Análise semântica (continuação)

Função, interação com o compilador Tabela de símbolos Análise semântica

Prof. Thiago A. S. Pardo

Estrutura geral de um compilador programa-fonte analisador léxico Tabela de símbolos analisador sintático analisador semântico Manipulação de erros Tabela de palavras e gerador de código intermediário símbolos reservados otimizador de código gerador de código programa-alvo dados de saída entrada

Gramática de atributos

- Gramática de atributos
 - Método usualmente utilizado para formalização semântica
 - Conjunto de atributos e regras semânticas para uma gramática
 - Cada regra sintática/gramatical pode ter regras semânticas associadas
 - Atributos associados aos símbolos gramaticais
 - Por exemplo, valor e escopo
 - □ x.valor, x.escopo
 - Regras semânticas que manipulam os atributos
 - Por exemplo, regra para somar os atributos valores de duas variáveis
 - □ x:=a+b, cuja regra é x.valor:=a.valor+b.valor

3

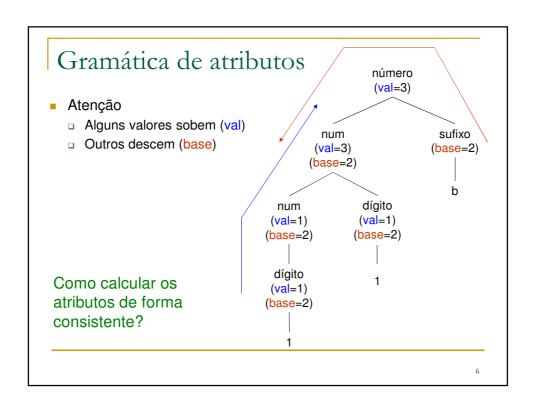
Gramática de atributos

Considerando a gramática abaixo

número \rightarrow num sufixo sufixo \rightarrow b | d num \rightarrow num dígito | dígito dígito \rightarrow 0|1|2|3|4|5|6|7|8|9

Grar	mática	de	atri	butos

Regras gramaticais	Regras semânticas
número → num sufixo	número.val = num.val
	num.base = sufixo.base
sufixo → b	sufixo.base = 2
sufixo → d	sufixo.base = 10
$num_1 \rightarrow num_2 digito$	num ₁ .val =
	if dígito.val = erro or num ₂ .val=erro then erro
	else num ₂ .val * num ₁ .base + dígito.val
	num ₂ .base = num ₁ .base
	dígito.base = num ₁ .base
num → dígito	num.val = dígito.val
	dígito.base = num.base
dígito → 0	dígito.val = 0
dígito → 1	dígito.val = 1
dígito → 2	dígito.val =
	if dígito.base=2 then erro else 2

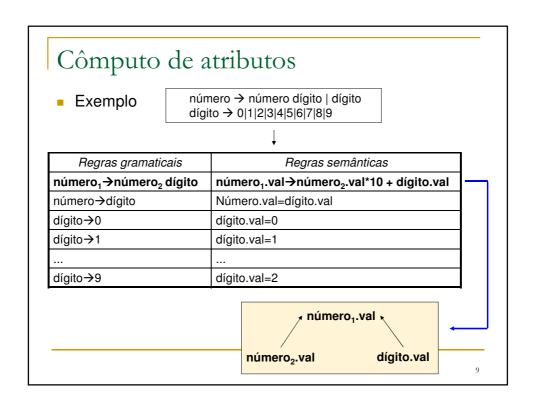


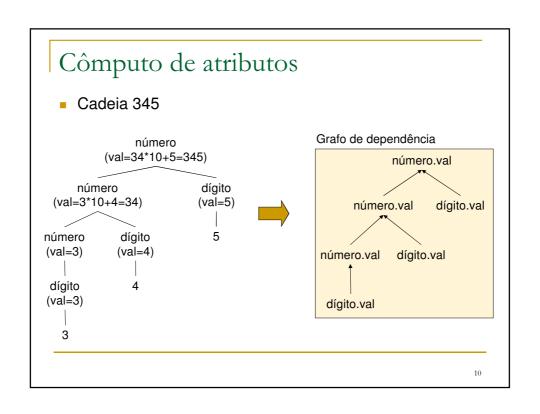
- Com base na árvore sintática explícita
 - Grafos de dependência
 - Compilador de mais de uma passagem
- Ad hoc
 - Análise semântica "comandada" pela análise sintática
 - Compilador de uma única passagem

7

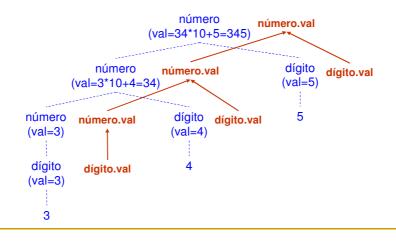
Cômputo de atributos

- Grafos de dependência
 - Especificam a ordem de cômputo dos atributos de cada regra gramatical em uma árvore sintática
 - Portanto, um grafo associado a cada regra gramatical
 - Para uma cadeia da linguagem, tem-se um grafo composto por todos os subgrafos





Grafo amarrado à árvore sintática



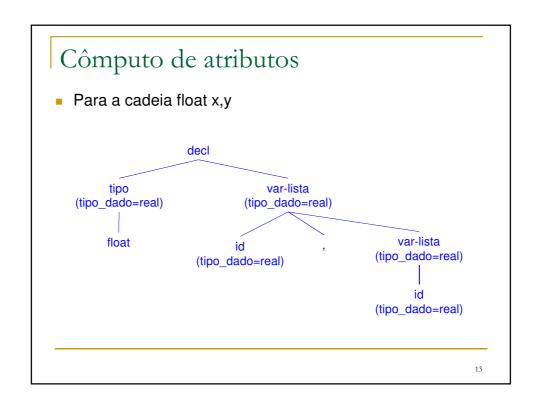
11

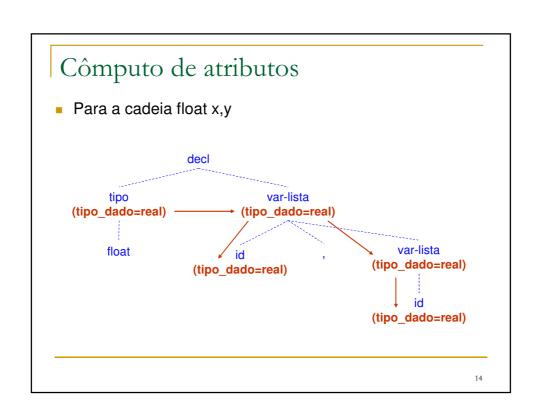
Cômputo de atributos

Considere a gramática abaixo e sua gramática de atributos

```
decl → tipo var-lista
tipo → int | float
var-lista → id, var-lista | id
```

Regras gramaticais	Regras semânticas		
decl → tipo var-lista	var-lista.tipo_dado = tipo.tipo_dado		
tipo → int	tipo.tipo_dado = integer		
tipo → float	tipo.tipo_dado = real		
$var-lista_1 \rightarrow id, var-lista_2$	id.tipo_dado = var-lista ₁ .tipo_dado		
	var-lista ₂ .tipo_dado = var-lista ₁ .tipo_dado		
var-lista → id	id.tipo_dado=var-lista.tipo_dado		





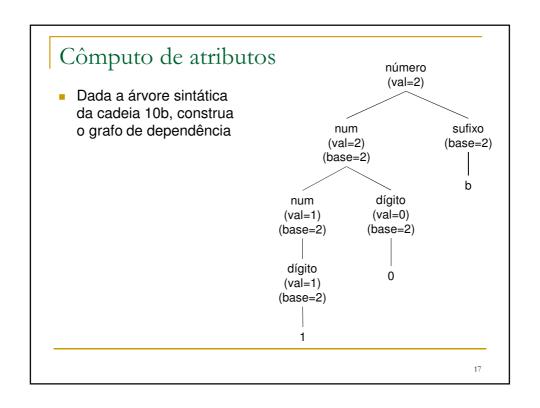
 Exercício: considere a gramática abaixo e sua gramática de atributos

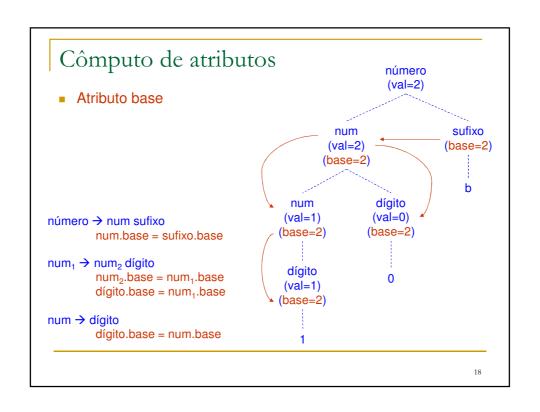
> número \rightarrow num sufixo sufixo \rightarrow b | d num \rightarrow num dígito | dígito dígito \rightarrow 0|1|2|3|4|5|6|7|8|9

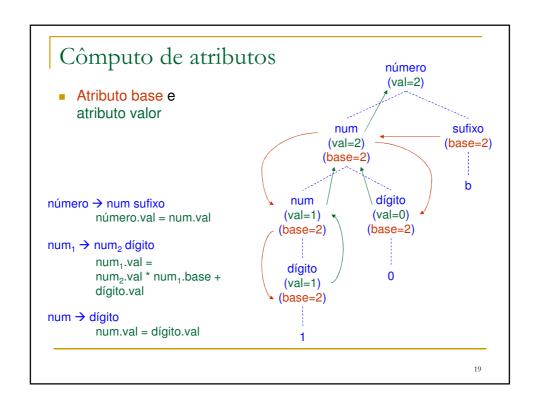
> > 15

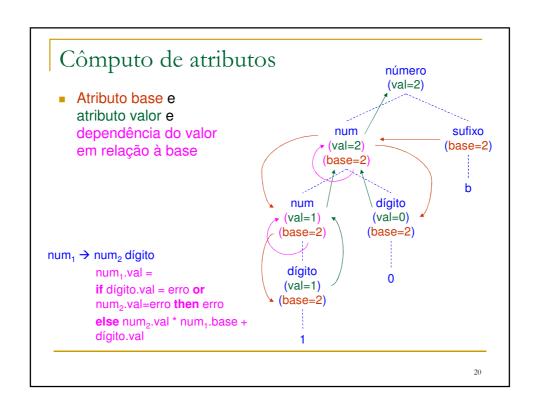
Cômputo de atributos

Regras gramaticais	Regras semânticas
número → num sufixo	número.val = num.val
	num.base = sufixo.base
sufixo → b	sufixo.base = 2
sufixo → d	sufixo.base = 10
num₁ → num₂ dígito	num ₁ .val =
	if dígito.val = erro or num ₂ .val=erro then erro
	else num ₂ .val * num ₁ .base + dígito.val
	num_2 .base = num_1 .base
	dígito.base = num ₁ .base
num → dígito	num.val = dígito.val
	dígito.base = num.base
dígito → 0	dígito.val = 0
dígito → 1	dígito.val = 1
dígito → 2	dígito.val =
1	if dígito.base=2 then erro else 2









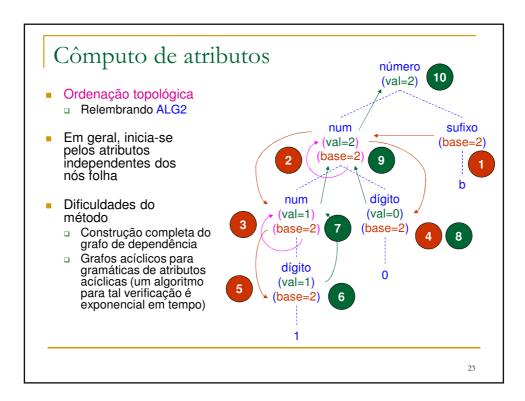
Dois tipos de atributos

- Atributos herdados: dependências de pais para filhos e/ou entre irmãos (p.ex., atributo base da gramática anterior)
- Atributos sintetizados: dependências apontam de filhos para pais (p.ex., atributo val da gramática anterior)
 - Uma gramática que só tem atributos sintetizados é denominada gramática S-atribuída

21

Cômputo de atributos

- Ordem de cômputo dos atributos
 - Os atributos que n\u00e3o dependem de outros atributos devem ser computados primeiro, logicamente
 - Opções
 - Ordenação topológica do grafo de dependências: método de árvore de análise sintática
 - Manualmente, determinado pelo projetista do compilador: método baseado em regras



- Método baseado em regras
 - Adotado em praticamente todos os compiladores
 - O projetista do compilador analisa a gramática de atributos e seus grafos de dependência e determina a ordem de cômputo dos atributos
 - Em geral, n\u00e3o \u00e9 muito complicado de se fazer
 - Para cada regra gramatical, definição do percurso realizado no trecho correspondente na árvore sintática
 - Possibilidades: percurso <u>em-ordem</u>, <u>pré-ordem</u>, <u>pós-ordem</u> ou arbitrário

- Exemplos de percursos para uma árvore binária (relembrando ALG1)
 - Percurso em-ordem
 - Filho esquerdo, raiz (nó pai, isto é, lado esquerdo da regra em foco), filho direito
 - Percurso pré-ordem
 - Raiz, filho esquerdo, filho direito
 - Percurso pós-ordem
 - Filho esquerdo, filho direito, raiz

25

Cômputo de atributos

 Exemplo: sub-rotina para computar o atributo tipo_dado da gramática abaixo

decl → tipo var-lista tipo → int | float var-lista → id, var-lista | id

Regras gramaticais	Regras semânticas		
decl → tipo var-lista	var-lista.tipo_dado = tipo.tipo_dado		
tipo → int	tipo.tipo_dado = integer		
tipo → float	tipo.tipo_dado = real		
var-lista ₁ → id, var-lista ₂	id.tipo_dado = var-lista ₁ .tipo_dado		
	var-lista ₂ .tipo_dado = var-lista ₁ .tipo_dado		
var-lista → id	id.tipo_dado=var-lista.tipo_dado		

 Exemplo: sub-rotina para computar o atributo tipo_dado da gramática abaixo

```
procedure AvalTipo(T: nó árvore);
begin
   case nó de T of
          AvalTipo(tipo):
          var-lista.tipo_dado:=tipo.tipo_dado;
          AvalTipo(var-lista);
                                                                       pré-ordem
          if filho de T=int then T.tipo dado:=inteiro
          else T.tipo dado:=real;
    var-lista:
          atribui T.tipo_dado a primeiro filho de T
          if terceiro filho de T não é NIL then
                    atribui T.tipo_dado a terceiro filho;
                    AvalTipo(terceiro filho de T);
                                                            decl → tipo var-lista
    end;
                                                            tipo → int | float
end:
                                                            var-lista → id, var-lista | id
```

Cômputo de atributos

- Opcionalmente, valores de atributos podem ser associados a parâmetros ou valores de retorno de subrotinas de cômputo de atributos em vez de serem armazenados nos nós de uma árvore sintática
 - Interessante para a situação em que muitos atributos são usados apenas temporariamente ou como suporte para cômputo de outros atributos
 - Normalmente, atributos herdados são passados via parâmetros e atributos sintetizados via valor de retorno

Exemplo

número → num sufixo sufixo → b | d num → num dígito | dígito dígito → 0|1|2|3|4|5|6|7|8|9

29

Cômputo de atributos

Regras gramaticais	Regras semânticas
número → num sufixo	número.val = num.val
	num.base = sufixo.base
sufixo → b	sufixo.base = 2
sufixo → d	sufixo.base = 10
num₁ → num₂ dígito	num ₁ .val =
	<pre>if digito.val = erro or num₂.val=erro then erro</pre>
	else num ₂ .val * num ₁ .base + dígito.val
	num_2 .base = num_1 .base
	dígito.base = num₁.base
num → dígito	num.val = dígito.val
	dígito.base = num.base
dígito → 0	dígito.val = 0
dígito → 1	dígito.val = 1
dígito → 2	dígito.val =
	if dígito.base=2 then erro else 2

```
function AvalComBase(T: nó_árvore; base: inteiro): inteiro;
var temp, temp2: inteiro;
begin
   case nó de T of
        número:
                 temp:=AvalComBase(filho à direita de T,0);
                 return AvalComBase(filho à esquerda de T,temp);
         num:
                 temp:=AvalComBase(filho à esquerda de T,base);
                 if filho à direita de T não é NIL then
                          temp2:=AvalComBase(filho à direita de T,base);
                          if temp<>erro and temp2<>erro then
                                   return base*temp+temp2
                          else return erro;
                 else return temp;
                                                             número → num sufixo
         sufixo:
                                                             sufixo \rightarrow b | d
                                                             num → num dígito | dígito
                 if filho de T=b then return 2
                                                             dígito \rightarrow 0|1|2|3|4|5|6|7|8|9
                 else return 10;
         dígito:
                 if base=2 and filho de T>1 then return erro
                 else return numval(filho de T);
   end;
end;
```

- Com base na árvore sintática explícita
 - Grafos de dependência
 - Compilador de mais de uma passagem
- Ad hoc
 - Análise semântica "comandada" pela análise sintática
 - Compilador de uma única passagem

- Estruturas de dados externas
 - Em vez de se armazenar os atributos na árvore sintática ou de manipulá-los via parâmetros e valores de retornos, os atributos podem ser armazenados em estruturas separadas
 - Variáveis globais
 - Listas
 - Tabelas
 - Em compilação, a tabela de símbolos é utilizada, em geral

33

Tabela de símbolos

- Estrutura principal da compilação
- Captura a sensitividade ao contexto e as ações executadas no decorrer do programa
- Pode estar atrelada a todas as etapas da compilação
- Permite a realização da análise semântica
- Fundamental na geração de código

- Permite saber durante a compilação de um programa o tipo e o valor de seus elementos (números e identificadores), escopo destes, número e tipo dos parâmetros de um procedimento, etc.
 - Cada token tem atributos/informações diferentes associadas

Cadeia	Token	Categoria	Tipo	Valor	
i	id	var	integer	1	
fat	id	proc	-	-	
2	num	-	integer	2	

35

Tabela de símbolos

- Exemplo de atributos de identificador de variável
 - Tipo de variável (inteira, real, etc.), nome da variável, endereço na memória, escopo (programa principal, função, etc.), etc.
- Para vetor, ainda seriam necessários atributos de <u>tamanho</u> do vetor, o valor de seus limites, etc.

- Principais operações efetuadas
 - Inserir: armazena na tabela informações fornecidas pelas declarações no programa
 - Busca: recupera da tabela informações de um elemento declarado no programa quando esse elemento é utilizado
 - Remover: remove (ou torna inacessível) da tabela informações sobre um elemento declarado que não se mostra mais necessário no programa
- As especificidades dessas operações são dependentes da linguagem de programação em questão

37

Tabela de símbolos

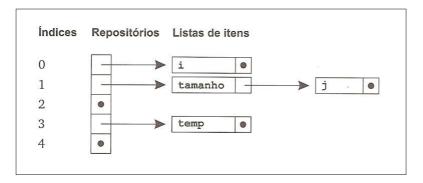
- A tabela é acessada pelo compilador sempre que um elemento é mencionado no programa
 - Verificar ou incluir sua declaração
 - Verificar seu tipo, escopo ou alguma outra informação
 - Atualizar alguma informação associada ao identificador (por exemplo, valor)
 - Remover um elemento quando este não se faz mais necessário ao programa

- Estrutura da tabela de símbolos: determinada pela eficiência das operações de inserir, verificar e remover
- Várias possibilidades
 - Implementação
 - Estática
 - Dinâmica: melhor opção
 - Estrutura
 - Listas, matrizes
 - Árvores de busca (por exemplo, B e AVL)
 - Acesso
 - Sequencial, busca binária, etc.
 - Hashing: opção mais eficiente
 - O elemento do programa é a chave e a função hash indica sua posição na tabela de símbolos
 - Necessidade de tratamento de colisões

39

Tabela de símbolos

 Exemplo de hashing com resolução de colisões para a inclusão dos identificadores i, j, tamanho e temp



- Questões de projeto
 - Tamanho da tabela: tipicamente, de algumas centenas a mil "linhas"
 - Dependente da forma de implementação
 - Na implementação dinâmica, não é necessário se preocupar tanto com isso
 - Uma única tabela para todas as declarações ou várias tabelas, sendo uma para cada tipo de declaração (constantes, variáveis, tipos, procedimentos e funções)
 - Por quê?

41

Tabela de símbolos

- Questões de projeto
 - Tamanho da tabela: tipicamente, de algumas centenas a mil "linhas"
 - Dependente da forma de implementação
 - Na implementação dinâmica, não é necessário se preocupar tanto com isso
 - Uma única tabela para todas as declarações ou várias tabelas, sendo uma para cada tipo de declaração (constantes, variáveis, tipos, procedimentos e funções)
 - Diferentes declarações têm diferentes informações/atributos (por exemplo, variáveis não têm número de argumentos, enquanto procedimentos têm)

- Representação de escopo de identificadores do programa: várias tabelas ou uma única tabela com a identificação do escopo (como um atributo ou por meio de listas ligadas, por exemplo) para cada identificador
 - Tratamento de escopo
 - Inserção de identificadores de mesmo nome, mas em níveis diferentes
 - □ Remoção de identificadores cujos escopos deixaram de existir
 - Em geral, na maioria das linguagens de programação, aplica-se a "regra do aninhamento mais próximo"

43

Tabela de símbolos

- Possibilidades para tratamento de escopos
 - Inclusão de um campo a mais na tabela de símbolos indicando o nível da variável no programa
 - Controle do nível durante a compilação do programa
 - □ Quando se chama um procedimento (ou função), faz-se nível=nível+1
 - □ Quando se sai de um procedimento (ou função), faz-se nível=nível-1
 - Associação das variáveis locais a um procedimento (ou função) à entrada da tabela para o procedimento (ou função) por meio, por exemplo, de uma lista encadeada
 - Atenção: para a checagem de tipos, deve-se saber quantos são e quais são os parâmetros de um procedimento (ou função) na tabela de símbolos
 - Tabelas diferentes para diferentes escopos

- Tratamento de escopo
 - Como diferenciar variáveis globais de locais
 - Tratamento de variáveis de mesmo nome, mas de escopos diferentes

```
program meu_prog
   procedure meu_proc(x: integer)
   var y: real
   begin
      read(y);
      x:=x+y
   end;
var x, y: integer
begin
   read(y);
   x:=x*y
end.
```

45

Exercício

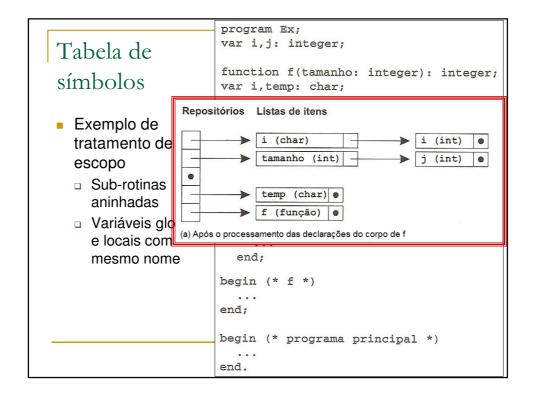
Gere a tabela de símbolos para o programa abaixo

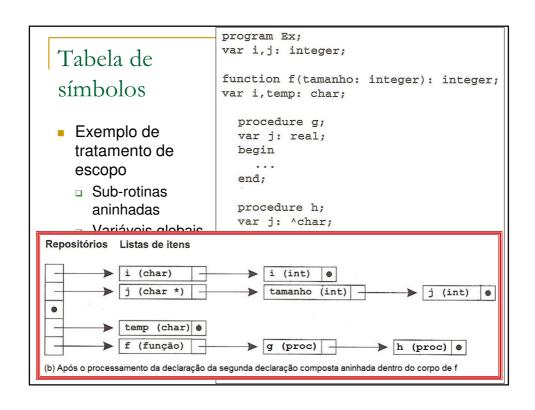
```
program meu_prog
  procedure meu_proc(x: integer)
  var y: real
  begin
    read(y);
    x:=x+y
  end;
var x, y: integer
begin
  read(y);
  x:=x*y
end.
```

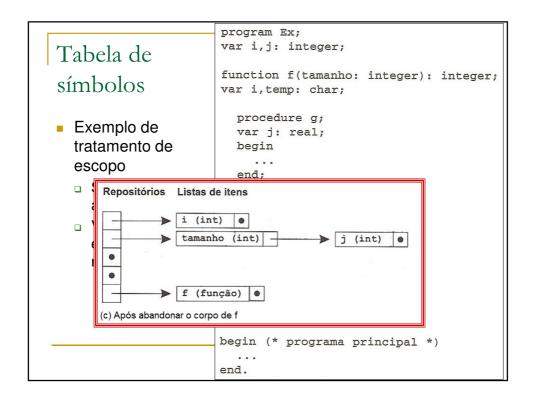
```
program Ex;
                       var i,j: integer;
Tabela de
                       function f(tamanho: integer): integer;
símbolos
                       var i, temp: char;
                         procedure g;
Exemplo de
                         var j: real;
  tratamento de
                         begin
                           . . .
  escopo
                         end;
  Sub-rotinas
     aninhadas
                         procedure h;
                         var j: ^char;

    Variáveis globais

                         begin
     e locais com
                         end;
     mesmo nome
                       begin (* f *)
                         . . .
                       end;
                       begin (* programa principal *)
                       end.
```





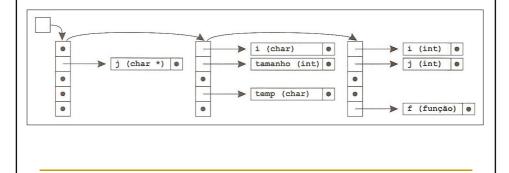


- Comportamento de <u>pilha</u>
 - Insere as declarações mais recentes, ocultando as antigas
 - Remove as mais recentes, voltando ao escopo anterior
 - Acesso às mais recentes

51

Tabela de símbolos

- Alternativa para tabela anterior: tabelas separadas para cada escopo
 - Mudar o escopo requer apenas a mudança do ponteiro



 A tabela de símbolos pode ser utilizada para armazenar as palavras reservadas e símbolos especiais da linguagem, podendo dispensar o uso da tabela de palavras e símbolos reservados

53

Tabela de símbolos

- Descritores
 - Registros (campos) que formam a tabela de símbolos
 - Armazenam as informações dos números e identificadores
- Diferentes identificadores têm diferentes descritores
 - Tem que se levar isso em consideração no projeto da tabela de símbolos para sua otimização
 - Possibilidade de se usar uniões para o caso de uma mesma tabela para tudo

- Inserção de elementos na tabela
 - Associação de regras semânticas às regras gramaticais
 - Verificar se o elemento já não consta na tabela
- Busca de informação na tabela
 - Realizada antes da inserção
 - Busca de informações para análise semântica
- Remoção de elementos da tabela
 - Tornar inacessíveis dados que não são mais necessários (por exemplo, após o escopo ter terminado)
 - Linguagens que permitem estruturação em blocos

55

Tabela de símbolos

- As sub-rotinas de inserção, busca e remoção podem ser inseridas diretamente na gramática de atributos
 - Explicitamente, via chamadas de sub-rotinas de manipulação da tabela
 - Compilação em mais de uma passagem, se árvore sintática é a base para a análise
 - Opcionalmente, compilação de <u>uma única passagem</u>, sendo a gramática de atributos utilizada apenas para a descrição da semântica

Exemplo: decl → tipo var-lista

tipo \rightarrow int | float

var-lista → id, var-lista | id

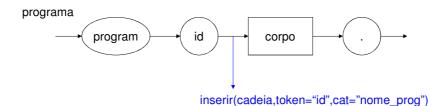
Regras gramaticais	Regras semânticas		
decl → tipo var-lista	var-lista.tipo_dado = tipo.tipo_dado		
tipo → int	tipo.tipo_dado = integer		
tipo → float	tipo.tipo_dado = real		
var-lista₁ → id, var-lista₂	id.tipo_dado = var-lista ₁ .tipo_dado var-lista ₂ .tipo_dado = var-lista ₁ .tipo_dado If busca(id)=FALSE then inserir(id,id.tipo_dado) else ERRO("identificador já declarado")		
var-lista → id	id.tipo_dado=var-lista.tipo_dado If busca(id)=FALSE then inserir(id,id.tipo_dado) else ERRO("identificador já declarado")		

Tabela de símbolos

- As sub-rotinas de inserção, busca e remoção podem ser inseridas diretamente na análise sintática
 - □ Solução *ad hoc*
 - Compilação de uma única passagem

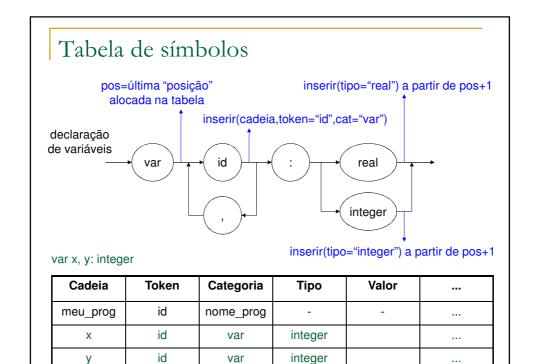


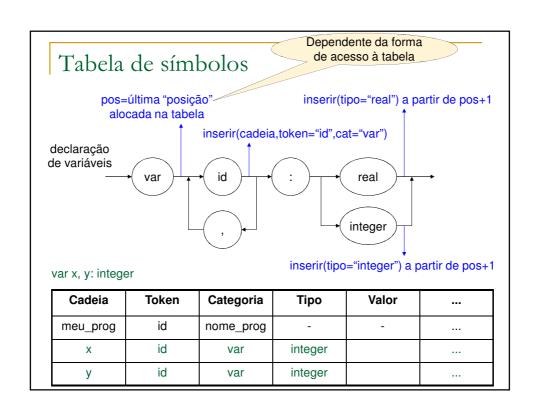
- Inserção de elementos na tabela
 - Declaração, principalmente

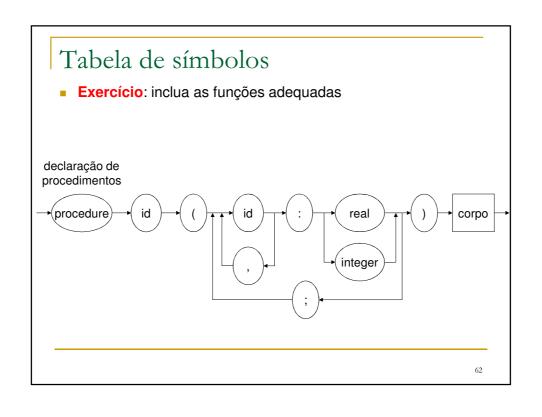


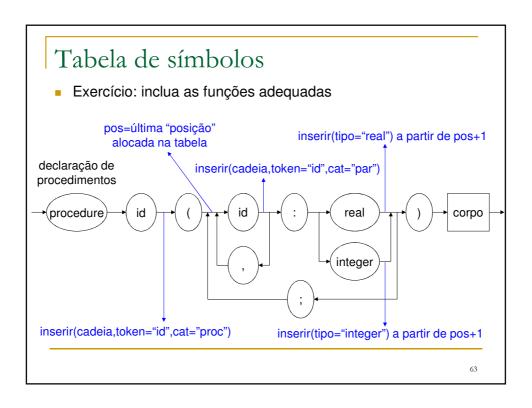
program meu_prog ...

Cadeia	Token	Categoria	Tipo	Valor	
meu_prog	id	nome_prog	-	-	





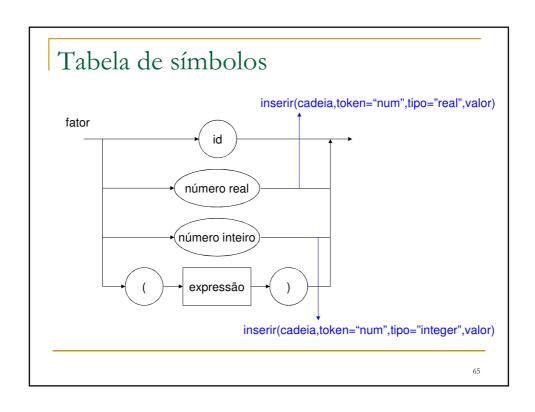




Exercício: inclua as funções adequadas

procedure meu_proc(a: integer; b,c: real) ...

Cadeia	Token	Categoria	Tipo	Valor	
meu_prog	id	nome_prog	-	-	
Х	id	var	integer		
У	id	var	integer		
meu_proc	id	proc	-	-	
а	id	par	integer		
b	id	par	real		
С	id	par	real		



i:=1

Cadeia	Token	Categoria	Tipo	Valor	
meu_prog	id	nome_prog	-	-	
Х	id	var	integer		
у	id	var	integer		
meu_proc	id	proc	-	-	
а	id	par	integer		
b	id	par	real		
С	id	par	real		
1	num	-	integer	1	

Exemplo de procedimento

```
inserir(cadeia,token="id",cat="nome_prog")
 programa
                               id
              program
                                            corpo
procedimento programa(Seg)
        se (simbolo=program) então obtem simbolo(cadeia,simbolo)
        senão ERRO(Seg+{id});
        se (simbolo=id) então
                inserir(cadeia, "id", "nome_prog")
                obtem_simbolo(cadeia,simbolo)
        senão ERRO(Seg+P(corpo));
        corpo(Seg+{.});
        se (simbolo=simb_ponto) então obtem_simbolo(cadeia,simbolo)
        senão ERRO(Seg);
fim
```

67

Tabela de símbolos

- Busca de informação
 - □ Sempre que um elemento do programa é utilizado
 - fator e comando
 - Verifica-se se foi declarado, seu tipo, etc.

