UNIVERSITÀ DEGLI STUDI DI UDINE

Dipartimento di Matematica e Informatica Corso di Laurea Magistrale in Informatica

PROGETTO LC2 PARTE 3 GRUPPO 2

Rosario Maione 105606 - maione.rosario@spes.uniud.it Matteo Bernardon 106027 - bernardon.matteo@spes.uniud.it Mihai Daniel Iftimie 105576 - iftimie.mihaidaniel@spes.uniud.it

Analisi di Lasso e assunzioni

Il linguaggio Lasso

Di seguito vengono elencati, per ogni punto critico della sintassi concreta di Lasso, come sono stati rappresentati i vari costrutti nella nostra grammatica in modo da rendere di più facile lettura il resto del documento e i file della realizzazione, aiutando anche a capire le varie scelte progettuali fatte:

• Le variabili vengono dichiarate con la sintassi

```
var(c :: integer = 8)
```

e vengono richiamate anteponendo il carattere \$ all'identificatore.

- Quando in un *if* si usa *else* questo non prende nessuna condizione, mentre quando si usa l'*else if* questo viene usato in lasso dando all'*else* una condizione fra parentesi tonde.
- Gli array vengono rappresentati con la sintassi

```
array(4,'c',"casa")
```

Da notare come un array in Lasso sia polimorfo.

• Il ciclo while viene usato con la sintassi

```
while(exp) => \{corpo\}
```

• Come costrutto iterativo lasso non prevede il for ma il loop, che viene usato con la sintassi

```
loop(int) => \{corpo\}.
```

- \bullet Non essendo supportati i puntatori sono stati introdotti seguendo la sintassi e la logica dei puntatori di C.
- Lasso non pervede le modalità di passaggio di parametri value, valueresult e constant, quindi sono state introdotte inserendo la modalità nella dichiarazione della variabile prima dell'identificatore, con la seguente sintassi

```
define\ pippo::char(constant\ pluto::char) => \{...\}
```

Se la modalità non viene indicata il parametro ha di default la modalità di passaggio per valore.

Assunzioni fatte

Di seguito vengono riportate le assunzioni fatte per quanto riguarda vincoli semantici e sintattici:

- Le procedure write e read prendono un solo parametro per volta.
- Le funzioni vengono identificate solo dall'identificatore e non dalla firma
- Le dichiarazioni di variabili avvengono sempre una alla volta(in modo da utilizzare la sitassi con *var* di Lasso).
- Può essere dichiarata una funzione all'interno del corpo di un'altra funzione, in questo caso però la funzione interna potrà essere richiamata solo localmente al blocco della funzione esterna.
- Una funzione e una variabile possono avere lo stesso identificatore, questo perchè hanno modalità di chiamata diverse.
- Non viene effettuato alcun casting automatico tra tipi, questo vuol dire che il casting su operatori aritmetici e relazionali viene effettuato (infatti questi hanno più tipi supportati) ma non viene effettuato sugli operatori di assegnamento, come * = o / =.

Grammatica

Di seguito viene presentata la grammatica di Lasso:

Parole riservate e simboli

Le parole riservate usate sono le seguenti:

```
False
         True
                     array
block
         boolean
                     case
char
         const
                     decimal
define
         else
                     goto
if
         integer
                     iterate
         loop_abort
                     loop_continue
loop
         match
main
                     not
pointer read
                     return
string
         val
                     valres
var
         void
                     while
write
```

I simboli usati sono i seguenti:

Struttura semantica

```
\langle Program \rangle ::= main => { \langle ListLine \rangle } 
 \langle ListLine \rangle ::= \epsilon 
 | \langle Line \rangle \langle ListLine \rangle 
 \langle Boolean \rangle ::= True 
 | False
```

```
\langle RExpr \rangle ::= \langle RExpr1 \rangle
                                \langle RExpr \rangle \mid \mid \langle RExpr1 \rangle
                                  \langle RExpr2 \rangle
\langle RExpr1 \rangle
                       ::=
                                   \langle RExpr1 \rangle && \langle RExpr2 \rangle
\langle RExpr2 \rangle
                                \langle RExpr3 \rangle
                       ::=
                                  not \langle RExpr3 \rangle
\langle RExpr3 \rangle
                                \langle RExpr4 \rangle
                       ::=
                                   \langle RExpr4 \rangle == \langle RExpr4 \rangle
                                   \langle RExpr4 \rangle != \langle RExpr4 \rangle
                                   \langle RExpr4 \rangle < \langle RExpr4 \rangle
                                   \langle RExpr4 \rangle \le \langle RExpr4 \rangle
                                   \langle RExpr4 \rangle > \langle RExpr4 \rangle
                                   \langle RExpr4 \rangle >= \langle RExpr4 \rangle
\langle RExpr4 \rangle
                                \langle RExpr5 \rangle
                       ::=
                                   \langle RExpr4 \rangle + \langle RExpr5 \rangle
                                   \langle RExpr4 \rangle - \langle RExpr5 \rangle
\langle RExpr5 \rangle
                                  \langle RExpr6 \rangle
                                   \langle RExpr5 \rangle * \langle RExpr6 \rangle
                                   \langle RExpr5 \rangle / \langle RExpr6 \rangle
                                   ⟨RExpr5⟩ % ⟨RExpr6⟩
\langle RExpr6 \rangle
                        ::=
                                  \langle RExpr7 \rangle
                                  -\langle RExpr7\rangle
\langle RExpr7 \rangle
                        ::=
                                \langle RExpr8 \rangle
                                   \langle FunCall \rangle
\langle RExpr8 \rangle
                       ::= \langle RExpr9 \rangle
                                  * \langle LExpr \rangle
                                  & \langle LExpr \rangle
\langle RExpr9 \rangle
                                \langle RExpr10 \rangle
                                   \langle Integer \rangle
                                   \langle Char \rangle
                                   \langle String \rangle
                                   \langle Double \rangle
                                   \langle Boolean \rangle
                                  array (\langle ListRExpr \rangle)
                         ::= (\langle RExpr \rangle)
\langle RExpr10 \rangle
                                     \langle LExpr \rangle
\langle FunCall \rangle ::= \langle Ident \rangle (\langle ListRExpr \rangle)
\langle ListRExpr \rangle ::=
                                        \langle RExpr \rangle
                                        \langle RExpr \rangle , \langle ListRExpr \rangle
                   ::= \langle LExpr1 \rangle
\langle LExpr \rangle
                               ++ \langle LExpr1 \rangle
                                --\langle LExpr1\rangle
```

```
\langle LExpr1 \rangle ::= \langle LExpr2 \rangle
                            \langle LExpr2 \rangle ++
                             \langle LExpr2 \rangle --
\langle LExpr2 \rangle ::= (\langle LExpr \rangle)
                            \langle BLExpr \rangle
\langle BLExpr \rangle ::= \$ \langle Ident \rangle
                              (Ident) (RExpr) 
\langle Decl \rangle ::= var (\langle Ident \rangle :: \langle Type \rangle)
                       var (\langle Ident \rangle :: \langle Type \rangle = \langle RExpr \rangle)
                       define \langle Ident \rangle :: \langle Type \rangle ( \langle ListParameter \rangle ) => { \langle ListLine \rangle
                       return \langle RExpr \rangle }
                       define \langle Ident \rangle :: void (\langle ListParameter \rangle) => \{\langle ListLine \rangle\}
\langle Type \rangle ::= boolean
                        char
                        decimal
                        integer
                        string
                        array
                        pointer \langle Type \rangle
\langle ListParameter \rangle ::= \epsilon
                                        \langle Parameter \rangle
                                        \langle Parameter \rangle, \langle ListParameter \rangle
\langle Parameter \rangle ::= \langle Modality \rangle \langle Ident \rangle :: \langle Type \rangle
\langle Modality \rangle ::= \epsilon
                               val
                               const
                               valres
\langle Line \rangle ::= block => \{ \langle ListLine \rangle \}
                       \langle Ident \rangle ( \langle ListRExpr \rangle )
                       \langle JumpStmt \rangle
                       \langle IterStmt \rangle
                       \langle SelectionStmt \rangle
                       \langle LExpr \rangle \langle Assignment-op \rangle \langle RExpr \rangle
                       \langle Decl \rangle
                       write (\langle RExpr \rangle)
                       read ( \langle RExpr \rangle )
                       \langle Ident \rangle:
                       goto (Ident)
\langle Assignment-op \rangle ::=
                                         /=
\langle JumpStmt \rangle ::= loop_abort
                                 loop_continue
```

Type System

Assiomi e insiemi usati

```
\begin{aligned} & \text{Tipi} & \overline{\Gamma \vdash x : T} \text{ con } x = T \in Type \\ & \text{booleani} & \overline{\Gamma \vdash Bool : boolean} \\ & \text{Identificatore} & \overline{\Gamma \vdash Ident \not\in Dom(\Gamma)} \\ & Type = \{boolean, char, decimal, integer, string, array, pointerType\} \\ & Bool = \{True, False\} \\ & Modality = \{\epsilon, val, valres, const\} \\ & JumpStmp = \{loop\_abort, loop\_continue\} \end{aligned}
```

Regole

Le regole vengono presentate ruotate di 90 gradi per migliorare la leggibilità una volta stampata la relazione

Vuoto $\overline{\Gamma \vdash \phi}$ $\Gamma \vdash ListLine$ Program $\overline{\Gamma} \vdash \mathsf{main} => ListLine$ $\Gamma \vdash RExpr1 : boolean$ $\Gamma \vdash RExpr2 : boolean$ Or $\Gamma \vdash RExpr1 || RExpr2 : boolean$ $\Gamma \vdash RExpr1 : boolean$ $\Gamma \vdash RExpr2 : boolean$ And $\Gamma \vdash RExpr1\&\&RExpr2:boolean$ $\Gamma \vdash RExpr : boolean$ Not $\Gamma \vdash \mathsf{not}RExpr:boolean$ $\Gamma \vdash RExpr1 : T1$ $\Gamma \vdash RExpr2 : T2$ $T1 = T2 \in Type$ == $\Gamma \vdash RExpr1 == RExpr2 : boolean$ $\Gamma \vdash RExpr1 : T1$ $\Gamma \vdash RExpr2 : T2$!= $T1 = T2 \in Type$ $\Gamma \vdash RExpr1! = RExpr2 : boolean$ $\Gamma \vdash RExpr1 : T1$ $\Gamma \vdash RExpr2 : T2$ $conT1 < T2se \in \{integer, decimal\}$ < $\Gamma \vdash RExpr1 < RExpr2 : boolean$ $con T1 = T2se \in char$ $\Gamma \vdash RExpr1:T1$ $\Gamma \vdash RExpr2 : T2$ $conT1 \leq T2se \in \{integer, decimal\}$ $\leq =$ $\Gamma \vdash RExpr1 \le RExpr2 : boolean$ $con T1 = T2se \in char$ $\Gamma \vdash RExpr1:T1$ $\Gamma \vdash RExpr2 : T2$ $conT1 \leq T2se \in \{integer, decimal\}$ > $\Gamma \vdash RExpr1 > RExpr2 : boolean$

 $con T1 = T2se \in char$

>=	$\frac{\Gamma \vdash RExpr1 : T1 \qquad \Gamma \vdash RExpr2 : T2}{\Gamma \vdash RExpr1 >= RExpr2 : boolean}$	$conT1 \le T2se \in \{integer, decimal\}$ $con\ T1 = T2se \in char$
+	$\frac{\Gamma \vdash RExpr1: T1 \qquad \Gamma \vdash RExpr2: T2}{\Gamma \vdash RExpr1 + RExpr2: T2}$	$T1 \leq T2 \in \{integer, decimal\} \text{ con } integer < decimal$
-	$\frac{\Gamma \vdash RExpr1: T1 \qquad \Gamma \vdash RExpr2: T2}{\Gamma \vdash RExpr1 - RExpr2: T2}$	$T1 \leq T2 \in \{integer, decimal\} \text{ con } integer < decimal$
*	$\frac{\Gamma \vdash RExpr1: T1 \qquad \Gamma \vdash RExpr2: T2}{\Gamma \vdash RExpr1 * RExpr2: T2}$	$T1 \leq T2 \in \{integer, decimal\} \text{ con } integer < decimal$
/	$\frac{\Gamma \vdash RExpr1: T1 \qquad \Gamma \vdash RExpr2: T2}{\Gamma \vdash RExpr1/RExpr2: decimal}$	$T1, T2 \in \{integer, decimal\}$
%	$\frac{\Gamma \vdash RExpr1 : integer \qquad \Gamma \vdash RExpr2 : integer}{\Gamma \vdash RExpr1\% RExpr2 : integer}$	
negativo	$\frac{\Gamma \vdash RExpr: T}{\Gamma \vdash -RExpr: T}$	$T \in \{integer, decimal\}$
$\mathrm{LExpr} =$	$\frac{\Gamma \vdash LExpr: T1 \qquad \Gamma \vdash RExpr: T2}{\Gamma \vdash LExpr = RExpr: T1}$	$T1 = T2 \in Type$
LExpr *=	$\frac{\Gamma \vdash LExpr : T1 \qquad \Gamma \vdash RExpr : T2}{\Gamma \vdash LExpr *= RExpr : T1}$	$T1 = T2 \in \{integer, decimal\}$
$\operatorname{LExpr} + =$	$\frac{\Gamma \vdash LExpr: T1 \qquad \Gamma \vdash RExpr: T2}{\Gamma \vdash LExpr+ = RExpr: T1}$	$T1 = T2 \in \{integer, decimal\}$
$\mathrm{LExpr} \ / =$	$\frac{\Gamma \vdash LExpr: T1 \qquad \Gamma \vdash RExpr: T2}{\Gamma \vdash LExpr/= RExpr: T1}$	$T1 = T2 \in \{integer, decimal\}$

LExpr -=	$\frac{\Gamma \vdash LExpr: T1 \qquad \Gamma \vdash RExpr: T2}{\Gamma \vdash LExpr-= RExpr: T1}$
preincremento	$\frac{\Gamma \vdash LExpr: T1 \qquad \Gamma \vdash RExpr: T2}{\Gamma \vdash LExpr = + + RExpr: T1}$
predecremento	$\frac{\Gamma \vdash LExpr: T1 \qquad \Gamma \vdash RExpr: T2}{\Gamma \vdash LExpr =RExpr: T1}$
postincremento	$\frac{\Gamma \vdash LExpr: T1 \qquad \Gamma \vdash RExpr: T2}{\Gamma \vdash LExpr = RExpr + + : T1}$
postdecremento	$\frac{\Gamma \vdash LExpr: T1 \qquad \Gamma \vdash RExpr: T2}{\Gamma \vdash LExpr = RExpr: T1}$
()	$\frac{\Gamma \vdash LExpr: T1 \qquad \Gamma \vdash RExpr: T2}{\Gamma \vdash LExpr = (RExpr): T1}$
BLExpr	$\frac{\Gamma \vdash Ident}{\Gamma \vdash \$Ident}$
BLExpr array	$\frac{\Gamma \vdash Ident \qquad \Gamma \vdash RExpr: integer}{\Gamma \vdash \$Ident(RExpr)}$
LExpr BLExpr	$\frac{\Gamma \vdash LExpr: T1 \qquad \Gamma \vdash BLExpr: T2}{\Gamma \vdash LExpr = BLExpr: T1}$
Funcall	$\frac{\Gamma \vdash Ident \qquad \Gamma \vdash ListRExpr}{\Gamma \vdash Ident(ListRExpr) : T}$
	$\Gamma \vdash LErm \cdot T1$ $\Gamma \vdash Funcall \cdot T2$

 $\Gamma \vdash LExpr = Funcall : T1$

RExpr Funcall

$$T1 = T2 \in \{integer, decimal\}$$

$$T1 = T2 \in Type$$

$$T \in Type, T \text{ rappresenta il tio di ritorno della funzione}$$

 $T1 = T2 \in Type$

```
\Gamma \vdash LExpr : Type
puntatore *
                            \Gamma \vdash *LExpr : pointerType
                                   \Gamma \vdash LExpr : Type
puntatore &
                            \Gamma \vdash \&LExpr: pointerType
                            \Gamma \vdash Ident
                                                  \Gamma \vdash T : Type
var decl
                             \Gamma \vdash \mathsf{var}(Ident :: T) : Type
                            \Gamma \vdash Ident
                                                  \Gamma \vdash T : T1
                                                                          \Gamma \vdash RExpr : T2
var decl ass
                                                                                                                                 T1 = T2 \in Type
                                    \Gamma \vdash \mathsf{var}(Ident :: T = (RExpr)) : T1
                                                             \Gamma \vdash ListParameter
                                                                                         \Gamma \vdash ListLine
                                                                                                              \Gamma \vdash RExpr:T2
                             \Gamma \vdash Ident
fundecl
                                                                                                                                 T1 = T2 \in T
                                       \Gamma \vdash \mathsf{define} Ident :: T(ListParameter) => \{Listline \mathsf{return} RExpr\} : Type
                                             \Gamma \vdash ListParameter
procdecl
                             \overline{\Gamma \vdash \mathsf{define}Ident::\mathsf{void}(ListParameter)} = > \{Listline\}
                            \Gamma \vdash Ident
                                                  \Gamma \vdash Modality
                                                                              \Gamma \vdash T : Type
Parameter
                                        \Gamma \vdash ModalityIdent :: T : Type
                                    \Gamma \vdash ListLine
Block
                            \Gamma \vdash \mathsf{block} => ListLine
                            \Gamma \vdash RExpr : boolean
                                                                   \Gamma \vdash ListLine
If
                                  \Gamma \vdash \mathsf{if}(RExpr) => \{ListLIne\}
                                                                   \Gamma \vdash ListLine
                            \Gamma \vdash RExpr : boolean
                                                                                               \Gamma \vdash Else
If else
                                         \Gamma \vdash \mathsf{if}(RExpr) => \{ListLIneElse\}
                              \Gamma \vdash ListLine
else
                            \Gamma \vdash \mathsf{else} ListLine
```

```
\Gamma \vdash RExpr : boolean
                                                                \Gamma \vdash ListLine
else if
                                    \Gamma \vdash \mathsf{else}(RExpr)ListLine
                          \Gamma \vdash RExpr : boolean
                                                                \Gamma \vdash ListLine
                                                                                          \Gamma \vdash Else
else if finale
                                           \Gamma \vdash \mathsf{else}(RExpr)ListLineElse
                          \Gamma \vdash RExpr : Type
                                                            \Gamma \vdash Case
match
                           \overline{\Gamma \vdash \mathsf{match}(RExpr)} => \{Case\}
                             \Gamma \vdash ListLine
case
                          \overline{\Gamma \vdash \mathsf{case} ListLine}
                          \Gamma \vdash ListRExpr
                                                         \Gamma \vdash ListLine
case listline1
                             \Gamma \vdash \mathsf{case}(ListRExpr)ListLine
                          \Gamma \vdash ListRExpr
                                                        \Gamma \vdash ListLine
                                                                                   \Gamma \vdash Case
case listline2
                                   \Gamma \vdash \mathsf{case}(ListRExpr)ListLineCase
                          \Gamma \vdash RExpr : boolean
                                                             \Gamma \vdash ListLine
while
                             \Gamma \vdash \mathsf{while}(RExpr) => \{ListLine\}
                          \Gamma \vdash RExpr : integer
                                                                \Gamma \vdash ListLine
loop
                              \Gamma \vdash \mathsf{loop}(RExpr) => \{ListLine\}
                          \Gamma \vdash RExpr : array
                                                             \Gamma \vdash ListLine
iterate
                           \Gamma \vdash \mathsf{iterate}(RExpr) => \{ListLine\}
                             \Gamma \vdash RExpr : T
                                                                                                                          T \in \{integer, decimal, char, string\}
write
                          \Gamma \vdash \mathsf{write}(RExpr)
                            \Gamma \vdash RExpr : T
```

read

 $\Gamma \vdash \mathsf{read}(RExpr)$

 $T \in \{integer, decimal, char, string\}$

Label $\frac{\Gamma \vdash Ident}{\Gamma \vdash Ident:}$

 $\frac{\Gamma \vdash Ident}{\Gamma \vdash \mathsf{goto}Ident}$

Descrizione della soluzione

Per ottenere un primo prototipo è stato usato BNFC, raffinando poi manualmente i file prodotti automaticamente dal tool. Di seguito vengono descritti solo i file presentati nella soluzione più significativi e/o modificati dopo la generazione automatica:

• lasso.cf

File usato come input di BNFC. Contiene la descrizione della sintassi astratta nella forma richiesta.

• Lexlasso.x

Contiene il lexer con tutti i token necessari.

• Parlasso.y

Contiene il parser con l'albero sintattico attributato. All'interno vengono implementati i controlli di semantica statica, il type checker e le istruzioni per generare il Three Address Code. Nei prossimi capitoli verranno descritti in dettaglio tutti gli elementi significativi di questo file.

• Printlasso.hs

Pretty printer, modulo di supporto utile a mostrare il programma di input dopo la scomposizione effettuata dall'intero processo. Rispetto a quello generato dal BNFC, non adatto alla nostra grammatica, sono state apportate delle modifiche per rendere la presentazione più chiara e fedele all'input. Abbiamo riscontrato problemi con il pretty printer di default, questo perchè il nostro linguaggio usa caratteri come il \$ e non mette i ; alla fine della riga.

• Testlasso.hs

File principale che contiene il main con le direttive per ottenere il risultato delparser, lanciare il pretty printer e mostrare il TAC.

• DataStructures.hs

Modulo di supporto contenente le strutture dati per gli attributi e le funzioni di controllo per il type checker, il pretty printer per il TAC, le strutture dati e le funzioni di controllo per il TAC.

Makefile

Oltre che a compilare con il comando make, è possibile lanciare i test preparate lanciando il comando make test associandolo ad un numero da 1 a 6 (i.e. make test1)

Type Checker

Attributi per il Type Checker:

• vEnvLoc e vEnvGlob :: [VarEnv]

```
Con VarEnv = | Var Ident Type Scope
```

Const Ident Type Scope

Ambienti, locale e globale, per le variabili

• fEnvLoc e fEnvGlob :: [FunEnv]

```
Con FunEnv =
| Fun Ident Type [Type] Scope
| Proc Ident [Type]
```

Ambienti, locale e globale, per le funzioni

• lEnvLoc e lEnvGlob :: [LabEnv]

```
Con LabEnv =
| Lab Ident Scope
| Got Ident String Scope
```

Ambienti, per label e goto

• typeAtt :: Type

Attributo per gestire il tipo di un nodo

• typeAttList :: [Type]

Attributo che raccoglie una lista di attributi, usato ad esempio nel passaggio di parametri.

• noType :: Bool

Attributo per la gestione degli errori di tipo, se il valore è True significa che il nodo o ha un problema sul tipo o ce l'ha uno dei suoi figli. Questa propagazione dell'errore sul tipo serve a mostrare in output l'errore più interno di ogni singola riga, che è molto più significativo del più esterno.

• trueLexpr :: Bool

Attributo usato per sapere se a un nodo di tipo left-expression viene assegnato qualche valore

• inLoop :: Bool

Attributo usato per sapere se il nodo si trova in un ciclo o meno

Three Address Code

Attributi per il Three Address Code:

• code :: [StructTAC]

Con StructTAC =

| NullOp String String

UnOp String String String

| BinOp String String String

UncondJump Int

CondJump String Int

| RelCondJump String String Int

FunDecl String String Int

FunCall String String Int [String]

ArrayLeft String String String

| ArrayRight String String String

LabelT Int

Return String

| LabelProg String

GotoProg String

Attributo che contiene la struttura dati per generare il codice intermedio

• addr :: String

Contiene il nome delle variabili temporanee

• listAddr :: [String]

Lista con nomi di variabili temporanee, viene usata per poter dichiarare i parametri prima di richiamare una procedura o una funzione

• countDown :: (int,int)

Coppia di interi in cui il primo elemento conta le nuove variabili introdotte nel codice intermedio, il secondo invece conta le label introdotte

• countUp :: (int,int)

Stesso meccanismo di countDown ma necessario al passaggio di attributi tra nodi.

Il TAC presentato non è completo, per ragioni di tempo alcune istruzioni non vengono tradotte in codice intermedio. Per altre invece è stata implementata tutta la logica di generazione e inserimento nella struttura dati, ma non siamo riusciti a completare l'ultima sessione di debugging. Abbiamo inoltre riscontrato difficoltà più serie, come il fatto di aver seguito la sintassi di Lasso per quanto riguarda gli array polimorfi, questa scelta si è rivelata solo in un secondo momento sbagliata, questo perchè la gestione di più

tipi, e il conseguente calcolo dello spazio per ogni posizione dell'array, è risultata molto complessa.