Gramática Livre-de-Contexto

Esdras Lins Bispo Jr. bispojr@ufg.br

Linguagens Formais e Autômatos Bacharelado em Ciência da Computação

22 de janeiro de 2018





Plano de Aula

- Revisão
 - Gramáticas Livre-de-Contexto
- ② Gramáticas Livre-de-Contexto
- 3 Ambiguidade
 - Forma Normal de Chomsky





Sumário

- Revisão
 - Gramáticas Livre-de-Contexto
- ② Gramáticas Livre-de-Contexto
- 3 Ambiguidade
 - Forma Normal de Chomsky





Método mais poderoso de descrição de linguagens;





- Método mais poderoso de descrição de linguagens;
- Linguagens como $\{0^n1^n|n\geq 0\}$ podem agora ser descritas;





- Método mais poderoso de descrição de linguagens;
- Linguagens como $\{0^n 1^n | n \ge 0\}$ podem agora ser descritas;
- Descrição de linguagens com estruturas recursivas;





- Método mais poderoso de descrição de linguagens;
- Linguagens como $\{0^n 1^n | n \ge 0\}$ podem agora ser descritas;
- Descrição de linguagens com estruturas recursivas;
- GLCs geram Linguagens Livres-de-Contexto (LLCs).





- Método mais poderoso de descrição de linguagens;
- Linguagens como $\{0^n 1^n | n \ge 0\}$ podem agora ser descritas;
- Descrição de linguagens com estruturas recursivas;
- GLCs geram Linguagens Livres-de-Contexto (LLCs).

Aplicações...

• Especificação e compilação de linguagens de programação.





$$A \rightarrow 0A1$$

$$A \rightarrow B$$

$$B \to \#$$





$$A \rightarrow 0A1$$

$$A \rightarrow B$$

$$B \rightarrow \#$$

Uma gramática consiste de...

Regras de substituição (também chamadas de produções);





$$A \rightarrow 0A1$$

$$A \rightarrow B$$

$$B \rightarrow \#$$

Uma gramática consiste de...

- Regras de substituição (também chamadas de produções);
- Símbolos que são chamados de variáveis;





$$A \rightarrow 0A1$$

$$A \rightarrow B$$

$$B \rightarrow \#$$

Uma gramática consiste de...

- Regras de substituição (também chamadas de produções);
- Símbolos que são chamados de variáveis;
- Símbolos que são chamados de terminais;





$$A \rightarrow 0A1$$

$$A \rightarrow B$$

$$B \rightarrow \#$$

Uma gramática consiste de...

- Regras de substituição (também chamadas de produções);
- Símbolos que são chamados de variáveis;
- Símbolos que são chamados de terminais;
- Uma variável é designada como variável inicial.





Como utilizar uma gramática?

• Escreva a variável inicial. Ela é a variável no lado esquerdo da primeira regra, a menos que especificado em contrário;





Como utilizar uma gramática?

- Escreva a variável inicial. Ela é a variável no lado esquerdo da primeira regra, a menos que especificado em contrário;
- Encontre uma variável que esteja escrita e uma regra que comece com essa variável. Substitua a variável escrita pelo lado direito dessa regra;





Como utilizar uma gramática?

- Escreva a variável inicial. Ela é a variável no lado esquerdo da primeira regra, a menos que especificado em contrário;
- Encontre uma variável que esteja escrita e uma regra que comece com essa variável. Substitua a variável escrita pelo lado direito dessa regra;
- Repita o passo 2 até que n\u00e3o reste nenhuma vari\u00e1vel.





 G_1 gera a cadeia 000#111

Α







Α

)

1





G_1 gera a cadeia 000#111

F

Α

00 A 11





G_1 gera a cadeia 000#111

0 A 1 00 A 11 000 A 111





G_1 gera a cadeia 000#111

 A

 0
 A
 1

 00
 A
 11

 000
 A
 111

 000
 B
 111





G_1 gera a cadeia 000#111

```
      A

      0
      A
      1

      00
      A
      11

      000
      A
      111

      000
      B
      111

      000
      #
      111
```





G_1 gera a cadeia 000#111

```
A
0
A
1
00
A
11
000
A
111
000
B
111
000
\#
111
```

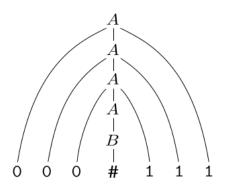




G_1 gera a cadeia 000#111











 O conjunto de todas as cadeias geradas por uma GLC constitui a linguagem da gramática;





- O conjunto de todas as cadeias geradas por uma GLC constitui a linguagem da gramática;
- $L(G_1) = \{0^n \# 1^n | n \ge 0\}$





- O conjunto de todas as cadeias geradas por uma GLC constitui a linguagem da gramática;
- $L(G_1) = \{0^n \# 1^n | n \ge 0\}$
- Qualquer linguagem que possa ser gerada por uma GLC é chamada de linguagem livre-do-contexto (LLC);





- O conjunto de todas as cadeias geradas por uma GLC constitui a linguagem da gramática;
- $L(G_1) = \{0^n \# 1^n | n \ge 0\}$
- Qualquer linguagem que possa ser gerada por uma GLC é chamada de linguagem livre-do-contexto (LLC);
- Podemos abreviar várias regras em apenas uma só como $A \to 0A1$ e $A \to B$ em $A \to 0A1|B$.





G_2 : fragmento da língua inglesa

```
 \langle \text{SENTENCE} \rangle \rightarrow \langle \text{NOUN-PHRASE} \rangle \langle \text{VERB-PHRASE} \rangle   \langle \text{NOUN-PHRASE} \rangle \rightarrow \langle \text{CMPLX-NOUN} \rangle | \langle \text{CMPLX-NOUN} \rangle \langle \text{PREP-PHRASE} \rangle   \langle \text{VERB-PHRASE} \rangle \rightarrow \langle \text{CMPLX-VERB} \rangle | \langle \text{CMPLX-VERB} \rangle \langle \text{PREP-PHRASE} \rangle   \langle \text{PREP-PHRASE} \rangle \rightarrow \langle \text{PREP} \rangle \langle \text{CMPLX-NOUN} \rangle   \langle \text{CMPLX-NOUN} \rangle \rightarrow \langle \text{ARTICLE} \rangle \langle \text{NOUN} \rangle   \langle \text{CMPLX-VERB} \rangle \rightarrow \langle \text{VERB} \rangle | \langle \text{VERB} \rangle \langle \text{NOUN-PHRASE} \rangle   \langle \text{ARTICLE} \rangle \rightarrow \text{a} \mid \text{the}   \langle \text{NOUN} \rangle \rightarrow \text{boy} \mid \text{girl} \mid \text{flower}   \langle \text{VERB} \rangle \rightarrow \text{touches} \mid \text{likes} \mid \text{sees}   \langle \text{PREP} \rangle \rightarrow \text{with}
```





G_2 : fragmento da língua inglesa

Cadeias em $L(G_2)$

- a boy sees
- the boy sees a flower
- a girl with a flower likes the boy





Derivação de "a boy sees"

```
\langle SENTENCE \rangle \Rightarrow \langle NOUN-PHRASE \rangle \langle VERB-PHRASE \rangle
\Rightarrow \langle CMPLX-NOUN \rangle \langle VERB-PHRASE \rangle
\Rightarrow \langle ARTICLE \rangle \langle NOUN \rangle \langle VERB-PHRASE \rangle
\Rightarrow a \langle NOUN \rangle \langle VERB-PHRASE \rangle
\Rightarrow a boy \langle VERB-PHRASE \rangle
\Rightarrow a boy \langle CMPLX-VERB \rangle
\Rightarrow a boy \langle VERB \rangle
\Rightarrow a boy sees
```





DEFINIÇÃO 2.2

Uma gramática livre-do-contexto é uma 4-upla (V, Σ, R, S) , onde

- 1. V é um conjunto finito denonimado de as variáveis,
- Σ é um conjunto finito, disjunto de V, denonimado de os terminais,
- 3. R é um conjunto finito de *regras*, com cada regra sendo uma variável e uma cadeia de variáveis e terminais, e
- **4.** $S \in V$ é a variável inicial.





Sumário

- Revisão
 - Gramáticas Livre-de-Contexto
- ② Gramáticas Livre-de-Contexto
- 3 Ambiguidade
 - Forma Normal de Chomsky





 G_1

$$A \rightarrow 0A1$$

$$A \rightarrow B$$

$$B \to \#$$





G_1

$$A \rightarrow 0A1$$

$$A \rightarrow B$$

$$B \rightarrow \#$$

Descrição formal

 $G_1 = (V, \Sigma, R, S)$ de forma que:

- $V = \{A, B\}$:
- $\Sigma = \{0, 1, \#\};$
- $R = \{A \to 0A1, A \to B, B \to \#\};$
- S = A.





 G_3

$$G_3=(\{S\},\{a,b\},R,S)$$
. O conjunto de regras R é: $S \to aSb \mid SS \mid \epsilon$





G_3

$$G_3=(\{S\},\{a,b\},R,S)$$
. O conjunto de regras R é: $S o aSb \mid SS \mid \epsilon$

G_4

$$G_4 = (V, \Sigma, R, S)$$
, de forma que:
$$V = \{\langle EXPR \rangle, \langle TERM \rangle, \langle FACTOR \rangle \};$$

$$\Sigma = \{a, +, \times, (,)\};$$
 R é o conjunto de regras:
$$\langle EXPR \rangle \rightarrow \langle EXPR \rangle + \langle TERM \rangle \mid \langle TERM \rangle \\ \langle TERM \rangle \rightarrow \langle TERM \rangle \times \langle FACTOR \rangle \mid \langle FACTOR \rangle \\ \langle FACTOR \rangle \rightarrow (\langle EXPR \rangle) \mid a$$

$$S = \langle EXPR \rangle.$$





Árvores Sintáticas

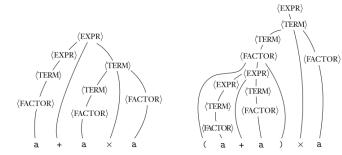


FIGURA 2.5 Árvores sintáticas para as cadeias a+axa e (a+a)xa





• Considere a união de GLCs;





- Considere a união de GLCs;
- Se a linguagem for regular, construa a GLC a partir do AFD correspondente;





- Considere a união de GLCs;
- Se a linguagem for regular, construa a GLC a partir do AFD correspondente;
- Considere subcadeias "ligadas";





- Considere a união de GLCs;
- Se a linguagem for regular, construa a GLC a partir do AFD correspondente;
- Considere subcadeias "ligadas";
- Considere estruturas recursivas.





Sumário

- Revisão
 - Gramáticas Livre-de-Contexto
- ② Gramáticas Livre-de-Contexto
- 3 Ambiguidade
 - Forma Normal de Chomsky





Gramática G₅

```
\langle \text{ EXPR } \rangle \rightarrow \langle \text{ EXPR } \rangle + \langle \text{ EXPR } \rangle \mid \langle \text{ EXPR } \rangle \times \langle \text{ EXPR } \rangle
\langle \text{ EXPR } \rangle \rightarrow (\langle \text{ EXPR } \rangle) \mid a
```





Gramática G₅

$$\langle \text{ EXPR } \rangle \rightarrow \langle \text{ EXPR } \rangle + \langle \text{ EXPR } \rangle \mid \langle \text{ EXPR } \rangle \times \langle \text{ EXPR } \rangle$$

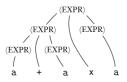
 $\langle \text{ EXPR } \rangle \rightarrow (\langle \text{ EXPR } \rangle) \mid a$

Cadeia $a + a \times a$

Esta gramática gera a + a × a ambiguamente.







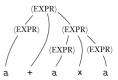


FIGURA 2.6

As duas árvores sintáticas para a cadeia $\mathtt{a+axa}$ na gramática G_5





G_4

$$\langle EXPR \rangle \rightarrow \langle EXPR \rangle + \langle TERM \rangle \mid \langle TERM \rangle$$

 $\langle TERM \rangle \rightarrow \langle TERM \rangle \times \langle FACTOR \rangle \mid \langle FACTOR \rangle$
 $\langle FACTOR \rangle \rightarrow (\langle EXPR \rangle) \mid a$





G_4

$$\langle EXPR \rangle \rightarrow \langle EXPR \rangle + \langle TERM \rangle \mid \langle TERM \rangle$$

 $\langle TERM \rangle \rightarrow \langle TERM \rangle \times \langle FACTOR \rangle \mid \langle FACTOR \rangle$
 $\langle FACTOR \rangle \rightarrow (\langle EXPR \rangle) \mid a$

Porém...

G₄ não é ambígua!





G_2

```
 \langle \text{SENTENCE} \rangle \rightarrow \langle \text{NOUN-PHRASE} \rangle \langle \text{VERB-PHRASE} \rangle 
 \langle \text{NOUN-PHRASE} \rangle \rightarrow \langle \text{CMPLX-NOUN} \rangle | \langle \text{CMPLX-NOUN} \rangle \langle \text{PREP-PHRASE} \rangle 
 \langle \text{VERB-PHRASE} \rangle \rightarrow \langle \text{CMPLX-VERB} \rangle | \langle \text{CMPLX-VERB} \rangle \langle \text{PREP-PHRASE} \rangle 
 \langle \text{PREP-PHRASE} \rangle \rightarrow \langle \text{PREP} \rangle \langle \text{CMPLX-NOUN} \rangle 
 \langle \text{CMPLX-NOUN} \rangle \rightarrow \langle \text{ARTICLE} \rangle \langle \text{NOUN} \rangle 
 \langle \text{CMPLX-VERB} \rangle \rightarrow \langle \text{VERB} \rangle | \langle \text{VERB} \rangle \langle \text{NOUN-PHRASE} \rangle 
 \langle \text{ARTICLE} \rangle \rightarrow \text{a | the } 
 \langle \text{NOUN} \rangle \rightarrow \text{boy | girl | flower } 
 \langle \text{VERB} \rangle \rightarrow \text{touches | likes | sees } 
 \langle \text{PREP} \rangle \rightarrow \text{with }
```





G_2

```
 \langle \text{SENTENCE} \rangle \rightarrow \langle \text{NOUN-PHRASE} \rangle \langle \text{VERB-PHRASE} \rangle 
 \langle \text{NOUN-PHRASE} \rangle \rightarrow \langle \text{CMPLX-NOUN} \rangle | \langle \text{CMPLX-NOUN} \rangle \langle \text{PREP-PHRASE} \rangle 
 \langle \text{VERB-PHRASE} \rangle \rightarrow \langle \text{CMPLX-VERB} \rangle | \langle \text{CMPLX-VERB} \rangle \langle \text{PREP-PHRASE} \rangle 
 \langle \text{PREP-PHRASE} \rangle \rightarrow \langle \text{PREP} \rangle \langle \text{CMPLX-NOUN} \rangle 
 \langle \text{CMPLX-NOUN} \rangle \rightarrow \langle \text{ARTICLE} \rangle \langle \text{NOUN} \rangle 
 \langle \text{CMPLX-VERB} \rangle \rightarrow \langle \text{VERB} \rangle | \langle \text{VERB} \rangle | \langle \text{VENB} \rangle \langle \text{NOUN-PHRASE} \rangle 
 \langle \text{ARTICLE} \rangle \rightarrow \text{a} | \text{the} 
 \langle \text{NOUN} \rangle \rightarrow \text{boy} | \text{girl} | \text{flower} 
 \langle \text{VERB} \rangle \rightarrow \text{touches} | \text{likes} | \text{sees} 
 \langle \text{PREP} \rangle \rightarrow \text{with}
```

Porém...

 G_2 é ambígua!

(e.g. a cadeia "the girl touches the boy with the flower")





Ambiguidade[']

Definição 2.7

Uma cadeia ω é derivada **ambiguamente** na gramática livre-do-contexto G se ela tem duas ou mais derivações mais à esquerda diferentes. A gramática G é **ambígua** se ela gera alguma cadeia ambiguamente.





Definicão 2.7

Uma cadeia ω é derivada **ambiguamente** na gramática livre-do-contexto G se ela tem duas ou mais derivações mais à esquerda diferentes. A gramática G é **ambígua** se ela gera alguma cadeia ambiguamente.

Existem linguagens que são inerentemente ambíguas!





Por quê utilizá-la?

- É uma forma simplificada de escrever uma GLC;
- Facilita a construção de algoritmos para GLC.





Por quê utilizá-la?

- É uma forma simplificada de escrever uma GLC;
- Facilita a construção de algoritmos para GLC.

Definição 2.8

Uma gramática livre-do-contexto está na **forma normal de Chomsky** se toda regra é da forma

$$A \rightarrow BC$$

$$A \rightarrow a$$

em que a é qualquer terminal e A, B e C são quaisquer variáveis - exceto que B e C não podem ser a variável inicial.





Por quê utilizá-la?

- É uma forma simplificada de escrever uma GLC;
- Facilita a construção de algoritmos para GLC.

Definição 2.8

Uma gramática livre-do-contexto está na **forma normal de Chomsky** se toda regra é da forma

$$A \rightarrow BC$$

$$A \rightarrow a$$

em que a é qualquer terminal e A, B e C são quaisquer variáveis - exceto que B e C não podem ser a variável inicial.

Adicionalmente...

Permite-se a regra $S \to \epsilon$, em que S é a variável inicial.



Teorema 2.9

Qualquer linguagem livre-de-contexto é gerada por uma gramática livre-do-contexto na forma normal de Chomsky.





$$G_6$$
 $S o ASA \mid aB$
 $A o B \mid S$
 $B o b \mid \epsilon$





G_6

$$S \rightarrow ASA \mid aB$$

$$A \rightarrow B \mid S$$

$$B \rightarrow b \mid \epsilon$$

???

Como converter G_6 para a forma normal de Chomsky?





1. Introduzir uma nova variável inicial

$$S \rightarrow ASA \mid aB$$

$$A \rightarrow B \mid S$$

$$B \rightarrow b \mid \epsilon$$





1. Introduzir uma nova variável inicial

$$S \rightarrow ASA \mid aB$$

$$A \rightarrow B \mid S$$

$$B \rightarrow b \mid \epsilon$$

$$S_0 \rightarrow S$$

$$S \rightarrow ASA \mid aB$$

$$A \rightarrow B \mid S$$

$$B \rightarrow b \mid \epsilon$$





2. Remover regras ϵ (B)

$$S_0 \rightarrow S$$

$$S \rightarrow ASA \mid aB$$

$$A \rightarrow B \mid S$$

$$B \rightarrow b \mid \epsilon$$





2. Remover regras ϵ (B)

$$S_0 \rightarrow S$$

$$S \rightarrow ASA \mid aB$$

$$A \rightarrow B \mid S$$

$$B \rightarrow b \mid \epsilon$$

$$S_0 \rightarrow S$$

$$S \rightarrow ASA \mid aB \mid a$$

$$A \rightarrow B \mid S \mid \epsilon$$

$$B \rightarrow b$$





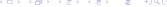
2. Remover regras ϵ (A)

$$S_0 \rightarrow S$$

 $S \rightarrow ASA \mid aB \mid a$
 $A \rightarrow B \mid S \mid \epsilon$







2. Remover regras ϵ (A)

$$S_0 \rightarrow S$$

 $S \rightarrow ASA \mid aB \mid a$
 $A \rightarrow B \mid S \mid \epsilon$
 $B \rightarrow b$

$$S_0 \rightarrow S$$

 $S \rightarrow ASA \mid aB \mid a \mid SA \mid AS \mid S$
 $A \rightarrow B \mid S$
 $B \rightarrow b$





$$S_0 \rightarrow S$$

 $S \rightarrow ASA \mid aB \mid a \mid SA \mid AS \mid S$
 $A \rightarrow B \mid S$
 $B \rightarrow b$





$$S_0 \rightarrow S$$

 $S \rightarrow ASA \mid aB \mid a \mid SA \mid AS \mid S$
 $A \rightarrow B \mid S$
 $B \rightarrow b$

$$S_0 \rightarrow S$$

 $S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$
 $A \rightarrow B \mid S$
 $B \rightarrow b$





$$S_0 o S$$

$$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$$

$$A \rightarrow B \mid S$$

$$B \rightarrow b$$





$$S_0 o {\color{red} S}$$

$$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$$

$$A \rightarrow B \mid S$$

$$B \rightarrow b$$

$$S_0 \rightarrow ASA \mid aB \mid a \mid SA \mid AS$$

$$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$$

$$A \rightarrow B \mid S$$

$$B \rightarrow b$$





$$S_0 \rightarrow ASA \mid aB \mid a \mid SA \mid AS$$

$$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$$

$$A \rightarrow B \mid S$$

$$B \rightarrow b$$





$$S_0
ightarrow ASA \mid aB \mid a \mid SA \mid AS$$

 $S
ightarrow ASA \mid aB \mid a \mid SA \mid AS$

$$A \rightarrow B \mid S$$

$$B \rightarrow b$$

$$S_0 \rightarrow ASA \mid aB \mid a \mid SA \mid AS$$

$$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$$

$$A \rightarrow S \mid b$$

$$B \rightarrow b$$





$$S_0 \rightarrow ASA \mid aB \mid a \mid SA \mid AS$$

 $S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$$A \rightarrow 5 \mid b$$

$$B \rightarrow b$$





$$S_0 \rightarrow ASA \mid aB \mid a \mid SA \mid AS$$

 $S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$

$$A \rightarrow 5 \mid b$$

$$B \rightarrow b$$

$$S_0 \rightarrow ASA \mid aB \mid a \mid SA \mid AS$$

$$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$$

$$A \rightarrow b \mid ASA \mid aB \mid a \mid SA \mid AS$$

$$B \rightarrow b$$





6. Acrescentar variáveis e regras adicionais

$$S_0 \rightarrow ASA \mid aB \mid a \mid SA \mid AS$$

 $S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$
 $A \rightarrow b \mid ASA \mid aB \mid a \mid SA \mid AS$
 $B \rightarrow b$





6. Acrescentar variáveis e regras adicionais

$$S_0
ightarrow ASA \mid aB \mid a \mid SA \mid AS$$

 $S
ightarrow ASA \mid aB \mid a \mid SA \mid AS$
 $A
ightarrow b \mid ASA \mid aB \mid a \mid SA \mid AS$
 $B
ightarrow b$

$$S_0 \rightarrow CA \mid DB \mid a \mid SA \mid AS$$

 $S \rightarrow CA \mid DB \mid a \mid SA \mid AS$
 $A \rightarrow b \mid CA \mid DB \mid a \mid SA \mid AS$
 $B \rightarrow b$
 $C \rightarrow AS$
 $D \rightarrow a$





Gramática Livre-de-Contexto

Esdras Lins Bispo Jr. bispojr@ufg.br

Linguagens Formais e Autômatos Bacharelado em Ciência da Computação

22 de janeiro de 2018



