Definire finitamente un linguaggio

Esempio (1): palindrome = "parola che letta da sx a dx è uguale a sé stessa letta da dx a sx"

es: abba amoroma

(Madam I'm Adam)

in senso esteso

(I topi non avevano nipoti)

$A = \{a, b\}$ $L = \{\varepsilon, a, b, aa, bb, aba, bab, aaa, bbb, abba, baab, aaaa, bbbb, \dots\}$

Una palindrome può essere

- o la stringa ε
- oppure a
- oppure b
- oppure a "palindrome" a
- oppure b "palindrome" b

Backus-Naur Form (BNF)

$$\langle P \rangle ::= \varepsilon \mid a \mid b \mid a \langle P \rangle a \mid b \langle P \rangle b$$

può essere oppure

Come grammatica:

$$P \to \varepsilon \mid a \mid b \mid aPa \mid bPb$$

Definizione ricorsiva in cui

- P è detto "simbolo nonterminale"
- a, b sono i "simboli terminali"

Alternativamente: (insieme di assiomi e regole d'inferenza)

$$\frac{}{\varepsilon \in L(P)} \qquad \frac{}{a \in L(P)} \qquad \frac{}{b \in L(P)}$$

$$\frac{w \in L(P)}{awa \in L(P)} \qquad \frac{w \in L(P)}{bwb \in L(P)}$$

$L(P)$ = "linguaggio generato a partire dal nonterminale P"
è il più piccolo insieme generato da questi assiomi e regole d'inferenza

Esempio (2): Espressioni aritmetiche formate a partire dalle variabili "a" e "b" con gli operatori $*$, $+$ e le parentesi (,)

Una expr può essere
- la variabile a,
- oppure la variabile b,
- oppure expr $*$ expr,
- oppure expr $+$ expr,
- oppure (expr)

BNF:

$$\langle E \rangle ::= a \mid b \mid \langle E \rangle * \langle E \rangle \mid \langle E \rangle + \langle E \rangle \mid (\langle E \rangle)$$

Grammatica

$E \to a$

$E \to b$

$E \to E * E$

$E \to E + E$

$E \to (E)$

più brevemente

$$E \to a \mid b \mid E * E \mid E + E \mid (E)$$

Come derivare una stringa?

$$P \to \varepsilon \mid a \mid b \mid aPa \mid bPb$$

Si parte da P e si applicano le produzioni

Ad es: abba è una palindrome perché

$$\underbrace{P \Rightarrow}_{P \to aPa} a\underline{P}a \underbrace{\Rightarrow}_{P \to bPb} a\underline{bPb}a \underbrace{\Rightarrow}_{P \to \varepsilon} abba \quad \text{cioè } P \overset{*}{\Rightarrow} abba$$

$$\underbrace{P \Rightarrow}_{P \to bPb} bPb \underbrace{\Rightarrow}_{P \to bPb} bbPbb \underbrace{\Rightarrow}_{P \to a} bbabb \quad \text{cioè } P \overset{*}{\Rightarrow} bbabb$$

$$\cfrac{\cfrac{\varepsilon \in L(P)}{bb = b\varepsilon b \in L(P)}}{abba = ab\varepsilon ba \in L(P)}$$

$$\cfrac{\cfrac{a \in L(P)}{bab \in L(P)}}{bbabb \in L(P)}$$

Altro esempio

$$E \to a \mid b \mid E * E \mid E + E \mid (E)$$

Come dimostrare che $a + (a * b) \in L(E)$?

$$E \Rightarrow \underline{E} + E \Rightarrow a + \underline{E} \Rightarrow a + (\underline{E}) \Rightarrow a + (\underline{E} * E)$$

$$\Rightarrow a + (a * \underline{E}) \Rightarrow a + (a * b)$$

$$\dfrac{a \in L(E) \qquad \dfrac{\dfrac{a \in L(E) \qquad b \in L(E)}{a * b \in L(E)}}{(a * b) \in L(E)}}{a + (a * b) \in L(E)}$$

GRAMMATICHE

- Tante classi di grammatiche
  - regolari, libere (da contesto), dipendenti dal contesto, monotone, generali (o "a struttura di frase)
  - Tutte seguono lo stesso pattern, differenziandosi solo per come sono caratterizzate le <u>produzioni</u> (o regole)

  (Le vedremo in seguito; classificazione di Chomsky)
- Quelle più utili (rapporto tra espressività e facilità d'analisi) sono le "libere"

<u>Def</u> Una <u>grammatica libera</u> da contesto è una quadrupla (NT, T, R, S) dove
- NT è un insieme finito di simboli non-terminali (di solito lettere maiuscole $A, B, ..$)
- T è un insieme finito di simboli terminali (di solito lettere minuscole $a, b, c, ..$)
- $S \in NT$ è detto simbolo iniziale
- R è un insieme finito di produzioni (o regole) della forma
  $V \rightarrow w$ dove $V \in NT$ e $w \in (T \cup NT)^*$

Esempi

(1) $G = (\{S\}, \{a, b, +, *\}, S, R)$

con $R = \{ S \xrightarrow{1} a,$
$S \xrightarrow{2} b,$
$S \xrightarrow{3} S+S,$
$S \xrightarrow{4} S*S \}$

$$\boxed{S \to a \mid b \mid S+S \mid S*S}$$

N.B. R è un insieme: numeriamo le produzioni per dare loro un "nome", non per ordinarle.

(2) $G_1 = (\{S, A\}, \{a, b\}, S, R_1)$

con $R_1 = \left\{ \begin{array}{l} S \xrightarrow{1} aAb \\ A \xrightarrow{2} aAb \\ A \xrightarrow{3} \varepsilon \end{array} \right\}$

$$\boxed{\begin{array}{l} S \to aAb \\ A \to aAb \mid \varepsilon \end{array}}$$

$G_2 = (\{S, A, B\}, \{a, b\}, S, R_2)$

con $R_2 = \left\{ \begin{array}{l} S \xrightarrow{1} AB \\ A \xrightarrow{2} aA \\ A \xrightarrow{3} a \\ B \xrightarrow{4} bB \\ B \xrightarrow{5} b \end{array} \right\}$

$$\boxed{\begin{array}{l} S \to AB \\ A \to aA \mid a \\ B \to bB \mid b \end{array}}$$

$G_3 = (\{S\}, \{a, b\}, S, R_3)$

con $R_3 = \{ S \xrightarrow{1} aSb,$
$S \xrightarrow{2} ab \}$

$$\boxed{S \to aSb \mid ab}$$

N.B. Di solito, usiamo solo la forma per rappresentare una grammatica.

Esempio (3)

$$E \to I \mid E+E \mid E*E \mid -E \mid (E)$$
$$I \to a \mid b \mid Ia \mid Ib$$

$L(I) = \{a, b\}^+$ qualunque sequenza nonvuota di "a" e "b"

↳ identificatori (insieme infinito)

$L(E)$ = espressioni aritmetiche costruite sugli identificatori

(cioè I genera "il lessico" (identificatori) che E può usare per costruire le expr. aritmetiche)

DERIVAZIONI

Data $G = (NT, T, R, S)$ libera da contesto, diciamo che da v si deriva immediatamente w (o anche "v si riscrive in un passo in w"), e lo denotiamo con $v \Rightarrow w$, se

$$\frac{v = xAy \qquad (A \rightarrow z) \in R \qquad w = xzy}{v \Rightarrow w}$$

$x, y, z \in (T \cup NT)^*$

Diciamo che da v si deriva w (o anche "v si riscrive in w"), e lo denotiamo con $v \Rightarrow^* w$, se esiste una sequenza finita (eventualmente vuota) di derivazioni immediate

$$v \Rightarrow w_0 \Rightarrow w_1 \Rightarrow \dots \Rightarrow w$$

Cioè

$$\frac{}{v \Rightarrow^* v} \qquad\qquad \frac{v \Rightarrow^* w \qquad w \Rightarrow z}{v \Rightarrow^* z}$$

$\Rightarrow^*$ è la chiusura riflessiva e transitiva della relazione $\Rightarrow$

Esempi

(1) G $\quad S \to a \mid b \mid S+S \mid S*S$

$$S+\underline{S} \Rightarrow S+\underline{S}*S \Rightarrow \underline{S}+a*S \Rightarrow S*S+a*S$$

cioè $S+S \Rightarrow^* S*S+a*S$

(2) G_1 $\quad S \to aAb$
$\quad A \to aAb \mid \varepsilon$

$$\underline{A}A \Rightarrow a\underline{A}bA \Rightarrow ab\underline{A} \Rightarrow aba\underline{A}b \Rightarrow abab$$

cioè $AA \Rightarrow^* abab$

(3) G_2 $\quad S \to AB$
$\quad A \to aA \mid a$
$\quad B \to bB \mid b$

$$S \Rightarrow A\underline{B} \Rightarrow Ab\underline{B} \Rightarrow \underline{A}bb \Rightarrow abb$$

cioè $S \Rightarrow^* abb$

ma anche

$$S \Rightarrow \underline{A}B \Rightarrow a\underline{B} \Rightarrow ab\underline{B} \Rightarrow abb$$

cioè ci possono essere <u>più derivazioni diverse</u> che generano la stessa stringa

LINGUAGGIO GENERATO

<u>Def</u>. Il linguaggio generato da una grammatica $G = (NT, T, R, S)$ è l'insieme

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

solo terminali

si parte dal simbolo iniziale S

Data G, come faccio a determinare L(G)?
E a verificare se $w \in L(G)$?

- studieremo tecniche opportune che in alcuni casi saranno anche efficienti
- algoritmo "naif": partire da S e provare ad applicare in tutti i modi possibili le produzioni per trovare una derivazione che generi w (nondeterminismo ⇒ backtracking)

In alcuni casi, tale verifica è semplice

Esempi

$G_3 \quad S \to aSb \mid ab$

$$S \Rightarrow ab$$
$$S \Rightarrow aSb \Rightarrow aabb$$
$$aSb \Rightarrow aaSbb \Rightarrow aaabbb$$
$$aaSbb \Rightarrow aaaSbbb \ldots$$

È chiaro che

$$L(G_3) = \{a^n b^n \mid n \geq 1\}$$

$G_1 \quad S \to aAb$

$ A \to aAb \mid \varepsilon$

$$S \Rightarrow aAb \Rightarrow ab$$
$$aAb \Rightarrow aaAbb \Rightarrow aabb$$
$$aaAbb \Rightarrow aaaAbbb \Rightarrow aaabbb$$
$$aaaAbbb \Rightarrow a^4Ab^4 \ldots$$

È chiaro che

$$L(G_1) = \{a^n b^n \mid n \geq 1\}$$

N.B. • G_1 e G_3 sono due grammatiche equivalenti perché $L(G_1) = L(G_3)$

• In generale esistono infinite grammatiche diverse che generano lo stesso ling.

Esempi

G_2

$$S \to AB$$
$$A \to aA \mid a$$
$$B \to bB \mid b$$

$$L(G_2) = L(S) = L(A) \cdot L(B)$$

$$A \Rightarrow a$$
$$A \Rightarrow aA \Rightarrow aa$$
$$aA \Rightarrow aaA \Rightarrow aaa$$
$$aaA \Rightarrow aaaA \ldots$$

$$L(A) = \{ a^n \mid n \geq 1 \}$$

idem per $L(B) = \{ b^n \mid n \geq 1 \}$

$$\Rightarrow L(A) \cdot L(B) = \{ a^n b^m \mid n, m \geq 1 \}$$

$$\left.\begin{array}{l} S \to AB \\ A \to aA \mid \varepsilon \\ B \to Bb \mid \varepsilon \end{array}\right] G$$

$$L(G) = \{ a^n b^m \mid n, m \geq 0 \}$$

N.B. $\varepsilon \in L(G)$ perché $S \Rightarrow \underline{A}B \Rightarrow \underline{B} \Rightarrow \varepsilon$

Esempi

$S \to AB$

$A \to aAb \mid \varepsilon$

$B \to bB \mid \varepsilon$

$L(S) = L(A) \cdot L(B)$

dove $L(A) = \{ a^n b^n \mid n \geq 0 \}$

$L(B) = \{ b^n \mid n \geq 0 \}$

quindi $L(S) = \{ a^n b^{n+m} \mid n, m \geq 0 \}$

$= \{ a^n b^m \mid m \geq n \geq 0 \}$

All'incontrario, dato un linguaggio, come determinare una grammatica che lo genera?

$L = \{ a^n b^m \mid n \geq m \geq 0 \}$

$S \to AB$

$A \to \varepsilon \mid aA$

$B \to aBb \mid \varepsilon$

$L = \{ a^{2n+1} \mid n \geq 0 \} = \{ a, aaa, aaaaa, \ldots \}$

$S \to a \mid aaS$

$L = \{ a^{2n} b^{2m+1} \mid n, m \geq 0 \}$

$S \to aaS \mid bB$

$B \to \varepsilon \mid bbB$

$L = \{ ab, ac, ad \}$

$S \to ab \mid ac \mid ad$

oppure

$S \to aA$

$A \to b \mid c \mid d$

$L = \{ a^{2n} b^{m} c^{2m} \mid n \geq 0, m \geq 1 \}$

$S \to aaS \mid bBcc$

$B \to \varepsilon \mid bBcc$

$L = \{ a^{n} b^{k} c^{n} \mid n \geq 0, k \geq 1 \}$

$S \to aSc \mid B$

$B \to b \mid bB$

$L = \{ a^{n} b^{n} c^{m} d^{m} \mid n, m \geq 0 \}$

$S \to AB$

$A \to aAb \mid \varepsilon$

$B \to cBd \mid \varepsilon$

$L = \{ a^n b^m c^m d^n \mid n, m \geq 0 \}$

$S \to aSd \mid bAc \mid \varepsilon$
$A \to bAc \mid \varepsilon$

$L = \{ w \in \{a,b\}^* \mid \exists x \in \{a,b\}^*. \; w = xaa \}$

$S \to aS \mid bS \mid aa$

$L = \{ w \in \{(,)\}^* \mid$ in w tutte le parentesi sono bilanciate $\}$

$S \to \varepsilon \mid (S) \mid SS$

$L = \{ w \in \{a,b\}^* \mid$ in w occorrono tante "a" quante "b" $\}$

$S \to \varepsilon \mid aSb \mid bSa \mid SS$

$L = \{ ww^R \mid w \in \{a,b\}^* \}$

palindromi di lunghezza pari

w^R è definita:
$\varepsilon^R = \varepsilon$
$(aw)^R = w^R a$

$S \to \varepsilon \mid aSa \mid bSb$

$L = \{ a^{2n} b^n \mid n \geq 0 \}$

$S \to aaSb \mid \varepsilon$

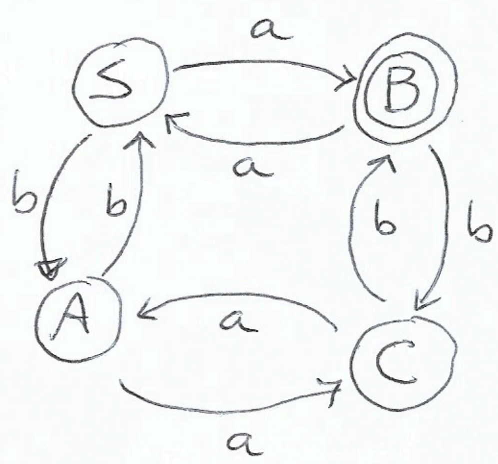
$L = \{ w \in \{a,b\}^* \mid$ in w le "a" siano dispari e le "b" pari $\}$

$S \to aB \mid bA$

$B \to aS \mid bC \mid \varepsilon$

$A \to aC \mid bS$

$C \to aA \mid bB$



S = a pari / b pari

B = a dispari / b pari

C = a dispari / b dispari

A = a pari / b dispari

$L = \{ a^n b^m c^p \mid n = m + p,\ n, m, p \geq 0 \}$

$S \to aSc \mid aBb \mid \varepsilon$

$B \to aBb \mid \varepsilon$

Ultimo esempio (Hard!)

$$L = \{ a^n b^m a^p b^q \mid n, m, p, q \geq 0, \; n+m = p+q \}$$

$$S \to \varepsilon \mid aSb \mid aAa \mid bBb \mid bCa$$
$$A \to \varepsilon \mid aAa \mid bCa$$
$$B \to \varepsilon \mid bBb \mid bCa$$
$$C \to \varepsilon \mid bCa$$

Esercizio

$$L = \{ a^n b^m b^n a^m \mid n, m \geq 0 \}$$

$$G = ?$$