝Repeatedly read from in , trim off leading and trailing space, output to out , until end is read and output.]]]]

new sort ⌜plan in? text ⌜plan out! text ⌝Repeatedly read from in until end is read, and output the sorted texts and end to out.]]]]

We can feed the output from trim to the input of sort by defining a channel for the purpose. If the original input comes from keys , and the final output goes to screen , then

new pipe? text!”". trim keys pipe. sort pipe screen. old pipe

Even better:

new pipe? text!”". trim keys pipe || sort pipe screen. old pipe

If sort wants input before it is available from trim , sort waits. If trim sends output before sort wants it, it is buffered.

An implementation may provide plans for a variety of languages. For example, it may provide a plan named Python , with one text parameter, whose execution is that of the Python fragment represented by the argument. It may provide asm , whose execution is that of the assembly-language program represented by the argument.

ProTem considers object orientation to be a programming style, rather than a programming-language style, or collection of language features. Object-oriented programming (as a style of programming) can be done in ProTem. Data structures, and the functions and procedures that access and update them, can be defined together in one dictionary. If many objects of the same type are wanted, the type can be parameterized as appropriate, defined once, and used many times.

To execute a program stored on someone else's computer, just invoke that remote program using its full address (computername_programname). For efficiency, it might be best to compile that remote program for your own computer and run it locally. Any nonlocal names (variables, channels, and so on) refer to entities on the computer where the program is compiled.
The keyword else is redundant; it could be deleted everywhere it occurs, reducing the number of keywords to 7. But I think it is useful redundancy, for both error detection and readability.

**Intentionally Omitted Features**

Each of the following omitted features is a syntactic convenience, and it's no trouble to add to the language. But they make the language larger, and that's a cost. And they move away from the form needed for verification. So they are not included in ProTem.

**assert**

| assert \( x \leq y \) | means | if \( \neg (x \leq y) \) "assert failure". stop |

**item assignment**

| \( S \triangleleft 3 := 5 \) | means | \( S := S \triangleleft 3 \triangleright 5 \) |
| \( L 3 := 5 \) | means | \( L := 3 \triangleright 5 \mid L \) |
| \( L 3 4 := 5 \) | means | \( L := (3;4) \triangleright 5 \mid L \) |

**name grouping**

| new \( x, y \) : int:= 0 | means | new \( x \) : int:= 0 || new \( y \) : int:= 0 |
| old \( x, y \) | means | old \( x \) || old \( y \) |
| \( (a, b: \text{nat} \to a+b) \) | means | \( \langle a: \text{nat} \to \langle b: \text{nat} \to a+b \rangle \rangle \) |
| plan \( a, b: \text{nat} \ [x:= a+b] \) | means | plan \( a: \text{nat} \ [\text{plan} b: \text{nat} \ [x:= a+b] \] |
| \( x, y:= 0 \) | means | \( x:= 0 \mid y:= 0 \) |

**looping constructs**

| while \( n > 0 \ [n:= n-1] \) | means | while \( [\text{if } n > 0 \ [n:= n-1. \ \text{while}]] \) |
| repeat \( [n:= n-1] \) until \( n=0 \) | means | repeat \( [n:= n-1. \ \text{if } \neg(n=0) \ [\text{repeat}]] \) |
| loop \( [n:= n-1. \ \text{if } n=0 \ [\text{exit}] \ [m:= m+1]] \) | means | loop \( [n:= n-1. \ \text{if } \neg(n=0) \ [m:= m+1. \ \text{loop}]] \) |

**return from program**

| if \( n=0 \) [return]. restOfProgram | means | if \( \neg(n=0) \ [\text{restOfProgram}] \) |

The assignment \( L := 3 \triangleright 5 \mid L \) should be compiled the same as \( L 3 := 5 \) would be if it were included in ProTem; the list \( L \) should not be copied. The same for string item assignment. In the loop

| while \( [\text{if } n > 0 \ [n:= n-1. \ \text{while}]] \) |

the last-action (tail recursive) call should be compiled as a branch (jump) instruction, with no stack activity, the same as a while-loop would be if it were included. Omitting string and list item assignment and special looping constructs should not cost execution time.

Dictionary and program parameters and arguments were considered and rejected.

| plan simplename _ \[ program ] | plan, parameter is dictionary |
| program dictionaryname | plan, dictionary argument |
| plan simplename \[ program ] | plan, parameter is program |
| program program | plan, program argument |

As a direct counterpart to the Unix cd command, we considered

| open dictionaryname |
| close dictionaryname |

to allow names in that dictionary to be referred to without stating the dictionary. For example, if we have dictionary \( abc \), and within it names \( x \) and \( y \), we can refer to these names as \( abc_x \) and \( abc_y \). By saying

| open abc |

we can then refer to them as just \( x \) and \( y \). But the interaction between open and scope is complex, and we can shorten names by synonym definition, so we left out open and close.
There is no frame construct in ProTem, but escape serves the same purpose.

In some languages there is a module or object construct for the purpose of grouping together related definitions. In ProTem, dictionaries serve that purpose.

Do we ever want the correcting pattern without an echo? For a password, we usually want to output some character, like •, for each key press, and program the corrections, rather than using the correcting pattern (see Read Password). If there is a reason anyone wants the correcting pattern without an echo, we can easily allow it, but it will complicate our LL(1/2) grammar.

Two binary operators ∆ (nand) and ∇ (nor) are missing. They aren't wanted very often, there aren't good keyboard substitutes for them, and they are easily synthesized:

\[ x \Delta y = \neg (x \land y) \]
\[ x \nabla y = \neg (x \lor y) \]

**Implementation Philosophy**

Ideally, an implementation checks whether the text presented to it represents a program, and issues an error message if it does not. That check should include determining whether every variable assignment is to a value that is included in the type of the variable. That determination is most helpful if it can be made before execution; but if not, it is still helpful if it can be made during an execution attempt.

While not an error, there are also expressions that cannot or should not be evaluated further. That presents an implementation problem, but not a semantic problem. For example,

\[ ! -3 \]

prints −3

ProTem does not evaluate the application of the negation operator – to the operand 3 (see Number Representation); it just prints the operator and operand. Similarly

\[ ! 1/0 \]

should print 1/0

\[ ![0; 1] 2 \]

should print [0; 1] 2

\[ ![r: rat \rightarrow 5] (1/0) \]

should print 5

\[ ! 1/0 = 1/0 \]

should print \( \top \)

\[ ![0; 1] 2 = [0; 1] 2 \]

should print \( \top \)

Due to the difficulty of implementation, it is permissible for an implementation to behave differently.

No programming language has ever been, or will ever be, implemented entirely. Every programming language is infinite; every implementation is finite. There is always a program too big for the implementation. There is a multitude of size limitations: the parse stack might overflow, the dictionary (symbol table) might be too small, the forward branch fixup list might be exceeded, and so on. It would be ugly to define a programming language by listing all the size limitations of programs. And it would be counter-productive because it would exclude implementations that can accommodate larger programs.

Whenever a program exceeds a size limitation, the implementation should not say “Error: limitation exceeded.”, because the program is not in error. The implementation should say “Apology: this implementation is too limited to accommodate your program.”. An “error” message tells a programmer to correct the error; there is no other option. An “apology” message gives the programmer 3 options: change the program to live within the limitation; change the implementation options to increase the limit that was exceeded; take the program to a different implementation.
Natural numbers and integers are usually limited to those that are representable in a specific number of bits, for example, 32 bits. This is a size limitation, just the same as other size limitations. It is more complicated and uglier to define arithmetic within finite limitations than to define the naturals and the integers. And it is counter-productive to do so, because it excludes an implementation with 64-bit arithmetic. As with other implementation limitations, numeric overflow should not get an “error” message; it should get an “apology” message.

Floating-point numbers and arithmetic should never be offered as a language feature. The programmer wants rational or real numbers and arithmetic, but may be willing to accept the floating-point approximation for the sake of efficiency. Floating-point, with a specific number of bits, is an implementation limitation. Any alternative to floating-point that increases the accuracy without taking too much time or space should be welcome.

ProTem is a rich programming system, offering many kinds of data and operators on data, and many ways to structure a computation. Some features may be difficult to implement. And some features may be of little use to most programmers. It may be a wise decision not to implement some features. For example, an implementer might decide that in a variable definition, the type must be one of

$$\text{nat} \ \text{int} \ \text{rat} \ \text{bin} \ \text{text} \ [n\text{type}]$$

where $n$ is a natural number and $\text{type}$ is any of these types just listed. An implementer may decide not to implement concurrent execution. No-one can complain that the complete language is not implemented, since it is impossible to completely implement any language. But ProTem is defined to allow all type expressions that make sense, and to allow concurrency, so the next implementation can accommodate programs that previous implementations could not accommodate.

**Predefined Names**

The predefined names are defined in the predefined scope, which is outside the persistent scope (see Scope). Predefined names can be redefined in more local scopes. In general, if a local definition redefines a name from a less local scope, the less local definition is covered, and inaccessible, because any local use of the name refers to the more local definition.

There is a predefined dictionary named `predefined` with all predefined names in it. It enables us to refer to a covered predefined name. For example, one of the predefined names is the imaginary number $i$ (a square root of $-1$). You may also want to define a local variable $i$. If you do, you can still refer to the predefined $i$ as `predefined_i` (unless you have also covered the predefined name `predefined` with a local definition). If predefined name $i$ is covered by a definition in a scope between the predefined scope and the local scope where you are working, you can get back the simple name $i$ as the predefined imaginary number in these three ways:

```
new i:= predefined_i  \ `constant definition
new i predefined_i    \ `synonym definition
new i:= 0&1            \ `constant definition
```

Dictionary `predefined` has read-permit-list `everyone` and write-permit-list only its creator (see Permit). Unless these permit lists are changed by its creator, only its creator can add new definitions into the predefined scope. For example,

```
new predefined_amaze = \(n: \text{nat} \rightarrow n+2\)
```

Likewise, only the `predefined` dictionary's creator can remove definitions from the predefined scope. For example,

```
old predefined_amaze
```
The command `esc n` is used to list the names in the predefined dictionary. The command `esc m` is used to get a description of a predefined name.

Here are the predefined names. Some definitions use § (those), defined in a Practical Theory of Programming. Each name is one of:

- **variable**: at present, there are no predefined variables
- **constant**: evaluated; not assignable
- **data**: unevaluated; evaluation upon use; not assignable
- **program**: unexecuted; execution upon use
- **channel**: reinitialized at the start of each session
- **unit**: unrelated to other predefined units
- **dictionary**: at present, the only predefined dictionary is predefined

---

### Definitions

- **abs**: `com→real data` Absolute value. `abs x = sqrt(re x^2 + im x^2)`.
- **all**: `com, char, bin, sall, [sall] data` All ProTem items.
- **arc**: `com→§(r: real → 0 ≤ r < 2*pi) data` The angle or arc of a complex number.
- **arccos**: `§(r: real → −1 ≤ r ≤ +1)→§(r: real → 0 < r < pi/2) data` A trigonometric function.
- **arcsin**: `§(r: real → −1 ≤ r ≤ +1)→§(r: real → 0 < r < pi/2) data` A trigonometric function.
- **arctan**: `real→§(r: real → 0 < r < pi/2) data` A trigonometric function.
- **await**: `program` A plan with one constant parameter of type `real*s`. If the argument represents the present or a future time, its execution does nothing but takes time until the instant given by the argument. If the argument represents the present or a past time, its execution does nothing and takes no time. See `time` and `wait` and `s`.
- **back**: `*nat→*nat data` If `i` is an item, `back (s; i) = s`.
- **bin**`:= T, ⊥ constant` The binary values.
- **bintext**: `bin→text constant` `bintext T = “T”` and `bintext ⊥ = “⊥”`.
- **bold**: `text→text data` Same text but in bold font.
- **ceil**: `real→int data` `r ≤ ceil r < r+1`.
- **char**: `data` The characters.
- **charnat**: `char→nat data` A one-to-one function with inverse `natchar`. The encoding might be extended ASCII or unicode. Character combinations, for example shift-option-a, also have numeric encodings.
- **click**: `char constant` The click character.
- **com**: `real&real data` The complex numbers.
- **cos**: `real→§(r: real → −1 ≤ r ≤ +1) data` A trigonometric function.
- **cosh**: `com→com data` A hyperbolic function.
- **cursor**: `nat; nat data` A data name whose value is the current cursor position.
- **delete**: `char constant` The delete or backspace character.
- **digit**: `char constant` The decimal digits.
- **div**: `real→§(r: real → r>0)→int data` `div a d` is the integer quotient when `a` is divided by `d`.
  
  \[(0 ≤ mod a d < d) ∧ (a = div a d × d + mod a d)\]

- **doubleclick**: `char constant` The doubleclick character.
- **e**: `2.718281828459045 (approximately) constant` The base of the natural logarithms.
- **end**: `char constant` The end-of-file character. It is greater than all other characters.
- **eval**: `text→*all data` If the argument represents a ProTem data expression, the evaluation is that of the represented data. It “unquotes” its argument. In `eval “x”`, the “x” refers to whatever `x` refers to at the location where `eval “x”` occurs. If the argument does not represent a ProTem data expression, the result is “error”.
- **even**: `int→bin data` A function that says whether its argument is even.
- **everyone**: `data` All identities.
exec  program  A plan with one text parameter. If the argument represents a ProTem program, the execution is that of the represented program. It “unquotes” its argument. If applied to “x:= x+1”, the “x” refers to whatever x refers to at the location where exec “x:= x+1” occurs.

If the argument does not represent a ProTem program, execution displays an error message.

exp: com→com  data  The exponential function. exp x = e^x.

false:=  \perp  constant  A binary value.

find: all→*all→nat  data  If i is an item in string S, then find i S is the index of its first occurrence; if not, then find i S = ∞.

fit: int→text→text  data  If i≥0 then fit i t is a text of length i obtained from t by either chopping off excess characters from the right end or by extending t with spaces on the right end. If i≤0 then fit i t is a text of length –i obtained from t by either chopping off excess characters from the left end or by extending t with spaces on the left end.

floor: real→int  data  floor r ≤ r < 1 + floor r

form: nat→nat→(nat+1)→real→text  data  Format a real number. form d e w r is a text representing real r with the final digit rounded. d is the number of digits after the decimal point; if d=0 the point is omitted. e is the number of digits in the exponent; if e>0 the decimal point will be placed after the first significant digit; if e=0 the ^^ is omitted and the decimal point will be placed as necessary. w is the total width; if w is greater than necessary, leading blanks are added; if w is less than sufficient, the text contains blobs.

form 4 1 10 pi = “3.1416^0”  form 2 0 6 (–pi) = “–3.14”

form 0 0 3 5 “5”  form 0 0 3 (–5) = “–5”  form 0 0 2 123 = “••”.

g  unit  Representing mass in grams.
i:= 0&1  constant  An imaginary number: a square root of –1.

im: com→real  data  The imaginary part of a complex number.
infinity:= ∞  constant  An infinite number, greater than all other numbers.

int = nat, –nat  data  The integers.

italic: text→text  data  Same text but in italic font.

keys?  text “ ”  channel  To the program that monitors key presses, it is an output channel; to all other programs, it is an input channel.

lb: $r$ real→r>0 → real  data  The binary (base 2 ) logarithm.

ln: $r$ real→r>0 → real  data  The natural or Napierian (base e ) logarithm.

log: $r$ real→r>0 → real  data  The common (base 10) logarithm.

m  unit  Representing distance in meters.

match: *all→*all→nat  data  If pattern occurs within subject, then match pattern subject is the index of its first occurrence. If not, then match pattern subject = ∞ subject.

maxint: int  constant  The maximum representable integer (machine dependent).

maxnat: nat  constant  The maximum representable natural (machine dependent).

microphone?  *sound!  silence  channel  To the microphone, it is an output channel; to all other programs, it is an input channel.

minint: int  constant  The minimum representable integer (machine dependent).

mod: real → $r$ real→r>0 → real  data  mod a d is the remainder when a is divided by d.

(0 ≤ mod a d < d) ∧ (a = div a d × d + mod a d)

movie = *picture  data  A string of pictures.

nat = 0,..∞  data  The natural numbers.

natchar: nat→char  data  A one-to-one function with inverse charnat. The encoding might be extended ASCII or unicode. Character combinations, for example shift-option-a, also have numeric encodings.

nil  constant  The empty string.

nl: char  constant  The new line character or next line character or return character.

null  constant  The empty bunch.
numtext: com→text data A text representation of a number. See also form.
odd: int→bin data A function that says whether its argument is odd.
ok program A program whose execution does nothing and takes no time.
ord = real, char, bin, all, *ord, [ord] data The ordered type, for which < > ≤ ≥ are defined.
pi := 3.141592653589793 (approximately) constant The ratio of a circle's circumference to its diameter.
picture = [[x*[y*(0..z)]]] data where x is the number of pixels in the horizontal dimension, y is the number in the vertical dimension, and z is the number of pixel values.
point data A function that applies to a list and gives its deep domain (a bunch of strings of indexes).
    It is a signal to the implementation that the strings in it will be used only as indexes to the list. It can therefore be implemented as a memory address (pointer).
pre: char→char constant The character predecessor function.
prefixed dictionary A dictionary containing all prefixed names.
printer? text "" channel To the printer, it is an input channel; to all other programs, it is an output channel.
rndNat: nat→nat→nat data A reasonably uniform function, dependent on a hidden variable, over the interval from (including) the first argument to (excluding) the second argument.
rndNatInit program A plan with one constant natural parameter. Its execution assigns a hidden variable to the natural value.
rndNatNext program Its execution assigns a hidden variable to the next value in a random sequence.
rndReal: real→real→real data A reasonably uniform function, dependent on a hidden variable, over the interval between the arguments.
rndRealInit program A plan with one constant real parameter. Its execution assigns a hidden variable to the real value.
rndRealNext program Its execution assigns a hidden variable to the next value in a random sequence.
rat = int/(nat+1) data The rational numbers.
re: com→real data The real part of a complex number.
real data The real numbers.
round: real→int data r–0.5 ≤ round r < r+0.5
s unit Representing time in seconds.
screen? text "" channel To the screen, it is an input channel; to all other programs, it is an output channel.
session: text data All keystrokes on channel keys since the start of a session.
sign: real → (–1, 0, 1) data The sign of a real number.
silence: sound data The silent sound.
sin: real → §{r: real → –1 ≤ r ≤ 1} data A trigonometric function.
sinh: com→com data A hyperbolic function.
sort: *ord→*ord data Sorts in nondecreasing order.
sound data The sounds.
speaker? *sound! silence channel To the speaker, it is an input channel; to all other programs, it is an output channel.
sqrt: com→com data The principal square root.
stop program Its execution does nothing and takes forever so that no computation can follow.
sub: *all→nat→nat→*all→*all data sub s n m t is a string formed from string s by replacing the substring from index n to index m with string t. The substring being replaced s\(n;..m)\) does not have to be the same length as the string t replacing it. If n=m this is insertion. If
\[ t = \text{nil} \text{ this is deletion. } \]
\[ \text{sub } s \text{ n m t } = s(0;..n); t; s(m;..\leftrightarrow s) \]
\[ \text{subst: } * \text{all} \rightarrow \text{all} \rightarrow \text{all} \quad \text{data} \quad \text{sub } s \text{ x y } \text{ is a string formed from string } s \text{ by replacing all occurrences of item } x \text{ with item } y . \]

\[ \text{suc: char} \rightarrow \text{char constant} \text{ The character successor function.} \]
\[ \text{tab: char constant} \text{ The tab character.} \]
\[ \text{tan: real} \rightarrow \text{real data} \text{ A trigonometric function.} \]
\[ \text{tanh: com} \rightarrow \text{com data} \text{ A hyperbolic function.} \]
\[ \text{text } = * \text{char data} \]
\[ \text{textbin: text} \rightarrow \text{(bin, “error”) data} \text{ If the argument represent a binary value, possibly preceded by space, tab, and new line characters, possibly followed by space, tab, and new line characters, the result is the represented binary value. Otherwise the result is “error”.} \]
\[ \text{textid: text} \rightarrow \text{everyone data} \text{ A not-invertible injective function.} \]
\[ \text{textnum: text} \rightarrow \text{(com, “error”) data} \text{ If the argument represents a number, possibly preceded by space, tab, and new line characters, possibly followed by space, tab, and new line characters, the result is the represented number. Otherwise the result is “error”.} \]
\[ \text{texttime: text} \rightarrow \text{(real, “error”) data} \text{ Given the time in seconds since 2000 January 1 at 0:00 UTC (the midnight that begins 2000 January 1 at longitude 0), and a time zone, the result is a readable text. Times before then are negative. For example, texttime “1947 September 16 at 14:24:32.5 UTC–5” = –68675727.5 × s . Otherwise the result is “error”.} \]
\[ \text{trim: text} \rightarrow \text{text data} \text{ A text formed from the argument by removing all leading and trailing space, tab, and new line characters.} \]
\[ \text{true } = \top \text{ constant} \text{ A binary value.} \]
\[ \text{wait program} \text{ A plan with one constant parameter of type \emph{real}s . If the argument is nonnegative, its execution does nothing and takes the time given by the argument. If the argument is nonpositive, its execution does nothing and takes no time. See \emph{await} and \emph{time} and \emph{s}.} \]

**Example Programs**

**Portation Simulation**

\[
\text{new simport ` a program to simulate portation} \\
\text{[`input: keys time} \\
\text{`output: screen} \\
\text{`use: ceil m nat nil nl numtext point rat real s sqrt textnum} \\
\text{`call: await stop} \\
\text{` Distance between control boxes is always 1 m.} \\
\text{` Merges do not overlap, so there's at most 1 corresponding box on the merging portway.} \\
\text{` Each divergence has a left branch and a right branch; there's no straight.} \\
\text{` Leading to a divergence, boxes record only one square speed.} \\
\]
` start of definitions

**new** km:= 1000×m. **new** h:= 60×60×s. ` kilometer and hour

**new** maxaccel:= 1.5×m/s. ` maximum deceleration = –maxaccel
**new** speedlimit:= 60×km/h. ` speed limit is 60 km/h everywhere
**new** cushion:= 1×s. ` reaction time for all porters
**new** impatience:= 10/s. ` acceleration factor

**new** maxdistance:= ceil(speedlimit^2 / (2×maxaccel)). ` max search distance ahead
**new** numporters:= 120.
**new** numboxes:= 7480.
**new** visualDelayTime:= 0.5×s. ` for human viewing

**new** porter. ` so porter can be referenced before it is defined

**new** box: [numboxes] *((“ahead left”, “ahead right”, “behind left”, “behind right”) → point box
|“beside” → point box
|“above” → point porter, numporters
|{“horizontal”, “vertical”) → nat}] ` box position on screen
:= [numboxes] * ((“ahead left”, “ahead right”, “behind left”, “behind right”) → 0
|“beside” → 0
|“above” → numporters ` indicates no porter above
|{“horizontal”, “vertical”) → 0 ]].

**new** porter: [numporters] * (“below” → point box ` what’s beneath
|“arrival time” → real×s ` arrival time at this box
|“speed” → real×m/s )] ` current speed
:= [numporters] * (“below” → 0
|“arrival time” → 0×s
|“speed” → 0×m/s )].

**new** draw [plan b: nat [plan c: “grey”, “blue”, “red” [UNFINISHED]]]. `end of draw
` draws a box at screen position (box b “horizontal”) (box b “vertical”) of color c.
` “grey” means no porter present, “blue” means porter present, “red” means crash
` UNFINISHED because graphical output has not yet been designed

` end of definitions, start of initialization

for b:= 0;..numboxes
[ [ ! “What box is ahead-left of box ”; b; “?” . ?!
   box:= (b; “ahead left”) → ?? | (??; “behind left”) → b | box.
! “What box is ahead-right of box ”; b; “?” . ?!
   box:= (b; “ahead right”) → ?? | (??; “behind right”) → b | box.
! “What box is beside box ”; b; “?” . ?!
   box:= (b; “beside”) → ?? | box.
! “What are the horizontal and vertical coordinates of box ”; b; “?” . ?!
   box:= (b; “horizontal”) → ?? | box.
?!. box:= (b; “vertical”) → ?? | box.
draw b “grey”). ` default color; may be changed below
for p:= 0;..numporters
  !“Porter “; p; “ is over what box? ”. ?!
  porter:= (p; “below”) → ?? | porter. box:= (?; “above”) → p | box.
draw (??) “blue”].

randNatInit 123456789. ` initialize a random number generator

` end of initialization, start of simulation

infiniteLoop
  ` the time of the start of each iteration of the infiniteLoop
  new p: point porter:= 0. ` p:= the porter that arrived at its current position first
  new t: realxs. ` t is a time, initially an infinite time
  for q:= 0;..numporters [if porter q “arrival time” < t [t:= porter q “arrival time”. p:= q]].
  old t.

  new b:= porter p “below”. ` the box below porter p
  new bb:= box b “beside”. ` the box beside b; if none then bb=b
  new boxesToDo: *[point box; natxms]:= nil.
    ` queue of boxes to be explored; their distances ahead of porter p
    ` queue is sorted by increasing distance ahead
    ` difference between any two distances in the queue is at most 1

  ` initialize boxesToDo
  if bb = b [boxesToDo:= nil]
  else [if box bb “above” = numporters [boxesToDo:= nil]]
    else [if porter (box bb “above”) “speed” < porter p “speed” [boxesToDo:= nil]]
      else [boxesToDo:= [bb; 0×m]]]]).
  boxesToDo:= boxesToDo; [box b “ahead left”; 1×m].
  if box b “ahead left” ≠ box b “ahead right” [boxesToDo:= boxesToDo; [box b “ahead right”; 1×m]].
  old b. old bb.

  new accel: real×m/s/s:= maxaccel. ` acceleration for porter p

  ` using boxesToDo calculate accel for porter p

nextBox [new b:= (boxesToDo∅) 0. `the box we are looking at
  new d:= (boxesToDo∅) 1. `its distance ahead of porter p
  boxesToDo:= boxesToDo(1;..↔boxesToDo).
  if d≤maxdistance
    [new desiredspeed = `according to porter pa
      ⟨pa: (point porter, numporters) →
        pa=numporters ⊨{ speedlimit
          =\( \sqrt{\text{pa} “speed”^2 + 2×\text{maxaccel}×\text{d} + (\text{maxaccel}×\text{cushion})^2} \)
          − \text{maxaccel}×\text{cushion} \) ∧ speedlimit}.]
      accel:= ( ( desiredspeed (box b “above”)
          ∧ desiredspeed (porter (box b “beside”) “above”)))
− porter p “speed”) × impatience) ∨ − maxaccel ∧ maxaccel.

if box b “above” = numporters = porter (box b “beside”) “above”
    ™ add boxes ahead to queue and continue
    boxesToDo:= boxesToDo; [box b “ahead left”; d+1×m].
    if box b “ahead left” ≠ box b “ahead right”
        [[boxesToDo:= boxesToDo; [box b “ahead right”; d+1×m]].
    nextBox
else [if ↔ boxesToDo > 0 [nextBox]]].
old boxesToDo.

` using accel, move porter p ahead one box
new b: point box:= porter p “below”.
box:=(b; “porter”) → numporters \ box. draw b “grey”.
randNatNext.
b:= box b (randNat 0 2 = 0 => “ahead left” = “ahead right”).
if box b “porter” < numporters [draw b “red”. stop] ™ crash
    porter:= (p; “below”) → b \ porter. box:=(b; “above”) → p \ box. draw b “blue”.
old b.
new speed:= sqrt (porter p “speed”^2 + 2×accel×m) × speedlimit.
porter:= (p; “arrival time”) → porter p “arrival time” + 2×m/(porter p “speed” + speed)
    \ (p; “speed”) → speed
    \ porter.

old speed. old accel. old p. ™ these olds aren't really necessary

await (time??+visualDelayTime).
infiniteLoop ]] ™ end of simport

Quote Notation Lengths

™ program to compare quote notation lengths with numerator/denominator lengths

™ output: screen
™ use: bin div even nat numtext odd

new shl = ⟨n: nat → ⟨m: nat → ` shift n left m places; n×2^m
result r: nat:= n [for i:= 0..m [r:= r×2]]⟩⟩.

new shr = ⟨n: nat → ⟨m: nat → ` shift n right m places; floor (n×2^m–m) or div n (2^m)
result r: nat:= n [for i:= 0..m [r:= div r 2]]⟩⟩.

new gcd = ⟨a: (nat+1) → ⟨b: (nat+1) → ` greatest common divisor of a and b
a=b ⊨ a = a<b ⊨ gcd a (b–a) ⊨ gcd (a–b) b⟩⟩.

new norm [plan num:: nat+1 [plan denom:: nat+1 \ normalize num/denom
    [new g:= gcd num denom. num:= num/g. denom:= denom/g]]]]

new count: nat:= 0. ™ number of examples
new qlen: nat := 0. ` total length of quote representations
new rlen: nat := 0. ` total length of numerator/denominator representations

for length := 1..15
  [for string := 0;..(shl 1 length) ` each string of that length
  [for quote := 0;..length ` each quote position (at least one bit to left of quote)
    [if even (shr string (length–1)) ≠ even (shr string (quote–1)) ` roll-normalized
      result repeatnorm: bin := ⊤
      [new len: nat := div (length–quote) 2. ` the length of the possibly repeating part
        trythislen [if len>0 ` 1 ≤ len ≤ (length–quote)/2
          [new extract = \i: nat → \l: nat → \index \ i \ length \ l
            shr string i – shl (shr string (i+l)) l \}.
          new ex := extract quote len.
          if ` the negative part is a repetition (twice or more) of ex
            result r: bin := ⊤
            [new i: nat := quote+len. ` i+len ≤ length
              iloop [new ey := extract i len.
                if ex=ey [i := i+len. ` i≤length
                  if i+len ≤ length [iloop]
                else [r := ⊥]]]
            else [r := ⊥]]
        else [repeatnorm := ⊥ else [len := len–1. trythislen]]]
  else [for point := 0;..length+1 ` each point position (right end, interior, left end)
    [if ` the rightmost bit is 1 or it's to the left of quote or point
      odd string v (quote=0) v (point=0)
    [` convert to numerator/denominator
      if num<0 [num := –num].
      new denom: nat := shl (shl 1 (length–quote) – 1) point.
      norm num denom.
      ` update statistics
      count := count+1. qlen := qlen+length.
      rlen := rlen+1. ` for the sign
      loop [num := div num 2. rlen := rlen+1.
        if num>0 [loop]].
      loop [denom := div denom 2. rlen := rlen+1.
        if denom>0 [loop]]]]]]]
  if denom>0 [loop]]))))]]]]]

! “In ”; count; “ examples, quote average length = ”;
qlen/count; “, num/denom average length = ”; rlen/count.

old shl. old shr. old gcd. old norm. old count. old qlen. old rlen

Huffman Codes

new Huffman ` a program to compute Huffman minimum redundancy prefix codes
[` input: keys
 ` output: screen
 ` use: back find nat nil nl point text textnum
new tree = [text], [tree; tree]. ` a binary tree with texts at the leaves
new forest: *
[nat; tree]= nil. ` the data structure. A string of trees, with a frequency for each tree

inputstart
[! “Enter a frequency, then a colon, then a message, then a new line, and repeat. ”;
“To end, just enter a new line.”; nl.
readloop
[?]!
if ↔?? = 0 ` Just new line was pressed.
[if ↔forest = 0 ` We haven't had any input yet. We need at least one
[if “Insufficient input. Try again.”. inputstart].
new c:= find “:” ??.
if c = ↔?? ![ “Bad format: no colon. Try again.”. readloop ].
new freq:= textnum (??(0;..c)).
if freq = “error” ![ “Bad frequency format. Try again.”. readloop ].
new message:= ??(c;..↔??).
` find where the new data goes in forest and put it there.
new i:= nat:= 0.
findloop ![if i = ↔forest ∨ (freq ≤ (forest\i)0) ` found where it goes
[forest:= forest\0;..i); [freq; [message]]; forest\i;..↔forest]\]
else ![i:= i+1. findloop]]. readloop]]).

` forest is now a nonempty string of pairs, each pair consisting of a frequency and a tree, each
` tree is a single leaf, each leaf is a list-text. They are in non-decreasing frequency order.
` For example: [3; [“a”]]; [4; [“b”]]; [9; [“c”]]; [12; [“d”]]; [15; [“e”]]; [20; [“f”]]

new here: nat:= 0. ` A new tree must be moved to position here or later.
loop ![if ↔forest ≥ 2
[` combine the first two trees into a new tree t
new t:= [(forest\0)0 + (forest\1)0; [(forest\0)1 ; (forest\1)1]].
` remove those first two trees from the forest
forest:= forest\2;..↔forest).
` put tree t into its place in the forest
innerloop ![if here = ↔forest ∨ (t 0 < (forest\here)0) ` we've found where it goes
[forest:= forest\0;..here); t; forest\here;..↔forest]. loop]
else ![here:= here+1. innerloop]]].

` forest is now a single pair consisting of the total of all frequencies and a code tree.
new t:= forest\1. ` the code tree
` Walk the tree, depth-first, printing leaves and their codes
new p: point t:= nil. ` a path within t starting at the root
new pt: text:= “”. ` same path as p but as a text for printing
loop ![if ~(*p: text) ` we are at a leaf
[! “code:”; pt: “; message:”; ~(*p: nl)]
else ![p:= p;0. pt:= pt;“0”; loop. p:= back p. pt:= back pt.
**Read Password**

` program to read a password, allowing corrections, displaying blobs
`input: keys
`output: screen
`use: char delete nl text textid
**new password: text: “”.**

```plaintext
pswd [!“Please enter password followed by return: ”].
read [?] char.
   if ??=nl [if password=“” [!“Empty password. Try again.”; nl. pswd]]
   else [!nl]]
   else [if ??=delete [if password≠“” [password:= password(0;..↔password–1). !delete]]
   else [password:= password; ??. !“•”].
read]]].

**new identity:= textid password. old password**
```

**Grammars**

**LL(1) Grammar**

In this grammar, for each nonterminal, every production except possibly the last begins with a different terminal. So director sets are not needed, and that's a special case of LL(1) that deserves its own name; perhaps LL(1/2). To parse a program, the parse stack begins with only the program nonterminal on it, and ends empty with no more input. However, ProTem functions as an operating system, parsing and executing each sequent in turn. So the parse stack begins with sequent on top, and . below it. When the stack is empty, the sequent is executed, the parse stack is reinitialized, and parsing resumes. A name control program is responsible for classifying names. For efficiency, the productions (except possibly the last) for each nonterminal should be placed in order of frequency. The following nonterminals can be eliminated by replacing them with their one production: program sequent data data6 data5 data4 data3 data1. This leaves the grammar with 31–8 = 23 nonterminals.

```plaintext
program sequent moreprogram
moreprogram . program empty
sequent phrase moresequent
moresequent || sequent empty
phrase new simplisticname afternewname
  old simplisticname compounder
  [ program ] arguments
  if data [ program ] elsepart
  case data [ program ] elsepart
for simplisticname := data [ program ]
plan simplisticname parameterkind data [ program ] arguments
! data
? inputafterq
```
simplename aftersimplename

afternewname : data := data
= data
:= data
? data ! data
[ program ]
_ afterunderscore
#1
simplename compounder
empty

afterunderscore simplename afternewname
empty

compounder _ simplename compounder
empty

elsepart else [ program ]
empty

parameterkind :
::!
?

aftersimplename [ program ]
compounder aftername

aftername := data
! data
! inputafterq
arguments

inputafterq ! echo
data afterpattern

afterpattern ! echo
empty

echo simplename compounder
empty

arguments number arguments
∞ arguments
text arguments
⊤ arguments
⊥ arguments
result simplename : data := data [ program ] arguments
{ data } arguments
[ data ] arguments
( data ) arguments
〈 simplename : data0 → data 〉 arguments
simplename compounder arguments
?? arguments
!! arguments
empty

data
  data6 moredata

moredata
  ⊨ data = data6 moredata
  empty

data6
  data5 moredata6

moredata6
  = data5 moredata6
  ≠ data5 moredata6
  < data5 moredata6
  > data5 moredata6
  ≤ data5 moredata6
  ≥ data5 moredata6
  : data5 moredata6
  empty

data5
  data4 moredata5

moredata5
  , data4 moredata5
  .. data4 moredata5
  | data4 moredata5
  ⊨ data = data4 moredata5
  empty

data4
  data3 moredata4

moredata4
  + data3 moredata4
  − data3 moredata4
  ;; data3 moredata4
  ; data3 moredata4
  .. data3 moredata4
  ` data3 moredata4
  empty

data3
  data2 moredata3

moredata3
  × data2 moredata3
  / data2 moredata3
  ∧ data2 moredata3
  ∨ data2 moredata3
  empty
LR(0) Grammar

The following grammar has no reduce-reduce choices and no shift-reduce choices. It has shift-shift choices. Such a grammar is commonly called LR(0), but it shouldn't be, because a shift action pushes an input symbol onto the parse stack, and therefore a shift action depends on the input symbol. It is a special case of LR(1) that deserves its own name, but not LR(0); perhaps LR(1/2).

To parse a program, the parse stack begins empty, and ends with only the program nonterminal on it and no more input. However, ProTem functions as an operating system, parsing and executing each sequent in turn. So the parse stack begins empty, and ends with . on top and sequent below it. The sequent is executed, the parse stack is reinitialized, and parsing resumes. A name control program is responsible for classifying names.
program sequent
  program . sequent

sequent phrase
  sequent || phrase

phrase
  new name : data := data
  new name := data
  new name = data
  new name [ program ]
  new name ? data ! data
  new name #1
  new name _
  new name name
  new name old name
  name := data
  name ! data!
  name ? data
  ? data
  name ? data ! name
  ? data ! name
  name ? data !
  ? data!
  name ? ! name
  ? ! name
  name ? !
  ? !
  simplename [ program ]
  if data [ program ]
  if data [ program ] else [ program ]
  case data [ program ]
  case data [ program ] else [ program ]
  for simplename := data [ program ]
  [ program ]
  plan

plan
  plan simplename : data [ program ]
  plan simplename :: data [ program ]
  plan simplename ? data [ program ]
  plan simplename ! data [ program ]
  plan data0
  name

data
  data6 ⊨ data = data
  data6

data6
  data6 = data5
data6 ≠ data5
data6 < data5
data6 > data5
data6 ≤ data5
data6 ≥ data5
data6 : data5
data5

data5 , data4
data5 .. data4
data5 | data4
data5 <= data > data4
data4

data4 ; data3
data4 .. data3
data4 ;; data3
data4 * data3
data4 + data3
data4 – data3
data3

data3 x data2
data3 / data2
data3 \ data2
data3 \ data2

data2 + data2
– data2
e data2
⇔ data2
# data2
~ data2
□ data2
\ data2
* data2
data1 * data2
data1 \ data2
data1 \ data2
data1 \ data2
data1 \ data2
data1 ^ data2
data1 \ data2

data1 data0
data1 @ data0
data1 %
data1 & data0
name ??
name !!
data0

data0   number
      ∞
text
⊤
⊥
[ data ]
{ data }
( data )
〈 simplename : data0 → data 〉
result simplename : data := data [ program ]
??
!!
name

name   simplename
name _ simplename

Acknowledgements

The first public mention of ProTem was


Theo Norvell wrote an MSc thesis in 1988 titled “Expressions, Types, and Data Structures in ProTem”. Hugh Redelmeier acted as design consultant and critic in 1990. Brian Parkinson found a bug in the implementation in 1990. The design of ProTem has been improved since then, and the old implementation is now out-of-date. A new implementation is partly written.