Grammatiche e linguaggi liberi

es. $L_{pal} = \{ w \in \Sigma^* \mid w = w^R \}$

Supponiamo $\Sigma = \{0, 1\}$ e che L_{pal} sia regolare.

Abbia il suo DFA n stati.

Allora $0^n 1 0^n$ è accettata. 0^n però passa per un loop, che omesso

non lascia la stringa palindroma.

L_{pal} può essere definito induttivamente:

base: $\varepsilon, 0, 1 \in L_{pal}$

passo induttivo: $a \in L_{pal} \Rightarrow 0a0, 1a1 \in L_{pal}$

L_{pal} è una CFG (grammatica libera), dove 0 e 1 sono detti

TERMINALI, a è VARIABILE.

Def. Una grammatica libera di contesto (CFG, "context-free grammar")

è una quadrupla (V, T, P, S), dove

- V sono le variabili (o nonterminali, o categorie sintattiche)
- T sono i terminali
- P sono le regole di produzione

- S è il simbolo iniziale

P è un insieme di regole della forma $A \to \alpha$ con $A \in V$, $\alpha \in (V \cup T)^*$

es. $G \mid L(G) = \{a^n b^m c^m d^{2n} \mid n \geq 0,\ m \geq 0\}$

$$A \begin{cases} P \to I \mid aPd^2 \\ I \to \varepsilon \mid bIc \end{cases} \qquad G = \{\{P, I\}, \{a, b, c, d\}, A, P\}$$

Def. (derivazione) $G = (V, T, P, S)$ CFG, $A \in V$, $\{\alpha, \beta\} \in (V \cup T)^*$ $\wedge$ $A \to \gamma \in P \iff (\alpha A \beta \underset{G}{\Rightarrow} \alpha \gamma \beta)$.

Talvolta la G e' omessa se ovvia. Diremo che da $\alpha A \beta$ si deriva $\alpha \gamma \beta$.

oss. $\alpha \underset{lm}{\Rightarrow} \beta$ se si sostituisce sempre a sinistra (left most), $\alpha \underset{rm}{\Rightarrow} \beta$ (right most).

Def. $L(G)$, con G CFG, si chiama linguaggio libero dal contesto (CFL) generato da G.

es. $\underset{(A)}{L(G_{pal})} = \underset{(B)}{\{ w \in \Sigma^* \mid w = w^R \}}$, $\Sigma = \{0,1\}$

induz. sulla lunghezza

(A)⊇(B)

- **base:** $\varepsilon \in L(G_{pal})$ perché $P \to \varepsilon$
- **passo induttivo:** $w \in L(G_{pal}) \Rightarrow 0w0, 1w1 \in L(G_{pal})$ perché $(P \overset{*}{\Rightarrow} w \Rightarrow 0w0 \;\wedge\; P \overset{*}{\Rightarrow} w \Rightarrow 1w1) \Longrightarrow (P \overset{*}{\Rightarrow} 0w0 \;\wedge\; P \overset{*}{\Rightarrow} 1w1)$.

induz. sulla derivazione

(A)⊆(B)

- **base:** $P \to \varepsilon | 0 | 1$, quindi solo palindromi
- **passo induttivo:** $x \in L(G_{pal}) \Rightarrow x = x^R$. $(x \Rightarrow 0x0 | 1x1) \Rightarrow$ $\Rightarrow (0x0)^R = 0x^R0 = 0x0 \;\wedge\; (1x1)^R = 1x1 \Rightarrow$ $\Rightarrow 0x0, 1x1 \in B$.

$\boxtimes$

Def. $S \overset{*}{\Rightarrow} \alpha \iff \alpha$ è una FORMA SENTENZIALE

$S \underset{lm}{\Rightarrow} \alpha$ — sinistra

" — destra

oss. $L(G) = \{ w \in T^* \mid w \text{ forma sentenziale} \}$

Alberi sintattici

(i) ogni nodo interno è una variabile

(ii) ogni foglia è un terminale

(iii) la frontiera, o prodotto, è la stringa ottenuta da sinistra a destra delle foglie.

Un albero può tuttavia essere ambiguo (i.e. $\exists\, w \in L(G) \mid w$ ha più di un albero sintattico di cui è il prodotto).

es.

$$P \begin{cases} E \to T \mid E+T \\ T \to F \mid T * F \\ F \to I \mid (E) \\ I \to a \mid aI \end{cases}$$

← queste produzioni eliminano l'ambiguità dell'albero sintattico

Def. Un linguaggio si dice inerentemente ambiguo se ogni sua grammatica è ambigua in almeno un albero sintattico.

es. $L = \underbrace{\{a^n b^n c^m d^m,\ n,m \geq 1\}}_{(A)} \cup \underbrace{\{a^n b^m c^m d^n,\ n,m \geq 1\}}_{(B)}$

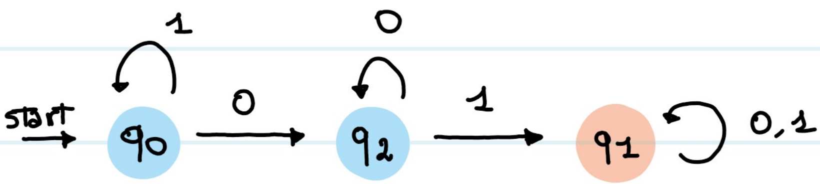
(A) $P_1 \to A_1B$, $A_1 \to ab \mid aA_1b$, $B \to cd \mid cBd$

(B) $P_2 \to aA_2d \mid aP_2d$, $A_2 \to bc \mid bA_2c$

$P \to P_1 \mid P_2$

OSS. Da un automa posso sempre estrarre una grammatica detta **REGOLARE**.

es.



$$\Big(\overbrace{\{q_0, q_1, q_2\}}^{Q}, \overbrace{\{0,1\}}^{\Sigma}, P, \underbrace{\{q_0\}}_{start}\Big)$$

$$P \begin{cases} q_0 \to 1q_0 \mid 0q_2 \qquad q_1 \to 0q_1 \mid 1q_1 \mid \varepsilon \\ q_2 \to 0q_2 \mid 1q_1 \end{cases}$$

Forma normale di Chomsky

Sono nella forma normale di Chomsky le grammatiche che ammettono solo produzioni nella forma $A \to \alpha \vee A \to BC$.

Teorema Sia m il cammino più lungo sull'albero sintattico per una forma normale di Chomsky sulla stringa w, allora $|w| \leq 2^{m-1}$.

base $m = 1$. Allora sicuramente $|w| \leq 1$.

passo induttivo Allora w discende secondo una produzione $A \to BC$. Per il passo induttivo, i prodotti di B e C hanno massimo 2^{m-2} terminali. Quindi w ha massimo $2^{m-2} + 2^{m-2} = 2^{m-1}$ terminali. $\blacksquare$

Pumping lemma

Teorema Sia G una CFG con n produzioni: allora $\exists\, w$ t.c.:

(i) $w = abcde$, $bd \neq \varepsilon$

(ii) $|bcd| \leq m$

(iii) ab^icd^ie è una forma sentenziale di G $\forall i \in \mathbb{N}$.

Proprietà dei CFL

Siano L e U due CFL, allora:

(i) $L \cup U$ è un CFL.

(ii) L^* è un CFL.

(iii) L^R è un CFL.

(iv) L^+ è un CFL.