# Tutorato di Automi e Linguaggi Formali

Esercizi Aggiuntivi: Grammatiche Context-Free e Forma Normale di Chomsky

**Gabriel Rovesti**
Corso di Laurea in Informatica - Università degli Studi di Padova

Tutorato Supplementare - 15-04-2025

## 1 Esercizi Aggiuntivi

**Esercizio 1.** Dimostrare che l'intersezione di un linguaggio context-free e un linguaggio regolare è un linguaggio context-free.

*Suggerimento*: Utilizzare un automa a pila per riconoscere il linguaggio context-free e un DFA per il linguaggio regolare.

**Soluzione.** Sia $L_1$ un linguaggio context-free riconosciuto da un PDA $P = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$ e $L_2$ un linguaggio regolare riconosciuto da un DFA $M = (Q', \Sigma, \delta', q'_0, F')$.

Possiamo costruire un PDA $P'$ che riconosce $L_1 \cap L_2$ simulando contemporaneamente $P$ e $M$ come segue:

$$P' = (Q \times Q', \Sigma, \Gamma, \delta'', (q_0, q'_0), Z_0, F \times F')$$

dove $\delta''$ è definito come:

$$\delta''((q, q'), a, Z) = \{((p, p'), \gamma) \mid (p, \gamma) \in \delta(q, a, Z) \text{ e } p' = \delta'(q', a)\}$$
$$\delta''((q, q'), \varepsilon, Z) = \{((p, q'), \gamma) \mid (p, \gamma) \in \delta(q, \varepsilon, Z)\}$$

Questo PDA simula simultaneamente $P$ su $L_1$ e $M$ su $L_2$. Accetta solo quando entrambi accettano, quindi riconosce esattamente $L_1 \cap L_2$.

Poiché $P'$ è un PDA ben definito, $L_1 \cap L_2$ è un linguaggio context-free.

**Esercizio 2.** Data la grammatica $G$:

$$S \to aSbS \mid bSaS \mid \varepsilon$$

a) Dimostrare che $L(G) = \{w \in \{a, b\}^* \mid n_a(w) = n_b(w)\}$, dove $n_a(w)$ e $n_b(w)$ rappresentano il numero di occorrenze di $a$ e $b$ in $w$.

b) Convertire $G$ in una grammatica in Forma Normale di Chomsky.

**Soluzione.** **a) Dimostrazione che** $L(G) = \{w \in \{a,b\}^* \mid n_a(w) = n_b(w)\}$

Dimostriamo prima che $L(G) \subseteq \{w \in \{a,b\}^* \mid n_a(w) = n_b(w)\}$.

Per induzione sulla lunghezza della derivazione:

- Base: $S \Rightarrow \varepsilon$. Chiaramente $n_a(\varepsilon) = n_b(\varepsilon) = 0$.
- Ipotesi induttiva: Supponiamo che per ogni stringa $w \in L(G)$ generata con al più $k$ passi, abbiamo $n_a(w) = n_b(w)$.
- Passo induttivo: Consideriamo una derivazione di lunghezza $k+1$. Questa derivazione inizia con $S \Rightarrow aSbS$ o $S \Rightarrow bSaS$.

  Nel caso $S \Rightarrow aSbS \Rightarrow^* aw_1bw_2$, dove $w_1, w_2 \in L(G)$ sono generati in al più $k$ passi, abbiamo:

  $$n_a(aw_1bw_2) = 1 + n_a(w_1) + n_a(w_2)$$
  $$n_b(aw_1bw_2) = 1 + n_b(w_1) + n_b(w_2)$$

  Per l'ipotesi induttiva, $n_a(w_1) = n_b(w_1)$ e $n_a(w_2) = n_b(w_2)$, quindi $n_a(aw_1bw_2) = n_b(aw_1bw_2)$.

  Il caso $S \Rightarrow bSaS$ è analogo.

Per dimostrare che $\{w \in \{a,b\}^* \mid n_a(w) = n_b(w)\} \subseteq L(G)$, procediamo per induzione sulla lunghezza di $w$:

- Base: Se $|w| = 0$, allora $w = \varepsilon$ e $\varepsilon \in L(G)$ mediante la produzione $S \to \varepsilon$.
- Passo induttivo: Sia $w$ una stringa con $n_a(w) = n_b(w)$ e $|w| > 0$. Possiamo scrivere $w = xay$ dove $a$ è il primo simbolo $a$ in $w$. Poiché $n_a(w) = n_b(w)$, esiste almeno un $b$ in $w$.

  Se $y$ contiene un $b$, possiamo scrivere $y = zbt$ dove $z$ non contiene $b$. Allora $w = xazbt$. Definiamo $w_1 = xz$ e $w_2 = t$. Si può dimostrare che $n_a(w_1) = n_b(w_1)$ e $n_a(w_2) = n_b(w_2)$. Per l'ipotesi induttiva, $w_1, w_2 \in L(G)$. Quindi $w = aw_1bw_2$ può essere generato mediante $S \Rightarrow aSbS \Rightarrow^* aw_1bw_2$.

  Un caso simile si applica se il primo simbolo di $w$ è $b$.

Quindi, $L(G) = \{w \in \{a,b\}^* \mid n_a(w) = n_b(w)\}$.

**b) Conversione in Forma Normale di Chomsky**

Passo 1: Aggiungere una nuova variabile iniziale $S_0$.

$$S_0 \to S$$
$$S \to aSbS \mid bSaS \mid \varepsilon$$

Passo 2: Eliminare le $\varepsilon$-regole. Poiché $S \to \varepsilon$, sostituiamo $S$ con $\varepsilon$ nelle altre produzioni:

$$S_0 \to S \mid \varepsilon$$
$$S \to aSbS \mid aS \mid SbS \mid bS \mid bSaS \mid bS \mid SaS \mid aS \mid ab \mid ba$$

Semplificando le regole duplicate:

$$\begin{aligned}
S_0 &\to S \mid \varepsilon \\
S &\to aSbS \mid aS \mid SbS \mid bS \mid bSaS \mid SaS \mid ab \mid ba
\end{aligned}$$

Passo 3: Eliminare le regole unitarie. Da $S_0 \to S$, sostituiamo con tutte le produzioni di $S$:

$$\begin{aligned}
S_0 &\to aSbS \mid aS \mid SbS \mid bS \mid bSaS \mid SaS \mid ab \mid ba \mid \varepsilon \\
S &\to aSbS \mid aS \mid SbS \mid bS \mid bSaS \mid SaS \mid ab \mid ba
\end{aligned}$$

Passo 4: Trasformare le regole nella forma corretta. Introduciamo nuove variabili:

$$\begin{aligned}
S_0 &\to A_1B_1 \mid A_1S \mid SB_1S \mid B_1S \mid B_1SA_1S \mid SA_1S \mid A_1B_1 \mid B_1A_1 \mid \varepsilon \\
S &\to A_1B_1S \mid A_1S \mid SB_1S \mid B_1S \mid B_1SA_1S \mid SA_1S \mid A_1B_1 \mid B_1A_1 \\
A_1 &\to a \\
B_1 &\to b
\end{aligned}$$

Continuiamo introducendo variabili aggiuntive per le produzioni con più di due simboli:

$$\begin{aligned}
S_0 &\to A_1B_1 \mid A_1S \mid C_1S \mid B_1S \mid B_1D_1 \mid E_1S \mid A_1B_1 \mid B_1A_1 \mid \varepsilon \\
S &\to A_1F_1 \mid A_1S \mid C_1S \mid B_1S \mid B_1D_1 \mid E_1S \mid A_1B_1 \mid B_1A_1 \\
A_1 &\to a \\
B_1 &\to b \\
C_1 &\to SB_1 \\
D_1 &\to SA_1 \\
E_1 &\to SA_1 \\
F_1 &\to B_1S
\end{aligned}$$

Possiamo semplificare ulteriormente eliminando produzioni duplicate. Il risultato finale è una grammatica in Forma Normale di Chomsky per $L(G)$.

**Esercizio 3.** Sia $L$ un linguaggio context-free. Dimostrare che il linguaggio $\text{HALF}(L) = \{x \mid xy \in L \text{ e } |x| = |y|\}$ è context-free.

**Soluzione.** Per dimostrare che $\text{HALF}(L)$ è context-free, costruiamo una grammatica che lo genera.

Sia $G = (V, \Sigma, R, S)$ una grammatica in Forma Normale di Chomsky che genera $L$. Costruiamo una nuova grammatica $G' = (V', \Sigma, R', S')$ dove:

$V' = V \cup \{S'\} \cup \{[A, B] \mid A, B \in V\}$

Le regole in $R'$ sono:

- $S' \to [S, A]$ per ogni $A \in V$
- $[A, B] \to [C, E][D, F]$ per ogni regola $A \to CD$ e $B \to EF$ in $R$
- $[A, B] \to a$ per ogni regola $A \to a$ in $R$ e $B \in V$

L'intuizione dietro questa costruzione è che la coppia $[A, B]$ genera la prima metà di una stringa che può essere generata a partire da $A$ nella grammatica originale, a condizione che la seconda metà possa essere generata a partire da $B$.

Possiamo dimostrare per induzione che $[A, B] \Rightarrow^* x$ in $G'$ se e solo se esistono stringhe $y, z$ tali che $A \Rightarrow^* xy$ in $G$, $B \Rightarrow^* z$ in $G$, e $|x| = |yz|/2$.

Da ciò segue che $S' \Rightarrow^* x$ in $G'$ se e solo se esiste una stringa $y$ tale che $xy \in L$ e $|x| = |y|$, ovvero $x \in \text{HALF}(L)$.

Quindi, $G'$ genera $\text{HALF}(L)$, e pertanto $\text{HALF}(L)$ è context-free.

**Esercizio 4.** In linguistica computazionale, molte lingue naturali possono essere approssimate con grammatiche context-free. Progettare una grammatica context-free semplificata per riconoscere frasi basilari in italiano con la struttura:

[soggetto] [verbo] [complemento oggetto] [complemento di termine]

- Il soggetto e il complemento oggetto sono formati da articolo + sostantivo
- Il complemento di termine è formato da preposizione + articolo + sostantivo
- Il verbo è un singolo termine

*Esempio*: "Lo studente legge il libro allo studente."

**Soluzione.** Ecco una grammatica context-free semplificata per riconoscere frasi basilari in italiano:

$$
\begin{aligned}
\langle\text{FRASE}\rangle &\to \langle\text{SOGGETTO}\rangle\langle\text{VERBO}\rangle\langle\text{OGGETTO}\rangle\langle\text{TERMINE}\rangle \\
\langle\text{SOGGETTO}\rangle &\to \langle\text{ARTICOLO}\rangle\langle\text{SOSTANTIVO}\rangle \\
\langle\text{OGGETTO}\rangle &\to \langle\text{ARTICOLO}\rangle\langle\text{SOSTANTIVO}\rangle \\
\langle\text{TERMINE}\rangle &\to \langle\text{PREPOSIZIONE}\rangle\langle\text{ARTICOLO}\rangle\langle\text{SOSTANTIVO}\rangle \\
\langle\text{ARTICOLO}\rangle &\to \text{il} \mid \text{lo} \mid \text{la} \mid \text{i} \mid \text{gli} \mid \text{le} \\
\langle\text{SOSTANTIVO}\rangle &\to \text{studente} \mid \text{libro} \mid \text{professore} \mid \text{esame} \\
\langle\text{VERBO}\rangle &\to \text{legge} \mid \text{studia} \mid \text{scrive} \mid \text{supera} \\
\langle\text{PREPOSIZIONE}\rangle &\to \text{a} \mid \text{al} \mid \text{allo} \mid \text{alla}
\end{aligned}
$$

Con questa grammatica, possiamo generare frasi come:

- "Lo studente legge il libro allo studente."
- "Il professore scrive l'esame allo studente."
- "Lo studente supera l'esame al professore."

Per convertire questa grammatica in Forma Normale di Chomsky, dovremmo:

1. Introdurre una nuova variabile iniziale
2. Eliminare le regole unitarie (in questo caso non ce ne sono)
3. Trasformare le regole con più di due simboli non terminali

4. Sostituire i terminali nelle regole con più di un simbolo

Ad esempio, la regola:

$$\langle \text{TERMINE} \rangle \rightarrow \langle \text{PREPOSIZIONE} \rangle \langle \text{ARTICOLO} \rangle \langle \text{SOSTANTIVO} \rangle$$

Diventerebbe:

$$\begin{aligned} \langle \text{TERMINE} \rangle &\rightarrow \langle \text{PREPOSIZIONE} \rangle \langle \text{X} \rangle \\ \langle \text{X} \rangle &\rightarrow \langle \text{ARTICOLO} \rangle \langle \text{SOSTANTIVO} \rangle \end{aligned}$$

dove $\langle \text{X} \rangle$ è una nuova variabile introdotta.