UNIVERSITÀ
DEGLI STUDI
DI TORINO

Proprietà dei linguaggi liberi e relazione con i linguaggi regolari

a.a. 2018-2019

Bibliografia

Le slide seguenti presentano argomenti trattati nel corso in modo sintetico.

Gli studenti che intendessero approfondirli possono far riferimento al cap. 7 del testo:

Automi, Linguaggi e Calcolabilità,
J. E. Hopcroft, R. Motwani, J. D. Ullman

Proprietà dei linguaggi context-free

Proprietà Pumping.

Ogni linguaggio libero dal contesto soddisfa una proprietà caratteristica, chiamata pumping, simile alla proprietà dei linguaggi regolari con lo stesso nome. In presenza di un falso linguaggio libero, l'uso di tale proprietà permette di ottenere una contraddizione.

Proprietà di chiusura.

Operazioni che applicate a linguaggi liberi dal contesto forniscono linguaggi liberi, cioè operazioni rispetto alle quali la classe dei linguaggi liberi è chiusa.

Derivazioni ricorsive

Una derivazione è *ricorsiva* se un non terminale produce in almeno un passo una stringa che contiene il non terminale stesso, cioè se è della forma $A \Rightarrow^+ \alpha A \beta$, con α e β non entrambe vuote.

- *ricorsiva sinistra* se $\alpha = \varepsilon$, cioè $A \Rightarrow^+ A\beta$,
- *ricorsiva destra* se $\beta = \varepsilon$, cioè $A \Rightarrow^+ \alpha A$.

Il *non terminale A è detto ricorsivo.*

Esempio:

a) $G = <\{S\}, \{a, b\}, \{S \rightarrow aSb \mid ab\}, S>$

$S \Rightarrow aSb$ S in un passo si riscrive con una stringa che contiene S stesso: S è una variabile ricorsiva.

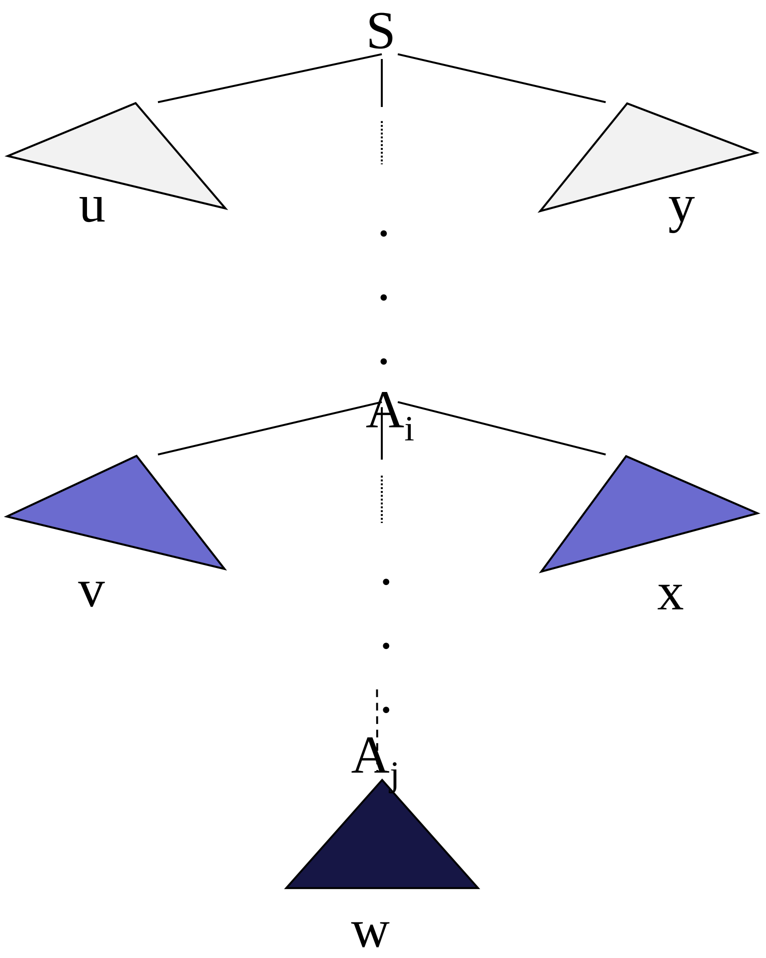
b) $G = <\{S\}, \{a, b\}, \{S \rightarrow aA \mid ab, A \rightarrow Sb\}, S>$

$S \Rightarrow aA \Rightarrow aSb$ S in due passi si riscrive con una stringa che contiene S stesso: S è una variabile ricorsiva.

Proprietà pumping dei linguaggi liberi

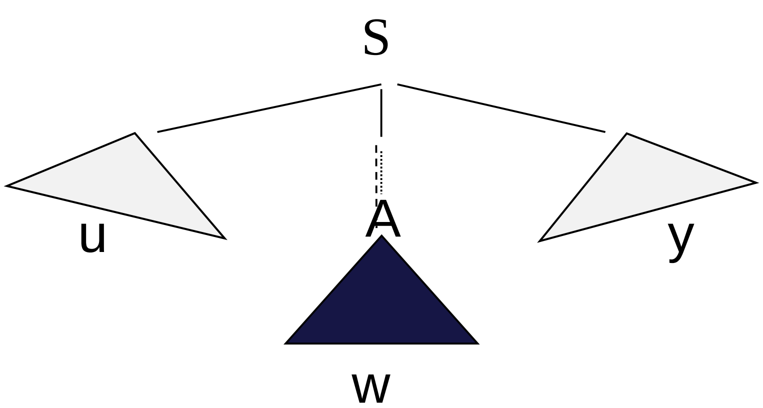
Consideriamo un linguaggio L generato da una grammatica con n variabili.

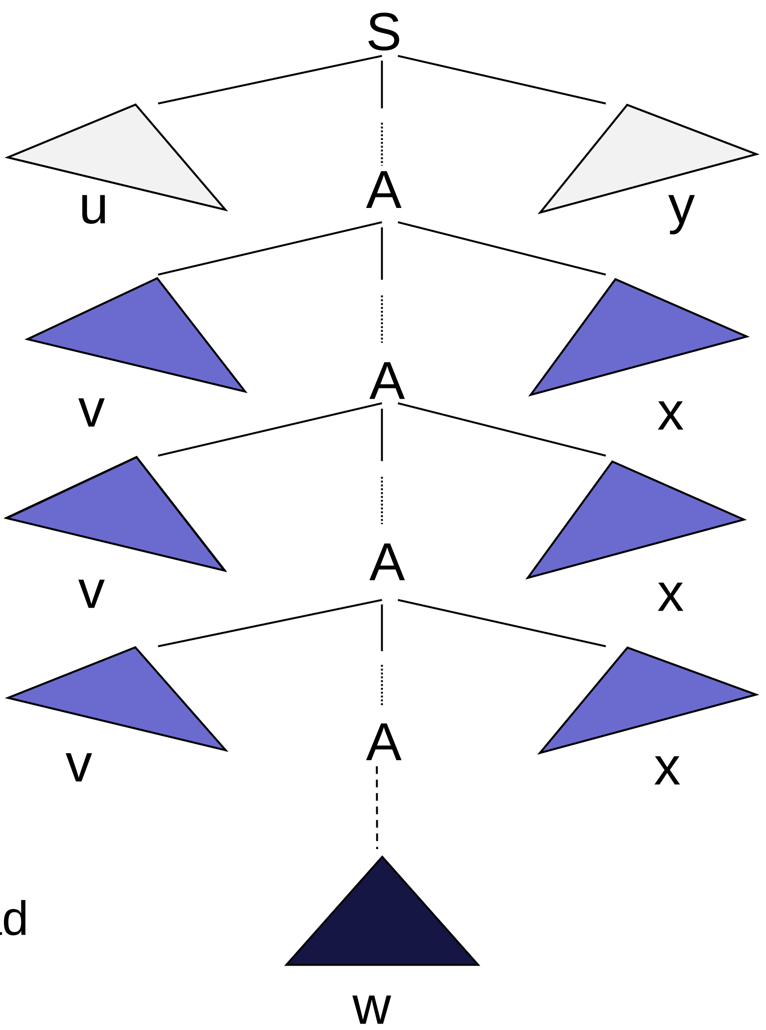
Se la stringa $z \in L$ è ottenuta con una derivazione di lunghezza $\geq n$, l'albero sintattico di z ha un simbolo non terminale che si ripete in un cammino dalla radice ad una foglia, cioè la grammatica presenta una variabile ricorsiva A.
Supponiamo $A_i = A_j$.
Chiamiamo w il prodotto del sottoalbero con radice A_j e vwx il prodotto del sottoalbero con radice A_i



Proprietà pumping dei linguaggi liberi

Poiché $A_i = A_j$, possiamo rimpiazzare il sottoalbero di A_i con quello di A_j, ottenendo quindi uwy, che ancora appartenere a L.





Oppure possiamo allungare il percorso tra A_i e A_j, ottenendo, ad esempio, la parola uvvvwxxxy, ancora generata da L.

Proprietà pumping dei linguaggi liberi

Proprietà pumping per i linguaggi liberi

Per ogni linguaggio context-free L esiste una costante caratteristica k, che dipende solo da L, tale che ogni frase $z \in L$, di lunghezza maggiore o uguale a k: $|z| \geq k$, si può scrivere come la concatenazione di cinque sottostringhe xuwvy con:

- $|uwv| \leq k$,
- $uv \neq \varepsilon$
- $xu^iwv^iy \in L$ per $i \geq 0$.

Proprietà di chiusura

Siano L e M due linguaggi context-free. Allora i seguenti linguaggi sono context-free:

Unione: $L \cup M$

Concatenazione: $L.M$

Chiusura: L^*

Inversione: $L^R = \{w^R \mid w \in L\}$

I linguaggi context-free non sono invece chiusi per complemento, differenza e intersezione

Sono chiusi per intersezione con i linguaggi regolari.

Proprietà di chiusura

La chiusura rispetto a unione, concatenazione e chiusura di Kleene è facilmente provabile.

Se si hanno due grammatiche libere che generano i linguaggi L e M rispettivamente, è possibile costruire una grammatica libera per i linguaggi $L \cup M$, $L.M$ e L^*.

Allo stesso modo è provabile la chiusura rispetto alla operazione di inversione:

Se il linguaggio L è generato dalla grammatica $G = <V, \Sigma, \mathbb{P}, S>$, il linguaggio L^R è generato dalla grammatica

$$G_R = <V, \Sigma, \mathbb{P}^R, S>$$

dove $\mathbb{P}^R$ è l'insieme delle produzioni ottenute da quelle di $\mathbb{P}$ scrivendone al contrario i corpi.

Esercizi

- Fornire le grammatiche dei linguaggi unione, concatenazione, chiusura di Kleene e inversione dei linguaggi generati dalle grammatiche di cui sono dati i seguenti insiemi di produzioni:

  $P1 = \{S \rightarrow SaS \mid B,\ B \rightarrow a\}$

  $P2 = \{S \rightarrow SaAb \mid c,\ A \rightarrow AdS \mid S\}$

  $P3 = \{S \rightarrow Ab \mid cS,\ A \rightarrow aA \mid \varepsilon\}$

- Costruire un automa a pila che riconosca l'intersezione dei linguaggi: $\{wbw^R \mid w \in \{r, v\}\}$ e $\{(rv)^n b(vr)^m \mid n, m \geq 0\}$

Relazione tra linguaggi liberi e linguaggi regolari

Un linguaggio regolare è anche libero dal contesto.

I linguaggi regolari formano però un sottinsieme proprio dei linguaggi liberi e per ogni linguaggio regolare si può costruire una grammatica le cui produzioni hanno una forma particolare.

Una grammatica è *unilineare* se le sue regole hanno la forma:

$A \rightarrow wB$ $\quad w \in T^*$ e $B \in V \cup \{\varepsilon\}$ *unilineare destra*

$L = \{(01)^n / n > 0\}$ a) $S \rightarrow 0A$, $A \rightarrow 1S \mid 1$ b) $S \rightarrow 01S \mid 01$

$A \rightarrow Bw$ $\quad w \in T^*$ e $B \in V \cup \{\varepsilon\}$ *unilineare sinistra*

$L = \{(01)^n / n > 0\}$ a) $S \rightarrow A1$, $A \rightarrow S0 \mid 0$ b) $S \rightarrow S01 \mid 01$

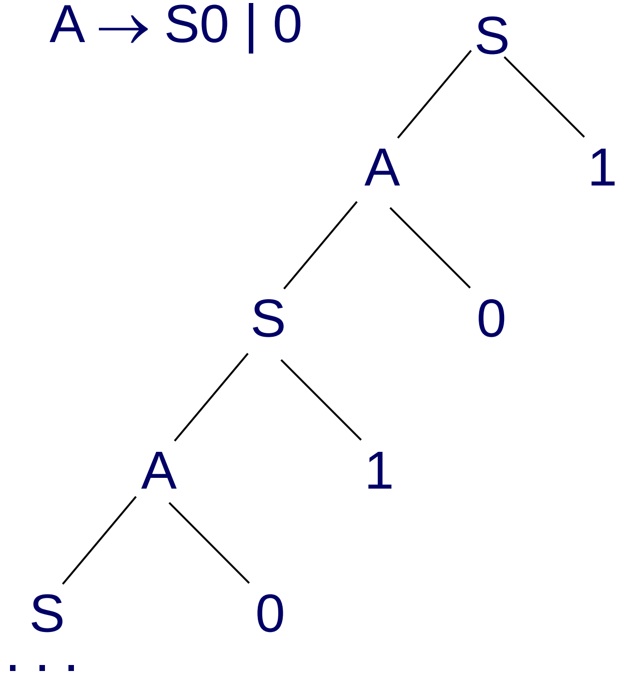
Nota: In una grammatica unilineare non si possono mescolare produzioni del primo e del secondo tipo.

Grammatiche unilineari

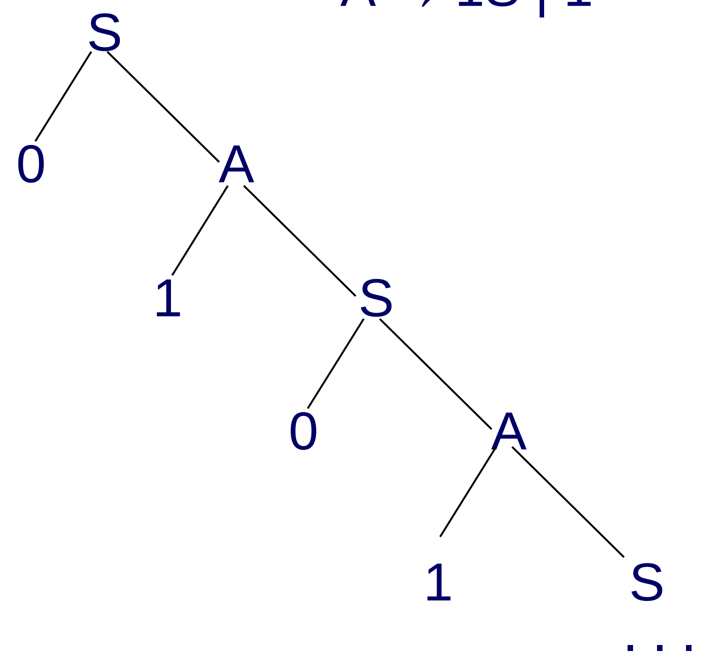
Ogni grammatica unilineare destra è equivalente a una grammatica unilineare sinistra, e viceversa.

Come sono gli alberi di derivazione di queste grammatiche ?

$S \to A1$
$A \to S0 \mid 0$



$S \to 0A$
$A \to 1S \mid 1$



Limiti delle grammatiche context-free

Le grammatiche libere dal contesto:

1. non sono in grado di generare stringhe con tre o più gruppi di caratteri ripetuti un numero di volte in relazione tra loro (ad esempio un numero uguale di volte, come $\{a^n b^n c^n \mid n > 0\}$).

2. non sanno generare coppie con lo stesso numero di simboli, se le coppie sono “intrecciate”.

   Esempio: $L = \{0^i 1^j 2^i 3^j \mid i, j \geq 1\}$.

3. non sanno ripetere una stringa di lunghezza arbitraria, se la stringa è su un alfabeto di cardinalità maggiore di 1.

   Esempio: $L = \{ww \mid w \in \{0,1\}^*\}$.

Limiti delle grammatiche context-free

Il seguente linguaggio non è libero dal contesto, cioè non esiste nessuna grammatica context-free che lo generi.

$L = \{0^n 1^n 2^n \mid n > 0\}$

Una grammatica che genera L è la seguente, in cui le riscritture non sono definite per le variabili, ma per le variabili “in un contesto”.

$< \{S, B, C\}, \{0, 1, 2\}, \mathbb{P}, S >$ dove

$$
\begin{aligned}
\mathbb{P} = \{ & S \rightarrow 0\, S\, BC \mid 01C, \\
& CB \rightarrow BC, \\
& 1B \rightarrow 11, \\
& 1C \rightarrow 12, \\
& 2C \rightarrow 22\}
\end{aligned}
$$

Classificazione di Chomsky

Chomsky ha introdotto una classificazione “storica” dei linguaggi, distinguendoli in tipo 3 (regolari), tipo 2 (liberi dal contesto), tipo 1 (contestuali) e tipo 0 (ricorsivamente enumerabili).

Le grammatiche che generano i diversi tipi di linguaggi differiscono per la forma delle produzioni.

- I linguaggi di “***tipo 0***” sono generati da grammatiche le cui produzioni $\alpha \rightarrow \beta$, con $\alpha, \beta \in (T \cup V)^*$, soddisfano l’unico vincolo $\alpha \neq \varepsilon$.
- Le grammatiche che generano i linguaggi di “***tipo 1***” hanno come vincolo ulteriore sulla forma delle produzioni $|\alpha| \leq |\beta|$.
- Le grammatiche di “***tipo 2***” sono le grammatiche context-free, le cui produzioni soddisfano il vincolo: $\alpha \in V$ e $\beta \in (T \cup V)^*$.
- I linguaggi di “***tipo 3***” o regolari sono generati dalle grammatiche unilineari.

Classificazione di Chomsky