

The PREV Language Specification

Boštjan Slivnik

August 25, 2016

1 Lexical structure

The programs in the PREV programming language are written in 7-bit ASCII character set (all other characters are invalid). A single character LF denotes the end of line regardless of the text file format.

The (groups of) symbols of the PREV programming language are:

- **Symbols:**

`+ & = : , }]) . / == >= > <= < @ % * != ! { [(| - ^`

- **Constants:**

- *Boolean constants:* `true false`
- *Integer constants:* An integer constant is a sequence of decimal digits optionally prefixed by a sign, i.e., “+” or “-”, denoting a 64-bit signed integer, i.e., from the interval $[-2^{63}, 2^{63} - 1]$.
- *Character constants:* A character constant consists of a single character name within single quotes. A character name is either a character with an ASCII code from the interval $[32, 126]$ (but not a backslash, a single or a double quote) or an escape sequence. An escape sequence starts with a backslash character followed by a backslash (denoting a backslash), a single quote (denoting a single quote), a double quote (denoting a double quote), “`\t`” (denoting TAB), or “`\n`” (denoting LF).
- *String constants:* A string constant is a possibly empty finite sequence of character names within double quotes. A character name is either a character with an ASCII code from the interval $[32, 126]$ (but not a backslash, a single or a double quote) or an escape sequence. An escape sequence starts with a backslash character followed by a backslash (denoting a backslash), a single quote (denoting a single quote), a double quote (denoting a double quote), “`\t`” (denoting TAB), or “`\n`” (denoting LF).
- *Pointer constant:* `null`
- *Void constant:* `none`

- **Type names:**

`boolean integer char string void`

- **Keywords:**

`arr else end for fun if then ptr rec typ var where while`

- **Identifiers:** An identifier is a nonempty finite sequence of letters, digits and underscores that starts with a letter or an underscore.

Additionally, the source might include the following:

- **White space:** Characters CR, LF, TAB, or space.
- **Comments:** A comment is a sequence of character that starts with an octothorpe character, i.e., “#”, and ends with the LF character (regardless of the text file format).

To break the source file into individual symbols, the first-match-longest-match rule must be used.

2 Syntactic structure

The concrete syntax of the PREV programming language is defined by an LR(1) grammar. Nonterminal and terminal symbols are written in *italic* and typewriter fonts, respectively. Terminal symbols IDENTIFIER, INTEGER, BOOLEAN, CHAR and STRING denote (all) identifiers, integer constants, boolean constants, character constants and string constants, respectively. The start symbol of the grammar is *Program*. The LR(1) grammar contains the following productions:

Program \rightarrow *Expression*
Expression \rightarrow *AssignmentExpression*
Expression \rightarrow *Expression* where *Declarations* end
Expressions \rightarrow *Expression*
Expressions \rightarrow *Expressions* , *Expression*
AssignmentExpression \rightarrow *DisjunctiveExpression*
AssignmentExpression \rightarrow *DisjunctiveExpression* = *DisjunctiveExpression*
DisjunctiveExpression \rightarrow *ConjunctiveExpression*
DisjunctiveExpression \rightarrow *DisjunctiveExpression* | *ConjunctiveExpression*
ConjunctiveExpression \rightarrow *RelationalExpression*
ConjunctiveExpression \rightarrow *ConjunctiveExpression* & *RelationalExpression*
RelationalExpression \rightarrow *AdditiveExpression*
RelationalExpression \rightarrow *AdditiveExpression* == *AdditiveExpression*
RelationalExpression \rightarrow *AdditiveExpression* != *AdditiveExpression*
RelationalExpression \rightarrow *AdditiveExpression* < *AdditiveExpression*
RelationalExpression \rightarrow *AdditiveExpression* > *AdditiveExpression*
RelationalExpression \rightarrow *AdditiveExpression* <= *AdditiveExpression*
RelationalExpression \rightarrow *AdditiveExpression* >= *AdditiveExpression*
AdditiveExpression \rightarrow *AdditiveExpression* + *MultiplicativeExpression*
AdditiveExpression \rightarrow *AdditiveExpression* - *MultiplicativeExpression*
AdditiveExpression \rightarrow *MultiplicativeExpression*
MultiplicativeExpression \rightarrow *MultiplicativeExpression* * *PrefixExpression*
MultiplicativeExpression \rightarrow *MultiplicativeExpression* / *PrefixExpression*
MultiplicativeExpression \rightarrow *MultiplicativeExpression* % *PrefixExpression*
MultiplicativeExpression \rightarrow *PrefixExpression*
PrefixExpression \rightarrow *PostfixExpression*
PrefixExpression \rightarrow + *PrefixExpression*
PrefixExpression \rightarrow - *PrefixExpression*
PrefixExpression \rightarrow ! *PrefixExpression*
PrefixExpression \rightarrow @ *PrefixExpression*
PrefixExpression \rightarrow [*Type*] *PrefixExpression*
PostfixExpression \rightarrow *AtomicExpression*
PostfixExpression \rightarrow *PostfixExpression* [*Expression*]
PostfixExpression \rightarrow *PostfixExpression* . IDENTIFIER
PostfixExpression \rightarrow *PostfixExpression* ^
AtomicExpression \rightarrow INTEGER

AtomicExpression \rightarrow BOOLEAN
AtomicExpression \rightarrow CHAR
AtomicExpression \rightarrow STRING
AtomicExpression \rightarrow null
AtomicExpression \rightarrow none
AtomicExpression \rightarrow IDENTIFIER *ArgumentsOpt*
AtomicExpression \rightarrow (*Expressions*)
AtomicExpression \rightarrow if *Expression* then *Expression* else *Expression* end
AtomicExpression \rightarrow for IDENTIFIER = *Expression* , *Expression* : *Expression* end
AtomicExpression \rightarrow while *Expression* : *Expression* end
ArgumentsOpt \rightarrow (
ArgumentsOpt \rightarrow erm
ArgumentsOpt \rightarrow (*Expressions*)
Declarations \rightarrow *Declaration*
Declarations \rightarrow *Declarations Declaration*
Declaration \rightarrow *TypeDeclaration*
Declaration \rightarrow *FunctionDeclaration*
Declaration \rightarrow *VariableDeclaration*
TypeDeclaration \rightarrow typ IDENTIFIER : *Type*
FunctionDeclaration \rightarrow fun IDENTIFIER (*ParametersOpt*) : *Type* *FunctionBodyOpt*
ParametersOpt \rightarrow erm
ParametersOpt \rightarrow *Parameters*
Parameters \rightarrow *Parameter*
Parameters \rightarrow *Parameters* , *Parameter*
Parameter \rightarrow IDENTIFIER : *Type*
FunctionBodyOpt \rightarrow erm
FunctionBodyOpt \rightarrow = *Expression*
VariableDeclaration \rightarrow var IDENTIFIER : *Type*
Type \rightarrow integer
Type \rightarrow boolean
Type \rightarrow char
Type \rightarrow string
Type \rightarrow void
Type \rightarrow arr [*Expression*] *Type*
Type \rightarrow rec { *Components* }
Type \rightarrow ptr *Type*
Type \rightarrow IDENTIFIER
Components \rightarrow *Component*
Components \rightarrow *Components* , *Component*
Component \rightarrow IDENTIFIER : *Type*

Note that the LR(1) grammar generates certain sentential forms which are prohibited by semantics.

3 Semantic rules

3.1 Namespaces

1. Names of types, functions, variables and parameters belong to one single global namespace.
2. Names of record components belong to record-specific namespaces, i.e., each record defines its own namespace containing names of its components.

3.2 Scope

1. where-expression

expr where *decl*₁ *decl*₂ ... *decl*_n end

creates a new scope:

- subexpression *expr* and declarations *decl*_{*i*}, for *i* = 1, 2, ..., *n*, belong to the new scope;
- names declared by declarations *decl*_{*i*}, for *i* = 1, 2, ..., *n*, are visible in the scope (unless they are redeclared within inner scopes).

2. Function declaration

fun ID (ID₁:*type*₁, ID₂:*type*₂, ..., ID_n:*type*_n) : *type* (= *expr*)_{opt}

creates a new scope:

- Function name ID, function type *type* and parameter types *type*_{*i*}, for *i* = 1, 2, ..., *n*, do not belong to the new scope.
- Parameter names ID_{*i*}, for *i* = 1, 2, ..., *n*, and function body *expr* (if specified) belong to the new scope and are visible in the entire new scope.

3.3 Constant integer expressions

Let $I = \{(-2^{63}) \dots (+2^{63} - 1)\}$. Semantic function

$$\llbracket \cdot \rrbracket_{\text{VAL}}: \mathcal{P} \rightarrow I$$

maps phrases of PREV to the integer values they denote. It is defined by the following rules:

$$\frac{\text{lexeme}(\text{INTEGER}) \in I}{\llbracket \text{INTEGER} \rrbracket_{\text{VAL}} = \text{lexeme}(\text{INTEGER})}$$

$$\frac{\llbracket \text{expr} \rrbracket_{\text{VAL}} = n \quad \text{op} \in \{+, -\}}{\llbracket \text{op expr} \rrbracket_{\text{VAL}} = \text{op } n} \quad \frac{\llbracket \text{expr}_1 \rrbracket_{\text{VAL}} = n_1 \quad \llbracket \text{expr}_2 \rrbracket_{\text{VAL}} = n_2 \quad \text{op} \in \{+, -, *, /, \%\}}{\llbracket \text{expr}_1 \text{ op expr}_2 \rrbracket_{\text{VAL}} = n_1 \text{ op } n_2}$$

In all other cases the value of $\llbracket \cdot \rrbracket_{\text{VAL}}$ is undefined (denoted by \perp).

3.4 Type system

A set

$$\begin{aligned} T = & \{ \text{boolean}, \text{integer}, \text{char}, \text{string}, \text{void} \} && \text{(atomic types)} \\ & \cup \{ \text{arr}(n \times \tau) \mid n \in I \wedge \tau \in T \} && \text{(arrays)} \\ & \cup \{ \text{rec}(\tau_1, \tau_2, \dots, \tau_n) \mid n > 0 \wedge \tau_1, \tau_2, \dots, \tau_n \in T \} && \text{(records)} \\ & \cup \{ \text{ptr}(\tau) \mid \tau \in T \} && \text{(pointers)} \\ & \cup \{ (\tau_1, \tau_2, \dots, \tau_n) \rightarrow \tau \mid n \geq 0 \wedge \tau_1, \tau_2, \dots, \tau_n \in T \} && \text{(functions)} \end{aligned}$$

denotes a set of all types of PREV. Two types are equal if they share the same structure.

Semantic function $\llbracket \cdot \rrbracket_{\text{TYP}}$

$$\llbracket \cdot \rrbracket_{\text{TYP}}: \mathcal{P} \rightarrow T$$

maps phrases of PREV to types.

Let function DECL maps a type, function, variable or parameter name to its declaration according to the namespace and scope rules specified above and let DECL _{τ} map a component name to its declaration within (a record) type τ . If a name is undefined in the current namespace or scope, functions DECL and DECL _{τ} return \perp . The semantic function $\llbracket \cdot \rrbracket_{\text{TYP}}$ is defined by the rules in the following subsections. In cases when no rule can be applied, the semantic function $\llbracket \cdot \rrbracket_{\text{TYP}}$ is left undefined (denoted by \perp).

3.4.1 Type expressions:

$$\begin{array}{c}
\frac{}{\llbracket \text{boolean} \rrbracket_{\text{TYP}} = \text{boolean}} \quad \frac{}{\llbracket \text{integer} \rrbracket_{\text{TYP}} = \text{integer}} \\
\\
\frac{}{\llbracket \text{char} \rrbracket_{\text{TYP}} = \text{char}} \quad \frac{}{\llbracket \text{string} \rrbracket_{\text{TYP}} = \text{string}} \quad \frac{}{\llbracket \text{void} \rrbracket_{\text{TYP}} = \text{void}} \\
\\
\frac{\llbracket \text{expr} \rrbracket_{\text{VAL}} = n \quad n > 0 \quad \llbracket \text{type} \rrbracket_{\text{TYP}} = \tau}{\llbracket \text{arr } [\text{expr}] \text{ type} \rrbracket_{\text{TYP}} = \text{arr}(n \times \tau)} \quad \frac{\llbracket \text{type} \rrbracket_{\text{TYP}} = \tau}{\llbracket \text{ptr type} \rrbracket_{\text{TYP}} = \text{ptr}(\tau)} \\
\\
\frac{\forall i: \llbracket \text{type}_i \rrbracket_{\text{TYP}} = \tau_i}{\llbracket \text{rec } \{ \text{name}_1 : \text{type}_1, \text{name}_2 : \text{type}_2, \dots, \text{name}_n : \text{type}_n \} \rrbracket_{\text{TYP}} = \text{rec}(\tau_1, \tau_2, \dots, \tau_n)} \\
\\
\frac{\llbracket \text{DECL}(\text{typ-name}) \rrbracket_{\text{TYP}} = \tau}{\llbracket \text{typ-name} \rrbracket_{\text{TYP}} = \tau}
\end{array}$$

3.4.2 Value expressions:

$$\begin{array}{c}
\frac{}{\llbracket \text{BOOLEAN} \rrbracket_{\text{TYP}} = \text{boolean}} \quad \frac{}{\llbracket \text{INTEGER} \rrbracket_{\text{TYP}} = \text{integer}} \\
\\
\frac{}{\llbracket \text{CHAR} \rrbracket_{\text{TYP}} = \text{char}} \quad \frac{}{\llbracket \text{STRING} \rrbracket_{\text{TYP}} = \text{string}} \\
\\
\frac{}{\llbracket \text{none} \rrbracket_{\text{TYP}} = \text{void}} \quad \frac{}{\llbracket \text{null} \rrbracket_{\text{TYP}} = \text{ptr}(\text{void})} \\
\\
\frac{\llbracket \text{expr} \rrbracket_{\text{TYP}} = \text{boolean}}{\llbracket ! \text{expr} \rrbracket_{\text{TYP}} = \text{boolean}} \quad \frac{\llbracket \text{expr} \rrbracket_{\text{TYP}} = \text{integer} \quad \text{op} \in \{+, -\}}{\llbracket \text{op expr} \rrbracket_{\text{TYP}} = \text{integer}} \\
\\
\frac{\llbracket \text{expr}_1 \rrbracket_{\text{TYP}} = \text{integer} \quad \llbracket \text{expr}_2 \rrbracket_{\text{TYP}} = \text{integer} \quad \text{op} \in \{+, -, *, /, \%\}}{\llbracket \text{expr}_1 \text{ op } \text{expr}_2 \rrbracket_{\text{TYP}} = \text{integer}} \\
\\
\frac{\llbracket \text{expr}_1 \rrbracket_{\text{TYP}} = \text{boolean} \quad \llbracket \text{expr}_2 \rrbracket_{\text{TYP}} = \text{boolean} \quad \text{op} \in \{\&, |\}}{\llbracket \text{expr}_1 \text{ op } \text{expr}_2 \rrbracket_{\text{TYP}} = \text{boolean}} \\
\\
\frac{\llbracket \text{expr}_1 \rrbracket_{\text{TYP}} = \tau_1 \quad \llbracket \text{expr}_2 \rrbracket_{\text{TYP}} = \tau_2 \quad \text{op} \in \{=, !=, <, >, <=, >=\} \quad \tau_1, \tau_2 \in \{\text{boolean}, \text{integer}, \text{char}\} \cup \{\text{ptr}(\tau) \mid \tau \in \mathbf{T}\} \quad \tau_1 = \tau_2}{\llbracket \text{expr}_1 \text{ op } \text{expr}_2 \rrbracket_{\text{TYP}} = \text{boolean}} \\
\\
\frac{\llbracket \text{expr}_1 \rrbracket_{\text{TYP}} = \tau_1 \quad \llbracket \text{expr}_2 \rrbracket_{\text{TYP}} = \tau_2 \quad \llbracket \text{expr}_1 \rrbracket_{\text{MEM}} = \text{true} \quad \tau_1, \tau_2 \in \{\text{boolean}, \text{integer}, \text{char}, \text{string}\} \cup \{\text{ptr}(\tau) \mid \tau \in \mathbf{T}\} \quad \tau_1 = \tau_2}{\llbracket \text{expr}_1 = \text{expr}_2 \rrbracket_{\text{TYP}} = \text{void}} \\
\\
\frac{\llbracket \text{DECL}(\text{var-name}) \rrbracket_{\text{TYP}} = \tau}{\llbracket \text{var-name} \rrbracket_{\text{TYP}} = \tau} \quad \frac{\llbracket \text{DECL}(\text{par-name}) \rrbracket_{\text{TYP}} = \tau}{\llbracket \text{par-name} \rrbracket_{\text{TYP}} = \tau} \\
\\
\frac{\forall i: \llbracket \text{expr}_i \rrbracket_{\text{TYP}} = \tau_i \quad \llbracket \text{DECL}(\text{fun-name}) \rrbracket_{\text{TYP}} = (\tau_1, \tau_2, \dots, \tau_n) \rightarrow \tau}{\llbracket \text{fun-name } (\text{expr}_1, \text{expr}_2, \dots, \text{expr}_n) \rrbracket_{\text{TYP}} = \tau}
\end{array}$$

$$\begin{array}{c}
\frac{\llbracket expr \rrbracket_{\text{typ}} = \tau \quad \llbracket \text{DECL}_{\tau}(comp\text{-}name) \rrbracket_{\text{typ}} = \tau'}{\llbracket expr.comp\text{-}name \rrbracket_{\text{typ}} = \tau'} \\
\\
\frac{\llbracket expr_1 \rrbracket_{\text{typ}} = \text{arr}(n \times \tau) \quad \llbracket expr_2 \rrbracket_{\text{typ}} = \text{integer}}{\llbracket expr_1[expr_2] \rrbracket_{\text{typ}} = \tau} \\
\\
\frac{\llbracket expr \rrbracket_{\text{typ}} = \tau \quad \llbracket expr \rrbracket_{\text{mem}} = \text{true}}{\llbracket @expr \rrbracket_{\text{typ}} = \text{ptr}(\tau)} \quad \frac{\llbracket expr \rrbracket_{\text{typ}} = \text{ptr}(\tau)}{\llbracket expr^{\wedge} \rrbracket_{\text{typ}} = \tau} \\
\\
\frac{\llbracket type \rrbracket_{\text{typ}} = \text{ptr}(\tau) \quad \llbracket expr \rrbracket_{\text{typ}} = \text{ptr}(\text{void})}{\llbracket [type] expr \rrbracket_{\text{typ}} = \text{ptr}(\tau)} \\
\\
\frac{\forall i: \llbracket expr_i \rrbracket_{\text{typ}} \neq \perp \quad \llbracket expr_n \rrbracket_{\text{typ}} = \tau}{\llbracket (expr_1, expr_2, \dots, expr_n) \rrbracket_{\text{typ}} = \tau} \\
\\
\frac{\llbracket cond\text{-}expr \rrbracket_{\text{typ}} = \text{boolean} \quad \llbracket then\text{-}body \rrbracket_{\text{typ}} \neq \perp \quad \llbracket else\text{-}body \rrbracket_{\text{typ}} \neq \perp}{\llbracket \text{if } cond\text{-}expr \text{ then } then\text{-}body \text{ else } else\text{-}body \text{ end} \rrbracket_{\text{typ}} = \text{void}} \\
\\
\frac{\llbracket var\text{-}name \rrbracket_{\text{typ}} = \text{integer} \quad \llbracket lo\text{-}expr \rrbracket_{\text{typ}} = \text{integer} \quad \llbracket hi\text{-}expr \rrbracket_{\text{typ}} = \text{integer} \quad \llbracket body \rrbracket_{\text{typ}} \neq \perp}{\llbracket \text{for } var\text{-}name : lo\text{-}expr, hi\text{-}expr : body \text{ end} \rrbracket_{\text{typ}} = \text{void}} \\
\\
\frac{\llbracket cond\text{-}expr \rrbracket_{\text{typ}} = \text{boolean} \quad \llbracket body \rrbracket_{\text{typ}} \neq \perp}{\llbracket \text{while } cond\text{-}expr : body \text{ end} \rrbracket_{\text{typ}} = \text{void}} \\
\\
\frac{\forall i: \llbracket decl_i \rrbracket_{\text{typ}} \neq \perp \quad \llbracket expr \rrbracket_{\text{typ}} = \tau}{\llbracket expr \text{ where } decl_1, decl_2, \dots, decl_n \text{ end} \rrbracket_{\text{typ}} = \tau}
\end{array}$$

3.4.3 Declarations:

$$\begin{array}{c}
\frac{\llbracket type \rrbracket_{\text{typ}} = \tau}{\llbracket \text{typ } typ\text{-}name : type \rrbracket_{\text{typ}} = \tau} \\
\\
\frac{\forall i: \llbracket decl_i \rrbracket_{\text{typ}} = \tau_i \quad \forall i: \tau_i \in \{\text{boolean}, \text{integer}, \text{char}, \text{string}\} \cup \{\text{ptr}(\tau) \mid \tau \in \mathbf{T}\} \quad \llbracket type \rrbracket_{\text{typ}} = \tau \quad \tau \in \{\text{boolean}, \text{integer}, \text{char}, \text{string}, \text{void}\} \cup \{\text{ptr}(\tau) \mid \tau \in \mathbf{T}\} \quad \llbracket expr \rrbracket_{\text{typ}} = \tau}{\llbracket \text{fun } fun\text{-}name(decl_1, decl_2, \dots, decl_n) : type (= expr)_{\text{opt}} \rrbracket_{\text{typ}} = (\tau_1, \tau_2, \dots, \tau_n) \rightarrow \tau} \\
\\
\frac{\llbracket type \rrbracket_{\text{typ}} = \tau}{\llbracket \text{var } var\text{-}name : type \rrbracket_{\text{typ}} = \tau} \quad \frac{\llbracket type \rrbracket_{\text{typ}} = \tau}{\llbracket par\text{-}name : type \rrbracket_{\text{typ}} = \tau} \quad \frac{\llbracket type \rrbracket_{\text{typ}} = \tau}{\llbracket comp\text{-}name : type \rrbracket_{\text{typ}} = \tau}
\end{array}$$

3.5 Memory objects

Semantic function $\llbracket \cdot \rrbracket_{\text{mem}}$

$$\llbracket \cdot \rrbracket_{\text{mem}}: \mathcal{P} \rightarrow \{\text{true}, \text{false}\}$$

maps phrases of PREV to boolean values: it maps a phrase to **true** if and only if it is a value expression and denotes an object in memory. It is defined by the following rules:

$$\begin{array}{c}
\frac{\llbracket \text{DECL}(\text{var-name}) \rrbracket_{\text{TYP}} = \tau}{\llbracket \text{var-name} \rrbracket_{\text{MEM}} = \text{true}} \quad \frac{\llbracket \text{DECL}(\text{par-name}) \rrbracket_{\text{TYP}} = \tau}{\llbracket \text{par-name} \rrbracket_{\text{MEM}} = \text{true}} \\
\\
\frac{\llbracket \text{expr} \rrbracket_{\text{TYP}} = \tau \quad \llbracket \text{DECL}_{\tau}(\text{comp-name}) \rrbracket_{\text{TYP}} = \tau'}{\llbracket \text{expr} . \text{comp-name} \rrbracket_{\text{MEM}} = \text{true}} \\
\\
\frac{\llbracket \text{expr}_1 \rrbracket_{\text{TYP}} = \text{arr}(n \times \tau) \quad \llbracket \text{expr}_2 \rrbracket_{\text{TYP}} = \text{integer}}{\llbracket \text{expr}_1 [\text{expr}_2] \rrbracket_{\text{MEM}} = \text{true}} \\
\\
\frac{\llbracket \text{expr} \rrbracket_{\text{TYP}} = \text{ptr}(\tau)}{\llbracket \text{expr}^{\wedge} \rrbracket_{\text{MEM}} = \text{true}}
\end{array}$$

In cases when no rule can be applied, the value of semantic function $\llbracket \cdot \rrbracket_{\text{MEM}}$ is **false**.