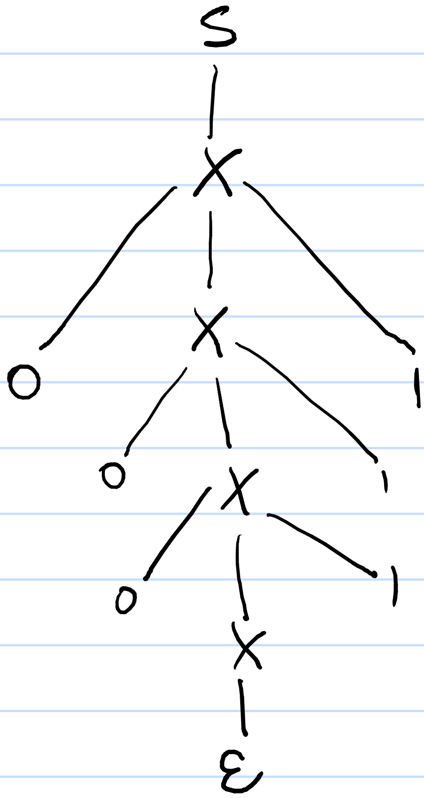
Ex:

$S \rightarrow X$

$X \rightarrow 0X1 \mid \varepsilon$

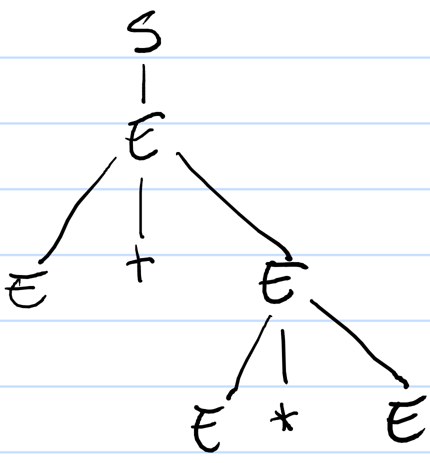
$w = 000111$

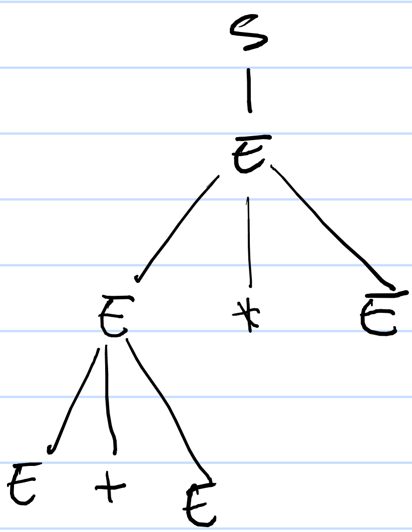


Ex: ID + ID * ID

$S \rightarrow E$

$E \rightarrow E+E \mid E*E \mid ID$





Árvores de análise sintática

- Diagramar uma linguagem

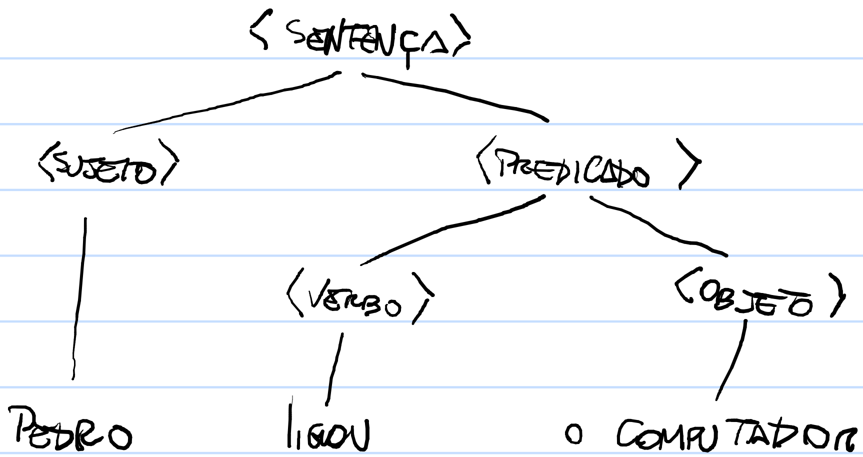
- Gramáticas geram frases com sintaxe correta

- Gostaríamos de analisar a estrutura de uma frase e determinar o seu significado.

Ex: Pedro ligou o computador

verbo objeto

<sujeito> predicado



- Cada bifurcação corresponde a uma regra da gramática

<sentença> → <suj> <pred>
<pred> → <verbo> <obj>
<suj> → Pedro
<verbo> → ligou
<obj> → o computador

DEF FORMAL

- ÁRVORES: GRAFOS SEM CICLOS

- AQUI: ORIENTADOS, SEM SETAS, "DE CIMA P/ BAIXO"

1) HÁ UM ÚNICO VÉRTICE CHAMADO RAIZ

2) OS SUCESSORES DE UM VÉRTICE SÃO TOTALMENTE ORDENADOS "DA ESQUERDA P/ DIREITA"
   ↳ PORQUE QUEREMOS ESCREVER DE ESQ. P/ DIR.

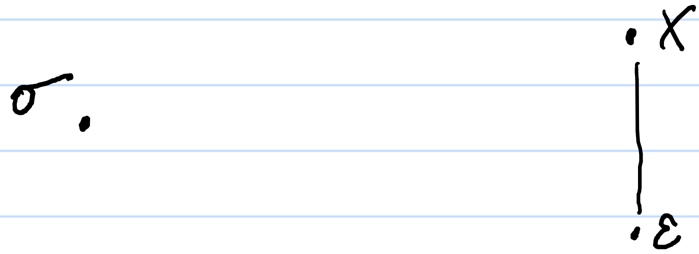
DEF: SEJA V' UM VÉRTICE DA ÁRVORE. HÁ UM ÚNICO CAMINHO QUE LIGA A RAIZ A V'. SEJA V UM VÉRTICE DESSE CAMINHO.

- V É ACCENDENTE DE V', E V' É DESCENDENTE DE V.
- SE V E V' SÃO SEPARADOS POR UMA ÚNICA ARESTA, ENTÃO V É PAI DE V', E V' É FILHO DE V
- FILHOS DO MESMO PAI SÃO IRMÃOS
- VÉRTICE SEM FILHO É UMA FOLHA
- VÉRTICE QUE NÃO É FOLHA É INTERIOR

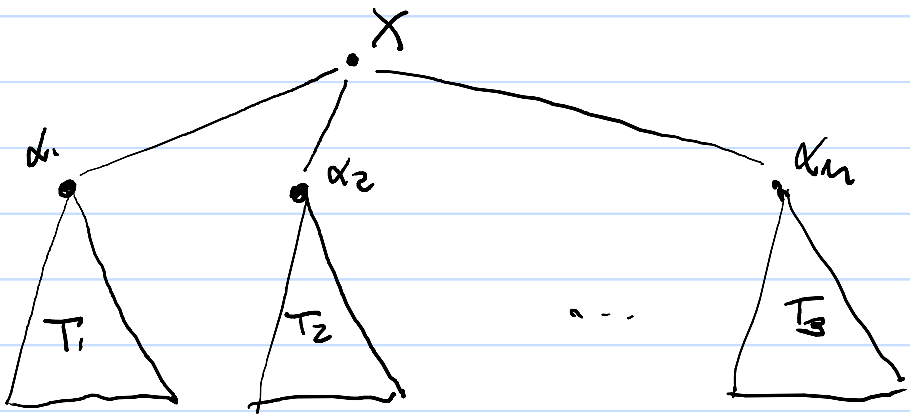
- A ORDENAÇÃO ENTRE IRMÃOS PODE SER ESTENDIDA PARA UMA ORDENAÇÃO DAS FOLHAS

DEF: ÁRVORES DE ANÁLISE SINTÁTICA. DEFINIDAS RECURSIVAMENTE

BÁSICAS: $\sigma \in T$, $X \in V$, e $X \to \varepsilon$ é uma regra



REGRAS DE COMBINAÇÃO: $T_1, \dots, T_n$ árvores gramaticais com rótulos $\alpha_1, \dots, \alpha_n$, resp. Se $X \to \alpha_1 \alpha_2 \cdots \alpha_n$ é uma regra de $\mathcal{G}$, então a árvore



OBS: Se τ possui vértice rotulado por $X \in V$, podemos substituir toda parte de τ descendente de X por outra com raiz X.

OBS2: Os únicos vértices rotulados por $T \cup \varepsilon$ são as folhas.

- Se $v \in \tau$ é rotulado por X e seus filhos por $\alpha_1, \dots, \alpha_n$, $X \to \alpha_1 \cdots \alpha_n$ é a REGRA ASSOCIADA

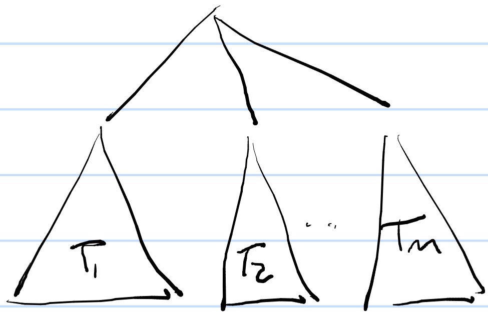
Colheita:

• Colheita de uma árvore (ou resultado)

Básicas

$$C(\bullet^{\sigma}) = \sigma \qquad c\left(\begin{matrix}X\\ |\\ \varepsilon\end{matrix}\right) = \varepsilon$$

Se

$T=$ 

, então $c(T) = c(T_1)\cdots c(T_m)$

• Como as folhas estão totalmente ordenadas, não há ambiguidade na colheita.

S-árvore

Def: Se $w \in L(G)$, então uma árvore de derivação p/ w é uma árvore com raiz S cuja colheita é w.

Def: Seja $w \in (T \cup V)^*$, $w = \alpha_1 \cdots \alpha_m$
Uma w-floresta $\mathcal{F}$ é uma seq. ordenada $T_1, \dots, T_m$ de árvores gramaticais, onde T_i é α_i-árvore.
A colheita de $\mathcal{F}$ é $c(\mathcal{F}) = c(T_1) \cdots c(T_2)$

• Como obter uma derivação (mais à esq.) a partir de uma árvore?

Alg.: Entrada: uma X-árvore τ, $X \in V$
Saída: uma derivação mais à esq. de $c(\tau)$
Alg 9.7 Pág 138

O alg funciona pois em cada passo removemos uma variável e a substituímos por sua regra.

Equivalência

PROPOSIÇÃO: SEJA X UMA VARIÁVEL DE UMA LCC G. SE EXISTE DERIVAÇÃO $X \Rightarrow^* w$, ENTÃO w É COLHEITA DE UMA X-ÁRVORE EM G.

PROVA: INDUÇÃO NO NÚMERO DE PASSOS p

BASE: $p=1$. ENTÃO $X \Rightarrow w$ E $w \in T^*$

PASSO DE INDUÇÃO: $p+1$ PRIMEIRO PASSO $X \Rightarrow v_1 \cdots v_m$, ONDE $v_i \in T \cup V$. ONDE $X \to v_1 \cdots v_m$ É REGRA DE G.

A DERIVAÇÃO CONTINUA COM $v_i \Rightarrow^* u_i$ ONDE $u_1 \cdots u_m = w$.
MAS ESSAS DERIVAÇÕES TÊM COMPRIMENTO NO MÁX p.
EXISTEM $t_1, \dots, t_m$ ÁRVORES COM $c(t_i) = u_i$ HI

TEO.: SEJA G UMA LCC E $w \in L(G)$. ENTÃO

1) EXISTE ÁRVORE DE DERIVAÇÃO CUJA COLHEITA É w
2) CADA ÁRVORE DE DERIVAÇÃO CUJA COLHEITA É w CORRESPONDE A UMA ÚNICA DERIVAÇÃO MAIS À ESQ. (RESP. À DIR) DE w.

Ambiguidade:

- Árvores gramaticais nos ajudam a interpretar corretamente uma frase

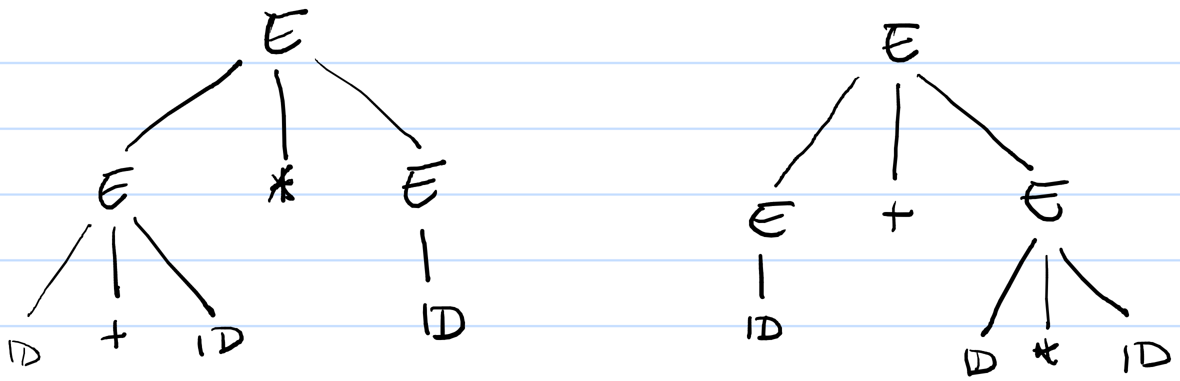
- Algumas frases admitem interpretações distintas

Ex: A seguir veio uma mãe com uma criança empurrando um carrinho.

Quem empurra o carrinho?

Def: A gramática G é ambígua se existe palavra $w \in L(G)$ que admite duas árvores de derivação distintas.

Ex: ID + ID * ID



- Uma sol. é criar uma nova gramática que gera a mesma linguagem, mas que não seja ambígua.
  - ↳ Introduzindo novas variáveis e regras

Ex:

$E \rightarrow E + F$

$E \rightarrow F$

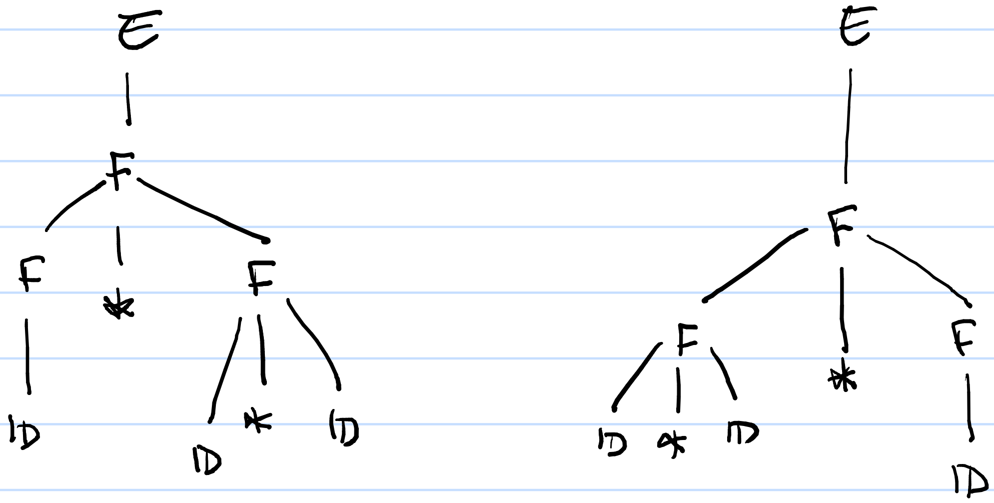
$F \rightarrow F * F$

$F \rightarrow (E)$

$F \rightarrow ID$

Evita o problema com ID + ID * ID, mas ainda é ambígua.

Ainda é ambígua pois



Geram a mesma palavra ID * ID * ID

• Para resolver isso podemos introduzir uma nova variável.

G'_{EXP}

$E \to E + T \mid T$

$T \to T * F \mid F$ FATOR

$F \to (E) \mid ID$ TERMO

<u>Ex</u>: G''_{EXP}

$S \to E$

$E \to (E)R \mid VR$

$R \to +E \mid *E \mid \varepsilon$

$V \to ID$

Não é ambígua pois se fixarmos $X \in V$, a regra $X \to w$ é caracterizada pelo primeiro terminal.

Prova: Derivação mais à esq., n-ésimo passo: $S \Rightarrow^* uXv \Rightarrow w$