

## Linguagens e Gramáticas Livre do Contexto

- Agora será considerada uma classe de linguagens maior que as regulares, uma vez que a maioria das linguagens não são regulares;
- Linguagens Livre do Contexto (LCC) possuem um papel importante para as linguagens naturais desde 1950 e para compiladores desde 1960;
- Gramáticas Livre do Contexto (GLC) constituem a base da sintaxe BNF;
- Recentemente, LLC têm sido usadas para descrever formatos de documentos via DTD (*document-type definition*) usado no XML (*extensible markup language*).

## GLC – Exemplo Informal

- Considere  $L_{\text{pal}} = \{w \in \Sigma^* \mid w = w^R\}$ . Por exemplo otto, madamimadam  $\in L_{\text{pal}}$ .
- Seja  $L = \{0,1\}$  e supõe-se que  $L_{\text{pal}}$  é regular. Seja  $n$  dado pelo lema do bombeamento. Então  $0^n 1 0^n \in L_{\text{pal}}$ . Ainda lendo  $0^n$  o AF deve fazer um loop; ao omitir-se o loop, tem-se que a palavra não pertence a linguagem. Contradição!!
- Define-se  $L_{\text{pal}}$  recursivamente:
  - Base:  $\epsilon, 0, 1$  são palíndromos;
  - Passo Indução: Se  $w$  é um palíndromo, então tb são  $0w0$  e  $1w1$ ;Nada mais é um palíndromo.

## GLC – Exemplo

- GLC é um mecanismo formal de definições de linguagens livre do contexto. Por exemplo  $L_{\text{pal}}$ .
  1.  $P \rightarrow \epsilon$
  2.  $P \rightarrow 0$
  3.  $P \rightarrow 1$
  4.  $P \rightarrow 0P0$
  5.  $P \rightarrow 1P1$
- 0 e 1 são símbolos terminais;
- P é uma variável (não-terminal);
- P é o símbolo inicial desta gramática;
- 1-5 são regras de produção.

## GLC – Definição Formal

- Uma gramática livre do contexto é uma quádrupla:

$$G = (V, T, P, S)$$

Onde:

- V é um conjunto finito de variáveis;
- T é um conjunto finito de símbolos terminais;
- P é um conjunto finito de produções da forma  $A \rightarrow \alpha$ , onde A é uma variável e  $\alpha \in (V \cup T)^*$ .
- S é a variável inicial.

## Exemplos GLC

- Exemplo 1 (palíndromos):**  
 $G_{pal} = (\{P\}, \{0, 1\}, A, P)$ , onde  
 $A = \{P \rightarrow \varepsilon, P \rightarrow 1, P \rightarrow 0, P \rightarrow 0P0, P \rightarrow 1P1\}$ . Ou, de maneira similar,  
 $A = \{P \rightarrow \varepsilon \mid 1 \mid 0 \mid 0P0 \mid 1P1\}$ .
- Exemplo 2 (expressões regulares):** ER's sobre o alfabeto  $\{0, 1\}$  pode ser definidas pela  
 $G_{exreg} = (\{E\}, \{0, 1\}, A, E)$ , onde  
 $A = \{E \rightarrow 0 \mid 1 \mid E.E \mid E+E \mid E^* \mid (E)\}$
- Exemplo 3 (expressões simples):**  $L((a+b)^*(a+b+0+1))$ . Tais expressões são definidas pela gramática:  
 $G = (\{E, I\}, \{+, x, (, ), a, b, 0, 1\}, P, E)$ , onde  
 $P = \{E \rightarrow I \mid E+E \mid ExE \mid (E), I \rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1\}$ .

## Derivações

- Inferência Recursiva: usando produções do corpo para cabeça;
- Derivações: usando produções da cabeça para o corpo.
- Exemplo de Inferência Recursiva:

|        | String        | Lang | Prod | String(s) used |
|--------|---------------|------|------|----------------|
| (i)    | a             | I    | 5    | -              |
| (ii)   | b             | I    | 6    | -              |
| (iii)  | b0            | I    | 9    | (ii)           |
| (iv)   | b00           | I    | 9    | (iii)          |
| (v)    | a             | E    | 1    | (i)            |
| (vi)   | b00           | E    | 1    | (iv)           |
| (vii)  | a + b00       | E    | 2    | (v), (vi)      |
| (viii) | (a + b00)     | E    | 4    | (vii)          |
| (ix)   | a * (a + b00) | E    | 3    | (v), (viii)    |

## Derivações

- Seja  $G = (V, T, P, S)$  uma GLC,  $A \in V$ ,  $\{\alpha, \beta\} \subset (V \cup T)^*$ , e  $A \rightarrow \gamma \in P$ .

Escreve-se:

$$\alpha A \beta \Rightarrow_G \alpha \gamma \beta, \text{ ou simplesmente, } \alpha A \beta \Rightarrow \alpha \gamma \beta$$

Isto é,  $\alpha A \beta$  deriva  $\alpha \gamma \beta$ .

**Definição Indutiva ( $\Rightarrow^*$ )** fecho reflexivo e transitivo de  $\Rightarrow$ .

- Base:** seja  $\alpha \in (V \cup T)^*$ . Então  $\alpha \Rightarrow^* \alpha$
- Indução:** se  $\alpha \Rightarrow^* \beta$  e  $\beta \Rightarrow \gamma$ , então  $\alpha \Rightarrow^* \gamma$

## Exemplo Derivação

- Derivação de  $a x (a + b00)$  a partir de E para a gramática:  
 $G = (\{E, I\}, \{+, x, (, ), a, b, 0, 1\}, P, E)$ , onde  
 $P = \{E \rightarrow I \mid E+E \mid ExE \mid (E), I \rightarrow a \mid b \mid Ia \mid Ib \mid I0 \mid I1\}$ .

$$\begin{aligned} E &\Rightarrow E x E \Rightarrow I x E \Rightarrow a x E \Rightarrow a x (E) \Rightarrow a x (E + E) \Rightarrow a x (I + E) \\ &\Rightarrow a x (a + E) \Rightarrow a x (a + I) \Rightarrow a x (a + I0) \Rightarrow a x (a + I00) \Rightarrow a x (a + b00) \end{aligned}$$

Obs.:

- A cada passo pode existir mais de uma possibilidade de derivação. Ex.:  $I x E \Rightarrow I x (E) \Rightarrow a x (E)$ .
- Nem todas as possibilidades levam a uma string particular. Ex.:  $E \Rightarrow E+E$  não leva à  $a x (a + b00)$ .

## Derivações Mais à Esquerda e Mais à Direita

- Derivação mais à esquerda ( $\Rightarrow_{lm}$ ): sempre substitui a variável mais à esquerda por uma das regras de produção; Ex.: derivação anterior.

- Derivação mais à direita ( $\Rightarrow_{rm}$ ): sempre substitui a variável mais à direita por uma das regras de produção; Ex.:

$$E \Rightarrow E x E \Rightarrow E x (E) \Rightarrow E x (E + E) \Rightarrow E x (E + I) \Rightarrow E x (E + I0) \\ \Rightarrow E x (E + I00) \Rightarrow E x (E + b00) \Rightarrow E x (I + b00) \Rightarrow E x (a + b00) \Rightarrow I x (a + b00) \Rightarrow a x (a + b00)$$

Pode-se concluir que  $E \Rightarrow_{rm}^* a x (a + b00)$

## Linguagem de uma Gramática

- Seja  $G = (V, T, P, S)$  a linguagem de  $G$  é:

$$L(G) = \{w \in T^* \mid S \Rightarrow_G^* w\}$$

Isto é, o conjunto de strings sobre  $T^*$  que são derivadas a partir do símbolo inicial. Se  $G$  é uma GLC, então  $L(G)$  é uma LLC.

Ex.:  $L(G_{pal})$  é uma linguagem livre do contexto.

**Definição (Forma Sentencial):** Seja  $G = (V, T, P, S)$  uma GLC, e  $\alpha \in (V \cup T)^*$ . Se:

$$S \Rightarrow^* \alpha$$

Diz-se que  $\alpha$  é uma **forma sentencial**.

## Formas Sentenciais

Se  $S \Rightarrow_{lm}^* \alpha$  diz-se que  $\alpha$  é uma **forma sentencial esquerda**.  
Se  $S \Rightarrow_{rm}^* \alpha$  diz-se que  $\alpha$  é uma **forma sentencial direita**.

Note que  $L(G)$  são **todas as formas sentenciais** que estão em  $T^*$ .

Exemplo 1:  $E x (I + E)$  é uma forma sentencial (que não é  $lm$  nem  $rm$ ), pois:

$$E \Rightarrow E x E \Rightarrow E x (E) \Rightarrow E x (E + E) \Rightarrow E x (I + E).$$

Exemplo 2:  $a x E$  é uma forma sentencial à direita, pois:

$$E \Rightarrow E x E \Rightarrow I x E \Rightarrow a x E$$

## Árvores de Derivação

- Se  $w \in L(G)$ , para alguma GLC, então  $w$  possui uma árvore de derivação, a qual informa a estrutura sintática de  $w$ ;
- $w$  poderia ser um programa, um documento XML, etc.
- Árvores de derivação constituem uma representação alternativa para derivações e inferências recursivas;
- Podem existir várias árvores de derivação para mesma string;
- Idealmente deveria existir somente uma árvore de derivação (a de estrutura "verdadeira") para cada string, i.e. a linguagem deveria ser livre de ambigüidades;
- Infelizmente, nem sempre é possível remover ambigüidades.

## Construção de Árvores de Derivação

Seja  $G = (V, T, P, S)$  uma GLC. Uma árvore é uma árvore de derivação para  $G$  se:

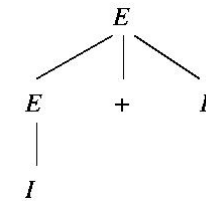
- Cada nó interior é rotulado por uma variável em  $V$ ;
- Cada folha é rotulada por um símbolo em  $V \cup T \cup \{\epsilon\}$ . Se uma folha é rotulada por  $\epsilon$ , então ela deve ser a única derivação "filha" de seu "pai".
- Se um nó interior é rotulado por  $A$ , e seus "filhos" da esquerda para direita possuem respectivamente os rótulos:  $X_1, X_2, \dots, X_k$ ,

Então  $A \rightarrow X_1 X_2 \dots X_k \in P$ .

## Exemplo 1

- Para a gramática (à esquerda) a árvore (à direita) é uma árvore de derivação que mostra a derivação de  $E \Rightarrow^* I+E$ :

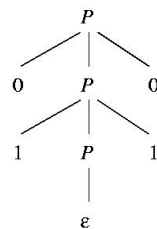
$E \rightarrow I$   
 $E \rightarrow E + E$   
 $E \rightarrow E \times E$   
 $E \rightarrow (E)$



## Exemplo 2

- Para a gramática (à esquerda) a árvore (à direita) é uma árvore de derivação que mostra a derivação de  $P \Rightarrow^* 0110$ :

$P \rightarrow \epsilon$   
 $P \rightarrow 0$   
 $P \rightarrow 1$   
 $P \rightarrow 0P0$   
 $P \rightarrow 1P1$



## A produção de uma Árvore de Derivação

- A produção de uma árvore de derivação é a string de folhas da esquerda para direita.

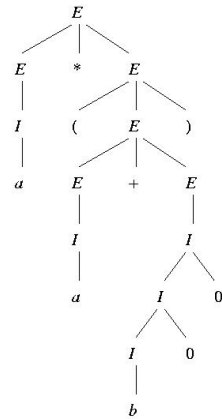
São importante aquelas árvores que:

- A produção é uma string terminal
- A raiz é rotulada pelo símbolo inicial

Observe que o conjunto de resultados destas árvores importantes constituem a linguagem da gramática.

### Exemplo 3

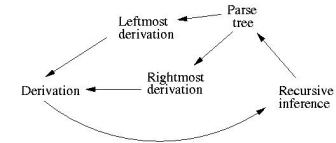
- Ao lado uma árvore (importante) de derivação que produz: **a x (a + b00)**



### Equivalência

Seja  $G = (V, T, P, S)$  uma GLC, e  $A \in V$ . Mostra-se a equivalência de:

1. É possível determinar por inferência recursiva que  $w$  está na linguagem de  $A$ ;
2.  $A \Rightarrow^* w$
3.  $A \Rightarrow_{lm}^* w$  e  $A \Rightarrow_{rm}^* w$
4. Existe uma árvore de derivação de  $G$  com raiz  $A$  e produção  $w$ .

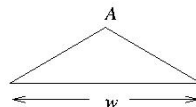


### De Inferências para Árvores

- (Teorema 1) Seja  $G = (V, T, P, S)$  uma GLC, e suponha que  $w$  está na linguagem da variável  $A$ . Então existe uma árvore de derivação para  $G$  com raiz  $A$  e produto  $w$ .

Prova (por indução no tamanho da inferência):

**Base:** (um passo de inferência) Então deve ter sido usado uma produção  $A \rightarrow w$ . A árvore de derivação correspondente é representada a seguir:

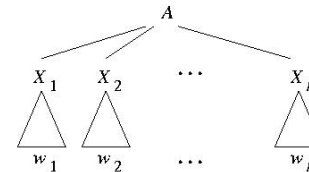


### Teorema 1 (cont.)

**Indução:** Suponha que  $w$  é inferida em  $n+1$  passos. Suponha que o último passo foi baseado na produção:  $A \rightarrow X_1 X_2 \dots X_k$ , onde  $X_i \in V \cup T$ .

Podemos particionar  $w$  como  $w_1 w_2 \dots w_k$  onde  $w_i = X_i$ , se  $X_i \in T$  ou  $w_i$  foi previamente inferida, estando na linguagem de  $X_i$  em no máximo  $n$  passos.

Pela hipótese de indução existem  $i$  árvores de derivação com raiz  $X_i$  e produção  $w_i$ . Então a seguir é representada uma árvore de derivação para  $G$  com raiz  $A$  e produção  $w$ :

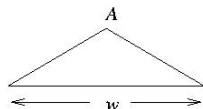


## De Árvores para Derivações

(Teorema 2) Seja  $G = (V, T, P, S)$  uma GLC, e suponha que existe uma árvore de derivação com raiz  $A$  e produção  $w$ . Então  $A \Rightarrow_{\text{lm}}^* w$  em  $G$ .

Prova (por indução na altura da árvore):

**Base:** A árvore de altura 1 é representada a seguir. Conseqüentemente  $A \rightarrow w$  e  $A \Rightarrow_{\text{lm}} w$ .

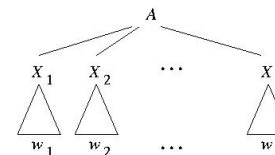


## Teorema 2 cont.

*Indução:* Altura  $n+1$ . A árvore é representada a seguir.

Então  $w = w_1 w_2 \dots w_k$ , onde:

1. Se  $X_i \in T$ , então  $w_i = X_i$
2. Se  $X_i \in V$ , então  $X_i \Rightarrow_{\text{lm}}^* w_i$  em  $G$  pela HI.



Mostra-se que  $A \Rightarrow_{\text{lm}}^* w$  provando por indução que:

$$\forall i, A \Rightarrow_{\text{lm}}^* w_1 w_2 \dots w_i X_{i+1} X_{i+2} \dots X_k$$

## Teorema 2 cont.

**Base:** Seja  $i = 0$ . Sabe-se que  $A \Rightarrow_{\text{lm}}^* X_1 X_2 \dots X_k$ .

**Hipótese de Indução:**  $A \Rightarrow_{\text{lm}}^* w_1 w_2 \dots w_{i-1} X_i X_{i+1} \dots X_k$

**Passo:**

(Caso 1)  $X_i \in T$ . Então,  $X_i = w_i$ , resultando  $A \Rightarrow_{\text{lm}}^* w_1 w_2 \dots w_{i-1} w_i X_{i+1} \dots X_k$

(Caso 2)  $X_i \in V$ . Pela HI existe uma derivação  $X_i \Rightarrow_{\text{lm}} \alpha_1 \Rightarrow_{\text{lm}} \alpha_2 \Rightarrow_{\text{lm}} \dots \Rightarrow_{\text{lm}} w_i$ . Então:

$$\begin{aligned} A &\Rightarrow_{\text{lm}}^* w_1 w_2 \dots w_{i-1} X_i X_{i+1} \dots X_k \Rightarrow_{\text{lm}}^* \\ &w_1 w_2 \dots w_{i-1} \alpha_1 X_{i+1} \dots X_k \Rightarrow_{\text{lm}}^* \\ &w_1 w_2 \dots w_{i-1} \alpha_2 X_{i+1} \dots X_k \Rightarrow_{\text{lm}}^* \\ &\dots \\ &w_1 w_2 \dots w_{i-1} w_i X_{i+1} \dots X_k \end{aligned}$$

## Exemplo

- A primeira sub-árvore, possui a derivação:

$$E \Rightarrow_{\text{lm}} I \Rightarrow_{\text{lm}} a$$

- A terceira sub-árvore possui a derivação:

$$\begin{aligned} E &\Rightarrow_{\text{lm}} (E) \Rightarrow_{\text{lm}} (E+E) \Rightarrow_{\text{lm}} (I+E) \Rightarrow_{\text{lm}} \\ &(a+E) \Rightarrow_{\text{lm}} (a+I) \Rightarrow_{\text{lm}} (a+I0) \Rightarrow_{\text{lm}} (a \\ &+ I00) \Rightarrow_{\text{lm}} (a+b00) \end{aligned}$$

Para a derivação de toda árvore começamos com  $E \Rightarrow_{\text{lm}} E$  x  $E$  e expandimos o primeiro  $E$  pela primeira derivação e o segundo  $E$  pela segunda derivação.

