The PREV'19 programming language

(academic year 2018/19)

Boštjan Slivnik

1 Lexical structure

Programs in the PREV'19 programming language are written in ASCII character set (no additional characters denoting post-alveolar consonants are allowed).

Programs in the PREV'19 programming language consist of the following lexical elements:

- Literals:
 - literals of type void: none
 - literals of type bool: true false
 - literals of type char:

An character with a character code in decimal range $\{32...126\}$ (from space to $\tilde{}$) enclosed in single quotes ($\dot{}$).

- literals of type int:

A nonempty finite string of digits (0...9) optionally preceded by a sign (+ or -).

- literals of pointer types: null
- string literals:

A possibly empty finite string of characters with character codes in decimal range $\{32...126\} \setminus \{34\}$ (from space to "but excluding double quote) enclosed in double quotes (").

• Symbols:

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

• Keywords:

arr bool char del do else end fun if int new ptr rec then typ var void where while

• Identifiers:

A nonempty finite string of letters (A...Z and a...z), digits (0...9), and underscores $(_)$ that (a) starts with either a letter or an underscore and (b) is not a keyword or a literal.

• Comments:

A string of characters starting with a hash (#) and extending to the end of line.

• White space:

Space, horizontal tab (HT), line feed (LF) and carriage return (CR). Line feed alone denotes the end of line within a source file. Horizontal tab is 8 spaces wide.

Lexical elements should be recognised from left to right using the longest match approach.

2 Syntax structure

The concrete syntax of the PREV programming language is defined by context free grammar with the start symbol source and productions

```
source \longrightarrow decl \{ decl \}
(program)
(type declaration)
                                       \longrightarrow typ identifier: type;
(variable declaration)
                              decl
                                       \longrightarrow var identifier: type;
                                       \longrightarrow fun identifier ([identifier: type {, identifer: type }]): type [=expr];
(function declaration)
                              decl
(atomic type)
                                       \longrightarrow void | bool | char | int
                              type
                                       \longrightarrow arr [expr] type
(array type)
                              type
                                       \longrightarrow rec (identifier: type { , identifier: type})
(record type)
                              type
(pointer type)
                                       \longrightarrow \mathtt{ptr}\ type
                              type
                                       \longrightarrow identifier
(named type)
                              type
                                       \longrightarrow (type)
(enclosed type)
                              type
(literal)
                                       \longrightarrow literal
                              expr
                                       \longrightarrow unop\ expr
(unary expression)
                              expr
(binary expression)
                                       \longrightarrow expr\ binop\ expr
                              expr
                                       \longrightarrow identifier
(variable access)
                              expr
                                       \longrightarrow identifier([expr\{,expr\}])
(function call)
                              expr
                                       \longrightarrow expr[expr]
(element access)
                              expr
(component access)
                              expr
                                       \longrightarrow expr.identifier
                                       \longrightarrow new (type)
(memory allocation)
                              expr
(memory deallocation)
                                       \longrightarrow \mathtt{del}\;(\mathit{expr})
                              expr
                                       \longrightarrow \{ stmt \{ stmt \} : expr [where decl \{ decl \}] \}
(compound expression) expr
                                       \longrightarrow (expr:type)
(typecast)
                              expr
(enclosed expression)
                              expr
                                      \longrightarrow (expr)
(expression)
                              stmt \longrightarrow expr;
(assignment)
                              stmt \longrightarrow expr = expr;
                                       \longrightarrow if expr then stmt \{stmt\} [else stmt \{stmt\}] end;
(conditional)
                              stmt
                                       \longrightarrow while expr do stmt \{stmt\} end;
(loop)
                              stmt
```

where *literal* denotes any literal, *unop* denotes an unary operator (any of !, +, -, \$ and @) and *binop* denotes a binary operator (any of $| , ^, & , ==, !=, <=, >=, <, >, +, -, *, / and %). In the grammar above, braces typeset as {} enclose sentential forms that can repeated zero or more times, brackets typeset as [] enclose sentential forms that can be present or not while braces and brackets typeset as {} and [] denote characters that are a part of the program text.$

Relational operators are non-associative, all other binary operators are left associative.

The precedence of operators is as follows:

```
THE LOWEST PRECEDENCE

# - (binary + and -)

the lowest precedence

(binary + and -)

(unary + and -)

THE HIGHEST PRECEDENCE
```

3 Semantics

3.1 Name binding

Namespaces. There are two kinds of 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.

Scopes. A new scope is created in two ways:

- 1. A compound expression creates a new scope. The scope starts right after { and ends just before }.
- 2. A function declaration creates a new scope. The name of a function, the types of parameters and the type of a result belong to the scope of the function declaration while the names of parameters and the expression denoting the function body (if present) belong to the new inner scope created by the function declaration.

All names declared within a given scope are visible in the entire scope unless hidden by a declaration in the nested scope. A name can be declared within the same scope at most once.

Let \mathcal{P} , \mathcal{I} , \mathcal{D} and \mathcal{T} be a set of all phrases, a set of all identifiers, a set of all declarations and a set of all types of the PREV'19 programming language, respectively.

The semantic function

$$\llbracket \cdot \rrbracket_{\mathrm{ENV}} : (\mathcal{P} \cup \mathcal{T}) \to (\mathcal{I} \to \mathcal{D})$$

maps a phrase of PREV'19 to an environment the phrase appears in.

The program itself appears in the empty environment $\mathcal{E}_0: \mathcal{I} \to \mathcal{D}$ which is undefined for all identifiers in \mathcal{I} , i.e., $\mathcal{E}_0(identifier) = \bot$ for each $identifier \in \mathcal{I}$.

(source):

(type and variable declaration):

$$\frac{ \llbracket \text{typ } \textit{identifier} : \textit{type} \, ; \rrbracket_{\text{ENV}} = \mathcal{E} }{ \llbracket \textit{type} \rrbracket_{\text{ENV}} = \mathcal{E} } \quad \frac{ \llbracket \text{var } \textit{identifier} : \textit{type} \, ; \rrbracket_{\text{ENV}} = \mathcal{E} }{ \llbracket \textit{type} \rrbracket_{\text{ENV}} = \mathcal{E} }$$

 $(function\ declaration)$:

(compound statement):

$$\begin{split} & [\![\{stmt_1 \dots stmt_n \colon expr \text{ where } decl_1 \dots decl_m\}]\!]_{\text{ENV}} = \mathcal{E} \\ & \forall i,j \in \{1 \dots m\} \colon \text{nm}(decl_i) \neq \text{nm}(decl_j) \\ & \mathcal{E}' = \mathcal{E} \rhd (\text{nm}(decl_1) \mapsto decl_1) \rhd \dots \rhd (\text{nm}(decl_m) \mapsto decl_m) \\ & \forall i \in \{1 \dots n\} \colon [\![stmt_i]\!]_{\text{ENV}} = \mathcal{E}' \quad \forall i \in \{1 \dots m\} \colon [\![decl_i]\!]_{\text{ENV}} = \mathcal{E}' \quad [\![expr]\!]_{\text{ENV}} = \mathcal{E}' \end{split}$$

(record type):

$$\frac{[\![\mathsf{rec}\;(identifier_1\!:type_1,\ldots,identifier_n\!:type_n)]\!]_{\mathrm{ENV}}=\mathcal{E}}{\forall i\in\{1\ldots n\}\!:[\![type_i]\!]_{\mathrm{ENV}}=\mathcal{E}}$$

(component access):

$$\begin{split} & & & \llbracket expr.id \rrbracket_{\text{ENV}} = \mathcal{E} \\ & & & \underbrace{ \llbracket expr \rrbracket_{\text{OFTYPE}} = \tau \quad \llbracket \tau \rrbracket_{\text{ENV}} = \mathcal{E}'}_{ \llbracket expr \rrbracket_{\text{ENV}} = \mathcal{E}} \end{split}$$

For all other phrases, the constituents of a phrase inherit the environment of the phrase itself.

Given a declaration, the auxiliary function nm returns the identifier declared by the declaration. The left associative operator \triangleright is defined as follows:

$$[\mathcal{E} \triangleright (identifier \mapsto decl)](identifier') = \begin{cases} decl & identifier' = identifier \\ \mathcal{E}(identifier') & \text{otherwise} \end{cases}$$

The semantic function

$$\llbracket \cdot
rbracket_{ ext{BIND}}: \mathcal{I} o \mathcal{D}$$

maps an identifier to its declaration. It is defined as

$$\frac{ [\mathit{identifier}]_{\mathrm{ENV}} = \mathcal{E} }{ [\mathit{identifier}]_{\mathrm{BIND}} = \mathcal{E}(\mathit{identifier}) }$$

whenever identifier appears in (named type), (variable access), (component access) or (function call).

3.2 Type system

Set

$$\mathcal{T}_{d} = \{ \mathbf{void}, \mathbf{bool}, \mathbf{char}, \mathbf{int} \}$$
 (atomic types)

$$\cup \{ \mathbf{arr}(n \times \tau) \mid n > 0 \wedge \tau \in \mathcal{T}_{d} \}$$
 (arrays)

$$\cup \{ \mathbf{rec}_{id_{1}, \dots, id_{n}}(\tau_{1}, \dots, \tau_{n}) \mid n > 0 \wedge \tau_{1}, \dots, \tau_{n} \in \mathcal{T}_{d} \}$$
 (records)

$$\cup \{ \mathbf{ptr}(\tau) \mid \tau \in \mathcal{T}_{d} \}$$
 (pointers)

denotes a set of all data types of PREV'19. Set

$$\mathcal{T} = \mathcal{T}_d \qquad \text{(data types)}$$

$$\cup \{ (\tau_1, \dots, \tau_n) \to \tau \mid n \ge 0 \land \tau_1, \dots, \tau_n, \tau \in \mathcal{T}_d \} \qquad \text{(functions)}$$

denotes a set of all types of PREV'19.

Two types are equal if they share the same structure.

Semantic functions

$$\llbracket \cdot \rrbracket_{\mathrm{ISTYPE}} : \mathcal{P} \to \mathcal{T} \quad \mathrm{and} \quad \llbracket \cdot \rrbracket_{\mathrm{OFTYPE}} : \mathcal{P} \to \mathcal{T}$$

maps phrases of PREV'19 to types. Function $[\cdot]_{ISTYPE}$ denotes the type described by a phrase, function $[\cdot]_{OFTYPE}$ denotes the type of a value described by a phrase.

Type expressions.

$$\frac{\llbracket type \rrbracket_{\text{ISTYPE}} = \tau \quad \text{val}(int) = n}{n > 0 \quad \tau \in \mathcal{T}_d \setminus \{ \text{void} \}} \\
\underline{\llbracket \text{arr} [int] type \rrbracket_{\text{ISTYPE}} = \text{arr}(n \times \tau)} \tag{T2}$$

$$\frac{[type_1]_{ISTYPE} = \tau_1 \dots [type_n]_{ISTYPE} = \tau_n}{\forall i \in \{1 \dots n\}: \tau_i \in \mathcal{T}_d \setminus \{\mathbf{void}\}} \\
\frac{[\mathbf{rec}(id_1: type_1, \dots, id_n: type_n)]_{ISTYPE} = \mathbf{rec}_{id_1, \dots, id_n}(\tau_1, \dots, \tau_n)}$$
(T3)

$$\frac{\llbracket type \rrbracket_{\text{ISTYPE}} = \tau \quad \tau \in \mathcal{T}_d}{\llbracket \mathbf{ptr} \ type \rrbracket_{\text{ISTYPE}} = \mathbf{ptr}(\tau)}$$
(T4)

$$\frac{[type]_{ISTYPE} = \tau}{[(type)]_{ISTYPE} = \tau}$$
(T5)

Value expressions.

$$\frac{\text{[none]}_{OFTYPE} = \mathbf{void}}{\text{[null]}_{OFTYPE} = \mathbf{ptr}(\mathbf{void})}$$
(v1)

$$\frac{\text{[bool]}_{\text{OFTYPE}} = \mathbf{bool}}{\text{[char]}_{\text{OFTYPE}} = \mathbf{char}} \frac{\text{[int]}_{\text{OFTYPE}} = \mathbf{int}}{\text{[int]}_{\text{OFTYPE}} = \mathbf{int}}$$
(v2)

$$\frac{\llbracket expr_1 \rrbracket_{\text{OFTYPE}} = \tau \quad \llbracket expr_2 \rrbracket_{\text{OFTYPE}} = \tau}{\tau \in \{ \mathbf{char}, \mathbf{int} \} \quad op \in \{+, -, *, /, \%\}}$$

$$\frac{\llbracket expr_1 \ op \ expr_2 \rrbracket_{\text{OFTYPE}} = \mathbf{int}}{\llbracket expr_1 \ op \ expr_2 \rrbracket_{\text{OFTYPE}} = \mathbf{int}}$$
(v5)

$$\frac{[expr_1]_{\text{OFTYPE}} = \tau \quad [expr_2]_{\text{OFTYPE}} = \tau}{\tau \in \{\text{char, int}\} \cup \{\text{ptr}(\tau) \mid \tau \in \mathcal{T}_d\} \quad op \in \{<=, >=, <, >\}}{[expr_1 \ op \ expr_2]_{\text{OFTYPE}} = \mathbf{bool}}$$
(v7)

$$\frac{\llbracket expr \rrbracket_{\text{OFTYPE}} = \tau \quad \tau \in \mathcal{T}_d \setminus \{ \mathbf{void} \}}{\llbracket \$ expr \rrbracket_{\text{OFTYPE}} = \mathbf{ptr}(\tau)} \quad \frac{\llbracket expr \rrbracket_{\text{OFTYPE}} = \mathbf{ptr}(\tau) \quad \tau \in \mathcal{T}_d \setminus \{ \mathbf{void} \}}{\llbracket @ expr \rrbracket_{\text{OFTYPE}} = \tau} \tag{v8}$$

$$\frac{\llbracket type \rrbracket_{\text{ISTYPE}} = \tau \quad \tau \in \mathcal{T}_d \setminus \{ \mathbf{void} \}}{\llbracket \text{new}(type) \rrbracket_{\text{OFTYPE}} = \mathbf{ptr}(\tau)} \qquad \frac{\llbracket expr \rrbracket_{\text{OFTYPE}} = \mathbf{ptr}(\tau) \quad \tau \in \mathcal{T}_d \setminus \{ \mathbf{void} \}}{\llbracket \text{del}(expr) \rrbracket_{\text{OFTYPE}} = \mathbf{void}} \tag{v9}$$

$$\underbrace{\llbracket expr_1 \rrbracket_{\text{OFTYPE}} = \mathbf{arr}(n \times \tau) \quad \llbracket expr_2 \rrbracket_{\text{OFTYPE}} = \mathbf{int}}_{\llbracket expr_1 \llbracket expr_2 \rrbracket_{\text{OFTYPE}} = \tau} \tag{v10}$$

$$\frac{\llbracket expr \rrbracket_{\text{OFTYPE}} = \mathbf{rec}_{id_1,\dots,id_n}(\tau_1,\dots,\tau_n) \quad identifier = id_i}{\llbracket expr.identifier \rrbracket_{\text{OFTYPE}} = \tau_i} \tag{v11}$$

$$[expr]_{OFTYPE} = \tau$$

$$[stmts: expr where decls]_{OFTYPE} = \tau$$
(v13)

$$\frac{[expr]_{OFTYPE} = \tau}{[(expr)]_{OFTYPE} = \tau}$$
(v15)

Statements.

$$\frac{\llbracket expr_1 \rrbracket_{\text{OFTYPE}} = \tau \quad \llbracket expr_2 \rrbracket_{\text{OFTYPE}} = \tau}{\tau \in \{ \mathbf{bool}, \mathbf{char}, \mathbf{int} \} \cup \{ \mathbf{ptr}(\tau) \mid \tau \in \mathcal{T}_d \}} \\
\frac{\llbracket expr_1 = expr_2 \rrbracket_{\text{OFTYPE}} = \mathbf{void}}{\lVert expr_1 = expr_2 \rrbracket_{\text{OFTYPE}} = \mathbf{void}} \tag{s1}$$

Declarations.

4 Lvalues

The semantic function

$$[\![\cdot]\!]_{\mathrm{ISADDR}} : \mathcal{P} \to \{\mathbf{true}, \mathbf{false}\}$$

denotes which phrases represent lvalues.

$$\frac{ \begin{tabular}{l} \begin{tabular}{l} \hline \end{tabular} & \hline \end{t$$

In all other cases the value of $[\![\cdot]\!]_{\mathrm{ISADDR}}$ equals $\mathbf{false}.$