Approfondimento 4.3

Dimostrazione del Teorema 4.12

Sia $G = (NT, T, R, S)$ una grammatica per $L = \mathcal{L}[G]$. Sia b il massimo fattore di ramificazione in un albero di derivazione di G (ovvero: b è il massimo numero di simboli che compaiono nella parte destra di una produzione di G). Possiamo certo supporre che $b \geq 2$. Per un semplice risultato di combinatoria, un albero con fattore di ramificazione b e altezza h ha al più b^h foglie (la radice ha altezza 0). Fissiamo allora $N = b^{|NT|+1}$; siccome $b \geq 2$, si ha $N > b^{|NT|}$. Pertanto, ogni albero di derivazione di una stringa di lunghezza $\geq N$ deve avere altezza *almeno* $|NT| + 1$.

Sia ora $z \in L$, con $|z| \geq N$, e consideriamo un albero di derivazione per z; se z ammette più di un albero, prendiamo quello (o uno tra quelli) con il minimo numero di nodi. Siccome $|z| \geq N$, l'albero ha altezza $\geq |NT| + 1$, cioè ha un cammino radice-foglia con almeno $|NT| + 2$ nodi. Prendiamo in considerazione uno di questi cammini lunghi. La foglia è un terminale, per cui sul cammino ci sono almeno $|NT| + 1$ nodi etichettati con non terminali. Per il solito principio della piccionaia, un'etichetta non terminale si deve ripetere. Cominciamo a risalire il cammino dalla foglia, e sia A il primo non terminale che si ripete (e si ripete certo nei primi $|NT| + 1$ nodi che si incontrano). La derivazione $S \Rightarrow^* z$ può allora essere decomposta in $S \Rightarrow^* uAy \Rightarrow^* uvAxy \Rightarrow^* uvwxy$, per opportune sottostringhe u, v, w, x, y. La Figura 16.7 illustra questa costruzione sull'albero di derivazione.

La figura dovrebbere convincere immediatamente che le stringhe v e x possono essere arbitrariamente pompate (assieme, cioè lo stesso numero di volte). Infatti si ha sia $A \Rightarrow^* w$ (con il sottoalbero ρ), sia $A \Rightarrow^* vwx$ (con il sottoalbero π che contiene a sua volta ρ). Siccome la grammatica è libera da contesto possiamo sostituire il piccolo ρ al posto di π, ottenendo $S \Rightarrow^* uAy \Rightarrow^* uwy = uv^0wx^0y$ (pompaggio zero); oppure possiamo sostituire una copia del grosso π al posto del suo sottoalbero ρ, ottenendo $S \Rightarrow^* uAy \Rightarrow^* uvAxy \Rightarrow^* uvvAxxy \Rightarrow^* uvvxxxy$ (pompaggio $k = 2$). E ovviamente possiamo ripetere questo procedimento quante volte vogliamo. Abbiamo dunque stabilito (iii): per ogni $k \geq 0$, $uv^kwx^ky \in L$.

Rimangono da verificare i vincoli sulle lunghezze delle stringhe. Per (ii) $|vx| \geq 1$: v e x non possono essere entrambe ϵ. Se lo fossero, infatti, l'albero con pompaggio 0 genererebbe ancora z ed avrebbe un numero di nodi strettamente inferiore a quello dell'albero di partenza (perché abbiamo rimosso almeno un nodo etichettato A). Ma questo non è possibile: abbiamo scelto un albero con il minimo numero di nodi che genera z. Per (i) $|vwx| \leq N$: abbiamo già osservato

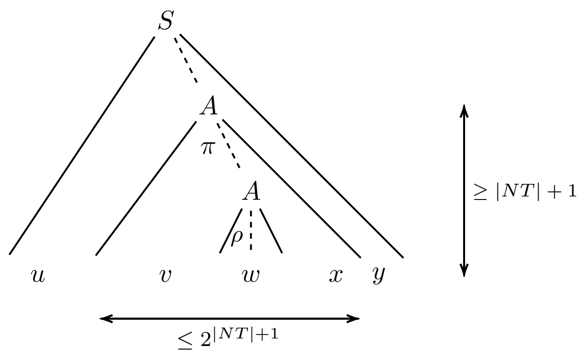

Figura 16.7 La dimostrazione del PL per i linguaggi liberi.

che l'occorrenza più alta di A (tra le due che consideriamo) genera vwx. Le due occorrenze stanno entrambe nel limite di $|NT| + 1$ nodi non terminali dalla foglia. Il sottoalbero radicato nella A superiore ha dunque altezza al più $|NT| + 1$, e dunque genera stringhe lunghe al più $b^{|NT|+1} = N$. $\square$