CC1004 - Modelos de Computação Teórica 15

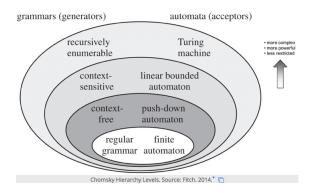
Ana Paula Tomás

Departamento de Ciência de Computadores Faculdade de Ciências da Universidade do Porto

Abril 2021

Linguagens independentes de contexto (LICs)

- Context-free grammars: Gramáticas independentes de contexto (GICs) ou livres de contexto
- Push-down automata: Autómatos de pilha



Exemplo 1

Seja L a linguagem de alfabeto $\{a,b\}$ das palavras obtidas por aplicação, uma ou mais vezes, das regras seguintes.

- (r1) $aaa \in L$
- (r2) $\alpha bb\beta \in L$, quaisquer que sejam $\alpha, \beta \in L$.

Podemos representar (i) e (ii) pelas duas **regras** seguintes

```
< Palavra> \rightarrow aaa < Palavra> \rightarrow < Palavra>  bb < Palavra>
```

as quais definem a **categoria gramatical** <*Palavra*>, dizendo que uma <*Palavra*> é aaa ou a justaposição de três sequências — uma qualquer <*Palavra*>, bb e uma qualquer <*Palavra*>.

Podemos gerar palavras por aplicação das regras de derivação:

$$<$$
 $Palavra> \Rightarrow <$ $Palavra> bb <$ $Palavra> \Rightarrow$ aaa bb $<$ $Palavra> \Rightarrow$ aaabb aaa

Exemplo 1

Seja L a linguagem de alfabeto $\{a,b\}$ das palavras obtidas por aplicação, uma ou mais vezes, das regras seguintes.

- (r1) $aaa \in L$
- (r2) $\alpha bb\beta \in L$, quaisquer que sejam $\alpha, \beta \in L$.

Podemos representar (i) e (ii) pelas duas regras seguintes

$$<$$
 Palavra $> \rightarrow$ aaa $<$ Palavra $> \rightarrow$ $<$ Palavra $>$ bb $<$ Palavra $>$

as quais definem a **categoria gramatical** <*Palavra*>, dizendo que uma <*Palavra*> é aaa ou a justaposição de três sequências — uma qualquer <*Palavra*>, bb e uma qualquer <*Palavra*>.

Podemos gerar palavras por aplicação das regras de derivação:

Exemplo 1

Seja L a linguagem de alfabeto $\{a,b\}$ das palavras obtidas por aplicação, uma ou mais vezes, das regras seguintes.

- (r1) $aaa \in L$
- (r2) $\alpha bb\beta \in L$, quaisquer que sejam $\alpha, \beta \in L$.

Podemos representar (i) e (ii) pelas duas regras seguintes

$$<$$
Palavra $> \rightarrow$ aaa
 $<$ Palavra $> \rightarrow$ $<$ Palavra $>$ bb $<$ Palavra $>$

as quais definem a **categoria gramatical** <*Palavra*>, dizendo que uma <*Palavra*> é aaa ou a justaposição de três sequências — uma qualquer <*Palavra*>, bb e uma qualquer <*Palavra*>.

Podemos gerar palavras por aplicação das regras de derivação:

$$<$$
 $Palavra> \Rightarrow <$ $Palavra> bb <$ $Palavra> \Rightarrow$ aaa bb $<$ $Palavra> \Rightarrow$ aaabb aaa

Alternativamente, como $L = \mathcal{L}((aaabb)^*aaa)$, podemos ainda definir L como sendo a **linguagem gerada** por aplicação das regras:

$$<$$
 Palavra $> \rightarrow$ aaa $<$ Palavra $> \rightarrow$ aaabb $<$ Palavra $>$

Exemplo 2

Seja $L_1 = \{0^{2k}1^k \mid k \in \mathbb{N}\}$. A linguagem L_1 é o menor subconjunto de $\{0, 1\}^*$ que satisfaz as duas condições (i) e (ii) seguintes.

- (i) $\varepsilon \in L_1$
- (ii) $\forall x \in L_1 \quad 00x1 \in L_1$

Equivalentemente, podemos dizer que L_1 é constituída pelas palavras que podem ser obtidas por aplicação, uma ou mais vezes, das regras (r_1) e (r_2) seguintes

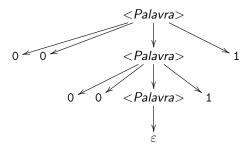
- (r_1) $\varepsilon \in L_1$
- (r_2) Qualquer que seja x, se $x \in L_1$ então $00x1 \in L_1$

À semelhança do exemplo anterior, podemos definir as palavras de L_1 como sendo da categoria <Palavra>, assim definida

$$<$$
 Palavra $> \rightarrow \varepsilon$ $<$ Palavra $> \rightarrow 00$ $<$ Palavra > 1

$$<$$
 Palavra $> \rightarrow \varepsilon$
 $<$ Palavra $> \rightarrow 00$ $<$ Palavra > 1

A palavra 000011 é uma <Palavra> que admite a árvore de derivação (ou árvore sintática) seguinte.



A derivação da palavra 000011

 $\langle Palavra \rangle \Rightarrow 00 \langle Palavra \rangle 1 \Rightarrow 0000 \langle Palavra \rangle 11 \Rightarrow 0000 \varepsilon 11 = 000011$

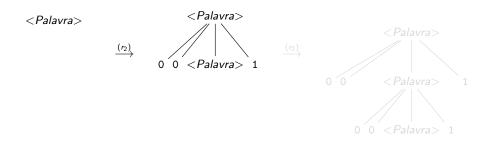
A árvore de derivação é assim construída



restando aplicar a regra < $Palavra> \rightarrow \varepsilon$.

Para simplificar representamos a árvore como um grafo não dirigido (com segmentos em vez de setas)

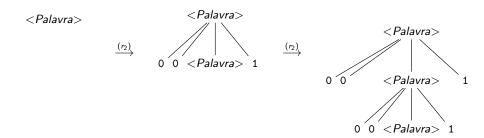
A árvore de derivação é assim construída



restando aplicar a regra < $Palavra> \rightarrow \varepsilon$.

Para simplificar representamos a árvore como um grafo não dirigido (com segmentos em vez de setas)

A árvore de derivação é assim construída



restando aplicar a regra < *Palavra* $> \rightarrow \varepsilon$.

Para simplificar representamos a árvore como um grafo não dirigido (com segmentos em vez de setas)

4□ > 4□ > 4□ > 4□ > 4□ > 1□

Exemplo 3

Seja F a linguagem de alfabeto $\Sigma = \{f, x, (,), ,\}$ assim definida indutivamente.

- (i) $x \in F$
- (ii) $f(\alpha_1, \alpha_2) \in F$, quaisquer que sejam $\alpha_1, \alpha_2 \in F$

Se definirmos as palavras da linguagem ${\it F}$ com categoria ${\it P}$, temos as regras:

$$\begin{array}{ccc} P & \to & \mathbf{x} \\ P & \to & \mathbf{f}(P,P) \end{array}$$

Notação abreviada para as regras

$$P \rightarrow x \mid f(P,P)$$

Alguns exemplos de palavras desta linguagem:

```
f(x,x)

f(f(x,x),x)

f(f(x,x),f(x,x))

f(f(x,x),f(f(x,x),x))
```

Exemplo 3

Seja F a linguagem de alfabeto $\Sigma = \{f, x, (,), ,\}$ assim definida indutivamente.

- (i) $x \in F$
- (ii) $f(\alpha_1, \alpha_2) \in F$, quaisquer que sejam $\alpha_1, \alpha_2 \in F$

Se definirmos as palavras da linguagem F com categoria P, temos as regras:

$$\begin{array}{ccc}
P & \to & \mathbf{x} \\
P & \to & \mathbf{f}(P,P)
\end{array}$$

Notação abreviada para as regras

$$P \rightarrow x \mid f(P,P)$$

Alguns exemplos de palavras desta linguagem:

$$f(x,x) f(f(x,x),x) f(f(x,x),f(x,x)) f(f(x,x),f(f(x,x),x))$$

Exemplo 3

Seja F a linguagem de alfabeto $\Sigma = \{f, x, (,), ,\}$ assim definida indutivamente.

- (i) $x \in F$
- (ii) $f(\alpha_1, \alpha_2) \in F$, quaisquer que sejam $\alpha_1, \alpha_2 \in F$

Se definirmos as palavras da linguagem F com categoria P, temos as regras:

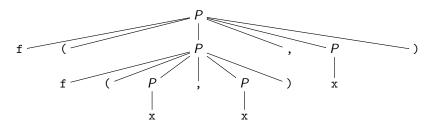
$$\begin{array}{ccc}
P & \to & \mathbf{x} \\
P & \to & \mathbf{f}(P,P)
\end{array}$$

Notação abreviada para as regras

$$P \rightarrow x \mid f(P,P)$$

Alguns exemplos de palavras desta linguagem:

A árvore de derivação (ou árvore sintática) de f(f(x,x),x):



f(f(x,x),x) admite várias derivações mas apenas esta árvore de derivação.
Derivação pela esquerdo:

$$P \Rightarrow \mathtt{f}(P,P) \Rightarrow \mathtt{f}(\mathtt{f}(P,P),P) \Rightarrow \mathtt{f}(\mathtt{f}(\mathtt{x},P),P) \Rightarrow \mathtt{f}(\mathtt{f}(\mathtt{x},\mathtt{x}),P) \Rightarrow \mathtt{f}(\mathtt{f}(\mathtt{x},\mathtt{x}),\mathtt{x})$$

Derivação pela direita

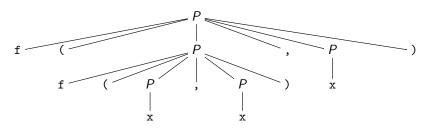
$$P \Rightarrow f(P,P) \Rightarrow f(P,x) \Rightarrow f(f(P,P),x) \Rightarrow f(f(P,x),x) \Rightarrow f(f(x,x),x)$$

Outra derivação

$$P \Rightarrow f(P,P) \Rightarrow f(P,x) \Rightarrow f(f(P,P),x) \Rightarrow f(f(x,P),x) \Rightarrow f(f(x,x),x)$$

◆ロト ◆母 ト ◆ 豊 ト ◆ 豊 ・ 夕 Q ②

A árvore de derivação (ou árvore sintática) de f(f(x,x),x):



f(f(x,x),x) admite várias derivações mas apenas esta árvore de derivação.

Derivação pela esquerda:

$$P \Rightarrow f(P,P) \Rightarrow f(f(P,P),P) \Rightarrow f(f(x,P),P) \Rightarrow f(f(x,x),P) \Rightarrow f(f(x,x),x)$$

Derivação pela direita

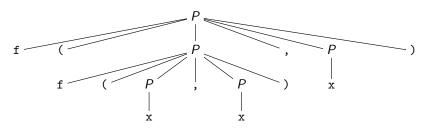
$$P \Rightarrow f(P,P) \Rightarrow f(P,x) \Rightarrow f(f(P,P),x) \Rightarrow f(f(P,x),x) \Rightarrow f(f(x,x),x)$$

Outra derivação

$$P \Rightarrow f(P,P) \Rightarrow f(P,x) \Rightarrow f(f(P,P),x) \Rightarrow f(f(x,P),x) \Rightarrow f(f(x,x),x)$$

- 4 日 ト 4 団 ト 4 邑 ト 4 邑 ト 9 Q C

A árvore de derivação (ou árvore sintática) de f(f(x,x),x):



f(f(x,x),x) admite várias derivações mas apenas esta árvore de derivação.

$$P \Rightarrow f(P,P) \Rightarrow f(f(P,P),P) \Rightarrow f(f(x,P),P) \Rightarrow f(f(x,x),P) \Rightarrow f(f(x,x),x)$$

$$P \Rightarrow f(P,P) \Rightarrow f(P,x) \Rightarrow f(f(P,P),x) \Rightarrow f(f(P,x),x) \Rightarrow f(f(x,x),x)$$

Outra derivação:

$$P \Rightarrow f(P,P) \Rightarrow f(P,x) \Rightarrow f(f(P,P),x) \Rightarrow f(f(x,P),x) \Rightarrow f(f(x,x),x)$$

Exemplo 4

```
< frase > \rightarrow < fn > < fv > < fn > < fv > < fn > \rightarrow  o gato < fn > \rightarrow  o rato < fn > \rightarrow  o peixe < fv > \rightarrow  come < fn >
```

A linguagem gerada a partir de *<frase>* contém as palavras (i.e., frases) *o gato come o peixe* mas também *o peixe come o gato*. Estão **sintaticamente corretas** embora nos possam parecer estranhas (semanticamente).

Além de < frase>, que seria o símbolo inicial da gramática, tem outras categorias, designadas por variáveis ou símbolos não terminais: < fn> e < fv>.

A linguagem gerada por uma gramática é o conjunto das palavras de alfabeto Σ que se podem derivar por aplicação das regras da gramática a **partir** do símbolo inicial. Os símbolos de Σ são os símbolos terminais da gramática. Σ não inclui as variáveis da gramática.

Exemplo 5

Que palavras de $\{(,),:,>,<,=,+,-,*,/,0,\ldots,9,a,b,c,\ldots,z,\cdot,;\}^*$ são geradas a partir de S por aplicação das regras seguintes?

```
S \rightarrow \text{if} \cdot (C) \cdot \text{then} \cdot \text{goto} \cdot N

ightarrow goto \cdot N
   \rightarrow V:=E
S \rightarrow \mathsf{stop}
E \rightarrow (EOE) \mid (-E) \mid V \mid N
O \rightarrow + | - | * | /
C \rightarrow VXF
X \rightarrow > | < | >= | <= | = | <>
N \rightarrow D \mid DN
D \rightarrow 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9
V \rightarrow L \mid LN \mid LV
L \rightarrow a|b|c|d|e|f|g|h
```

Noção de Gramática Independente de Contexto

Uma gramática independente de contexto é um quarteto

$$\mathcal{G} = (V, \Sigma, P, S)$$

em que V e Σ são conjuntos de símbolos, ambos finitos e não vazios, e tais que $V \cap \Sigma = \emptyset$, $S \in V$, e P é uma relação binária finita de V em $(V \cup \Sigma)^*$.

- V é o conjunto das variáveis (ou não terminais)
- ullet diz-se **símbolo inicial** de ${\mathcal G}$
- Σ é o alfabeto (conjunto dos símbolos terminais)
- P é um conjunto finito constituído pelas **produções** ou **regras**. Usualmente escreve-se $X \to w$ se $(X, w) \in P$.

A linguagem gerada por \mathcal{G} , que se denota por $\mathcal{L}(\mathcal{G})$, é constituída pelas palavras que se podem derivar a partir do seu símbolo inicial.

Gramática Independente de Contexto

As GICs são *independentes de contexto* porque as regras que definem os não terminais (i.e., as *categorias gramaticais*) são aplicadas sem estarem sujeitas a restrições introduzidas por algum contexto. Nas GICs, a parte esquerda de cada regra é um não terminal.

A forma das regras de produção determina a expressividade da gramática. Nas **gramáticas mais gerais** (ditas, de $Tipo\ 0$) as regras têm a forma:

$$\alpha \to \beta$$
 com $\alpha, \beta \in (V \cup \Sigma)^*, \alpha \neq \varepsilon$

Não iremos estudar gramáticas desse tipo, que definem o topo da hierarquia.

As **gramáticas regulares**, que geram linguagens regulares, são de Tipo 3. As **gramáticas independentes de contexto** são de Tipo 2.

$$\begin{array}{c} \bullet \;\; \mathcal{G}_1 = \big(\{\,T\}, \{0,1\}, \{\,T \to 0\,T, \;\; T \to 1\,T, \;\; T \to \varepsilon\}, \,T \big) \\ \\ \mathcal{L}(\mathcal{G}_1) = \{0,1\}^\star \end{array}$$

•
$$\mathcal{G}_2 = (\{U\}, \{0, 1\}, \{U \to 10U, U \to \varepsilon\}, U)$$

 $\mathcal{L}(\mathcal{G}_2) = \{10\}$

•
$$\mathcal{G}_3 = (\{U, Z\}, \{0, 1\}, \{Z \to 0U, U \to 10U, U \to \varepsilon\}, Z)$$

•
$$\mathcal{G}_4 = (\{Z, U\}, \{0, 1\}, \{Z \to 0U, U \to 1Z, U \to \varepsilon\}, Z)$$

$$\mathcal{L}(\mathcal{G}_4) = \{0\}\{10\}^* = \mathcal{L}(\mathcal{G}_3)$$

•
$$G_5 = (\{Z, U\}, \{0, 1\}, \{Z \to 0U, U \to 1Z, U \to \varepsilon\}, U)$$

$$\mathcal{L}(\mathcal{G}_5) = \{10\}^* = \mathcal{L}(\mathcal{G}_2)$$

•
$$\mathcal{G}_1 = (\{T\}, \{0, 1\}, \{T \to 0T, \ T \to 1T, \ T \to \varepsilon\}, T)$$

$$\mathcal{L}(\mathcal{G}_1) = \{0, 1\}^*$$

•
$$G_2 = (\{U\}, \{0,1\}, \{U \to 10U, U \to \varepsilon\}, U)$$

$$\mathcal{L}(\mathcal{G}_2) = \{10\}^\star$$

•
$$\mathcal{G}_3 = (\{U, Z\}, \{0, 1\}, \{Z \to 0U, U \to 10U, U \to \varepsilon\}, Z)$$

$$\mathcal{L}(\mathcal{G}_3) = \{0\}\{10\}^*$$

•
$$\mathcal{G}_4 = (\{Z, U\}, \{0, 1\}, \{Z \to 0U, U \to 1Z, U \to \varepsilon\}, Z)$$

$$\mathcal{L}(\mathcal{G}_4) = \{0\}\{10\}^* = \mathcal{L}(\mathcal{G}_3)$$

•
$$\mathcal{G}_5 = (\{Z, U\}, \{0, 1\}, \{Z \to 0U, U \to 1Z, U \to \varepsilon\}, U)$$

$$\mathcal{L}(\mathcal{G}_5) = \{10\}^* = \mathcal{L}(\mathcal{G}_2)$$

◆ロト ◆団ト ◆草ト ◆草ト 草 めるぐ

$$\begin{array}{c} \bullet \;\; \mathcal{G}_1 = (\{T\},\{0,1\},\{T \rightarrow \mathtt{0}T,\; T \rightarrow \mathtt{1}T,\; T \rightarrow \varepsilon\},T) \\ \\ \mathcal{L}(\mathcal{G}_1) = \{\mathtt{0},\mathtt{1}\}^\star \end{array}$$

•
$$G_2 = (\{U\}, \{0,1\}, \{U \to 10U, U \to \varepsilon\}, U)$$

$$\mathcal{L}(\mathcal{G}_2) = \{10\}^\star$$

•
$$\mathcal{G}_3 = (\{U, Z\}, \{0, 1\}, \{Z \to 0U, U \to 10U, U \to \varepsilon\}, Z)$$

$$\mathcal{L}(\mathcal{G}_3) = \{0\}\{10\}^\star$$

•
$$\mathcal{G}_4 = (\{Z, U\}, \{0, 1\}, \{Z \to 0U, U \to 1Z, U \to \varepsilon\}, Z)$$

$$\mathcal{L}(\mathcal{G}_4) = \{0\}\{10\}^* = \mathcal{L}(\mathcal{G}_3)$$

•
$$\mathcal{G}_5 = (\{Z, U\}, \{0, 1\}, \{Z \to 0U, U \to 1Z, U \to \varepsilon\}, U)$$

$$\mathcal{L}(\mathcal{G}_5) = \{10\}^* = \mathcal{L}(\mathcal{G}_2)$$

•
$$\mathcal{G}_1 = (\{T\}, \{0, 1\}, \{T \to 0T, \ T \to 1T, \ T \to \varepsilon\}, T)$$

$$\mathcal{L}(\mathcal{G}_1) = \{0, 1\}^*$$

•
$$G_2 = (\{U\}, \{0,1\}, \{U \to 10U, U \to \varepsilon\}, U)$$

$$\mathcal{L}(\mathcal{G}_2) = \{10\}^\star$$

•
$$G_3 = (\{U, Z\}, \{0, 1\}, \{Z \to 0U, U \to 10U, U \to \epsilon\}, Z)$$

$$\mathcal{L}(\mathcal{G}_3)=\{0\}\{10\}^\star$$

•
$$\mathcal{G}_4 = (\{Z, U\}, \{0, 1\}, \{Z \to 0U, U \to 1Z, U \to \varepsilon\}, Z)$$

$$\mathcal{L}(\mathcal{G}_4) = \{0\}\{10\}^* = \mathcal{L}(\mathcal{G}_3)$$

•
$$\mathcal{G}_5 = (\{Z, U\}, \{0, 1\}, \{Z \to 0U, U \to 1Z, U \to \varepsilon\}, U)$$

$$\mathcal{L}(\mathcal{G}_5) = \{10\}^* = \mathcal{L}(\mathcal{G}_2)$$



•
$$\mathcal{G}_1 = (\{T\}, \{0, 1\}, \{T \to 0T, \ T \to 1T, \ T \to \varepsilon\}, T)$$

$$\mathcal{L}(\mathcal{G}_1) = \{0, 1\}^*$$

•
$$G_2 = (\{U\}, \{0,1\}, \{U \to 10U, U \to \varepsilon\}, U)$$

$$\mathcal{L}(\mathcal{G}_2) = \{10\}^\star$$

•
$$G_3 = (\{U, Z\}, \{0, 1\}, \{Z \to 0U, U \to 10U, U \to \varepsilon\}, Z)$$

$$\mathcal{L}(\mathcal{G}_3)=\{0\}\{10\}^\star$$

•
$$\mathcal{G}_4 = (\{Z, U\}, \{0, 1\}, \{Z \to 0U, U \to 1Z, U \to \varepsilon\}, Z)$$

$$\mathcal{L}(\mathcal{G}_4) = \{0\}\{10\}^* = \mathcal{L}(\mathcal{G}_3)$$

•
$$\mathcal{G}_5 = (\{Z, U\}, \{0, 1\}, \{Z \to 0U, U \to 1Z, U \to \varepsilon\}, U)$$

$$\mathcal{L}(\mathcal{G}_5) = \{10\}^{\star} = \mathcal{L}(\mathcal{G}_2)$$



•
$$\mathcal{G}_6=(\{S\},\{\mathtt{a},\mathtt{b}\},\{S o\mathtt{aaa},\ S o\mathtt{aaabb}S\},S)$$

$$\mathcal{L}(\mathcal{G}_6)=\{\mathtt{aaabb}\}^\star\{\mathtt{aaa}\}$$

$$\begin{aligned} \textbf{G}_7 &= (\{S,C\},\{\mathtt{a},\mathtt{b}\},\{S\to\mathtt{a}S\mathtt{b},\,S\to C,\,C\to\mathtt{b}C\mathtt{a},\,C\to\mathtt{b}\},S) \\ \\ \mathcal{L}(\mathcal{G}_7) &= \{\mathtt{a}^n\mathtt{b}^{m+1}\mathtt{a}^m\mathtt{b}^n \mid \, m\in\mathbb{N},n\in\mathbb{N}\} \end{aligned}$$

•
$$\mathcal{G}_6=(\{S\},\{\mathtt{a},\mathtt{b}\},\{S o\mathtt{aaa},\ S o\mathtt{aaabb}S\},S)$$

$$\mathcal{L}(\mathcal{G}_6)=\{\mathtt{aaabb}\}^\star\{\mathtt{aaa}\}$$

$$\begin{array}{l} \bullet \;\; \mathcal{G}_7 = \left(\{S,C\}, \{\mathtt{a},\mathtt{b}\}, \{S \to \mathtt{a} S\mathtt{b}, \; S \to C, \; C \to \mathtt{b} C\mathtt{a}, \; C \to \mathtt{b}\}, S \right) \\ \\ \mathcal{L}(\mathcal{G}_7) = \left\{ \mathtt{a}^n \mathtt{b}^{m+1} \mathtt{a}^m \mathtt{b}^n \; \mid \; m \in \mathbb{N}, n \in \mathbb{N} \right\} \end{array}$$