



ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

Bài 10

Phân tích ngữ nghĩa

Nội dung

- **Những vấn đề ngữ nghĩa**
- **Bảng ký hiệu**
 - Luật về phạm vi ảnh hưởng của biến
 - Các sơ đồ dịch để xây dựng bảng ký hiệu
- **Kiểm tra kiểu (Type checking)**
 - Hệ thống kiểu trong ngôn ngữ lập trình
 - Đặc tả một bộ kiểm tra kiểu
 - Chuyển đổi kiểu

Phân tích ngữ nghĩa

- Tìm ra các lỗi sau giai đoạn phân tích cú pháp
 - Kiểm tra sự tương ứng về kiểu
 - Kiểm tra sự tương ứng giữa việc sử dụng hàm, biến với khai báo của chúng
 - Xác định phạm vi ảnh hưởng của các biến trong chương trình
- Phân tích ngữ nghĩa thường sử dụng cây cú pháp

Bảng ký hiệu và phạm vi

- Phạm vi là gì
- Quản lý phạm vi tĩnh và động
- Những vấn đề liên quan đến phạm vi
- Bảng ký hiệu
- Xây dựng bảng ký hiệu

Quản lý phạm vi

- Đây là vấn đề liên quan đến sự phù hợp giữa khai báo và sử dụng của mỗi định danh.
- Phạm vi ảnh hưởng của mỗi định danh là phần chương trình có thể truy cập tới định danh đó.
- Một định danh có thể tham chiếu các đối tượng khác nhau trong các phạm vi khác nhau của chương trình. Phạm vi của hai định danh giống nhau không được giao nhau

Phạm vi tĩnh và động

- Phần lớn các ngôn ngữ quản lý phạm vi theo kiểu tĩnh. Thông tin phạm vi chỉ phụ thuộc văn bản chương trình. Luật phạm vi gần nhất được áp dụng
- Một số ít ngôn ngữ cho phép quản lý phạm vi động, quản lý phạm vi khi thực hiện chương trình (Lisp, Snobol, Perl khi dùng một số từ khóa đặc biệt)
- Quản lý phạm vi động nghĩa là khi một ký hiệu được tham chiếu, chương trình dịch sẽ tham chiếu vào stack chứa các bản hoạt động để tìm ra thông tin về ký hiệu đó

Ví dụ

QL phạm vi tĩnh

```
1  const int b = 5;
2  int foo()
3  {
4      int a = b + 5;
5      return a;
6  }
7
8  int bar()
9  {
10     int b = 2;
11     return foo();
12 }
13
14 int main()
15 {
16     foo(); // returns 10
17     bar(); // returns 10
18     return 0;
19 }
```

QL phạm vi động

```
1  const int b = 5;
2  int foo()
3  {
4      int a = b + 5;
5      return a;
6  }
7
8  int bar()
9  {
10     int b = 2;
11     return foo();
12 }
13
14 int main()
15 {
16     foo(); // returns 10
17     bar(); // returns 7
18     return 0;
19 }
```

Những vấn đề về quản lý phạm vi

- Khi xét một khai báo chứa một định danh, liệu đã tồn tại định danh đó trong phạm vi hiện hành chưa?
- Khi sử dụng một định danh, liệu nó đã được khai báo chưa? Nếu nó đã được khai báo (theo luật phạm vi gần nhất) liệu khai báo có tương thích với sử dụng không?

Bảng ký hiệu (Symbol table)

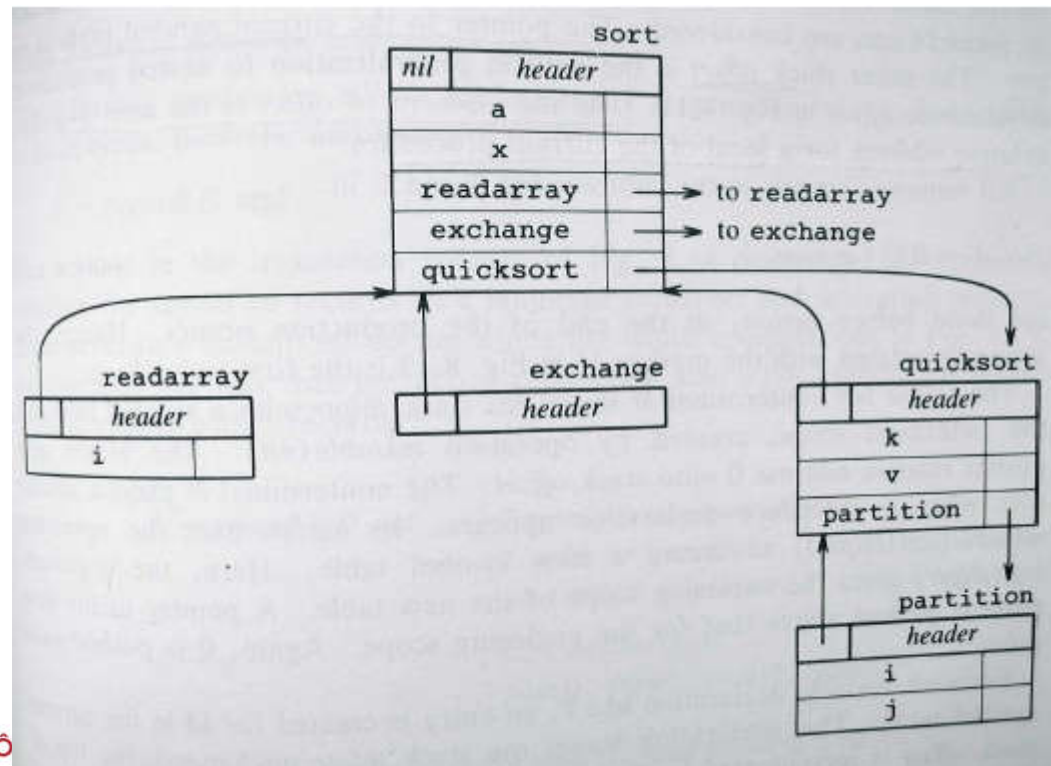
- Một cấu trúc dữ liệu cho phép theo dõi quan hệ hiện hành của các định danh (để kiểm tra ngữ nghĩa và sinh mã một cách hiệu quả)
- Phần quan trọng trong phân tích ngữ nghĩa là theo dõi các hằng/biến/kiểu/hàm/thủ tục xem có phù hợp với khai báo của chúng không?
- Khi thêm một định danh vào bảng ký hiệu, cần ghi lại thông tin trong khai báo của định danh đó.

Ngôn ngữ có cấu trúc khối

- Khối trong ngôn ngữ lập trình là tập các cấu trúc ngôn ngữ có chứa khai báo
- Một ngôn ngữ là có cấu trúc khối nếu
 - Các khối được lồng bên trong những khối khác
 - Phạm vi của khai báo trong mỗi khối là chính khối đó và các khối chứa trong nó
- Luật lồng nhau gần nhất
 - Cho nhiều khai báo của cùng một tên. Khai báo có hiệu lực là khai báo nằm trong khối gần nhất

Giải pháp nhiều bảng ký hiệu

- Các bảng cần được kết nối từ phạm vi trong ra phạm vi ngoài và ngược lại



Các hàm và thủ tục trong chương trình ví dụ sort

```
program sort;
  var a : array [0..10] of integer;
      x : integer;
  procedure readarray;
    var i : integer;
    begin ... end;
  procedure exchange(i, j : integer);
    begin x := a[i]; a[i] := a[j]; a[j] := x end;
  procedure quicksort(m, n : integer);
    var k, v : integer;
    function partition(y, z : integer) : integer
      var i, j : integer;
      begin ... exchange(i, j) ... end
    begin
      if (n > m) then begin
        i := partition(m, n);
        quicksort(m, i - 1);
        quicksort(i + 1, n)
      end
    end;
  begin
    ...
    quicksort(1, 9)
  end.
```

Xây dựng bảng ký hiệu

- Những thao tác cần thiết
- Vào phạm vi (enter scope): tạo ra một phạm vi mới trong các phạm vi lồng nhau
- Xử lý khai báo: Thêm một định danh vào bảng ký hiệu của phạm vi hiện hành
- Xử lý việc sử dụng định danh: Kiểm tra xem định danh có xuất hiện trong bảng ký hiệu của
 - Phạm vi hiện hành
 - Các phạm vi từ phạm vi hiện hành ra ngoài theo luật phạm vi gần nhất
- Ra khỏi phạm vi (exit scope): ra khỏi phạm vi hiện hành

Các luật về phạm vi lồng nhau

- Toán tử insert vào bảng ký hiệu không được ghi đè những khai báo trước
- Toán tử lookup vào bảng ký hiệu luôn luôn tham chiếu luật phạm vi gần nhất
- Toán tử delete chỉ được xóa khai báo gần nhất

Bảng ký hiệu hoạt động như một stack

Cấu trúc dữ liệu cho bảng ký hiệu

- Danh sách liên kết không sắp thứ tự: Thích hợp cho việc phân tích các chương trình sử dụng số lượng biến nhỏ
- Danh sách liên kết sắp thứ tự. Thao tác bổ sung tốn kém nhưng thao tác tìm kiếm lại nhanh chóng hơn
- Cây nhị phân tìm kiếm
- Bảng băm: thường được dùng nhất

Xây dựng trong giai đoạn phân tích cú pháp

- Chỉ có thể bắt đầu nhập thông tin vào bảng ký hiệu từ khi phân tích từ vựng nếu ngôn ngữ lập trình không cho khai báo tên trùng nhau.
- Nếu cho phép các phạm vi, bộ phân tích từ vựng chỉ trả ra tên của định danh cùng với token
 - Định danh được thêm vào bảng ký hiệu khi vai trò cú pháp của định danh được phát hiện

Khái niệm kiểm tra kiểu

- Kiểm tra xem chương trình có tuân theo các luật về kiểu của ngôn ngữ không
- Trình biên dịch quản lý thông tin về kiểu
- Việc kiểm tra kiểu được thực hiện bởi bộ kiểm tra kiểu (type checker), một bộ phận của trình biên dịch

Ví dụ về kiểm tra kiểu

- Toán tử % của C chỉ thực hiện khi các toán hạng là số nguyên
- Chỉ có mảng mới có chỉ số và kiểu của chỉ số phải nguyên
- Một hàm phải có một số lượng tham số nhất định và các tham số phải đúng kiểu

Kiểm tra kiểu

- Có hai phương pháp tĩnh và động
- Phương pháp áp dụng trong thời gian dịch là tĩnh
- Trong các ngôn ngữ như C hay Pascal, kiểm tra kiểu là tĩnh và được dùng để kiểm tra tính đúng đắn của chương trình trước khi nó được thực hiện
- Kiểm tra kiểu tĩnh cũng được sử dụng khi xác định dung lượng bộ nhớ cần thiết cho các biến
- Bộ kiểm tra kiểu được xây dựng dựa trên
 - Các biểu thức kiểu của ngôn ngữ
 - Bộ luật để định kiểu cho các cấu trúc

Biểu thức kiểu (Type Expression)

Biểu diễn kiểu của một cấu trúc ngôn ngữ

Một biểu thức kiểu là một kiểu dữ liệu chuẩn hoặc được xây dựng từ các kiểu dữ liệu khác bởi cấu trúc kiểu (*Type Constructor*)

1. Kiểu dữ liệu chuẩn (int, real, boolean, char) là biểu thức kiểu
2. Biểu thức kiểu có thể liên hệ với một tên. Tên kiểu là biểu thức
3. Cấu trúc kiểu được ứng dụng vào các biểu thức kiểu tạo ra biểu thức kiểu

Cấu trúc kiểu

(a) Mảng (*Array*). Nếu T là biểu thức kiểu thì $array(I, T)$ là biểu thức kiểu biểu diễn một mảng với các phần tử kiểu T và chỉ số trong miền I

Ví dụ : `array [10] of integer` có kiểu `array(1..10,int)`;

(b) Tích Descarter Nếu T_1 và T_2 là các biểu thức kiểu thì tích Descarter $T_1 \times T_2$ là biểu thức kiểu

(c) Bản ghi (*Record*) Tương tự như tích Descarter nhưng chứa các tên khác nhau cho các kiểu khác nhau,

Ví dụ

```
struct
{
  double r;
  int i;
}
Có kiểu ((r x double) x (i x int))
```

Cấu trúc kiểu (tiếp)

- (d) *Con trỏ*: Nếu T là biểu thức kiểu thì $\text{pointer}(T)$ là biểu thức kiểu
- (e) *Hàm* Nếu D là miền xác định và R là miền giá trị của hàm thì kiểu của nó được biểu diễn là
biểu_thức : $D : R$.

Ví dụ hàm của C

$\text{int } f(\text{char } a, b)$

Có kiểu: $\text{char} \times \text{char} : \text{int}$.

Hệ thống kiểu (Type System)

- Tập các luật để xây dựng các biểu thức kiểu trong những phần khác nhau của chương trình
- Được định nghĩa thông qua định nghĩa tựa cú pháp
- Bộ kiểm tra kiểu thực hiện một hệ thống kiểu
- Ngôn ngữ định kiểu mạnh: Chương trình dịch kiểm soát được hết các lỗi về kiểu

Đặc tả một bộ kiểm tra kiểu

- Ngôn ngữ đơn giản với mỗi tên được liên kết với một kiểu
- Văn phạm thuộc tính để khai báo biểu thức

$P \rightarrow D;E$

$D \rightarrow D;D \mid \text{id} : T$

$T \rightarrow \text{char} \mid \text{int} \mid \text{array}[\text{num}] \text{ of } T \mid \uparrow T$

$E \rightarrow \text{literal} \mid \text{num} \mid \text{id} \mid E \text{ mod } E \mid E[E] \mid E\uparrow$

Thuộc tính

- Thuộc tính là khái niệm trừu tượng biểu diễn một đại lượng bất kỳ , chẳng hạn một số, một xâu, một vị trí trong bộ nhớ .
...
- Thuộc tính được gọi là **tổng hợp** nếu giá trị của nó tại một nút trong cây được xác định từ giá trị của các nút con của nó.
- Thuộc tính **kế thừa** là thuộc tính tại một nút mà giá trị của nó được định nghĩa dựa vào giá trị nút cha và/hoặc các nút anh em của nó.

Định nghĩa tựa cú pháp (syntax directed definition)

Định nghĩa tựa cú pháp là dạng tổng quát của văn phạm phi ngữ cảnh để đặc tả cú pháp của ngôn ngữ vào.

- Mỗi ký hiệu của văn phạm liên kết với một tập thuộc tính ,
- Mỗi sản xuất $A \rightarrow \alpha$ liên hệ với một tập các quy tắc ngữ nghĩa để tính giá trị thuộc tính liên kết với những ký hiệu xuất hiện trong sản xuất. Tập các quy tắc ngữ nghĩa có dạng

$$b = f(c_1, c_2, \dots, c_n)$$

f là một hàm và b thoả một trong hai yêu cầu sau:

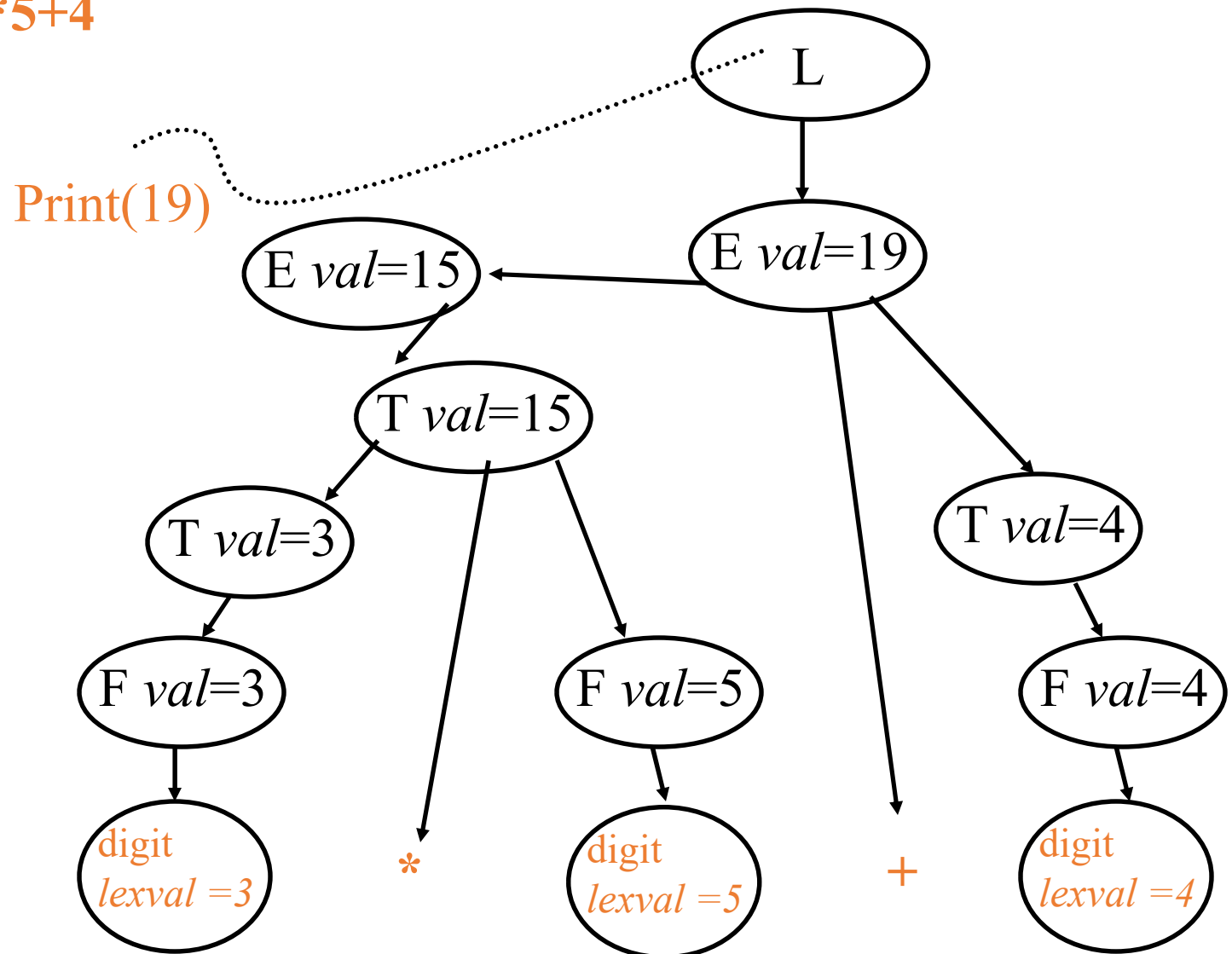
- b là một thuộc tính tổng hợp của A và c_1, \dots, c_n là các thuộc tính liên kết với các ký hiệu trong vế phải sản xuất $A \rightarrow \alpha$
- b là một thuộc tính thừa kế một trong những ký hiệu xuất hiện trong α , và c_1, \dots, c_n là thuộc tính của các ký hiệu trong vế phải sản xuất $A \rightarrow \alpha$

Ví dụ một định nghĩa tựa cú pháp với thuộc tính tổng hợp

Sản xuất	Quy tắc ngữ nghĩa
$L \rightarrow E \text{ return}$	Print (E.val)
$E \rightarrow E1 + T$	$E.val = E1.val + T.val$
$E \rightarrow T$	$E.val = T.val$
$T \rightarrow T1 * F$	$T.val = T1.val * F.val$
$T \rightarrow F$	$T.val = F.val$
$F \rightarrow (E)$	$F.val = E.val$
$F \rightarrow \text{digit}$	$F.val = \text{digit.Lexval}$

- Các ký hiệu E, T, F liên hệ với thuộc tính **tổng hợp** val
- Từ tố **digit** có thuộc tính tổng hợp lexval (Được bộ phân tích từ vựng đưa ra)
- Định nghĩa tựa cú pháp này nhằm tính ra giá trị của biểu thức xuất phát từ giá trị lưu trữ trong bảng ký hiệu của các nhân tử là số

Input $3*5+4$



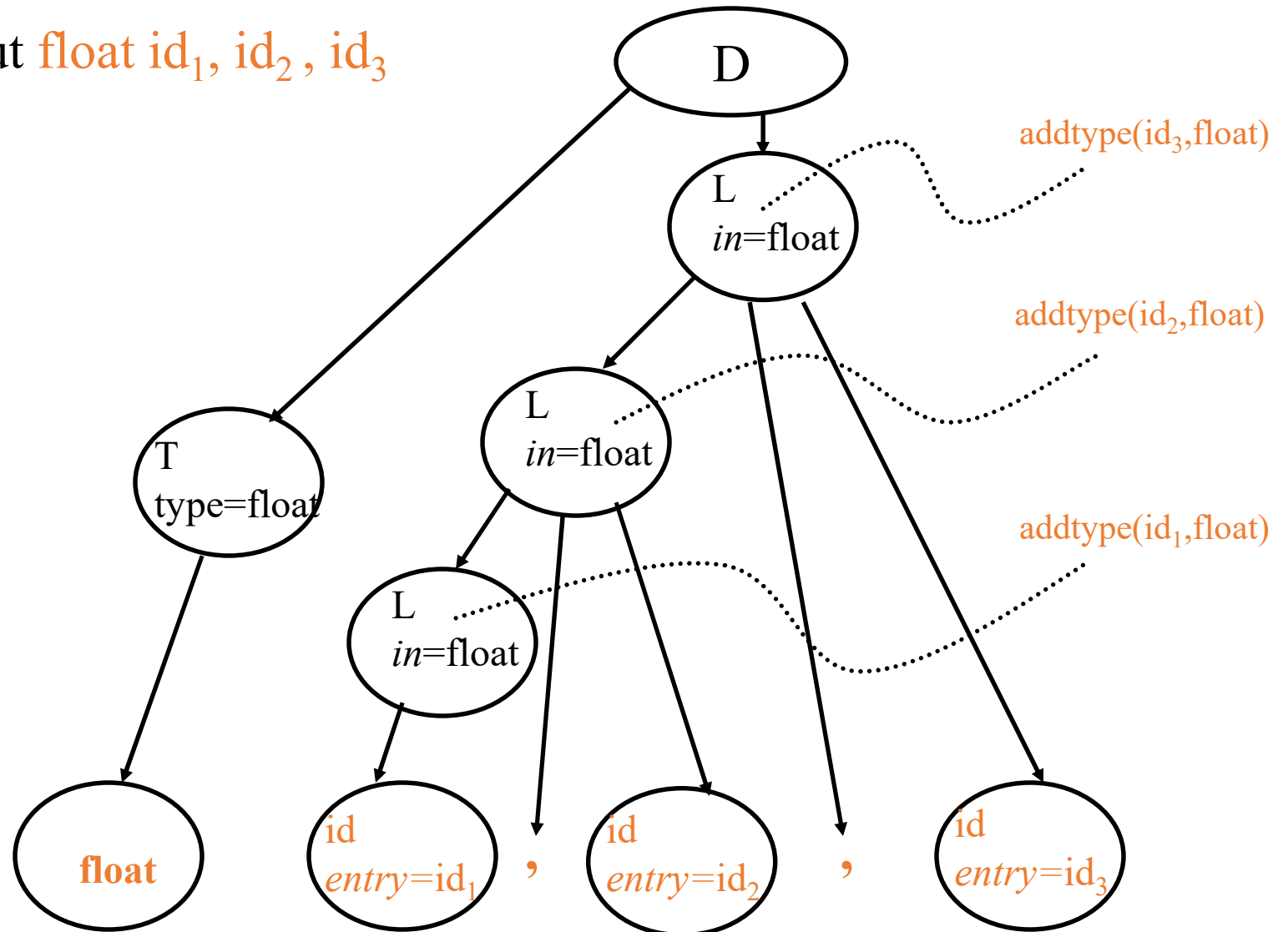
Ví dụ DTC với thuộc tính kế thừa

Sản xuất	Quy tắc ngữ nghĩa
$D \rightarrow T L$	$L.in = T.type$
$T \rightarrow int$	$T.type = integer$
$T \rightarrow float$	$T.type = float$
$L \rightarrow L_1, id$	$L1.in = L.in$ $addtype(id.entry, L.in)$
$L \rightarrow id$	$addtype(id.entry, L.in)$

- $L.in$ là thuộc tính kế thừa, chỉ kiểu của danh sách biến L
- Kiểu của danh sách biến L được tính từ kiểu của T
- Kiểu trong thuộc tính $L.in$ được truyền cho id cuối cùng và danh sách $L1$ gồm các id còn lại Thông tin về kiểu được lưu vào bảng ký hiệu
- Thông tin kiểu được lưu cho tất cả các biến trong danh sách

Ví dụ: Cây PTCPCCG cho ví dụ 2

Input **float** id_1, id_2, id_3



Ví dụ

Sản xuất	Quy tắc ngữ nghĩa
$L \rightarrow E \text{ return}$	Print (E.val)
$E \rightarrow E1 + T$	$E.val = E1.val + T.val$
$E \rightarrow T$	$E.val = T.val$
$T \rightarrow T1 * F$	$T.val = T1.val * F.val$
$T \rightarrow F$	$T.val = F.val$
$F \rightarrow (E)$	$F.val = E.val$
$F \rightarrow \text{digit}$	$F.val = \text{digit.Lexval}$

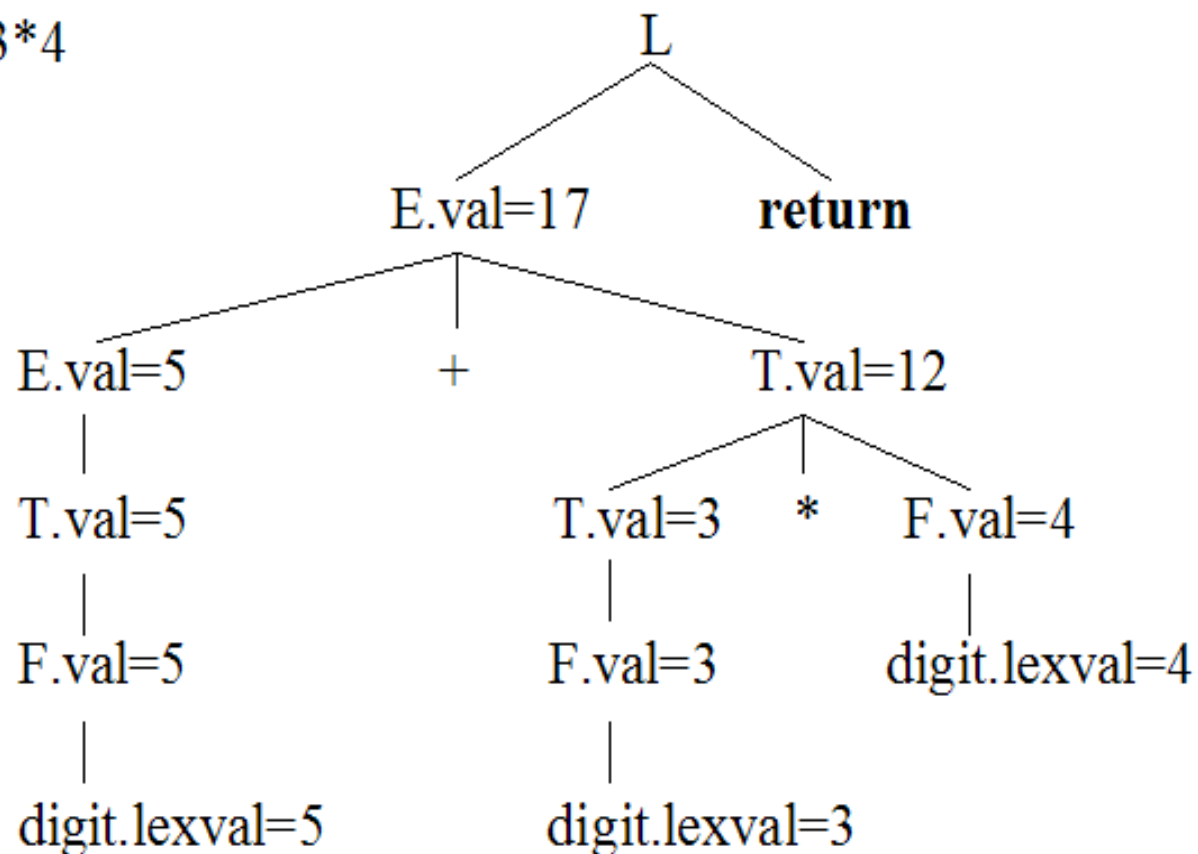
- Các ký hiệu E , T , F liên hệ với thuộc tính tổng hợp val
- Từ tố *digit* có thuộc tính tổng hợp lexval (Được bộ phân tích từ vựng đưa ra)

Cây phân tích cú pháp có chú giải

- Cây cú pháp có chỉ ra giá trị các thuộc tính tại mỗi nút được gọi là cây cú pháp có chú giải.

Ví dụ

Input: $5+3*4$



Bộ kiểm tra kiểu của định danh

SẢN XUẤT	QUY TẮC NGŨ NGHĨA
$D \rightarrow \text{id} : T$	$\text{addtype}(\text{id.entry}, T.\text{type})$
$T \rightarrow \text{char}$	$T.\text{type} := \text{char}$
$T \rightarrow \text{int}$	$T.\text{type} := \text{int}$
$T \rightarrow \uparrow T_1$	$T.\text{type} := \text{pointer}(T_1.\text{type})$
$T \rightarrow \text{array}[\text{num}] \text{ of } T_1$	$T.\text{type} := \text{array}(1..\text{num.val}, T_1.\text{type})$

Bộ kiểm tra kiểu của biểu thức

SẢN XUẤT	QUY TẮC NGŨ NGHĨA
$E \rightarrow \text{literal}$	$E.type := char$
$E \rightarrow \text{num}$	$E.type := int$
$E \rightarrow \text{id}$	$E.type := lookup(id.entry)$
$E \rightarrow E_1 \text{ mod } E_2$	$E.type := \text{if } E_1.type = int \text{ and } E_2.type = int$ then int else $type_error$
$E \rightarrow E_1[E_2]$	$E.type := \text{if } E_2.type = int \text{ and } E_1.type = array(s,t)$ then t else $type_error$
$E \rightarrow E_1 \uparrow$	$E.type := \text{if } E_1.type = pointer(t) \text{ then } t$ else $type_error$

Bộ kiểm tra kiểu của lệnh

SẢN XUẤT	QUY TẮC NGŨ NGHĨA
$S \rightarrow \text{id} := E$	$S.type := \text{if } \text{id}.type = E.type \text{ then } \text{void}$ $\text{else } \text{type_error}$
$S \rightarrow \text{if } E \text{ then } S_1$	$S.type := \text{if } E.type = \text{boolean} \text{ then } S_1.type$ $\text{else } \text{type_error}$
$S \rightarrow \text{while } E \text{ do } S_1$	$S.type := \text{if } E.type = \text{boolean} \text{ then } S_1.type$ $\text{else } \text{type_error}$
$S \rightarrow S_1; S_2$	$S.type := \text{if } S_1.type = \text{void} \text{ and } S_2.type = \text{void}$ $\text{then } \text{void}$ $\text{else } \text{type_error}$

Bộ kiểm tra kiểu của hàm

SẢN XUẤT	QUY TẮC NGŨ NGHĨA
$D \rightarrow \text{id} : T$	$\text{addtype}(\text{id.entry}, T.\text{type}); D.\text{type} := T.\text{type}$
$D \rightarrow D_1; D_2$	$D.\text{type} := D_1.\text{type} \times D_2.\text{type}$
$\text{Fun} \rightarrow \text{fun id}(D) : T; B$	$\text{addtype}(\text{id.entry}, D.\text{type}:T.\text{type})$
$B \rightarrow \{S\}$	
$S \rightarrow \text{id}(EList)$	$E.\text{type} := \text{if } \text{lookup}(\text{id.entry})=t_1 : t_2 \text{ and } EList.\text{type}=t_1$ then t_2 else type_error
$EList \rightarrow E$	$EList.\text{type} := E.\text{type}$
$EList \rightarrow EList, E$	$EList.\text{type} := EList_1.\text{type} \times E.\text{type}$

Hàm kiểm tra biểu thức kiểu tương đương

```
function sequiv(s, t): boolean;  
begin  
    if s và t là cùng kiểu dữ liệu chuẩn then  
        return true;  
    else if s = array(s1, s2) and t = array(t1, t2) then  
        return sequiv(s1, t1) and sequiv(s2, t2)  
    else if s = s1 x s2 and t = t1 x t2) then  
        return sequiv(s1, t1) and sequiv(s2, t2)  
    else if s = pointer(s1) and t = pointer(t1) then  
        return sequiv(s1, t1)  
    else if s = s1 → s2 and t = t1 → t2 then  
        return sequiv(s1, t1) and sequiv(s2, t2)  
    else  
        return false;  
end;
```

Chuyển đổi kiểu

- Kiểu của $x+i$ với
 x kiểu real
 i kiểu int

Khi dịch sang lệnh máy, phép cộng với kiểu real và kiểu int có mã lệnh khác nhau

- Tùy ngôn ngữ và bộ luật chuyển đổi sẽ quy đổi các toán hạng về một trong hai kiểu

Chuyển đổi kiểu trong biểu thức

SẢN XUẤT	QUY TẮC NGŨ NGHĨA
$E \rightarrow \text{num}$	$E.type := int$
$E \rightarrow \text{num.num}$	$E.type := real$
$E \rightarrow \text{id}$	$E.type := \text{lookup}(\text{id.entry})$
$E \rightarrow E_1 \text{ op } E_2$	$E.type :=$ if $E_1.type = int$ and $E_2.type = int$ then int else if $E_1.type = int$ and $E_2.type = real$ then $real$ else if $E_1.type = real$ and $E_2.type = int$ then $real$ else if $E_1.type = real$ and $E_2.type = real$ then $real$ else $type_error$