pumping lemma per CFL

La forma normale di Chomsky (CNF) per le CFG prevede che tutte le produzioni siano:

A -> a, con a in T

A-> BC, con B e C variabili

non c’è ε !!

Teorema. Per ogni CFG G, esiste una CFG G’ in CNF tale che L(G’)=L(G)-{ε} ripuliamo la CFG

--dai simboli inutili

--dalle produzioni unitarie, i.e. A->B

--dalla ε

un simbolo X è <u>utile</u> per una CFG se
$S \Rightarrow^* \alpha X \beta \Rightarrow^* w$

un simbolo è generatore se
$X \Rightarrow^* w$

è raggiungibile se
$S \Rightarrow^* \alpha X \beta$

prima di eliminano i non generatori e poi i non raggiungibili
restano solo i simboli utili

esempio:

S-> AB | a

A -> b

non generatori : B

resta S -> a e A -> b

ma A è diventato non raggiungibile, quindi resta

S -> a

Attenzione all’ordine

Teorema 7.2 data la CFG G, costruiamo G1 in 2 passi:

1) Otteniamo G2 da G eliminando da G le variabili non generatrici e le produzioni in cui queste variabili compaiono
2) Otteniamo G1 da G2 eliminando da G2 i simboli non raggiungibili

G1 contiene tutti e soli i simboli utili di G e quindi L(G1)=L(G)

Dimostrazione: (1) butta variabili e produzioni che non generano (2) butta variabili e produzioni che non sono raggiungibili da S, mai variabili utili

calcolo dei simboli generatori $\prod$:
-base: i simboli terminali sono generatori, $\prod$ =T
-induzione: aggiungiamo a $\prod$ tutte le variabili X
t.c. esiste X -> α in cui α ha solo simboli in $\prod$

fino a quando $\prod$ cresce

Calcolo dei simboli raggiungibili

Base: R contiene S

Induzione: Aggiungere ad R i simboli in β tale che A-> β per A in R

Nello stesso modo:

calcolo dell'insieme Z dei simboli che producono ε

base: $X \rightarrow \varepsilon$ mettiamo X in Z
induzione: aggiungiamo a Z ogni Y tale che Y-> α con i simboli di α tutti in Z

Per una qualsiasi CFG G possiamo costruire G' t.c. $L(G')=L(G)-\{\varepsilon\}$ e senza ε-produzioni

Idea della costruzione di G’:

Prendiamo una qualsiasi produzione p di G, A->X1…Xk, se m dei k Xi sono in Z, allora G’ ha 2^m produzioni ottenute da p annullando tutti i sottoinsiemi delle m Xi

non si considera m=k, infatti A è in Z

Le produzioni A -> ε sono eliminate

Esempio:

S-> AB

A-> a AA | ε

B-> bBB | ε

Z={A,B,S} quindi G’ ha

S-> AB | A | B

A->aAA | aA |aA | a

B-> bBB | bB | bB | b

La forma normale di Chomsky (CNF) per le CFG prevede che tutte le produzioni siano:

A -> a, con a in T

A-> BC, con B e C variabili

Teorema 7.17.
Negli alberi di derivazione di una CFG in CNF è vero che se il cammino più lungo è di lunghezza n, allora il prodotto w è t.c. $|w| <= 2^{n-1}$

Dimostrazione: per induzione su n
Base: n=1 l'albero consiste della radice e di una foglia etichettata con un terminale
Prodotto w = foglia quindi $|w|=1 = 2^{1-1}$

Induzione: altezza n. La radice usa A->BC
B e C sono radici di alberi di altezza al massimo n-1,

quindi i loro prodotti, per ipotesi induttiva, sono di lunghezza al più 2^{n-2}, e $2 * 2^{n-2} = 2^{n-1}$

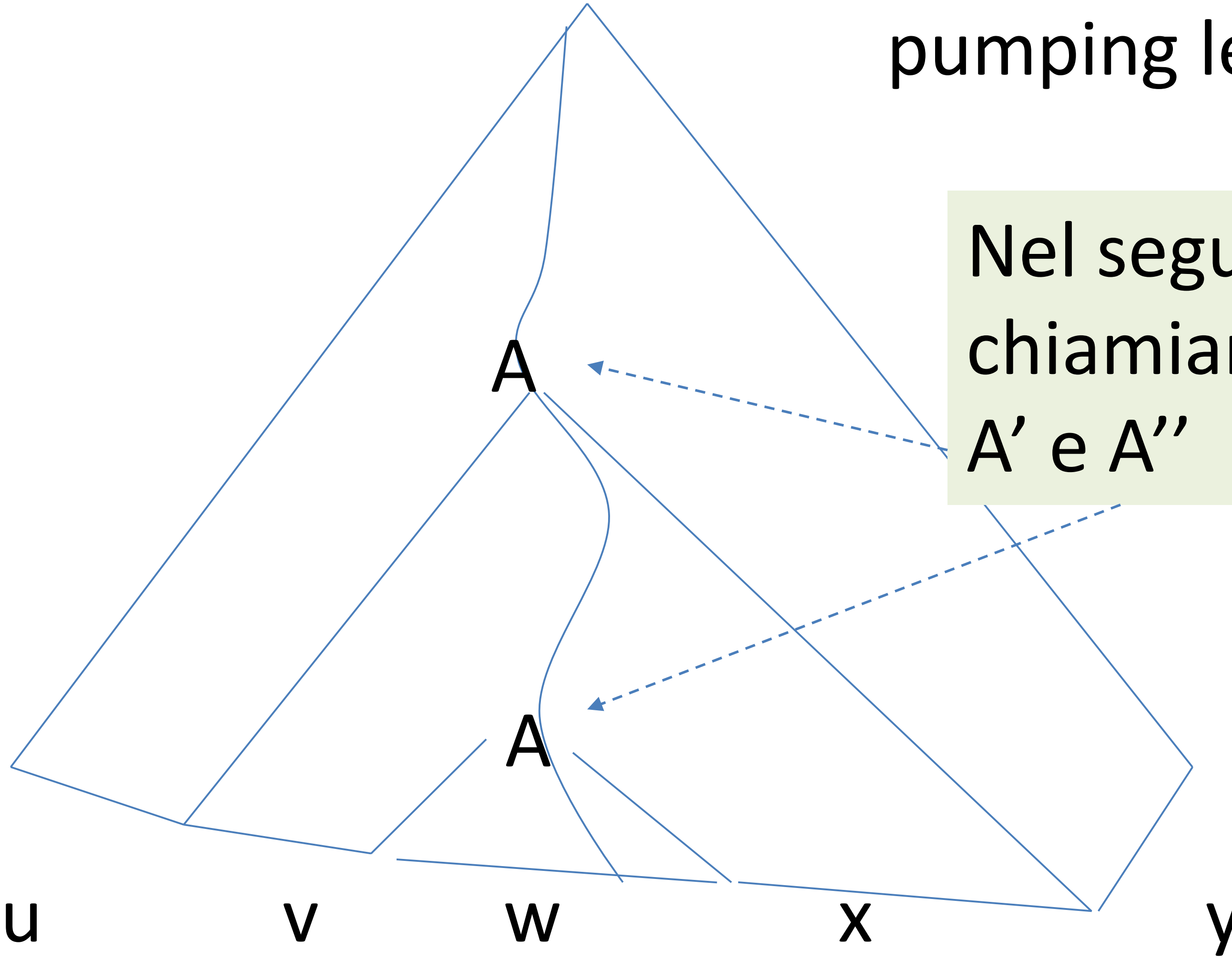
Conseguenza:

se m è il numero delle variabili della grammatica in CNF:
un albero con prodotto w t.c. $|w|=2^m$ deve avere un cammino di lunghezza >= m+1

e allora su quel cammino almeno una variabile ripete

Quindi nel seguito $n = 2^m$

pumping lemma



Nel seguito li chiamiamo A’ e A’’

$z = uvwxy$

con $|vwx| <= n$ e $vx \; != \varepsilon$

In $z = uvwxy$ deve essere vero che $|vwx| <= n$ e $vx != \varepsilon$:
Perché l'altezza di A' è <= m+1 (infatti in un albero di questa altezza c'è ripetizione) e quindi il suo prodotto xwx ha lunghezza <= $2^m=n$
Inoltre il cammino da A' a A'' è costruito con produzione X->BC e ogni nonterminale deve produrre qualche terminale (per la CNF) e quindi $|vx|>0$

Basta ora osservare che dall'albero della figura precedente possiamo generare un numero infinito di alberi che sono tutti della grammatica di partenza:

--in A' possiamo incollare l'albero con radice A'', ottenendo l'albero con prodotto uwy

--possiamo incollare in A'' l'albero con radice A', ottenendo il prodotto uv^2wx^2y

--dall'albero appena costruito, potremmo di nuovo incollare in A'' l'albero con radice A' ottenendo il prodotto uv^3wx^3y

--in questo modo possiamo ottenere alberi di derivazione con prodotto uv^iwx^iy per qualsiasi $i>=0$.

Il pumping Lemma serve (tra l'altro) a dimostrare che certi linguaggi non sono CF

Esempio: $L=\{0^n1^n2^n \mid n > 0\}$ non è CF

Per il pumping lemma, esiste n tale che ogni z più lunga di n è z=uvwxy. Scegliamo
$z= 0^n1^n2^n$ sappiamo che $|vwx| <= n$ e quindi o contiene degli 0 o contiene dei 2, ma non entrambi.

Supponiamo che vwx contenga 0 e non 2.
Quindi ripetendo v e x (non entrambi vuoti) aumento il n. di 0 e/o 1 mantenendo inalterato il n. di 2 e quindi otterrei una stringa che non è nel linguaggio L.