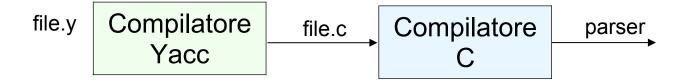
Yacc

- Generatore di parser LALR(1)
- YACC = "Yet Another Compiler Compiler" → sintomo di due fatti:
 - 1. Popularità dei generatori di parser in quegli anni
 - 2. Storicamente: fasi del compilatore intrecciate con l'analisi sintattica





Yacc (ii)

• Specifica Yacc: strutturalmente identica a Lex

Dichiarazioni
%%
Regole di traduzione
%%

Funzioni ausiliarie

Dichiarazioni black box (definizioni ausiliarie): %{ #include, costanti, variabili %}
 white box (token, ...)

• <u>Esempio</u>: calcolatore (interprete): parsing bottom-up → azioni di calcolo

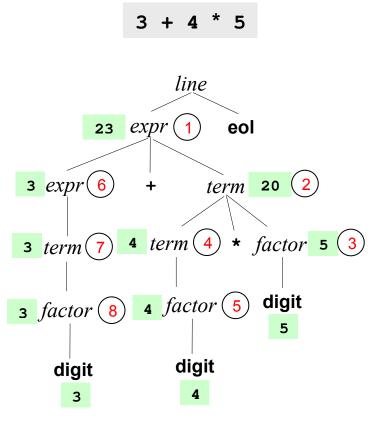
```
line → expr eol
expr \rightarrow expr + term \mid term
term \rightarrow term * factor \mid factor
factor \rightarrow (expr) \mid digit

Ricorsiva a sinistra
```

Yacc (iii)

```
line \rightarrow expr eol
expr \rightarrow expr + term \mid term
term \rightarrow term * factor | factor
factor \rightarrow (expr) \mid digit
```

```
8 {
                                                  Pila di triple (X,s,v)
                       #include <stdio.h>
                                                  v = attributo semantico
                      #include <ctype.h>
                      int yylex();
                      void yyerror();
                      용}
                      %token DIGIT
                      응응
                                  expr '\n' { printf("%d\n", $1); }
                      line
                                  expr '+' term { $$ = $1 + $3; }
                      expr
                                  term { $$ = $1; }
                                  term '*' factor { $$ = $1 * $3; }
                      term
                                  factor { $$ = $1; }
                                  '(' expr ')' { $$ = $2; }
                      factor
                                  DIGIT { $$ = $1; }
                       용용
                      int yylex()
                      { int c;
                        c = getchar();
                        if (isdigit(c)){
                          yylval = c - '0';
return(DIGIT);
                        }
                        return(c);
                      void yyerror(){fprintf(stderr, "Syntax error\n");}
                      void main(){yyparse();}
```



Yacc (iv)

1. **Dichiarazioni** (% { dichiarazioni C % } dichiarazioni di terminali (token) di G

2. Regole di traduzione = regole di produzione + azioni semantiche

$$A \rightarrow \alpha_{1} \mid \alpha_{2} \mid \dots \mid \alpha_{n}$$

$$\Rightarrow \qquad A : \alpha_{1} \{ azione 1 \} \\ \mid \alpha_{2} \{ azione 2 \} \\ \mid \alpha_{n} \{ azione n \} \\ \mid \vdots$$

- Assioma = primo nonterminale (default), o %start line
- 2 modi per riconoscere token \(\begin{aligned} \'+' \\ \DIGIT \end{aligned}
- 'c' = simbolo terminale 'c'
- Nonterminale = stringa di caratteri alfanumerici
- Alternative separate da
- Separazione di ogni gruppo di alternative + azioni semantiche da ;
- Azione semantica = frammento di codice C
- Pseudo-variabli per referenziare valori di attributi semantici (default: intero) (\$\frac{\f{\frac{\frac{\frac{\frac{\frac{\f{\frac{\frac{\frac{\frac{\frac{\fig}\fir\f{\frac{\frac{\frac{\frac{\frac{\frac{\frac{\frac{\f{\f

Yacc (v)

- yylval = variabile contenente il valore lessicale dei token → assegnata dal lexer (valore associato al terminale spostato sulla pila)
- Azione semantica eseguita nella riduzione \$\$ = f(\$1, \$2, ...)

```
expr : expr '+' term {$$ = $1 + $3;}
| term
;
azione di default: $$ = $1;
```

3. Funzioni ausiliarie = funzioni C necessarie per completare la funzione di parsing

In particolare $\begin{cases} \frac{yylex()}{yyerror()} \implies \text{ chiamate da } \frac{yyparse()}{yyerror()} \rightarrow \text{return} \end{cases} \rightarrow \text{return} \begin{cases} 0: \text{ ok} \\ 1: \text{ errore} \end{cases}$

Yacc (vi)

• Compilazione:

```
bison -dvg -o calc.c calc.y
cc -o calc calc.c
dot -Tpdf -o calc.pdf calc.gv
```

Opzioni:

-d (header): genera file.h = dichiarazioni delle informazioni esportabili (codifica dei simboli per Lex)

-V (verbose): genera file.output = descrizione testuale della tabella di parsing LALR(1)

-g (graphic): genera file.gv = rappresentazione dell'automa di parsing LALR(1) nel linguaggio dot

Yacc (vii)

G ambigua → conflitti → individuati da Yacc (opzione –v : mostra anche le soluzioni)

Se ∃ conflitti → consultare file.output per vedere conflitti soluzioni

- Regole Yacc per risoluzione dei conflitti:
 - 1. Spostamento/riduzione → scelto lo spostamento
 - 2. Riduzione/riduzione → scelta la <u>prima</u> regola di produzione (nel file)

Yacc (viii)

• Generalizzazione del tipo di valori computati dalle azioni semantiche (cioè: tipo delle pseudo-variabili, es. calcolatore per numeri <u>reali</u>)

```
%{
...
#define YYSTYPE float
...
%}
```

Definizione del tipo in un file separato:

 typedef ... TYPE;

```
#define YYSTYPE TYPE (nel file Yacc)
```

Esempio: Costruzione dell'albero sintattico:

```
typedef ... *PNODE;

puntatore al nodo dell'albero
```

Yacc (ix)

• Azioni semantiche embedded: quando necessario eseguire codice <u>prima</u> del riconoscimento completo di una produzione

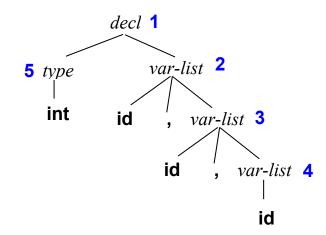
```
decl \rightarrow type \ var-list;

type \rightarrow int \mid float

var-list \rightarrow id, var-list \mid id
```

```
int a, b, c;
```

<u>Goal</u>: Analizzando gli identificatori in *var-list*, qualificare ogni **id** con il rispettivo tipo.

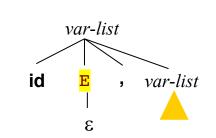


 $\mathbf{A} \to \varepsilon$ ridotto dopo l'azione su B

Interpretazione di Yacc delle azioni embedded:

{ azione embedded }

```
A : B E C;
E : { azione embedded };
ε-produzione
```



Tecnologie dei Linguaggi Artificiali

 \equiv

Costruzione Bottom-up dell'Albero (Semi) Concreto

```
program → stat-list

stat-list → stat stat-list | stat

stat → def-stat | assign-stat

def-stat → def id (def-list)

def-list → domain-decl, def-list | domain-decl

domain-decl → id : domain

domain → integer | string | boolean

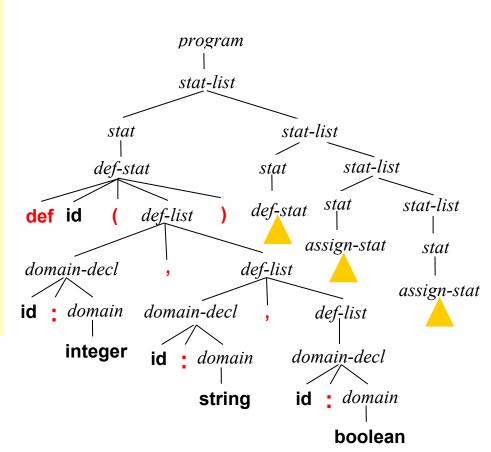
assign-stat → id := { tuple-list }

tuple-list → tuple-const tuple-list | \varepsilon

tuple-const → (simple-const-list)

simple-const-list → simple-const | boolconst
```

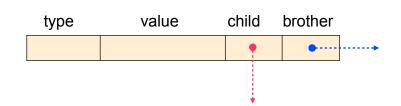
```
def R (A: integer, B: string, C: boolean)
def S (D: integer, E: string)
R := {(3, "alpha", true)(5, "beta", false)}
S := {(125, "sun")(236, "moon")}
```



Costruzione Bottom-up dell'Albero (Semi) Concreto (ii)

```
typedef union
{
    int ival;
    char *sval;
    enum {FALSE, TRUE} bval;
} Value;

typedef struct snode
{
    Typenode type;
    Value value;
    struct snode *child, *brother;
} Node;
```



zucchero sintattico: può mancare nell'albero

• Regole di "stato" del nodo:

Nonterminale

def.h

```
#include <stdio.h>
#include <stdlib.h>
typedef enum
   NPROGRAM,
   NSTAT LIST,
    NSTAT,
   NDEF STAT,
   NDEF LIST,
   NDOMAIN DECL,
    NDOMAIN,
   NASSIGN STAT,
   NTUPLE LIST,
   NTUPLE CONST,
   NSIMPLE CONST LIST,
    NSIMPLE CONST
} Nonterminal;
typedef enum
    T INTEGER,
    T STRING,
    T BOOLEAN,
    T INTCONST,
    T BOOLCONST,
    T STRCONST,
    T ID,
    T NONTERMINAL
} Typenode;
```

```
typedef union
{
    int ival;
    char *sval;
    enum {FALSE, TRUE} bval;
} Value;

typedef struct snode
{
    Typenode type;
    Value value;
    struct snode *child, *brother;
} Node;

typedef Node *Pnode;
```

```
char *newstring(char*);
int yylex();

Pnode nontermnode(Nonterminal),
    idnode(),
    keynode(Typenode),
    intconstnode(),
    strconstnode(),
    boolconstnode(),
    newnode(Typenode);

void treeprint(Pnode, int),
    yyerror();
```

lexer.lex

```
8 {
#include "parser.h"
#include "def.h"
int line = 1:
Value lexval;
용}
%option noyywrap
spacing
            ([\t])+
            [A-Za-z]
letter
digit
            [0-9]
intconst
            {digit}+
           \"([<sup>^</sup>\"])*\"
strconst
boolconst
            false true
id
            {letter}({letter}|{digit})*
sugar
            [(){}:,]
응응
{spacing}
\n
            {line++;}
def
            {return(DEF);}
integer
            {return(INTEGER);}
string
            {return(STRING);}
            {return(BOOLEAN);}
boolean
{intconst} {lexval.ival = atoi(yytext); return(INTCONST);}
{strconst} {lexval.sval = newstring(yytext); return(STRCONST);}
{boolconst} {lexval.bval = (yytext[0] == 'f' ? FALSE : TRUE);
             return(BOOLCONST);}
            { lexval.sval = newstring(yytext); return(ID); }
{id}
{sugar}
            {return(yytext[0]);}
":="
            {return(ASSIGN);}
            {return(ERROR);}
용용
```

```
char *newstring(char *s)
{
   char *p;

   p = malloc(strlen(s)+1);
   strcpy(p, s);
   return(p);
}
```

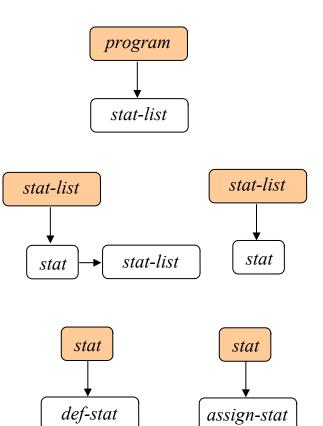
parser.h

```
enum yytokentype
{
...
    DEF = 258,
    INTEGER = 259,
    STRING = 260,
    BOOLEAN = 261,
    ID = 262,
    INTCONST = 263,
    STRCONST = 264,
    BOOLCONST = 265,
    ASSIGN = 266,
    ERROR = 267
};
```

parser.y

```
용 {
#include "def.h"
#define YYSTYPE Pnode
extern char *yytext;
extern Value lexval;
                      analizzatore lessicale
extern int line;
extern FILE *yyin;
Pnode root = NULL;
용}
%token DEF INTEGER STRING BOOLEAN ID INTCONST STRCONST BOOLCONST ASSIGN
%token ERROR
응 응
   $$
program : stat list {root = $$ = nontermnode(NPROGRAM);
                     $$->child = $1;}
stat list : stat stat list {$$ = nontermnode(NSTAT LIST);
                            $$->child = $1:
                            $1->brother = $2;}
          | stat {$$ = nontermnode(NSTAT LIST);
                  $$->child = $1;}
stat : def stat {$$ = nontermnode(NSTAT);
                 $$->child = $1;}
     assign stat {$$ = nontermnode(NSTAT);
                    $$->child = $1;}
```

 $program \rightarrow stat-list$ $stat-list \rightarrow stat stat-list \mid stat$ $stat \rightarrow def$ -stat \quad assign-stat



parser.y (ii)

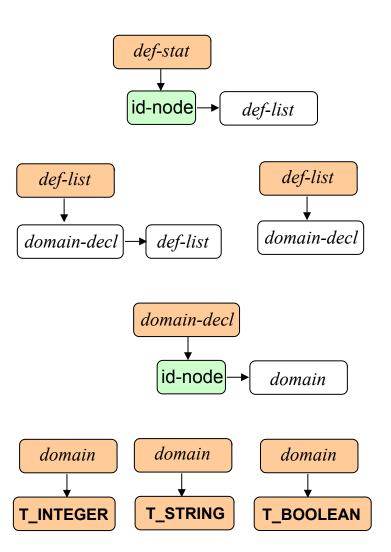
```
def-stat \rightarrow def id ( def-list )

def-list \rightarrow def-list, domain-decl | domain-decl

domain-decl \rightarrow id : domain

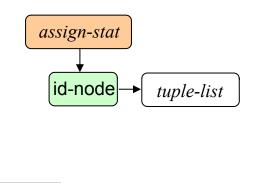
domain \rightarrow integer | string | boolean
```

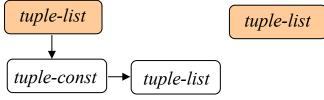
```
def stat : DEF
           ID {$$ = idnode();}
           '(' def list ')' {$$ = nontermnode(NDEF STAT);
                             $$->child = $3:
                             $3->brother = $5;}
def_list : domain_decl ',' def_list {$$ = nontermnode(NDEF LIST);
                                     $$->child = $1;
                                     $1->brother = $3;}
          domain decl {$$ = nontermnode(NDEF LIST);
                        $$->child = $1;}
domain decl : ID {$$ = idnode();}
              ':' domain {$$ = nontermnode(NDOMAIN DECL);
                          $$->child = $2;
                          2->brother = 4;
            ;
domain : INTEGER {$$ = nontermnode(NDOMAIN);
                  $$->child = keynode(T INTEGER);}
         STRING {$$ = nontermnode(NDOMAIN);
                 $$->child = keynode(T STRING);}
         BOOLEAN {$$ = nontermnode(NDOMAIN);
                  $$->child = keynode(T BOOLEAN);}
```

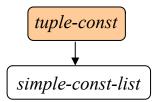


parser.y (iii)

```
assign-stat → id := { tuple-list }
tuple-list → tuple-const tuple-list | \varepsilon
tuple-const → ( simple-const-list )
```

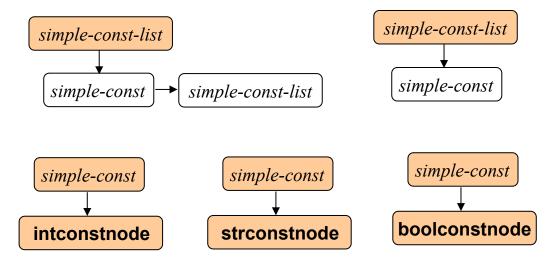






parser.y (iv)

 $simple-const-list \rightarrow simple-const$, $simple-const-list \mid simple-const$ $simple-const \rightarrow intconst \mid strconst \mid boolconst$



parser.y (v)

```
Pnode nontermnode(Nonterminal nonterm)
   Pnode p = newnode(T NONTERMINAL);
   p->value.ival = nonterm;
   return(p);
}
Pnode idnode()
   Pnode p = newnode(T ID);
   p->value.sval = lexval.sval;
   return(p);
Pnode keynode (Typenode keyword)
   return(newnode(keyword));
Pnode intconstnode()
   Pnode p = newnode(T INTCONST);
   p->value.ival = lexval.ival;
   return(p);
Pnode strconstnode()
   Pnode p = newnode(T STRCONST);
   p->value.sval = lexval.sval;
   return(p);
```

```
Pnode boolconstnode()
  Pnode p = newnode(T BOOLCONST);
  p->value.bval = lexval.bval;
  return(p);
Pnode newnode (Typenode tnode)
  Pnode p = malloc(sizeof(Node));
 p->type = tnode;
 p->child = p->brother = NULL;
  return(p);
int main()
  int result;
  yyin = stdin;
  if((result = yyparse()) == 0)
    treeprint(root, 0);
 return(result);
void yyerror()
  fprintf(stderr, "Line %d: syntax error on symbol \"%s\"\n",
          line, yytext);
 exit(-1);
```

tree.c

```
#include "def.h"
char* tabtypes[] =
"INTEGER",
"STRING",
"BOOLEAN",
"INTCONST",
"BOOLCONST",
"STRCONST",
"ID",
"NONTERMINAL"
};
char* tabnonterm[] =
  "PROGRAM",
  "STAT LIST",
  "STAT",
  "DEF STAT",
  "DEF LIST",
  "DOMAIN DECL",
  "DOMAIN",
  "ASSIGN STAT",
  "TUPLE LIST",
  "TUPLE CONST",
  "SIMPLE CONST LIST",
  "SIMPLE CONST"
};
```

```
void treeprint(Pnode root, int indent)
 int i;
 Pnode p;
 for(i=0; i<indent; i++)</pre>
   printf(" ");
 printf("%s", (root->type == T NONTERMINAL ? tabnonterm[root->value.ival] :
                                              tabtypes[root->type]));
 if(root->type == T ID | root->type == T STRCONST)
   printf(" (%s)", root->value.sval);
 else if(root->type == T INTCONST)
                                                                     valore lessicale
   printf(" (%d)", root->value.ival);
 else if(root->type == T BOOLCONST)
   printf(" (%s)", (root->value.ival == TRUE ? "true" : "false"));
 printf("\n");
 for(p=root->child; p != NULL; p = p->brother)
   treeprint(p, indent+1);
```

makefile

```
bup: lexer.o parser.o tree.o
    cc -q -o bup lexer.o parser.o tree.o
lexer.o: lexer.c parser.h def.h
    cc -q -c lexer.c
parser.o: parser.c def.h parser.dot
    cc -q -c parser.c
    dot -Tpdf -o parser.pdf parser.qv
tree.o: tree.c def.h
    cc -q -c tree.c
lexer.c: lexer.lex parser.y parser.h parser.c def.h
    flex -o lexer.c lexer.lex
parser.h: parser.y def.h
    bison -dvg -o parser.c parser.y
```

Esecuzione

```
def R (A: integer, B: string, C: boolean)
R := {(3, "alpha", true)(5, "beta", false)}
```

```
PROGRAM
   STAT LIST
       STAT
           DEF_STAT
                ID (R)
                DEF LIST
                    DOMAIN DECL
                        ID (A)
                        DOMAIN
                            INTEGER
                    DEF LIST
                        DOMAIN_DECL
                            ID (B)
                            DOMAIN
                                STRING
                        DEF_LIST
                            DOMAIN DECL
                                ID (C)
                                DOMAIN
                                    BOOLEAN
        STAT_LIST
            STAT
                ASSIGN STAT
                    ID (R)
                    TUPLE LIST
                        TUPLE CONST
                            SIMPLE CONST_LIST
                                SIMPLE CONST
                                    INTCONST (3)
                                SIMPLE CONST LIST
                                    SIMPLE CONST
                                        STRCONST ("alpha")
                                    SIMPLE CONST LIST
                                        SIMPLE CONST
                                            BOOLCONST (true)
                        TUPLE LIST
                            TUPLE CONST
                                SIMPLE CONST LIST
                                    SIMPLE CONST
                                        INTCONST (5)
                                    SIMPLE CONST LIST
                                        SIMPLE CONST
                                            STRCONST ("beta")
                                        SIMPLE CONST LIST
                                            SIMPLE CONST
                                                BOOLCONST (false)
                            TUPLE LIST
```

Costruzione Bottom-up dell'Albero Astratto

```
program → stat-list

stat-list → stat stat-list | stat

stat → def-stat | assign-stat

def-stat → def id (def-list)

def-list → domain-decl, def-list | domain-decl

domain-decl → id : domain

domain → integer | string | boolean

assign-stat → id := { tuple-list }

tuple-list → tuple-const tuple-list | \varepsilon

tuple-const → (simple-const-list)

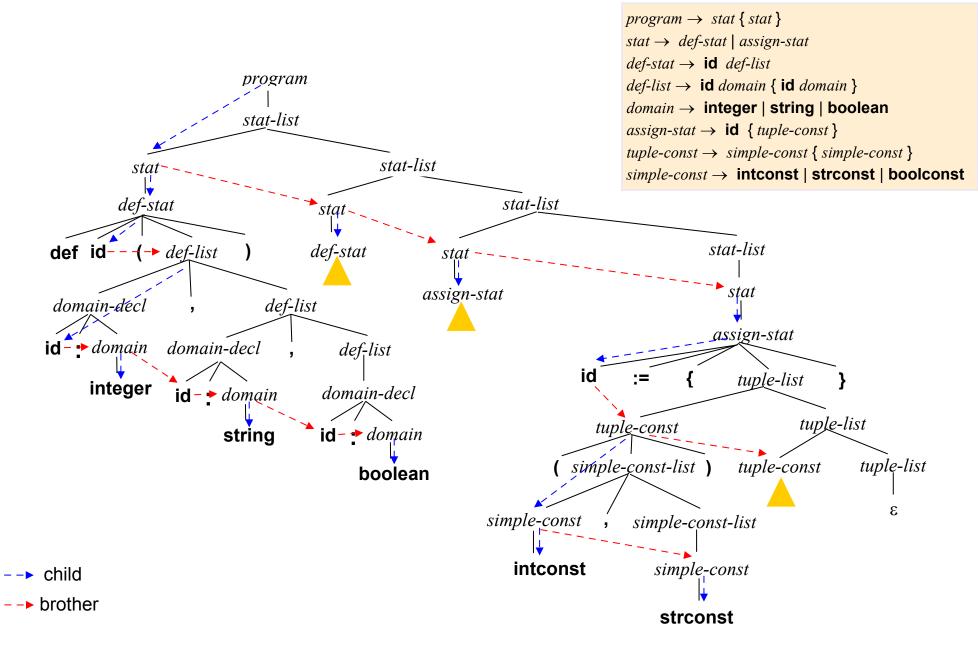
simple-const-list → simple-const | simple-const

simple-const → intconst | strconst | boolconst
```

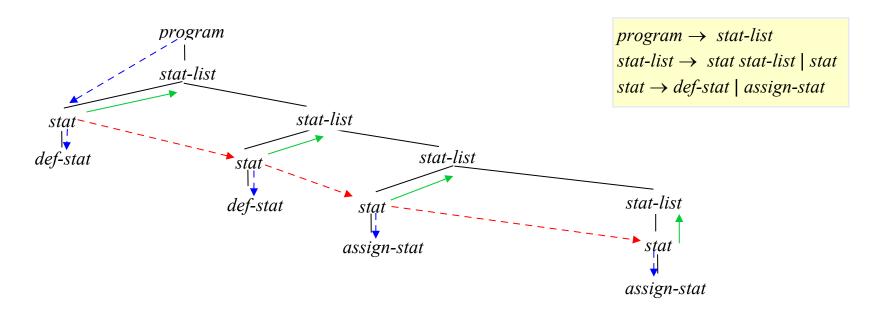
```
program \rightarrow stat \{ stat \}
stat \rightarrow def-stat \mid assign-stat
def-stat \rightarrow id def-list
def-list \rightarrow id domain \{ id domain \}
domain \rightarrow integer \mid string \mid boolean
assign-stat \rightarrow id \{ tuple-const \}
tuple-const \rightarrow simple-const \{ simple-const \}
simple-const \rightarrow intconst \mid strconst \mid boolconst
```

```
def R (A: integer, B: string, C: boolean)
def S (D: integer, E: string)
R := {(3, "alpha", true)(5, "beta", false)}
S := {(125, "sun")(236, "moon")}
```

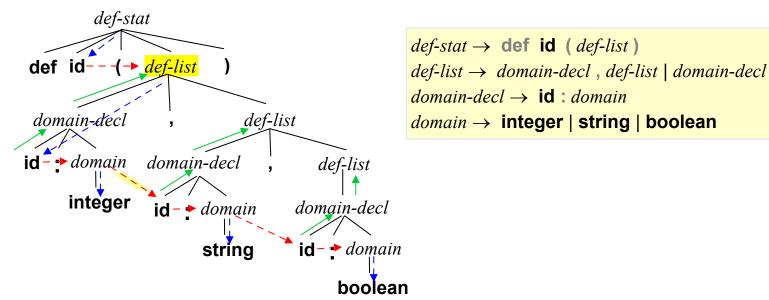
Costruzione Bottom-up dell'Albero Astratto (ii)



Costruzione Bottom-up dell'Albero Astratto (iii)



Costruzione Bottom-up dell'Albero Astratto (iv)



Costruzione Bottom-up dell'Albero Astratto (v)

```
id := { tuple-list }

tuple-const tuple-list

( simple-const-list ) tuple-const tuple-list

simple-const simple-const

strconst
```

```
assign\text{-}stat \rightarrow \text{id} := \{ tuple\text{-}list \}
tuple\text{-}list \rightarrow tuple\text{-}const tuple\text{-}list \mid \epsilon
tuple\text{-}const \rightarrow (simple\text{-}const\text{-}list)
simple\text{-}const\text{-}list \rightarrow simple\text{-}const, simple\text{-}const\text{-}list \mid simple\text{-}const
simple\text{-}const \rightarrow \text{intconst} \mid \text{strconst} \mid \text{boolconst}
```

Esecuzione

```
def R (A: integer, B: string, C: boolean)
def S (D: integer, E: string)
R := {(3, "alpha", true)(5, "beta", false)}
S := {(125, "sun")(236, "moon")}
```

```
PROGRAM
   STAT
       DEF STAT
           ID (R)
            DEF LIST
               ID (A)
               DOMAIN
                    INTEGER
               ID (B)
               DOMAIN
                    STRING
               ID (C)
                DOMAIN
                    BOOLEAN
   STAT
       DEF STAT
           ID (S)
           DEF LIST
               DOMAIN
                    INTEGER
               ID (E)
               DOMAIN
                    STRING
    STAT
       ASSIGN STAT
           ID (R)
           TUPLE CONST
               SIMPLE CONST
                   INTCONST (3)
                SIMPLE CONST
                    STRCONST ("alpha")
               SIMPLE CONST
                    BOOLCONST (true)
           TUPLE CONST
               SIMPLE CONST
                    INTCONST (5)
               SIMPLE CONST
                    STRCONST ("beta")
                SIMPLE CONST
                    BOOLCONST (false)
   STAT
       ASSIGN STAT
           ID (S)
           TUPLE CONST
               SIMPLE CONST
                    INTCONST (125)
               SIMPLE CONST
                    STRCONST ("sun")
           TUPLE CONST
               SIMPLE CONST
                    INTCONST (236)
               SIMPLE_CONST
                    STRCONST ("moon")
```