Gramáticas livre-de-contexto

Universidade Federal de Campina Grande – UFCG
Centro de Engenharia Elétrica e Informática – CEEI
Departamento de Sistemas e Computação – DSC
Professor: Andrey Brito
Período: 2023.2

Da aula passada: Definição formal

- Uma gramática é definida por G = <V, Σ, R, S>
  - V: conjunto das variáveis
  - Σ: conjunto dos terminais (símbolos do alfabeto da linguagem)
  - R: conjunto das regras que definem a gramática
  - S: variável inicial

Da aula passada: Que tipo de linguagens são geradas por gramáticas?

- Considere agora um tipo especial de gramática...

- Uma gramática $G = \langle V, \Sigma, R, S \rangle$ é linear (à direita) se suas produções forem da forma $A \rightarrow xB$, onde
  - $A \in V$
  - $B \in (V \cup \lambda)$
  - $x \in (\Sigma \cup \lambda)$

Exemplo 1:

$A \rightarrow 0B$
$B \rightarrow 0C$
$C \rightarrow 1$

Exemplo 2:

$A \rightarrow 0B$
$B \rightarrow 0C$
$C \rightarrow 1C \mid 0$

Da aula passada: Gramáticas Sensíveis ao Contexto

- Mais de um símbolo do lado esquerdo da regra

- Ou seja, certas trocas podem só estar habilitadas em determinados contextos

- Exemplo: G = <{S,B,C},{a,b,c},R,S>

  R: $S \rightarrow aBC \mid aSBC \mid \lambda$
  $CB \rightarrow BC$
  $aB \rightarrow ab$
  $bB \rightarrow b$
  $bC \rightarrow bc$
  $cC \rightarrow cc$
  ...

Da aula passada: $L = 0^n1^n2^n$

$S \rightarrow LDABCR$
$LDA \rightarrow LAAD$
$ADA \rightarrow AAD$
$ADB \rightarrow ABBD$
$BDB \rightarrow BBD$
$BDC \rightarrow BCCD$
$CDC \rightarrow CCD$
$DR \rightarrow ER$

$CE \rightarrow EC$
$BE \rightarrow EB$
$AE \rightarrow EA$
$LE \rightarrow LD$
$A \rightarrow 0$
$B \rightarrow 1$
$C \rightarrow 2$
$R \rightarrow \lambda$
$LD \rightarrow \lambda$

Note que você não conseguiria gerar algo diferente!

Mais restrições, mais poder...
Mais poder → mais linguagens

Gramática “livre de contexto” - Definição

- $G = (V, \Sigma, R, S)$
  - V é o conjunto de variáveis
  - Σ é o conjunto de terminais
  - R é um conjunto de regras do tipo $\alpha \rightarrow \beta$ onde $\alpha \in V$ e $\beta \in (V \cup \Sigma)^*$
  - $S \in V$ é o símbolo inicial

Regras do tipo $S \rightarrow 0S1$ são possíveis

- Toda vez que a regra for usada duas partes da palavra vão crescer de forma proporcional
- A seguinte gramática gera um linguagem que não é regular

$$S \rightarrow 0S1 \mid \lambda$$

Para gramaticas lineares à direita: $A \rightarrow xB$, onde

$A \in V$

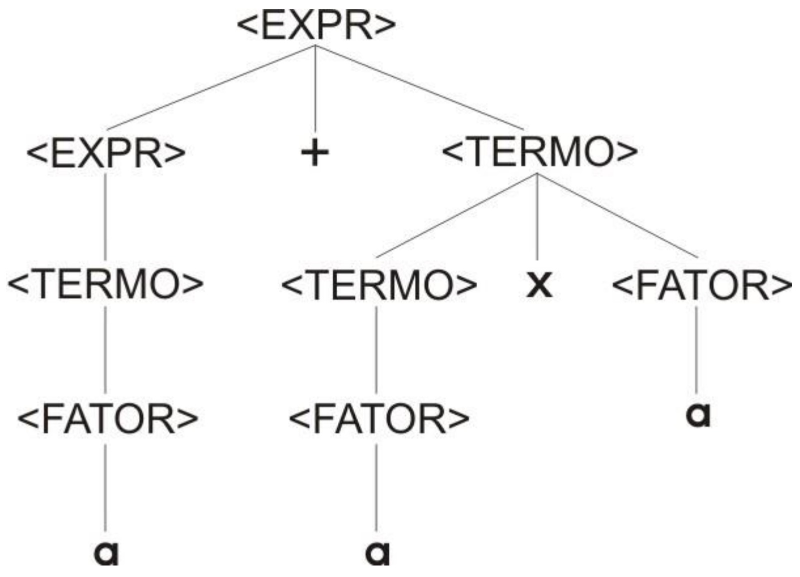
$B \in (V \cup \lambda)$

$x \in (\Sigma \cup \lambda)$

Exemplo de geração de palavra

- $a + a \times a$

<EXPR> $\rightarrow$ <EXPR> + <TERMO> | <TERMO>
<TERMO> $\rightarrow$ <TERMO> $\times$ <FATOR> | <FATOR>
<FATOR> $\rightarrow$ (<EXPR>) | a



Gramáticas livre-de-contexto

Formas normais

Linguagem “livre de contexto”

- G=<V, Σ, R, S> uma gramática livre de contexto
- Então:

$$L(G) = \{w \in \Sigma^* \mid S \Rightarrow^* w\}$$

- L é uma <u>linguagem livre de contexto</u> se, e somente se, existir uma gramática G livre-de-contexto tal que L = L(G)

Forma normal de Chomsky

- Limita os tipos de regra em uma gramática (mas sem mudar a sua capacidade!)
  - A variável inicial não aparece do lado direito e toda regra deve ser de uma das formas

    $A \rightarrow BC$

    $A \rightarrow a$

- Mas isso não é uma gramática linear?
  - Não! Tem duas variáveis!
  - Tente converter a seguinte gramática nesse formato: $S \rightarrow 0S1 \mid \#$

Forma normal de Chomsky

- Mas como chegar lá? Vamos remover os itens proibidos, tomando cuidado para não alterar a linguagem resultante!

- Ou seja, removeremos
  - O símbolo inicial do lado direito
  - Regras com mais de duas variáveis
  - Regras que misturam variáveis e terminais
  - Regras que levam uma variável à λ (a não ser para a variável inicial)

Forma normal de Chomsky

- Motivação: simplifica algoritmos que trabalham com gramáticas
  - Reconhecedores de linguagens, provas (bombeamento, autômatos)
  - Uma palavra de comprimento n precisa de **2*n -1 passos de derivação**

- Teorema: toda GLC possui uma GLC equivalente na forma normal de Chomsky

- Prova: podemos transformar uma GLC comum
  - Modificar todas as regras fora do padrão
  - Compensar as alterações para garantir que a mesma linguagem é gerada

Passo 1: símbolo inicial

- Adicione uma nova regra $S_0 \rightarrow S$
  - Assim garantimos que o símbolo inicial não apareça do lado direito

- Não precisa de compensação

Passo 2: regras com λ

- Precisam ser removidas (a não ser para o símbolo inicial): se existe A → λ, remova a regra

Passo 2: regras com λ

- Precisam ser removidas (a não ser para o símbolo inicial): se existe A → λ, remova a regra

- Compense a remoção
    - Se existe uma regra R → uAv, rescreva como R → uAv | uv
        - Obs: “u” pode conter terminais e outras variáveis
    - Se existe algo do tipo R → uAvAx, rescreva como R → uAvAx | uAvx | uvAx | uvx
    - Se existe uma regra do tipo R → A, adicione R → λ a não ser que esta regra tenha sido removida anteriormente (pois nesse caso já foi compensada)

Passo 3: regras com uma variável

- Removemos as regras do tipo A → B
- Compense
  - Se existe um B → U...
    - Adicione A → U...
    - A não ser que essa regra tenha sido removida anteriormente (i.e., não precisa de compensação)

Passo 3: regras com uma variável

- Removemos as regras do tipo A → B

Passo 4: apenas dois símbolos por regra

- Transforme todas as regras da forma $A \rightarrow u_1u_2...u_k$ onde $k \geq 3$

Passo 4: apenas dois símbolos por regra

- Transforme todas as regras da forma $A \rightarrow u_1 u_2 ... u_k$ onde $k \geq 3$
  - Seja u_i um símbolo terminal ou variável
  - Objetivo: máximo dois símbolos do lado direito
  - Resultado
    - $A \rightarrow u_1 A_1$
    - $A_1 \rightarrow u_2 A_2$
    - …
    - $A_{k-2} \rightarrow u_{k-1} u_k$

Passo 5: não misturar terminais e variáveis

Passo 5: não misturar terminais e variáveis (ou ter dois terminais)

- Se y é um terminal
  - Transforme $A \rightarrow yB$ em $A \rightarrow UB$
  - Crie $U \rightarrow y$

- Substituição simples, não precisa de compensação

- A mesma estratégia é aplicada quando temos dois terminais

Exemplo

- R: $S \rightarrow ASA \mid aB$

  $A \rightarrow B \mid S$

  $B \rightarrow b \mid \lambda$

Exemplo

$S \rightarrow ASA \mid aB$

$A \rightarrow B \mid S$

$B \rightarrow b \mid \lambda$

Exemplo: passo 1

- Removendo símbolo inicial do lado direito

$$S_0 \rightarrow S$$

$$S \rightarrow ASA \mid aB$$

$$A \rightarrow B \mid S$$

$$B \rightarrow b \mid \lambda$$

Exemplo: passo 2

- Removendo regras com λ

$$S_0 \rightarrow S$$

$$S \rightarrow ASA \mid aB$$

$$A \rightarrow B \mid S$$

$$B \rightarrow b \mid \lambda$$

Exemplo: passo 2

- Removendo regras com λ

$$S_0 \rightarrow S$$

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

$$A \rightarrow B \mid S \mid \lambda$$

$$B \rightarrow b \mid \lambda$$

Exemplo: passo 2

- Novamente: removendo regras com λ

  $S_0 \rightarrow S$

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

  $A \rightarrow B \mid S \mid \lambda$

  $B \rightarrow b$

Exemplo: passo 2

- Novamente: removendo regras com λ

$S_0 \rightarrow S$

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

$A \rightarrow B \mid S \mid \lambda$

$B \rightarrow b$

Exemplo: passo 2

- Novamente: removendo regras com λ

$S_0 \rightarrow S$

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

$A \rightarrow B \mid S \mid \lambda$

$B \rightarrow b$

Exemplo: passo 2

- Removidas regras com λ

$$S_0 \rightarrow S$$

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

$$A \rightarrow B \mid S$$

$$B \rightarrow b$$

Exemplo: passo 3

- Removendo regras com uma variável

$$S_0 \rightarrow S$$

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

$$A \rightarrow B \mid S$$

$$B \rightarrow b$$

Exemplo: passo 3

- Removendo regras com uma variável

$$S_0 \rightarrow S$$

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

$$A \rightarrow B \mid S$$

$$B \rightarrow b$$

Exemplo: passo 3

- Novamente: removendo regras com uma variável

$$S_0 \rightarrow S$$

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

$$A \rightarrow B \mid S$$

$$B \rightarrow b$$

Exemplo: passo 3

- Novamente: removendo regras com uma variável

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

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

$A \rightarrow B \mid S$

$B \rightarrow b$

Exemplo: passo 3

- Novamente: removendo regras com uma variável

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

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

$$A \rightarrow B \mid S$$

$$B \rightarrow b$$

Exemplo: passo 3

- Novamente: removendo regras com uma variável

$$S_0 \rightarrow ASA \mid aB \mid a \mid SA \mid AS$$
$$S \rightarrow ASA \mid aB \mid a \mid SA \mid AS$$
$$A \rightarrow B \mid S$$
$$B \rightarrow b$$

Exemplo: passo 3

- Novamente: removendo regras com uma variável

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

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

$$A \rightarrow B \mid S \mid b$$

$$B \rightarrow b$$

Exemplo: passo 3

- Novamente: removendo regras com uma variável

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

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

$$A \rightarrow S \mid b$$

$$B \rightarrow b$$

Exemplo: passo 3

- Novamente: removendo regras com uma variável

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

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

$$A \rightarrow S \mid b$$

$$B \rightarrow b$$

Exemplo: passo 3

- Novamente: removendo regras com uma variável

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

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

$$A \to S \mid b$$

$$B \to b$$

Exemplo: passo 3

- Novamente: removendo regras com uma variável

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

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

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

$B \rightarrow b$

Exemplo: passo 3

- Novamente: removendo regras com uma variável

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

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

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

$B \rightarrow b$

Exemplo: passo 3

- Novamente: removendo regras com uma variável

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

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

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

$$B \rightarrow b$$

Exemplo: passo 4

- Máximo dois símbolos do lado direito de uma regra

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

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

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

$$B \rightarrow b$$

Exemplo: passo 4

- Máximo dois símbolos do lado direito de uma regra

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

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

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

$B \rightarrow b$

Exemplo: passo 4

- Máximo dois símbolos do lado direito de uma regra

$S_0 \rightarrow ASA \mid AA_1 \mid aB \mid a \mid SA \mid AS$

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

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

$A_1 \rightarrow SA$

$B \rightarrow b$

Exemplo: passo 4

- Máximo dois símbolos do lado direito de uma regra

$$S_0 \rightarrow AA_1 \mid aB \mid a \mid SA \mid AS$$

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

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

$$A_1 \rightarrow SA$$

$$B \rightarrow b$$

Exemplo: passo 4

- Máximo dois símbolos do lado direito de uma regra

$S_0 \rightarrow AA_1 \mid aB \mid a \mid SA \mid AS$

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

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

$A_1 \rightarrow SA$

$B \rightarrow b$

Exemplo: passo 4

- Máximo dois símbolos do lado direito de uma regra

  $S_0 \to AA_1 \mid aB \mid a \mid SA \mid AS$

  $S \to ASA \mid AA_1 \mid aB \mid a \mid SA \mid AS$

  $A \to b \mid ASA \mid AA_1 \mid aB \mid a \mid SA \mid AS$

  $A_1 \to SA$

  $B \to b$

Exemplo: passo 4

- Máximo dois símbolos do lado direito de uma regra

$$S_0 \rightarrow AA_1 \mid aB \mid a \mid SA \mid AS$$

$$S \rightarrow AA_1 \mid aB \mid a \mid SA \mid AS$$

$$A \rightarrow b \mid AA_1 \mid aB \mid a \mid SA \mid AS$$

$$A_1 \rightarrow SA$$

$$B \rightarrow b$$

Exemplo: passo 4

- Máximo dois símbolos do lado direito de uma regra

$S_0 \rightarrow AA_1 \mid aB \mid a \mid SA \mid AS$

$S \rightarrow AA_1 \mid aB \mid a \mid SA \mid AS$

$A \rightarrow b \mid AA_1 \mid aB \mid a \mid SA \mid AS$

$A_1 \rightarrow SA$

$B \rightarrow b$

Exemplo: passo 5

- Não misturar símbolos terminais e variáveis

$S_0 \rightarrow AA_1 \mid aB \mid a \mid SA \mid AS$

$S \rightarrow AA_1 \mid aB \mid a \mid SA \mid AS$

$A \rightarrow b \mid AA_1 \mid aB \mid a \mid SA \mid AS$

$A_1 \rightarrow SA$

$B \rightarrow b$

Exemplo: passo 5

- Não misturar símbolos terminais e variáveis (ou ter dois terminais)

$$S_0 \rightarrow AA_1 \mid aB \mid a \mid SA \mid AS$$

$$S \rightarrow AA_1 \mid aB \mid a \mid SA \mid AS$$

$$A \rightarrow b \mid AA_1 \mid aB \mid a \mid SA \mid AS$$

$$A_1 \rightarrow SA$$

$$B \rightarrow b$$

Exemplo: passo 5

- Não misturar símbolos terminais e variáveis

$S_0 \rightarrow AA_1 \mid aB \mid a \mid SA \mid AS$

$S \rightarrow AA_1 \mid aB \mid a \mid SA \mid AS$

$A \rightarrow b \mid AA_1 \mid aB \mid a \mid SA \mid AS$

$A_1 \rightarrow AS$

$U \rightarrow a$

$B \rightarrow b$

Exemplo: passo 5

- Não misturar símbolos terminais e variáveis

$S_0 \rightarrow AA_1 \mid UB \mid a \mid SA \mid AS$

$S \rightarrow AA_1 \mid UB \mid a \mid SA \mid AS$

$A \rightarrow b \mid AA_1 \mid UB \mid a \mid SA \mid AS$

$A_1 \rightarrow AS$

$U \rightarrow a$

$B \rightarrow b$

Transformação de uma gramática para FNC

$S \rightarrow ASA \mid aB$

$A \rightarrow B \mid S$

$B \rightarrow b \mid \lambda$

$S_0 \rightarrow AA_1 \mid UB \mid a \mid SA \mid AS$

$S \rightarrow AA_1 \mid UB \mid a \mid SA \mid AS$

$A \rightarrow b \mid AA_1 \mid UB \mid a \mid SA \mid AS$

$A_1 \rightarrow SA$

$U \rightarrow a$

$B \rightarrow b$

Forma normal de Greibach

- Uma GLC $G = \langle V, \Sigma, R, S \rangle$ está na forma normal de Greibach se suas produções forem da forma $A \rightarrow xB$, onde
  - $A \in V$
  - $B \in (V \cup \lambda)^*$
  - $x \in \Sigma$

- Toda GLC pode ser escrita na FNG

- Uma vantagem: toda palavra de n símbolos é gerada em ***n*** derivações

Ambiguidade

Derivando uma palavra de formas diferentes

- Como gerar a palavra 0011 usando a seguinte gramática?

  S → AB
  A → 0A | λ
  B → 1B | λ

- Existe mais de uma forma de gerar a mesma palavra?

- São regras diferentes ou as mesmas regras? A árvore sintática é diferente?
  - Note que a sequência de derivações para cada variável é o que importa
  - (É isso que explica como a palavra foi gerada)

Derivação mais à esquerda

- Fazer primeiro a substituição das variáveis à esquerda
- Poderia ser também “à direita”, o importante é fixar uma ordem
- No caso anterior, as regras aplicadas são claramente as mesmas e a árvore não era realmente diferente

Árvore de derivação

- Mas nem sempre substituir mais à esquerda resolve...
  - S → 0S1 | 1S0 | SS | λ
  - Como gerar 010101 ?

Árvore de derivação

- Mas nem sempre substituir mais à esquerda resolve...
  - S → 0S1 | 1S0 | SS | λ
  - Como gerar 010101 ?
  - Aí sim, é ambiguidade!
- Lembre que a árvore sintática dá a interpretação da palavra, então ambiguidade é ruim para geração de um compilador
  - Não sabe qual foi a regra aplicada
- Qual o impacto de ambiguidade em expressões?