Teoria da Computação
Linguagens Livres de Contexto
Árvore de Derivação e Ambiguidade

Prof. Jefferson Magalhães de Morais

Aceitação de uma sentença

- Busca-se localizar uma **sequência de produções** que, quando aplicada à raiz da gramática, forneça como resultado a sentença fornecida para análise
- Sendo possível completar a derivação, diz-se que a sentença pertence à linguagem, caso contrário, ela não pertence à linguagem
- Na prática, é possível identificar uma grande quantidade de sequências distintas de derivação, sendo que todas elas resultam na mesma sentença analisada (escolhas arbitrárias de símbolo não-terminal a ser substituído)
- Costuma-se fixar **critérios de derivação de sentenças** (útil para construir reconhecedores sintáticos)

Derivações mais à esquerda e mais à direita

- A **derivação mais à esquerda** quando a substituição de um não-terminal, do lado direito da produção, ocorre mais à esquerda da cadeia. De forma análoga, define-se a **derivação mais à direita**
- **Exemplo**: considere as sequências de derivações para a sentença $a + a$

$$\{E \to T + E,\quad E \to T,\quad T \to F * T,\quad T \to F,\quad F \to (E),\quad F \to a\}$$

1. **Derivações mais à esquerda**

   $E \Rightarrow T + E \Rightarrow F + E \Rightarrow a + E \Rightarrow a + T \Rightarrow a + F \Rightarrow a + a$

2. **Derivações mais à direita**

   $E \Rightarrow T + E \Rightarrow T + T \Rightarrow T + F \Rightarrow T + a \Rightarrow F + a \Rightarrow a + a$

3. **Derivações arbitrárias**

   $E \Rightarrow T + E \Rightarrow F + E \Rightarrow F + T \Rightarrow a + T \Rightarrow a + F \Rightarrow a + a$

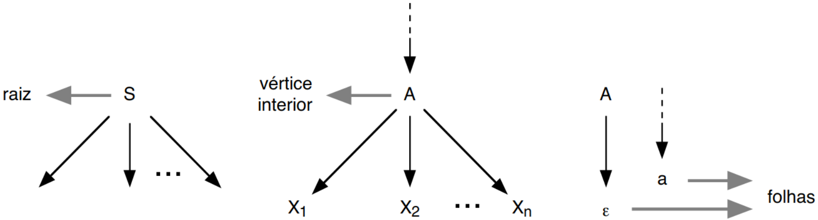
Árvores de derivação

- É um sistema de representação da **sequência de derivação** de uma palavra

- Auxilia na demonstração formal de teoremas e facilita a representação interna em compiladores e interpretadores

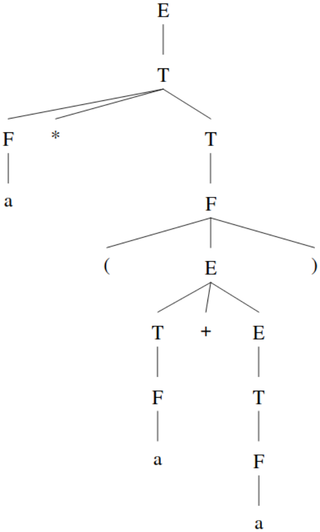
- A árvore é um grafo orientado, acíclico e com as propriedades
  1. Todo vértice é rotulado com um elemento de $V \cup \{\varepsilon\}$
  2. O rótulo da raiz é S
  3. Os rótulos de vértices internos são elementos de $N = V - \Sigma$
  4. Se um vértice tem o rótulo A, e $X_1, X_2, \ldots, X_n$ são descendentes diretos de A, ordenados da esquerda para a direita, então $A \Rightarrow X_1, X_2, \ldots, X_n$ deve pertencer ao conjunto P de regras
  5. O rótulo de um vértice folha é um símbolo terminal ou um símbolo vazio. O vazio deve ser descendente único de seu ancestral direto ($A \rightarrow \varepsilon$)

Representação das árvores de derivação

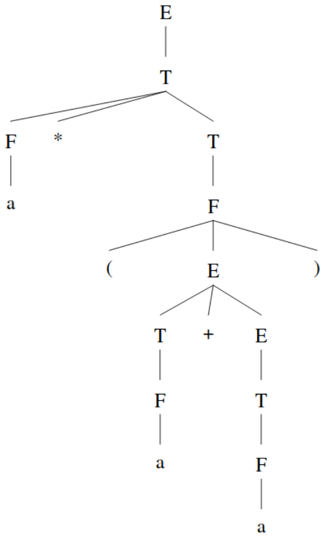


Árvores de derivação

- **Fronteira** é a cadeia formada pela concatenação dos símbolos que são folhas (da esquerda para a direita)
- A sentença ao lado é da forma $a * (a + a)$
- A árvore de derivação informa quais produções foram aplicadas, mas **não** em que ordem elas foram aplicadas
- Exemplo: na profundidade 2, quem foi derivado primeiro, F ou T?

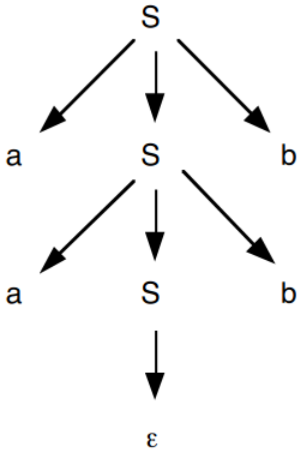


- Um única árvore de derivação pode representar derivações distintas de uma mesma palavra
  1. Derivações mais à esquerda
     $E \Rightarrow T \Rightarrow F * T \Rightarrow a * T \Rightarrow a * F \Rightarrow a * (E) \Rightarrow a * (T + E) \Rightarrow a * (F + E) \Rightarrow a * (a + E) \Rightarrow a * (a + T) \Rightarrow a * (a + F) \Rightarrow a * (a + a)$
  2. Derivações mais à direita
     $E \Rightarrow T \Rightarrow F * T \Rightarrow F * T \Rightarrow F * F \Rightarrow F * (E) \Rightarrow F * (T + E) \Rightarrow F * (T + T) \Rightarrow F * (T + F) \Rightarrow F * (T + a) \Rightarrow F * (F + a) \Rightarrow F * (a + a) \Rightarrow a * (a + a)$
  3. Derivações arbitrárias
     $E \Rightarrow T \Rightarrow F * T \Rightarrow F * F \Rightarrow a * F \Rightarrow a * (E) \Rightarrow a * (T + E) \Rightarrow a * (T + T) \Rightarrow a * (F + T) \Rightarrow a * (F + F) \Rightarrow a * (a + F) \Rightarrow a * (a + a)$



Exemplo

Considere a gramática livre de contexto $G = (\{S, a, b\}, \{a, b\}, \{S \to aSb \mid \varepsilon\}, S)$ que gera a linguagem $L(G) = \{a^n b^n \mid n \geq 0\}$ e a árvore de derivação para a sentença $aabb$



Ambiguidade

- Uma GLC é dita **ambígua** quando $\exists \alpha \in L(G), \alpha \in \Sigma^*$ que possua mais de uma sequência distinta de derivações exclusivamente à esquerda ou a direita
- Equivalentemente, GLC é ambígua quando uma mesma palavra pode ser associada a mais de uma árvore de derivação

- Em muitas aplicações, e.g. otimização de algoritmos de reconhecimento, é conveniente que a gramática usada seja não-ambígua

- Um GLC G é **não-ambígua** se $\forall w \in L(G), w \in \Sigma^*$ existir uma única sequência de derivações mais à esquerda e uma única sequência de derivações mais à direita que a gerem

Exemplo: considere a gramática das expressões aritméticas sobre $\{a, +, *, (,)\}$ utilizando apenas um símbolo não-terminal E e com as produções

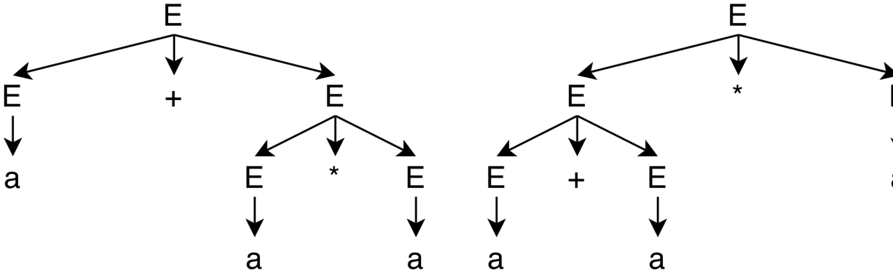
$$\begin{aligned} \{E &\rightarrow E + E, \\ E &\rightarrow E * E, \\ E &\rightarrow a, \\ E &\rightarrow (E)\} \end{aligned}$$

Pode-se perceber que a sentença $a + a * a$ pode ser derivada à esquerda ao menos de duas formas distintas

- $E \Rightarrow E + E \Rightarrow a + E \Rightarrow a + E * E \Rightarrow a + a * E \Rightarrow a + a * a$
- $E \Rightarrow E * E \Rightarrow E + E * E \Rightarrow a + E * E \Rightarrow a + a * E \Rightarrow a + a * a$

Ambiguidade

- $E \Rightarrow E + E \Rightarrow a + E \Rightarrow a + E * E \Rightarrow a + a * E \Rightarrow a + a * a$
- $E \Rightarrow E * E \Rightarrow E + E * E \Rightarrow a + E * E \Rightarrow a + a * E \Rightarrow a + a * a$



- $\therefore$ a gramática é ambígua

Ambiguidade

- **Teorema (Derivações à esquerda $\Leftrightarrow$ derivações à direita)**: Se $w \in L(G)$, com G sendo uma GLC, e existindo duas (ou mais) derivações mais à esquerda para w em G, então existem também, correspondentemente, duas (ou mais) derivações mais à direita para w em G

  Considere-se uma GLC G e a sequência de derivações mais à esquerda $S \Rightarrow^* \alpha_1 X \gamma$, com $\alpha_1 \in \Sigma^*, X \in N$, e $\gamma \in V^*$

  Considere que $X \Rightarrow^* \alpha_2$, com $\alpha_2 \in \Sigma^*$ e $\gamma \Rightarrow^* \alpha_3$, com $\alpha_3 \in \Sigma^*$, então $S \Rightarrow^* \alpha_1 X \gamma \Rightarrow^* \alpha_1 \alpha_2 \alpha_3$ e $\alpha_1 \alpha_2 \alpha_3 \in L(G)$

  Sejam $X \rightarrow \beta_1$ e $X \rightarrow \beta_2$ duas produções distintas de G, então existem duas derivações à esquerda distintas para cada cadeia $\alpha_1 \alpha_2 \alpha_3$

  - $S \Rightarrow^* \alpha_1 X \gamma \Rightarrow \alpha_1 \beta_1 \gamma \Rightarrow^* \alpha_1 \alpha_2 \gamma \Rightarrow^* \alpha_1 \alpha_2 \alpha_3$
  - $S \Rightarrow^* \alpha_1 X \gamma \Rightarrow \alpha_1 \beta_2 \gamma \Rightarrow^* \alpha_1 \alpha_2 \gamma \Rightarrow^* \alpha_1 \alpha_2 \alpha_3$

- De maneira análoga, se $\alpha_1\alpha_2\alpha_3 \in L(G)$ e $X \Rightarrow^* \alpha_2$, então existe uma sequência de derivações mais à direita $S \Rightarrow^* \mu X \alpha_3$, com $\mu \in V^*$ e $\mu \Rightarrow^* \alpha_1$

  Portanto, $S \Rightarrow^* \mu X \alpha_3 \Rightarrow^* \mu\alpha_2\alpha_3 \Rightarrow^* \alpha_1\alpha_2\alpha_3$

  Como, por hipótese, $X \Rightarrow \beta_1 \Rightarrow^* \alpha_2$ e $X \Rightarrow \beta_2 \Rightarrow^* \alpha_2$, então existem duas sequências de derivações mais à direita para a cadeia $\alpha_1\alpha_2\alpha_3$

  - $S \Rightarrow^* \mu X \alpha_3 \Rightarrow \alpha_1\beta_1\alpha_3 \Rightarrow^* \mu\alpha_2\alpha_3 \Rightarrow^* \alpha_1\alpha_2\alpha_3$
  - $S \Rightarrow^* \mu X \alpha_3 \Rightarrow \alpha_1\beta_2\alpha_3 \Rightarrow^* \mu\alpha_2\alpha_3 \Rightarrow^* \alpha_1\alpha_2\alpha_3$

Ambiguidade

- **Teorema (Múltiplas derivações $\Leftrightarrow$ múltiplas árvores)**: Seja G uma GLC. Para toda cadeia $w \in L(G)$, o número de sequências distintas de derivações mais à esquerda (e portanto mais à direita) é igual ao número de árvores de derivação distintas que representam w

- **Teorema (Múltiplas derivações $\Leftrightarrow$ única árvore)**: Seja G uma GLC não-ambígua. Para toda cadeia $w \in L(G)$, toda e qualquer sequência de derivações que produz w é representada através da mesma e única árvore de derivação

Ambiguidade: resumidamente

- Cada uma das cadeias pertencentes a uma linguagem gerada por uma GLC não-ambígua possuí uma única sequência de derivações mais à esquerda, uma única sequência de derivações mais à direita e uma única árvore de derivação
- Uma GLC é ambígua se for possível identificar, na linguagem por ela gerada, pelo menos uma cadeia que
  - possa ser derivada por **duas ou mais** sequências distintas de derivações mais à esquerda
  - possa ser derivada por **duas ou mais** sequências distintas de derivações mais à direita
  - possa ser representada por **duas ou mais** árvores de derivação distintas

Linguagens inerentemente ambíguas

- Raramente se diz que uma linguagem é ambígua, uma vez que uma mesma linguagem pode ser gerada por inúmeras gramáticas distintas (ambíguas e não-ambíguas)
- Entretanto, há casos em que se pode provar que uma **linguagem é inerentemente ambígua**, indicando que toda e qualquer gramática que gera a linguagem deve ser necessariamente ambígua
- **Exemplo**: $L = \{a^n b^n c^m d^m \mid n \geq 1, m \geq 1\} \cup \{a^n b^m c^m d^n \mid n \geq 1, m \geq 1\}$ é inerentemente ambígua

Exercício

- A linguagem $L = \{a^i b^j c^k \mid i = j \text{ ou } j = k\}$ é inerentemente ambígua. Apresente uma GLC e uma sentença $w \in L(G)$ com mais de uma derivação à esquerda

$$
\begin{aligned}
P = \{ & \\
& S \to C \mid B \\
& A \to aAb \mid \varepsilon \\
& C \to Cc \mid A \\
& B \to aB \mid D \\
& D \to bDc \mid \varepsilon\}
\end{aligned}
$$

A sentença $w = \varepsilon$ com duas derivações à esquerda

$S \Rightarrow C \Rightarrow A \Rightarrow \varepsilon$

$S \Rightarrow B \Rightarrow D \Rightarrow \varepsilon$