ANALISI SINTATTICA

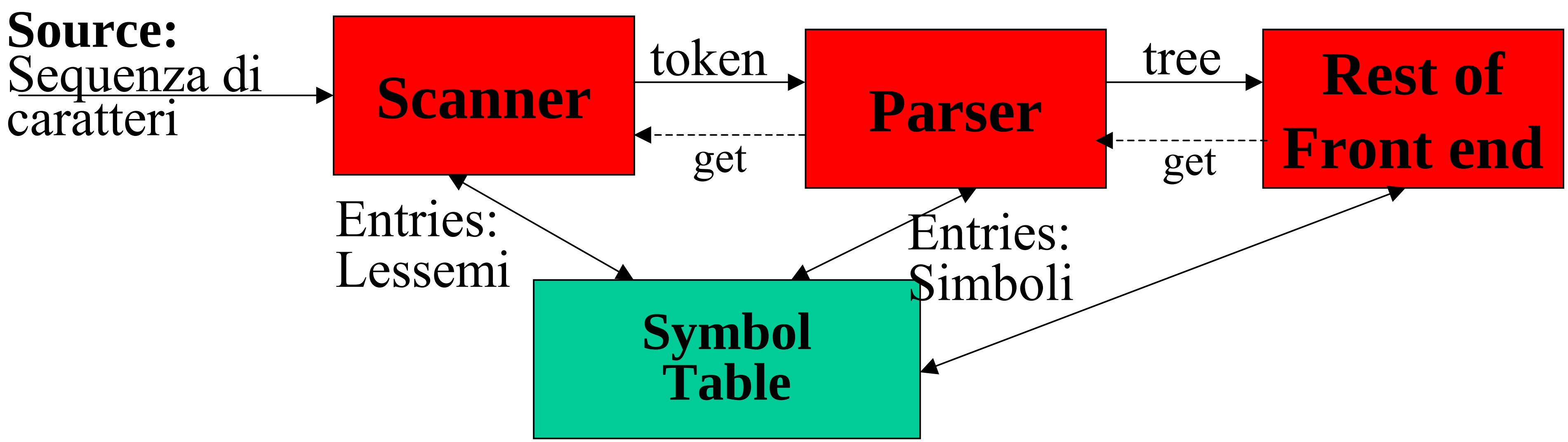
* Riconosce sequenze di tokens che formano strutture della sintassi del linguaggio

* Linguaggio sintattico piu' complesso di quello lessicale

$\{u^n v^n \mid n \geq 0\}$ non e' regolare

ma: num+(((id*id)+num)/id)

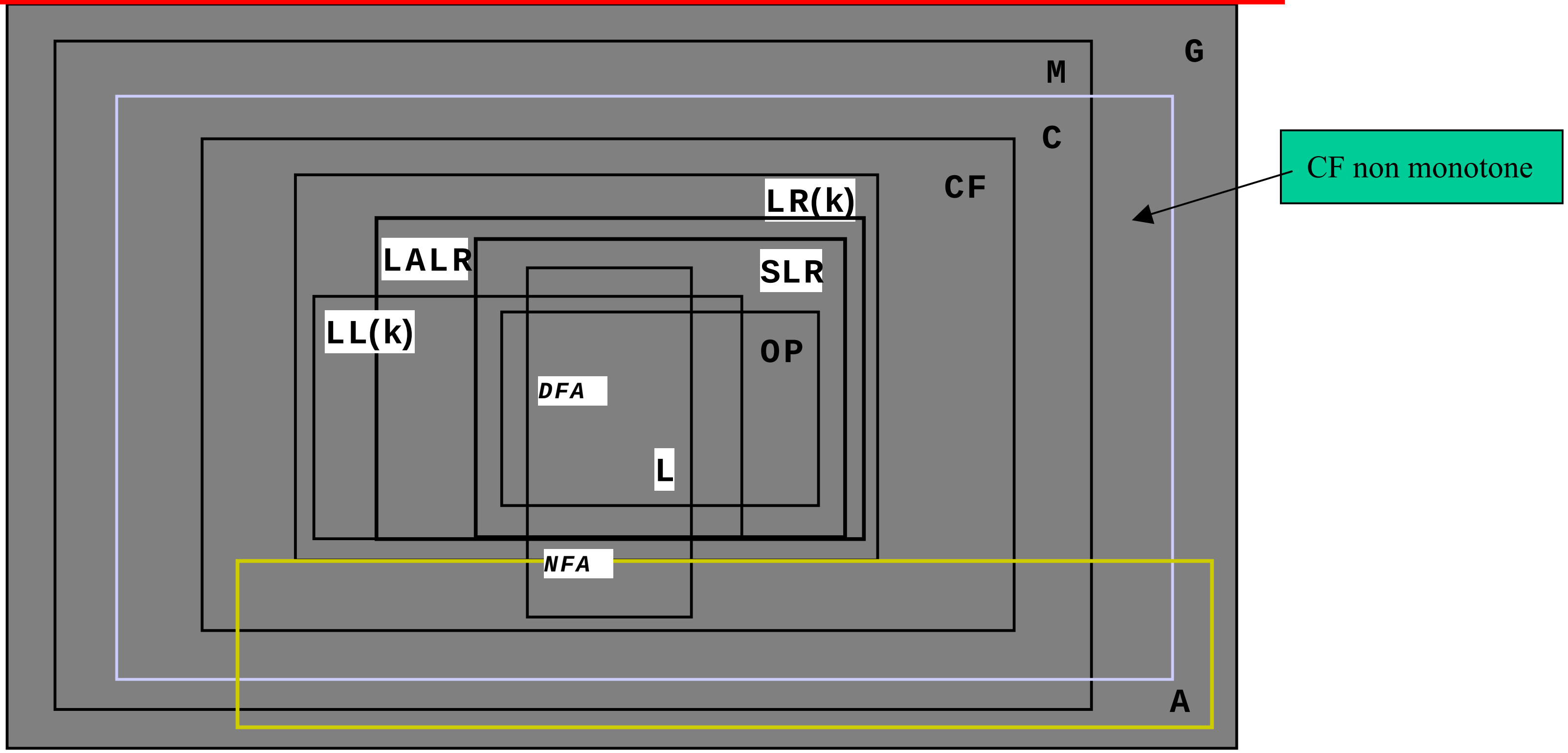
Una struttura una passata



Analisi Sintattica (Parser)
guidata dagli altri strumenti

Classificazione delle grammatiche (Chomsky)

Inclusione tra grammatiche

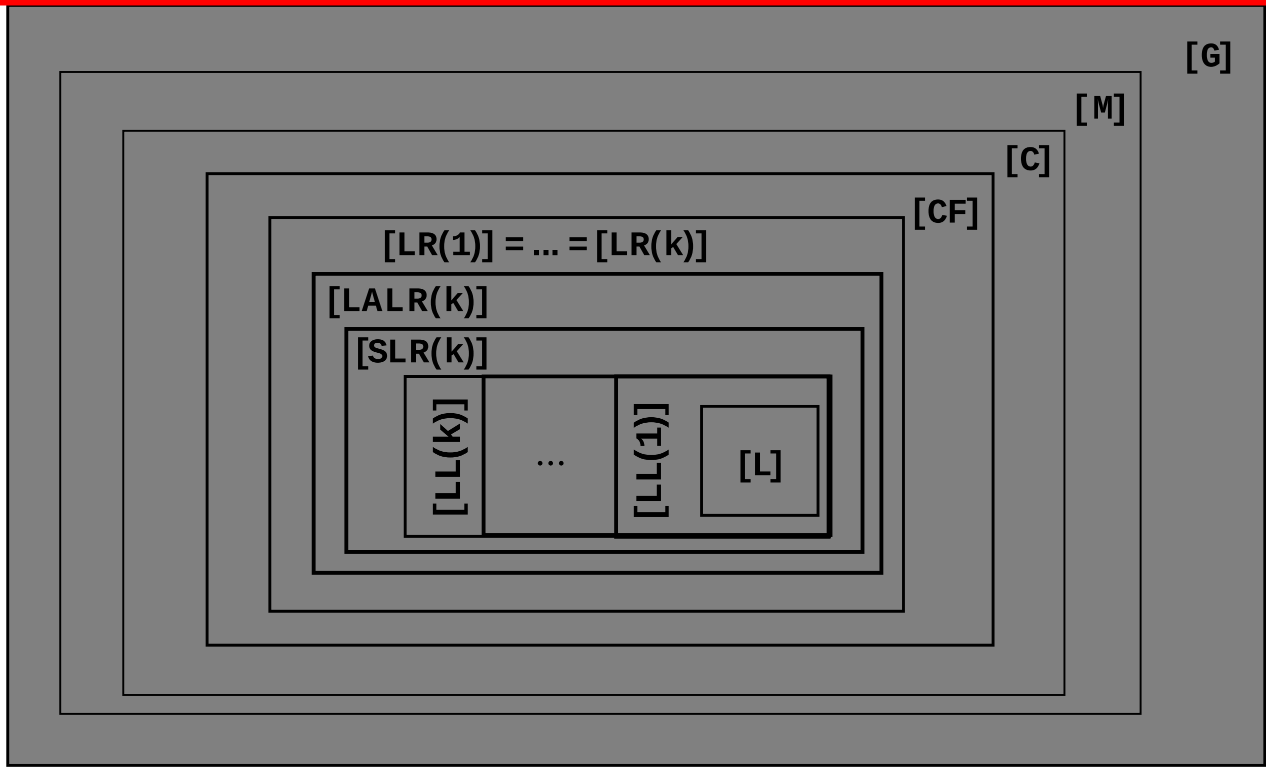


G = generali [linguaggi r.e. ma non ricorsivi]
A = ambigue
M = monotone [linguaggi ricorsivi]
C = contestuali [$u^n v^n z^n$]
CF = context free [$u^n v^n$]
LR(k) = bottom-up/k simboli
LALR(k) = bottom-up/k simboli con lookahead
SLR(k) = bottom-up/k simboli semplici
LL(k) = top-down/k simboli
OP = precedenza di operatore
L = lineari [linguaggi regolari: $u^n v^m$]
DFA = Automi Deterministici *NFA* = Automi NonDeterministici

grammatica
[linguaggio definito]

Classificazione dei Linguaggi

Inclusione tra Linguaggi



[G] = linguaggi r.e. ma non ricorsivi
[M] = linguaggi ricorsivi
[C] = linguaggi contestuali: $\{u^n v^n z^n \mid n \geq 0\}$
[CF] = linguaggi context free: $\{u^n v^m z^k \mid n,m,k \geq 0 \text{ and } (n=m \text{ or } m=k)\}$
[LR(k)] = bottom-up/k simboli: $\{u^m v^n \mid m>n \geq 0\}$
[LALR(k)] = bottom-up/k simboli con lookahead
[SLR(k)] = bottom-up/k simboli semplici
[LL(k)] = top-down/k simboli: $\{u^n v^n \mid n \geq 0\}$
[L] = lineari [linguaggi regolari: $\{u^n v^m \mid n \geq 0, m \geq 0\}$

DERIVAZIONE

$G = \langle V, \Sigma, s \in V, P \rangle$

relazione $=>_G$ su $(\Sigma \cup V)^* \times (\Sigma \cup V)^*$

$\alpha A \beta => \alpha \gamma \beta$ *sse* $A ::= \gamma \in P$

per $\alpha, \beta, \gamma \in (\Sigma \cup V)^*$

Omettiamo G da $=>_G$ intendendosi sempre chiara la grammatica G cui ci riferiamo

$=>*$: chiusura riflessiva e transitiva di $=>$

- $\alpha =>* \alpha$
- se $\alpha_1 => \alpha_2 => ... => \alpha_n$
  allora $\alpha_1 =>* \alpha_n$

$=>+$: chiusura transitiva di $=>$

se $\alpha_1 => \alpha_2 => ... => \alpha_n$
allora $\alpha_1 =>+ \alpha_n$

LINGUAGGIO GENERATO

$G = <V, \Sigma, s \in V, P>$

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

(dove $\Rightarrow^+$ e' ovviamente definita a partire $\Rightarrow_G$)

Ad esempio: le produzione di G siano

p1: E::= E+E
p2: E::= E*E
p3: E::= id

chi e' in V e chi in Σ ?
chi e' s ?

p1: E::= E+E
p2: E::= E*E
p3: E::= id

Esempio (continua)

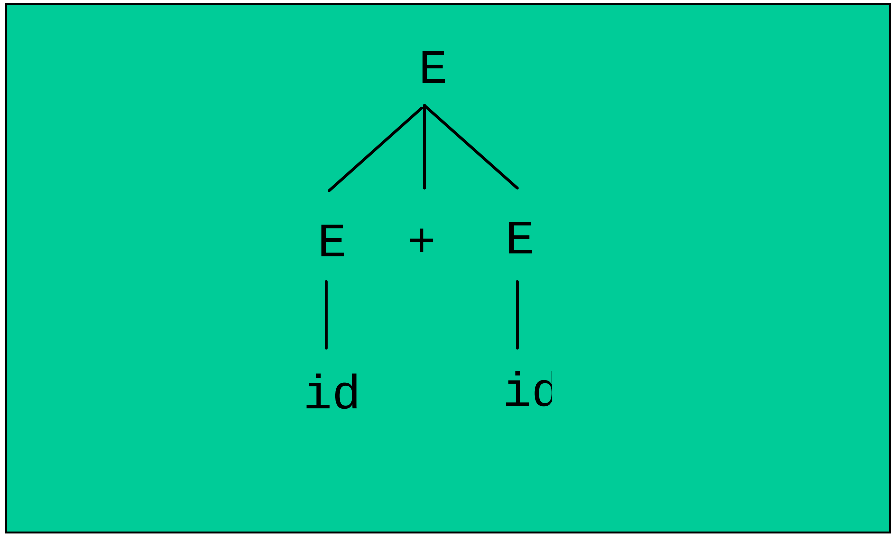
allora: id+id $\in$ L(E)
infatti:

$$E \Rightarrow_{(p1)} E+E \Rightarrow_{(p3)} id+E \Rightarrow_{(p3)} id+id$$

$$E \Rightarrow_{(p1)} E+E \Rightarrow_{(p3)} E+id \Rightarrow_{(p3)} id+id$$

PARSE TREE e AMBIGUITA'

parse tree = mostrano la derivazione senza
mostrare l'ordine delle derivazioni



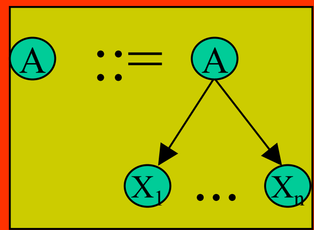
DERIVAZIONE (sul dominio T degli Alberi)

$G = \langle V, \Sigma, s \in V, P \rangle$

$(\Sigma \cup V)_T{}^* =$ minimo insieme tale che: $\forall a \in (\Sigma \cup V)$

- foglia: $\langle [a,-] \rangle \in (\Sigma \cup V)_T{}^*$
- $\forall \langle t1,\ldots,tn \rangle, \langle u1,\ldots,un \rangle \in (\Sigma \cup V)_T{}^*$
  - albero: $\langle [a, \langle t1,\ldots,tn \rangle] \rangle \in (\Sigma \cup V)_T{}^*$
  - foresta: $\langle t1,\ldots,tn, u1,\ldots,un \rangle \in (\Sigma \cup V)_T{}^*$

$A ::= x_1 \ldots x_n \in P$ *sse*



$\in P_T$

per $X_i \in (\Sigma \cup V)$

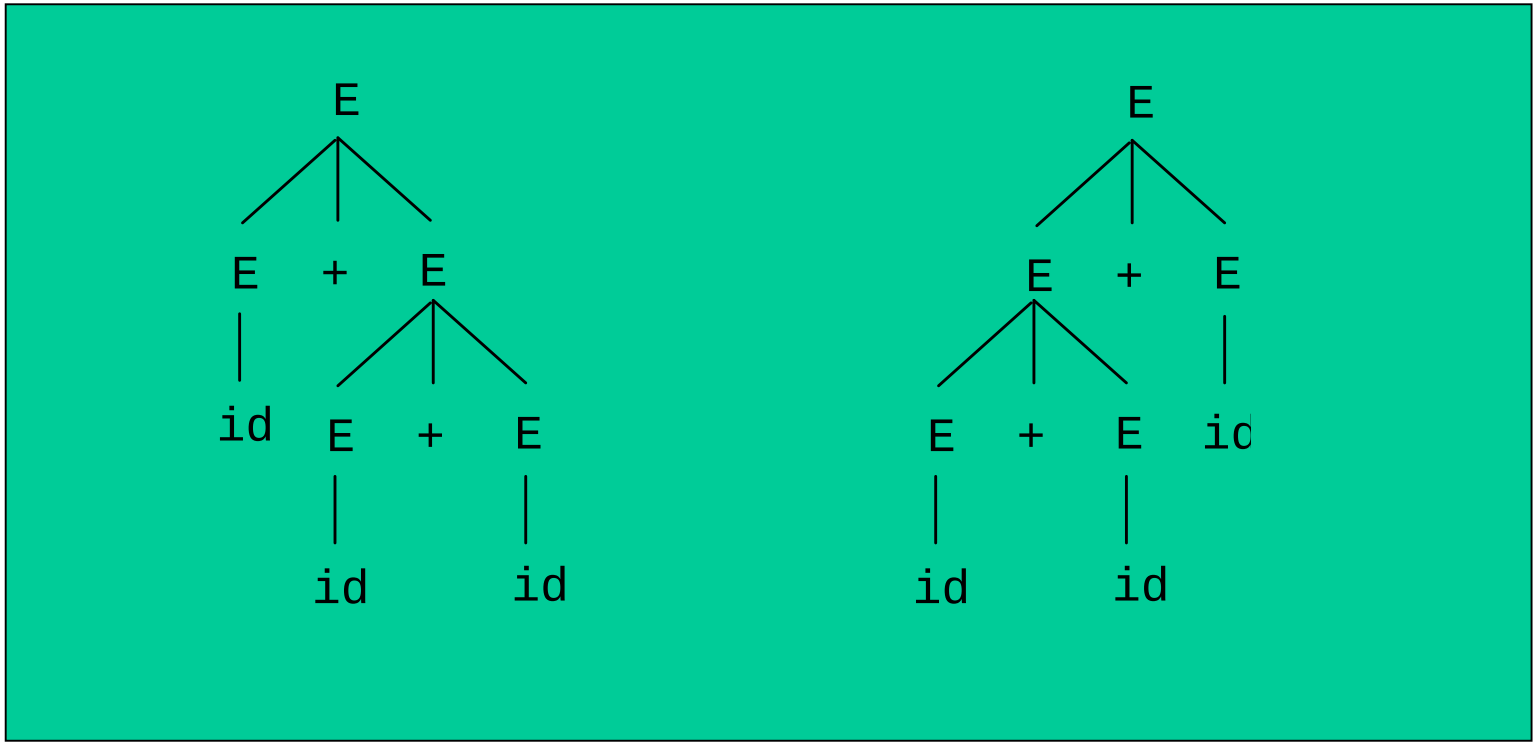
relazione $\Rightarrow$ su $(\Sigma \cup V)_T{}^* \times (\Sigma \cup V)_T{}^*$

$\alpha A \beta \Rightarrow \alpha \gamma \beta$ *sse* $A ::= \gamma \in P_T$

per $\alpha, \beta, \gamma \in (\Sigma \cup V)_T{}^*$ e $\langle t1,\ldots,tn \rangle \langle u1,\ldots,un \rangle = \langle t1,\ldots,tn,u1,\ldots,un \rangle$

p1: E::= E+E
p2: E::= E*E
p3: E::= id

ambiguita' = piu parse tree con stessa frontiera



Una ricetta per il linguaggio generato

$S ::= e_s$
$A_1 ::= e_1$
…
$A_n ::= e_n$

1) semiordinamento $\geq^G$ su A_i

$\forall i \geq 0, \quad e_i \equiv f(A_{i_1},\ldots,A_{i_{n_i}}) \quad \text{con } A_{i_1},\ldots,A_{i_{n_i}} \geq^G A_i$

Rimozione mutua ricorsione

$A_i ::= f(A_{i_1},\ldots,A_j,\ldots,A_{i_{n_i}}) \quad \text{con } A_{i_1},\ldots,A_{i_{n_i}} \geq^G A_i$
$A_j ::= g(A_{j_1},\ldots,A_i,\ldots,A_{j_{n_j}}) \quad \text{con } A_{j_1},\ldots,A_{j_{n_j}} \geq^G A_i \geq^G A_j$

$\Downarrow$

$A_i ::= f(A_{i_1},\ldots, g(A_{j_1},\ldots,A_i,\ldots,A_{j_{n_j}}),\ldots,A_{i_{n_i}})$
$A_j ::= g(A_{j_1},\ldots,A_i,\ldots,A_{j_{n_j}})$

Esempio

S::=u A B v | B u
A::= v u v | B
B::= u S v | u v u

$B \geq_G A \geq_G S$

S::=u A B v | B u
B::= u S v | u v u

S::=u A B v | B u
B::= u (u A B v | B u) v | u v u

S::=u A B v | B u
A::= v u v | B
B::= uu A B vv | u B uv | u v u

A::= v u v | B
B::= uu A B vv | u B uv | u v u

A::= v u v | B
B::= uu (v u v | B) B vv | u B uv | u v u

S::=u A B v | B u
A::= v u v | B
B::= uu vuv B vv | uu B B vv | u B uv | u v u

2) associazione di sottolinguaggi $L(A_i)$

$\forall i \geq 0, \quad L(A_i)=L(e_i)$

L(e) e' un'espressione su $\mathbf{2}^{\Sigma^*}$ contenente:

X (prodotti finiti)

$\cup$ (unione numerabile)

L(e) e' continua su $\mathbf{2}^{\Sigma^*}$

Quando $L(A_i)=L(e_i)$ e' ricorsiva: $L(e_i) \equiv E(L(A_i))$

Risolviamo equazioni ricorsive $X=E(X)$ nella variabile $X=L(A_i)$ su $\mathbf{2}^{\Sigma^*}$

$$X=\cup_{i\in \aleph} E(\perp)^{\mathbf{i}}$$

$$E(\perp)^{\mathbf{0}} = E(X\leftarrow \perp)$$

$$E(\perp)^{\mathbf{k+1}} = E(X\leftarrow E(\perp)^{\mathbf{k}})$$

Esempio

$S ::= u\ S \mid \varepsilon$ $\longrightarrow$ $X = \underbrace{\{u\} \times X \cup \{\lambda\}}_{E(X)}$

$E(X \leftarrow \bot)^{0} = \{u\} \times \bot \cup \{\lambda\} = \bot \cup \{\lambda\} = \{\lambda\}$

$E(X \leftarrow \bot)^{1} = \{u\} \times \{\lambda\} \cup \{\lambda\} = \{u, \lambda\}$

$E(X \leftarrow \bot)^{2} = \{u\} \times \{u, \lambda\} \cup \{\lambda\} = \{uu, u, \lambda\}$

$E(X \leftarrow \bot)^{3} = \{u\} \times \{uu, u, \lambda\} \cup \{\lambda\} = \{u^3, u^2, u, \lambda\}$

$\Downarrow$

$E(X \leftarrow \bot)^{n} = \{u^n, u^{n-1}, \ldots, u, \lambda\}$

$\Downarrow$

$L(S) = \{u^n \mid n \in \aleph\}$ $= u^*$

Esempio

$S ::= u\ S\ v \mid z$ $\longrightarrow$ $X = \underbrace{\{u\} \times X \times \{v\} \cup \{z\}}_{E(X)}$

$$E(X \leftarrow \perp)^{0} = \{u\} \times \perp \times \{v\} \cup \{z\} = \perp \cup \{z\} = \{z\}$$

$$E(X \leftarrow \perp)^{1} = \{u\} \times \{z\} \times \{v\} \cup \{z\} = \{uzv, z\}$$

$$E(X \leftarrow \perp)^{2} = \{u\} \times \{uzv, z\} \times \{v\} \cup \{z\} = \{u^2zv^2, uzv, z\}$$

$$\Downarrow$$

$$E(X \leftarrow \perp)^{n} = \{u^n z v^n, u^{n-1} z v^{n-1}, \ldots, z\}$$

$$\Downarrow$$

$$L(S) = \{u^n z v^n \mid n \geq 0\}$$

Esempio

$A ::= A+A$
$A ::= A*A$
$A ::= id$

$\Rightarrow X = \underbrace{\{x+y \mid x,y \in X\} \cup \{x*y \mid x,y \in X\} \cup \{id\}}_{E(X)}$

$$E(X \leftarrow \bot)^0 = \{x+y \mid x,y \in \bot\} \cup \{x*y \mid x,y \in \bot\} \cup \{id\}$$
$$= \bot \cup \bot \cup \{id\} = \{id\}$$

$$E(X \leftarrow \bot)^1 = \{x+y \mid x,y \in \{id\}\} \cup \{x*y \mid x,y \in \{id\}\} \cup \{id\}$$
$$= \{id+id\} \cup \{id*id\} \cup \{id\} = \{id,\ id+id,\ id*id\}$$

$$E(X \leftarrow \bot)^2 = \{x+y \mid x,y \in E(X \leftarrow \bot)^1\} \cup \{x*y \mid x,y \in E(X \leftarrow \bot)^1\} \cup \{id\}$$
$$= \{id,\ id+id,\ id*id, id+id+id,\ id+id*id,\ id*id+id,\ id*id*id, \ldots,\ id*id*id*id\}$$

$\Downarrow$

$$E(X \leftarrow \bot)^n = \{id,\ id\ t^k \mid t \in \{+id,\ *id\},\ k \in [1..2^n-1]\}$$

$\Downarrow$

$$L(A) = \{id\ t^n \mid t \in \{+id,\ *id\},\ n \in \aleph\}$$

TOP DOWN e BOTTOM UP PARSING

data una grammatica G ed una stringa w
di terminali dell'alfabeto dobbiamo dire se:

$$w \in L(G)$$

ovvero mostrare una derivazione di w

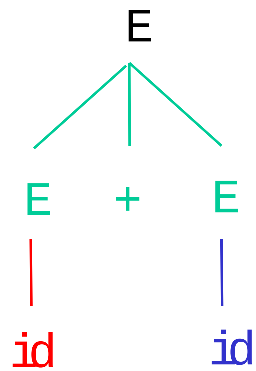
decidibile (grammatiche monotone/derivazioni ordinabili)

ma puo' essere inefficiente ($O(n^P)$)

P = scelte (produzioni) per simbolo

Top-down (semplice da costruire a mano, poche grammatiche)

$E \Rightarrow_l E+E \Rightarrow_l id+E \Rightarrow_l id+id$



primo non terminale da sinistra

prima produzione per esso

fallimento: si torna indietro per alternativa

Passo 1
Passo 2
Passo 3

Top-down = leftmost

$G = \langle V, \Sigma, s \in V, P \rangle$

Forma Sentenziale Sinistra (per G):LSF_G

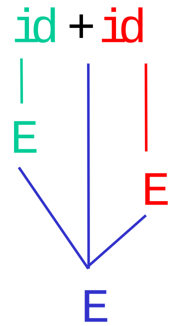
$\alpha\beta\gamma \in LSF_G$ **se e solo se** $s \underset{l}{\Rightarrow}^{+} \alpha\beta\gamma$

Solo Sentenziali Sinistre

$$L(G) = \{w \in \Sigma^* \mid s \Rightarrow^{+} w\}$$
$$= \{w \in \Sigma^* \mid s \underset{l}{\Rightarrow}^{+} w\}$$

Bottom-up (tecniche piu’ complesse, molte grammatiche ed anche piu’ linguaggi)

$E \Rightarrow_{r} E+E \Rightarrow_{r} E+id \Rightarrow_{r} id+id$



maniglia

riduzione

fallimento: si torna indietro per vera maniglia

Passo 1
Passo 2
Passo 3

Bottom-up = rightmost reversed

$G = \langle V, \Sigma, s \in V, P \rangle$

Forma Sentenziale Destra (per G): RSF_G

$\alpha\beta\gamma \in RSF_G$ **se e solo se** $s \;_r\!\Rightarrow^+ \alpha\beta\gamma$

Solo Sentenziali Destre

$$L(G) = \{w \in \Sigma^* \mid s \Rightarrow^+ w\}$$
$$= \{w \in \Sigma^* \mid s \;_r\!\Rightarrow^+ w\}$$

$B ::= \beta \in P$ e' **maniglia** di $\alpha\beta\gamma \in RSF_G$ **se e solo se** $\alpha B\gamma \in RSF_G$

Top-down