Grammatiche Libere

DEFINIZIONE: Una grammatica libera dal contesto è una QUADRUPLA $G=(V,T,P,S)$ dove:

- V è un insieme finito di variabili
- T è un alfabeto di simboli terminali
- P è un insieme finito di produzioni della forma $A \to \alpha$
  - $A \in V$ è detta testa della produzione
  - $\alpha \in (V \cup T)^*$ è detta corpo della produzione
- $S \in V$ è il simbolo iniziale della grammatica

Fissata una grammatica $G=(V,T,P,S)$ definiamo le DERIVAZIONI in uno o più passi come segue:

- $\alpha A \beta \Rightarrow_G \alpha \gamma \beta$ se $A \to \gamma \in P$
  $\alpha\gamma\beta$ deriva in un passo da $\alpha A \beta$ in G

- $\Rightarrow^*_G$ per la chiusura riflessiva e transitiva di $\Rightarrow_G$
  - $\alpha \Rightarrow^*_G \alpha$
  - se $\alpha \Rightarrow_G \beta$ e $\beta \Rightarrow^*_G \gamma$, allora $\alpha \Rightarrow^*_G \gamma$

  Quando $\alpha \Rightarrow^*_G \beta$ diciamo che β deriva in più passi da α in G

LINGUAGGIO GENERATO DA UNA GRAMMATICA

DEFINIZIONE: Data una grammatica G, il linguaggio generato da G è definito come:

$$L(G) \stackrel{def}{=} \{ w \in T^* \mid S \Rightarrow_G^* w \}$$

potere espressivo maggiore

TEOREMA: Per ogni linguaggio regolare L $\exists$ una grammatica libera G tale che $L(G) = L$

DIMOSTRAZIONE: sia $A = (Q, \Sigma, \delta, q_0, F)$ che riconosce L

definiamo $G = (Q, \Sigma, P, q_0)$

- se $q \in Q$ e $a \in \Sigma$ allora $q \to a\,\delta(q, a) \in P$
- se $q \in F$ allora $q \to \varepsilon \in P$

$$q_0 \Rightarrow^* w \iff \hat{\delta}(q_0, w) \in F$$

ALBERI SINTATTICI

CARATTERISTICHE:

- Ogni nodo interno ($\neq$ foglie) è etichettato con una variabile in V
- Ogni foglia è etichettata con una variabile in V o da un terminale in T o da ε
- Se una foglia è etichettata con ε è anche l'unico figlio del genitore
- Se un nodo interno è etichettato con A, i suoi figli sono etichettati (da SX a DX) con $X_1, X_2, \dots X_n$, allora $A \to X_1 X_2 \dots X_n$ è una produzione in P

Il PRODOTTO di un albero sintattico è la stringa ottenuta concatenando, da SX a DX, le etichette di tutte le foglie dell'albero

TEOREMA: $A \Rightarrow^*_G \alpha$ se e solo se $\exists$ un albero sintattico di G con radice A e prodotto α

DEFINIZIONE: DERIVAZIONI CANONICHE Una derivazione $X \Rightarrow^* \alpha$ si dice derivazione a sinistra se ad ogni passo viene riscritta la variabile più a sinistra. Viceversa per le derivazioni a destra. Rispettivamente le indichiamo con pedice "lm" e "rm" leftmost e rigthmost

ELIMINAZIONE DELLE AMBIGUITÀ

ESEMPIO: ESPRESSIONI ARITMETICHE

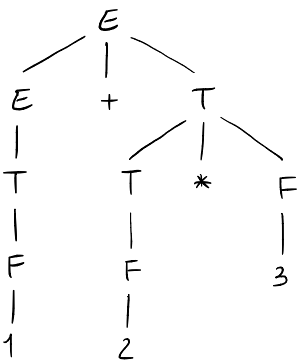
- Stratifico e sbilancio le espressioni
- Espressione = somma di termini
- Termine = prodotto di fattori
- Fattore = costante o espressione tra parentesi

$$(\{E,T,F\}, \{0,1,2,3,4,5,6,7,8,9,+,*,(,)\}, P, E)$$

con P insieme di produzioni:

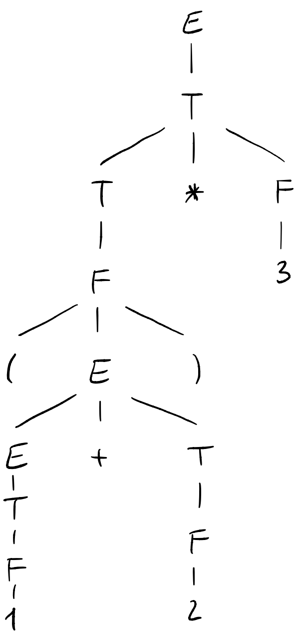
- $E \to T \mid E+T$ ← si creano espressioni associative sbilanciate verso sinistra
- $T \to F \mid T*F$
- $F \to 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9 \mid (E)$

Deriviamo la stringa $1+2*3$



Grazie alla stratificazione forzata non è più possibile ottenere un albero sintattico equivalente in quanto siamo costretti a scegliere prima il + o il *

Stringa $(1+2)*3$



NON ESISTE un algoritmo generale per disambiguare un linguaggio

esempio di LINGUAGGIO INERENTEMENTE AMBIGUO

$$L = \{ a^n b^n c^m d^m \mid n \geq 1, m \geq 1 \} \cup \{ a^n b^m c^m d^n \mid n \geq 1, m \geq 1 \}$$

In ogni grammatica che genera L ci sono sempre almeno due derivazioni canoniche distinte che generano una stringa di forma $a^n b^n c^n d^n$

$$\begin{aligned} S &\Rightarrow_{lm} AB \\ &\Rightarrow_{lm} aAbB \\ &\Rightarrow_{lm} aabbB \\ &\Rightarrow_{lm} aabbcBd \\ &\Rightarrow_{lm} aabbccdd \end{aligned} \qquad \begin{aligned} S &\Rightarrow_{lm} C \\ &\Rightarrow_{lm} aCd \\ &\Rightarrow_{lm} aaDdd \\ &\Rightarrow_{lm} aabDcd \\ &\Rightarrow_{lm} aabbccdd \end{aligned}$$