

Teoria dei Linguaggi

Indice

1. Lezione 12 [04/04]	3
1.1. Varianti di automi	3
1.1.1. Automi pesati	3
1.1.2. Automi probabilistici	3
2. Lezione 13 [09/04]	4
2.1. Varianti di automi	4
2.1.1. One-way VS two-way	4
2.1.2. Read-only VS read-write	4
2.1.3. Memoria esterna	5
2.2. Automi two-way	6
2.2.1. Esempi vari	6
2.2.2. Definizione formale	8
2.2.3. Potenza computazionale	9
3. Lezione 14 [11/04]	14
3.1. Simulazione	14
3.1.1. Problema di Sakoda & Sipser	14
3.2. Automi a pila	16
3.2.1. Versione non deterministica	16
3.2.2. Accettazione	19
3.2.3. Determinismo VS non determinismo	21
3.2.4. Trasformazioni	22
4. Lezione 15 [23/04]	24
4.1. Esempi	24
4.2. Equivalenza tra grammatiche di tipo 2 e automi a pila	24
4.2.1. Ripasso e introduzione	24
4.2.2. Da grammatica di tipo 2 ad automa a pila	27
4.2.3. Automa a pila a grammatica di tipo 2	30
4.3. Forme normali per le grammatiche context-free	30
4.3.1. FN di Greibach	30
5. Lezione 16 [30/04]	33
5.1. Fine dimostrazione	33
5.1.1. Forma normale per gli automi a pila	33
5.1.2. Dimostrazione	35
5.2. Forma normale di Chomsky	38
5.2.1. Definizione	38
5.2.2. Costruzione	38
5.2.2.1. Eliminazione delle ϵ -produzioni	39
5.2.2.2. Eliminazione delle produzioni unitarie	40
5.2.2.3. Eliminazione dei simboli inutili	41
5.2.2.4. Eliminazione dei terminali	41
5.2.2.5. Smontaggio delle produzioni	42
5.2.3. Esempio	42

1. Lezione 12 [04/04]

1.1. Varianti di automi

Per finire questa lezione infinita, vediamo qualche **variante** di automi.

1.1.1. Automi pesati

La prima variante che vediamo sono gli **automi pesati**. Essi associano ad ogni transizione un peso. Il **peso di una stringa** viene calcolato come la somma dei pesi delle transizioni che la stringa attraversa per essere accettata. Questo peso poi può essere usato in problemi di ottimizzazione, come trovare il cammino di peso minimo, ma questo ha senso solo su NFA.

1.1.2. Automi probabilistici

Un tipo particolare di automi pesati sono gli **automi probabilistici**, che come pesi sulle transizioni hanno la probabilità di effettuare quella transizione. Visto che parliamo di **probabilità**, i pesi sono nel range $[0, 1]$ e, dato uno stato, tutte le transizioni uscenti sommano a 1. In realtà, potremmo sommare a meno di 1 se nascondiamo lo stato trappola. Con questi automi possiamo chiederci con che probabilità accettiamo una stringa.

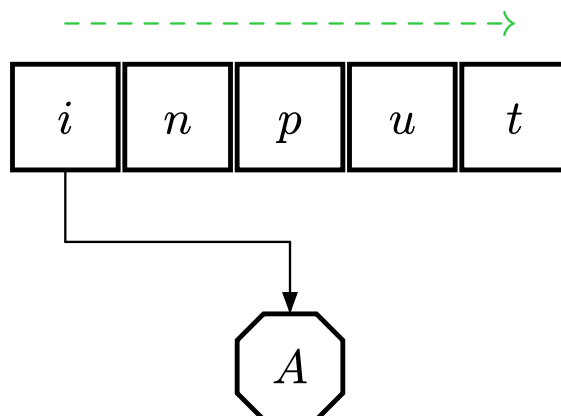
Questi automi li possiamo usare come **riconoscitori a soglia**: tutte le parole oltre una certa soglia le accettiamo, altrimenti le rifiutiamo.

Questi automi comunque non sono più potenti dei DFA: si può dimostrare che se la soglia λ è **isolata**, ovvero nel suo intorno non cade nessuna parola, allora possiamo trasformare questi automi probabilistici in DFA. Se la soglia non è isolata riusciamo a riconoscere una strana classe di linguaggi, che però ora non ci interessa.

2. Lezione 13 [09/04]

2.1. Varianti di automi

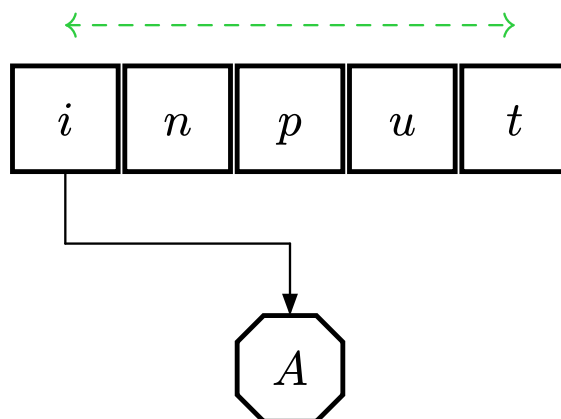
Come possiamo **rappresentare** un automa a stati finiti? Questa macchina è molto semplice: abbiamo un **nastro** che contiene l'input, esaminato da una **testina in sola lettura** che, spostandosi **one-way** da sinistra verso destra, permette ad un **controllo a stati finiti** di capire se la stringa in input deve essere accettata o meno.



La classe di linguaggi che riconosce un automa a stati finiti è la classe dei **linguaggi regolari**. Ma possiamo fare delle modifiche a questo modello? Se sì, che cosa possiamo cambiare?

2.1.1. One-way VS two-way

Se permettiamo all'automa di spostarsi da sinistra verso destra ma anche viceversa, andiamo ad ottenere gli **automi two-way**, che in base alla possibilità di leggere e basta o leggere e scrivere e in base alla lunghezza del nastro saranno in grado di riconoscere diverse classi di linguaggi.



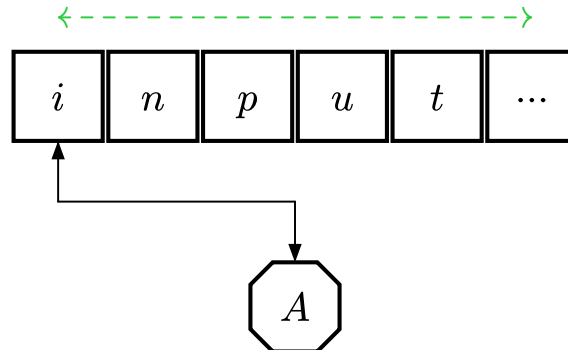
2.1.2. Read-only VS read-write

Se manteniamo l'automa one-way, rendere il nastro anche in lettura non modifica per niente il comportamento dell'automa: infatti, anche se scriviamo, visto che siamo one-way non riusciremo mai a leggere quello che abbiamo scritto.

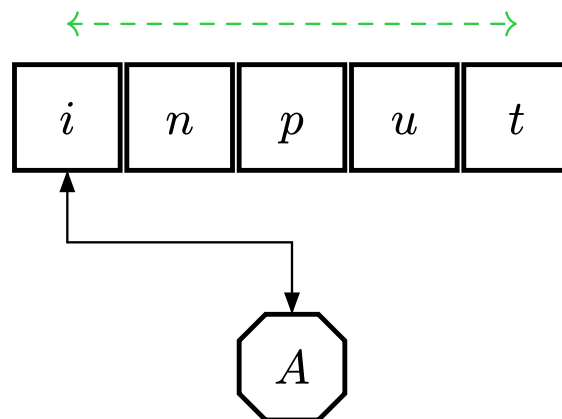
Consideriamo quindi un automa two-way che però mantiene la read-only del nastro: la classe che otteniamo è ancora una volta quella dei **linguaggi regolari**, e questo lo vedremo oggi.

Rendiamo ora la testina capace di poter scrivere sul nastro che abbiamo a disposizione. Ora, in base a come è fatto il nastro abbiamo due situazioni:

- se rendiamo il nastro illimitato oltre la porzione occupata dall'input, andiamo ad riconoscere i linguaggi di tipo 0, ovvero otteniamo una **macchina di Turing**:



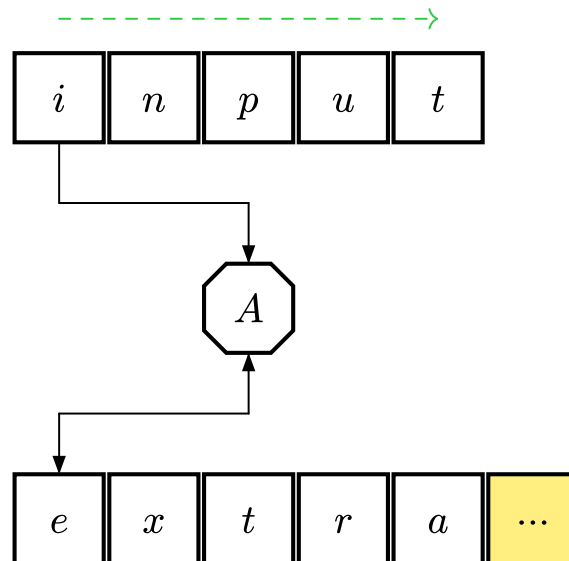
- se invece lasciamo il nastro grande quando l'input andiamo a riconoscere i linguaggi di tipo 1, ovvero otteniamo un **automa limitato linearmente**. Quest'ultima cosa vale perché nelle grammatiche di tipo 1 le regole di produzione non decrescono mai, e un automa limitato linearmente per capire se deve accettare cerca di costruire una derivazione al contrario, accorciando mano a mano la stringa arrivando all'assioma S :



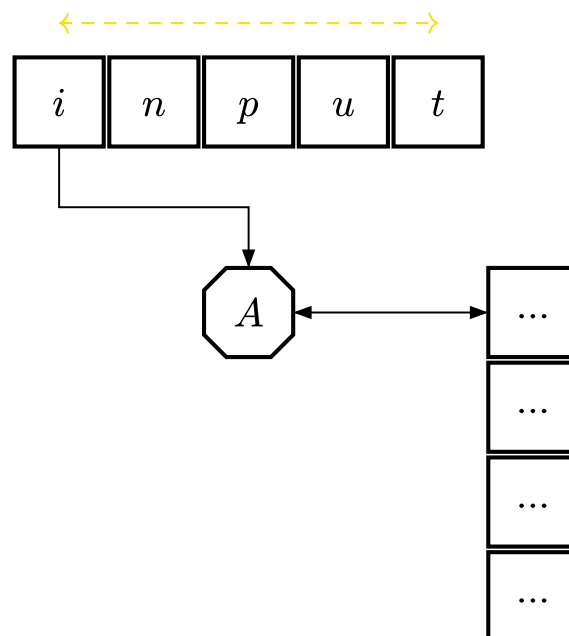
2.1.3. Memoria esterna

L'ultima modifica che possiamo pensare per queste macchine è l'aggiunta di una **memoria esterna**.

Dato un automa one-way con nastro read-only, se aggiungiamo un secondo nastro in read-write che funga da memoria esterna, otteniamo le due situazioni che abbiamo visto per gli automi two-way con possibilità di scrivere sul nastro di input.



Un caso particolare è se la memoria esterna è codificata come una **pila** illimitata, ovvero riesco a leggere solo quello che c'è in cima, allora andiamo a riconoscere i linguaggi di tipo 2, ottenendo quindi un **automa a pila**. Se passiamo infine ad un two-way con una pila diventiamo più potenti ma non sappiamo di quanto.



2.2. Automi two-way

Tra tutte queste varianti, fissiamoci sugli **automi two-way**, ovvero quelli che hanno il nastro in sola lettura e hanno la possibilità di andare avanti e indietro nell'input. Vediamo prima di tutto qualche linguaggio per il quale possiamo usare un automa two-way.

2.2.1. Esempi vari

Esempio 2.2.1.1: Riprendiamo l'operazione α . Dato L regolare, α era tale che

$$\alpha(L) = \{x \in \Sigma^* \mid xx \in L\}.$$

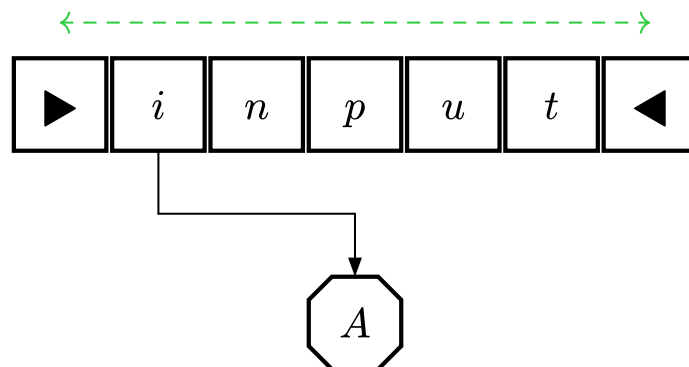
Abbiamo A un DFA che accetta L , come posso costruire un two-way per $\alpha(L)$? Potremmo leggere x la prima volta, ricordarci in che stato siamo arrivati, tornare indietro e poi ripartire a leggere x dallo stato nel quale eravamo arrivati e vedere se finiamo in uno stato finale.

Per fare ciò, ci serve sapere dove finisce il nastro: vedremo come fare tra poco.

Il numero di stati nel two-way è $3n$:

- n stati di A che usiamo per leggere x ;
- n stati che tengono traccia dello stato nel quale siamo arrivati con x e che ci permettono di ritornare all'inizio della stringa;
- n stati che fanno ripartire la computazione dallo stato nel quale siamo arrivati con x e controllano se finiamo in uno stato finale.

Abbiamo sollevato poco fa il problema: come facciamo a capire dove finisce il nastro? Andiamo a inserire dei **marcatori**, uno a sinistra e uno a destra, che delimitano la stringa. Se per caso arriviamo su un marcatore non possiamo andare oltre: possiamo solo rientrare sul nastro. In realtà, vedremo che in un particolare caso usciremo dai bordi.



Vediamo ancora un po' di esempi.

Esempio 2.2.1.2: Definiamo

$$L_n = (a + b)^* a (a + b)^{n-1}$$

il solito linguaggio dell' n -esimo simbolo da destra uguale ad una a .

Avevamo visto che con un NFA avevamo $n + 1$ stati, mentre con un DFA avevamo 2^n stati perché ci ricordavamo una finestra di n simboli. Ora diventa tutto più facile: ci spostiamo, ignorando completamente la stringa, sul marcatore di destra, poi contiamo n simboli e vediamo se accettare o rifiutare.

Come numero di stati siamo circa sui livelli dell'NFA, visto che dobbiamo solo scorrere la stringa per intero e poi tornare indietro di n .

Esempio 2.2.1.3: Definiamo infine

$$K_n = (a + b)^* a (a + b)^{n-1} a (a + b)^*$$

il linguaggio delle parole che hanno due a distanti n . Come lo scriviamo un two-way per K_n ?

Potremmo partire dall'inizio e scandire la stringa x . Ogni volta che troviamo una a andiamo a controllare n simboli dopo e vediamo se troviamo una seconda a :

- se sì, accettiamo;
- se no, torniamo indietro di $n - 1$ simboli per andare avanti con la ricerca.

Il numero di stati è:

- 1 che ricerca le a ;
- n stati per andare in avanti;
- 1 stato di accettazione;
- $n - 1$ stati per tornare indietro.

Ma allora il numero di stati è $2n + 1$.

Abbiamo trovato una buonissima soluzione per l'esempio precedente, ma se volessimo una soluzione alternativa che utilizza un automa sweeping? Ma cosa sono ste cose?

Un **automa sweeping** è un automa che non cambia direzione mentre si trova nel nastro, ma è un automa che rimbalza avanti e indietro sugli end marker. La soluzione che abbiamo trovato non usa automi sweeping perché se il simbolo a distanza n è una b noi invertiamo la direzione e torniamo indietro.

Esempio 2.2.1.4: Cerchiamo una soluzione che utilizzi un automa sweeping per K_n .

Supponiamo di numerare le celle del nastro da 1 a k . Partendo nello stato 1, andiamo a guardare tutte le celle a distanza n : se troviamo una a e poi subito dopo ancora una a accettiamo, altrimenti andiamo avanti fino a quando rimbalziamo sul marker, tornando indietro e andando sulla cella 2. Da qui facciamo ripartire la computazione, andando ogni volta avanti di una cella.

In generale, dalla cella $p \in \{1, \dots, n\}$ noi visitiamo tutte le celle $tn + p$.

Con questo approccio, il numero di stati è $O(n^2)$ perché dobbiamo muoverci di n simboli un numero n di passate. Possiamo farlo con un numero lineare di stati se facciamo la ricerca modulo n anche al ritorno, ma è questo è negli esercizi.

2.2.2. Definizione formale

Abbiamo visto come è costruito un automa two-way, ora vediamo la definizione formale. Definiamo

$$M = (Q, \Sigma, \delta, q_0, q_f)$$

un **2NFA** tale che:

- Q rappresenta l'insieme degli stati;

- Σ rappresenta l'**alfabeto** di input;
- q_0 rappresenta lo **stato iniziale**;
- δ rappresenta la **funzione di transizione**, ed è tale che

$$\delta : Q \times (\Sigma \cup \{\blacktriangleright, \blacktriangleleft\}) \longrightarrow 2^{Q \times \{+1, -1\}},$$

ovvero prende uno stato e un simbolo dell'alfabeto compresi gli end marker e ci restituisce i nuovi stati e che movimento dobbiamo fare con la testina. Ho dei **divieti**: se sono sull'end marker sinistro non ho mosse che mi portano a sinistra, idem ma specchiato su quello di destra con una piccola eccezione, che vediamo tra poco;

- q_f è lo **stato finale** e si raggiunge «passando» oltre l'end marker destro, unico caso in cui si può superare un end marker.

Con questo modello possiamo incappare in **loop infiniti**, che:

- nei DFA non ci fanno accettare;
- negli NFA magari indicano che abbiamo fatto una scelta sbagliata e c'era una via migliore.

Ci sono poi diverse modifiche che possiamo fare a questo modello, ad esempio:

- possiamo estendere le mosse con la **mossa stazionaria**, ovvero quella codificata con 0 che ci mantiene nella posizione nella quale siamo, ma possono essere eliminate con una coppia di mosse sinistra+destra o viceversa;
- possiamo utilizzare un **insieme di stati finali**;
- possiamo **non** usare gli end marker, rendendo molto difficile la scrittura di automi perché non sappiamo dove finisce la stringa.

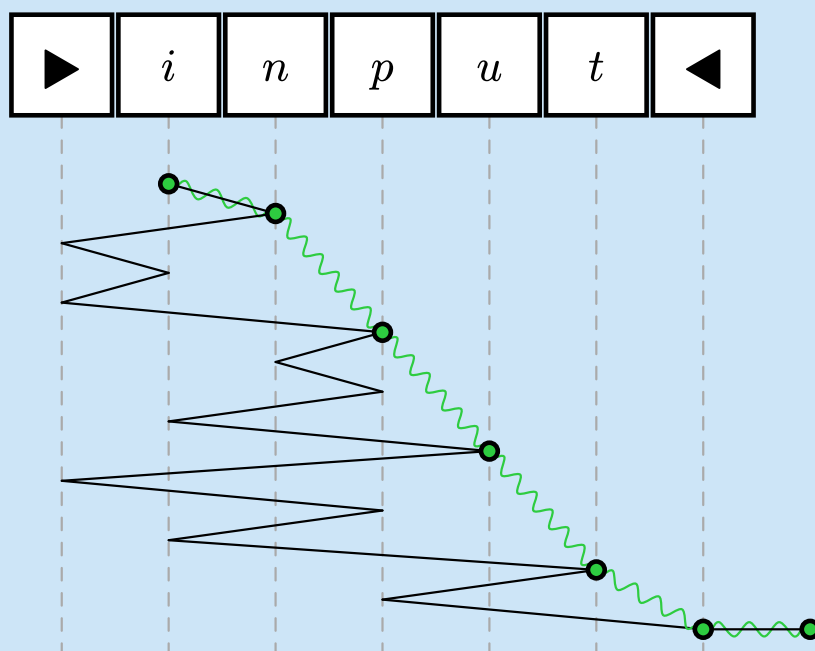
2.2.3. Potenza computazionale

Avere a disposizione un two-way sembra darci molta potenza, ma in realtà non è così: infatti, questi modelli sono equivalenti agli automi a stati finiti one-way, detti anche **1DFA**.

Teorema 2.2.3.1: Vale

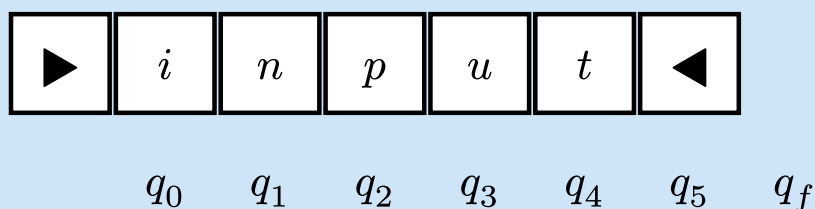
$$L(2DFA) = L(1DFA).$$

Dimostrazione 2.2.3.1.1: Abbiamo a disposizione un 2DFA nel quale abbiamo inserito un input che viene accettato. Vogliamo cambiare la computazione del 2DFA in una computazione di un 1DFA. Vediamo che stati vengono visitati nel tempo.



Prima di tutto, dobbiamo ricordarci che nei DFA non abbiamo end marker, quindi abbiamo solo l'input. Nell'automa two-way ci sono momenti dove entro nelle celle per la prima volta: nel grafico sopra sono segnati in verde. Chiamiamo questi stati $q_{i \geq 0}$.

Usiamo delle **scorciatoie**: visto che nel 1DFA non possiamo andare avanti a indietro, dobbiamo tagliare via le computazioni che tornano indietro e vedere solo in che stato esco.



Come vediamo, a me interessa sapere in che stato devo spostarmi a partire dalla mia posizione, evitando quello che viene fatto tornando all'indietro. Per tagliare le parti che tornano indietro usiamo delle **matrici**, molto simili a quelle della lezione precedente. Quelle matrici erano nella forma $M_w[p, q]$ che conteneva un 1 se e solo se partendo da q finivo in p leggendo w .

Le matrici che costruiamo ora sono nella forma

$$\tau_w : Q \times Q \longrightarrow [0, 1]$$

che mi vanno a definire il primo stato che incontriamo quando leggiamo un nuovo carattere della stringa.

Nella matrice abbiamo $\tau_w[p, q] = 1$ se e solo se esiste una sequenza di mosse che:

- inizia sul simbolo più a destra della porzione di nastro che contiene $(\blacktriangleright w)$ nello stato p ;

- termina quando la testina esce a destra dalla porzione di nastro considerata nello stato q .

Ad esempio, considerando l'esempio sopra, vale

$$\tau_{\text{inp}}[q_2, q_3] = 1.$$

Vediamo come ottenere induttivamente queste tabelle. Partiamo con $w = \varepsilon$: la porzione di nastro che stiamo considerando è formata solo da \blacktriangleright , ma non potendo andare a sinistra l'unica mossa che possiamo fare è andare a destra, quindi andare in un nuovo stato, ovvero

$$\tau_\varepsilon[p, q] = 1 \iff \delta(p, \blacktriangleright) = (q, +1).$$

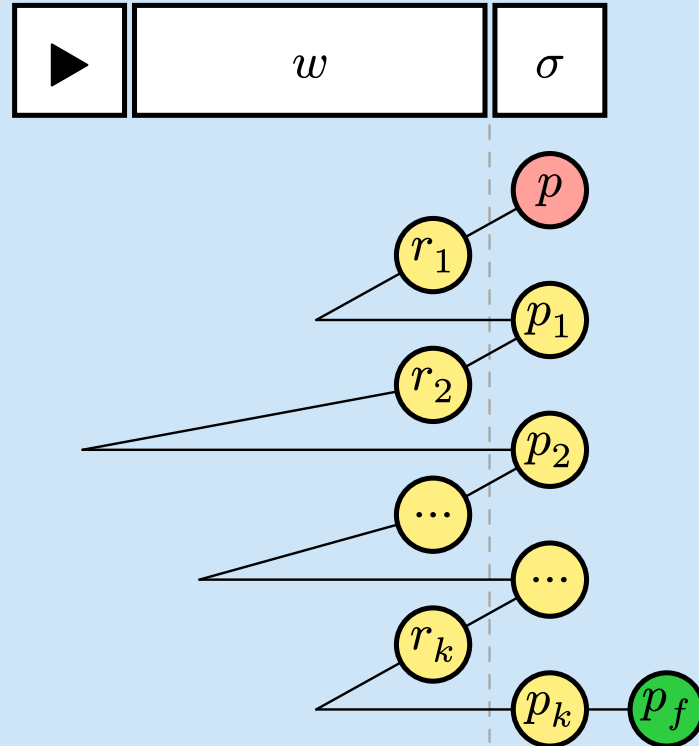
Supponiamo di aver calcolato la tabella di w , vediamo come costruire induttivamente la tabella di $w\sigma$, con $w \in \Sigma^*$ e $\sigma \in \Sigma$. Se vale

$$\delta(p, \sigma) = (q, +1)$$

la tabella è molto facile, perché sto subito uscendo dallo stato p , ovvero

$$\tau_{w\sigma}[p, q] = 1.$$

Se invece andiamo indietro dobbiamo capire cosa fare.



Ogni volta che da p_i torniamo indietro finiamo in uno stato r_{i+1} , che poi dopo un po' di giri finisce per forza in p_{i+1} . Andiamo avanti così, fino ad un certo p_k , dal quale usciamo e andiamo in q . In poche parole

$$\tau_{w\sigma}[p, q] = 1$$

se e solo se esiste una sequenza di stati

$$p_0, p_1, \dots, p_k, r_1, \dots, r_k \mid k \geq 0$$

tale che:

- $p_0 = p$, ovvero parto dallo stato p , per definizione;
- $\delta(p_{i-1}, \sigma) = (r_i, -1) \quad \forall i \in \{1, \dots, k\}$, ovvero in tutti i p tranne l'ultimo io torno indietro;
- $\tau_w[r_i, p_i] = 1$, ovvero da r_i giro in w e poi torno in p_i ;
- $\delta(p_k, \sigma) = (q, +1)$, ovvero esco fuori dal $w\sigma$.

Notiamo che se prendiamo $k = 0$ abbiamo la situazione precedente in cui uscivo direttamente. Inoltre, k è il numero massimo di stati del DFA perché se faccio ancora un giro in w dopo p_k vado in uno stato già visto ed entro in un loop infinito.

Notiamo una cosa **importantissima**: se due stringhe hanno la stessa tabella, ovvero $\tau_w = \tau_{w'}$, allora l'aggiunta di un qualsiasi carattere σ genera tabelle risultanti uguali, ovvero $\tau_{w\sigma} = \tau_{w'\sigma}$. Ma allora esiste una funzione

$$f_\sigma : [Q \times Q \rightarrow [0, 1]] \rightarrow [Q \times Q \rightarrow [0, 1]]$$

che genera una tabella a partire da una data, ed è tale che

$$\forall w \in \Sigma^* \quad \tau_{w\sigma} = f_\sigma(\tau_w).$$

In poche parole, la nuova tabella dipende solo da σ e non da w , e questa tabella è esattamente quella calcolata con i 4 punti messi sopra. Le tabelle, inoltre, sono tantissime ma sono un numero finito.

Siamo pronti per costruire il 1DFA che tanto stiamo bramando. Noi avevamo $M = (Q, \Sigma, \delta, q_0, F)$ che è un 2DFA, vogliamo costruire

$$M' = (Q', \Sigma, \delta', q'_0, F')$$

1DFA che sia equivalente a M . Esso è tale che:

- Q è l'**insieme degli stati** e lo usiamo tenere traccia dello stato nel quale siamo e della tabella che usiamo per calcolare lo stato successivo, ovvero

$$Q' = Q \times [Q \times Q \rightarrow [0, 1]];$$

- q'_0 è lo **stato iniziale** ed è la coppia

$$q'_0 = (q_0, \tau_\epsilon);$$

- δ' è la **funzione di transizione** che manda avanti l'automa, ovvero

$$\delta'((p, T), \sigma) = (q, T')$$

con:

- T' che mi dà indicazioni sullo stato nel quale arrivo con σ , che ho però appena letto, quindi $T' = f_\sigma(T)$;
- vale $T'[p, q] = 1$ perché io devo uscire in q partendo da p ;

- F' è l'**insieme degli stati finali**, ostico perché nel two-way abbiamo gli end marker, nel one-way non li abbiamo. Per accettare dovevo sfiorare l'end marker di destra e finire in q_f , ma questa informazione la ricavo dalla tabella del right marker, ovvero

$$F' = \{(q, T) \mid (f_{\blacktriangleleft}(T))[q, q_f] = 1\}.$$

Ma allora stiamo simulando un 2DFA con un 1DFA, ma gli 1DFA riconoscono la classe dei linguaggi regolari, quindi anche la classe degli automi a stati finiti two-way riconosce la classe dei linguaggi regolari. ■

Che considerazioni possiamo fare sul numero di stati? Sappiamo che:

- il numero di stati è $|Q| = n$;
- il numero di tabelle è $|[Q \times Q] \rightarrow [0, 1]| = 2^{n^2}$.

Ma allora il numero di stati è

$$|Q'| \leq n2^{n^2}.$$

Come vediamo, la simulazione è **poli-esponenziale**.

3. Lezione 14 [11/04]

3.1. Simulazione

La scorsa lezione abbiamo visto gli automi two-way e abbiamo dimostrato che hanno la stessa potenza computazionale degli automi a stati finiti. Avevamo visto la trasformazione da 2DFA a 1DFA, ma la stessa trasformazione può essere fatta per il passaggio da 2NFA a 1NFA.

Ma quanto costano queste trasformazioni?

Nel caso partissimo da un 2DFA e volessimo arrivare in un 1DFA, il costo in termini di stati è

$$\leq \dots,$$

mentre cambiando il punto di partenza con un 2NFA il salto diventa ancora peggiore:

$$\leq 2^n 2^{n^2} = 2^{n^2+n}.$$

Ma questo ce lo potevamo aspettare: abbiamo già un salto esponenziale da NFA a DFA, quindi ciao.

Ci sono due simulazioni che sono però molto particolari e importanti.

3.1.1. Problema di Sakoda & Sipser

La prima trasformazione che vediamo è quella da 2NFA a 2DFA: qua non possiamo usare la costruzione per sottoinsiemi perché ad un certo punto potresti avere il non determinismo su una mossa che però ti sposta la testina su due caratteri diversi della stringa, e questo non è possibile. Ci serve quindi una trasformazione alternativa, ma ci arriviamo dopo.

La seconda trasformazione è quella da 1NFA a 2DFA: questa trasformazione cerca di capire se, dando il two-way ad un automa deterministico, esso è capace di simulare il non determinismo.

Vediamo un paio di esempi.

Esempio 3.1.1.1: Definiamo

$$L_n = (a + b)^* a (a + b)^{n-1}$$

il classicissimo linguaggio dell' n -esimo carattere da destra pari ad una a .

Sappiamo che:

- esiste un 1NFA di $n + 1$ stati;
- esiste un 1DFA di 2^n stati.

Abbiamo visto un automa two-way per questo linguaggio, che usa poco più di n stati, quindi in questo caso riusciamo a togliere il non determinismo a basso costo.

Esempio 3.1.1.2: Definiamo

$$K_n = (a + b)^* a (a + b)^{n-1} a (a + b)^*$$

il solito linguaggio con due a a distanza n .

Avevamo visto che un 1NFA per questo linguaggio usava $n + 2$ stati, quindi una quantità lineare in n . Per un 2DFA abbiamo visto che esiste anche qui una soluzione lineare in n , quindi anche qui eliminiamo il non determinismo a basso costo.

Abbiamo visto due esempi che sembrano dare buone notizie, ma riusciamo a dimostrare che si riesce sempre a fare un 2DFA di n stati partendo da un 1NFA di n stati? Purtroppo, nessuno ci è mai riuscito.

Questi problemi sono i **problemi di Sakoda & Sipser**, ideati nel 1978 e che riguardano il costo della simulazioni di automi non deterministici one-way e two-way per mezzo di automi two-way deterministici, ovvero si chiedono se il movimento two-way aiuta nell'eliminazione del non determinismo.

Cosa sappiamo su questi problemi? Diamo qualche **upper** e **lower bound**.

Per il problema da 1NFA a 2DFA, si sfrutta la **costruzione per sottoinsiemi** per ottenere un 1DFA, che è anche un 2DFA che non torna mai indietro, ottenendo quindi un numero di stati

$$\leq 2^n.$$

Un lower bound per questo problema invece è

$$\geq n^2.$$

Per il problema da 2NFA a 2DFA, si fa un passaggio intermedio all'1NFA e poi al 1DFA, che come prima è anche 2DFA, quindi gli stati sono

$$\leq 2^{n^2+n}.$$

Il lower bound, invece, è lo stesso del problema precedente.

Ci sono casi particolari che hanno delle dimostrazioni precise:

- se utilizziamo dei 2DFA sweeping il costo per la trasformazione è **esponenziale**, ma questo non risolve il problema perché (???) ci sono automi non sweeping che per diventarlo hanno un salto esponenziale (???)
- se $\Sigma = \{a\}$:
 - ▶ se facciamo la trasformazione da 2NFA a 2DFA l'upper bound è

$$e^{O(\log^2(n))},$$

ovvero una funzione super polinomiale ma meno di una esponenziale. Inoltre, se si dimostra che esiste un lower bound super polinomiale, allora abbiamo dimostrato che

$$L = NL \text{ (???)};$$

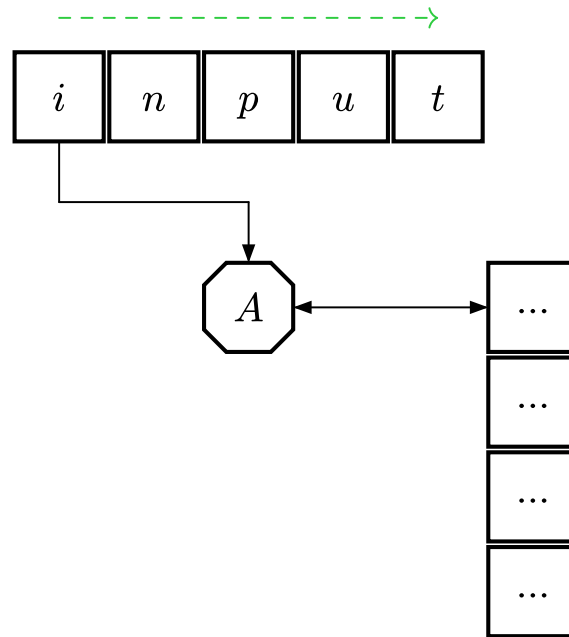
- ▶ se facciamo la trasformazione da 1NFA a 2DFA l'upper bound diventa esattamente n^2 , quindi la trasformazione fatta è ottimale.

Dei ricercatori hanno trovato degli **automi completi** per questi problemi, ovvero degli automi che permettono lo studio dei problemi solo su questi pochi automi scelti per poi far «arrivare» tutte le conseguenze a tutti gli altri automi. Scritto malissimo, sono tipo gli NP-completi.

Per ora, la **congettura** che circola tra la gente è che i costi siano **esponenziali nel caso peggiore**.

3.2. Automi a pila

Lasciamo finalmente stare gli automi a stati finiti per passare ad una nuova classe di riconoscitori: gli **automi a pila**. Essi sono praticamente degli automi a stati finiti con testina di lettura one-way ai quali viene aggiunta una **memoria infinita con restrizioni di accesso**, ovvero l'accesso avviene solo sulla cima della memoria, con politica LIFO.



Come vediamo, la parte degli automi a stati finiti ce l'abbiamo ancora, ma ora abbiamo una **memoria esterna**, che nell'immagine è sulla destra, che possiamo utilizzare con una politica di accesso LIFO. Per via di questa politica, questi automi sono anche detti **automi pushdown**, o **PDA**, perché quando inserisci qualcosa lo fai a spingere giù.

3.2.1. Versione non deterministica

Vediamo subito la definizione formale **non deterministica** dei PDA.

Sia M un PDA definito dalla tupla

$$(Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$$

tale che:

- Q è un **insieme finito non vuoto di stati**, che rappresenta il controllo a stati finiti;
- Σ è un **alfabeto finito non vuoto di input**;
- Γ è un **alfabeto finito non vuoto di simboli della pila**;
- δ è la **funzione di transizione**;
- $q_0 \in Q$ è lo **stato iniziale**;
- $Z_0 \in \Gamma$ è il **simbolo iniziale sulla pila**;
- $F \subseteq Q$ è un **insieme di stati finali**.

La **funzione di transizione** è definita come segue:

$$\delta : Q \times (\Sigma \cup \{\varepsilon\}) \times \Gamma \longrightarrow 2^{Q \times \Gamma^*}.$$

In poche parole, consideriamo lo **stato corrente**, il **simbolo sulla testina** o una ε -mossa e il **simbolo sulla cima della pila** per capire in che stato dobbiamo muoverci e che stringa andare ad inserire sulla pila. La lettura del carattere in cima alla pila lo va a **distruggere**.

Questa versione però non ci piace molto perché Γ^* è potenzialmente un **insieme infinito**, e non ci piace avere un insieme infinito di possibilità, quindi sostituiamo la definizione della funzione di transizione con questa analoga, ma molto migliore per noi:

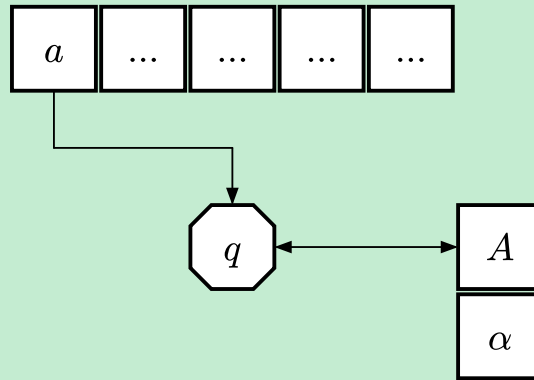
$$\delta : Q \times (\Sigma \cup \{\varepsilon\}) \times \Gamma \longrightarrow \text{PF}(Q \times \Gamma^*).$$

Con PF intendiamo l'**insieme delle parti finite**, ovvero un insieme finito di possibilità prese dall'insieme delle parti. Ora sì che la definizione ci piace.

Facciamo qualche esempio. Come convenzione useremo le **maiuscole** per i simboli della pila.

Esempio 3.2.1.1: Facciamo che la funzione di transizione sia definita in questo modo:

$$\delta(q, a, A) = \{(q_1, \varepsilon), (q_2, BCC)\}.$$

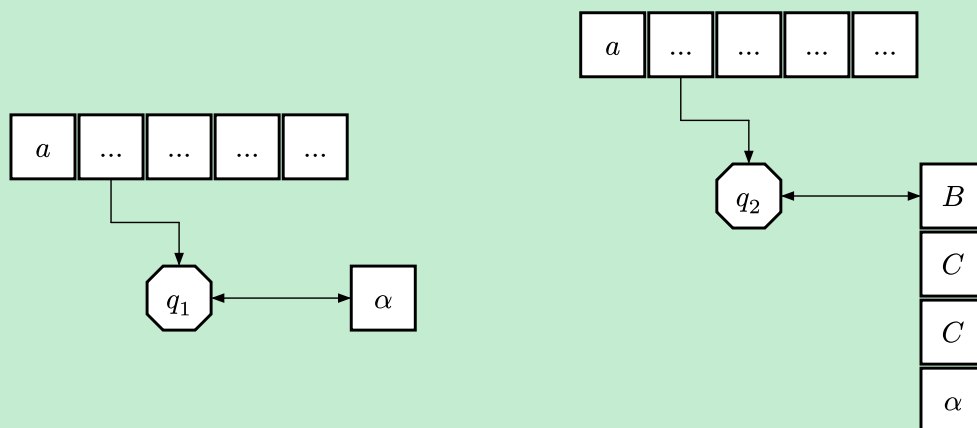


Con α nel disegno si intende una stringa in Γ^* perché oltre ad A potremmo avere altro.

Cosa vuol dire quella regola della funzione di transizione? Ci sta dicendo che se ci troviamo nello stato q , leggiamo a sul nastro leggiamo A sulla cima della pila, possiamo:

- andare in q_1 e non mettere altro sulla pila, praticamente consumando un simbolo in input;
- andare in q_2 e mettere sulla pila la stringa BCC .

Vediamo la rappresentazione dei due casi nei quali possiamo finire.

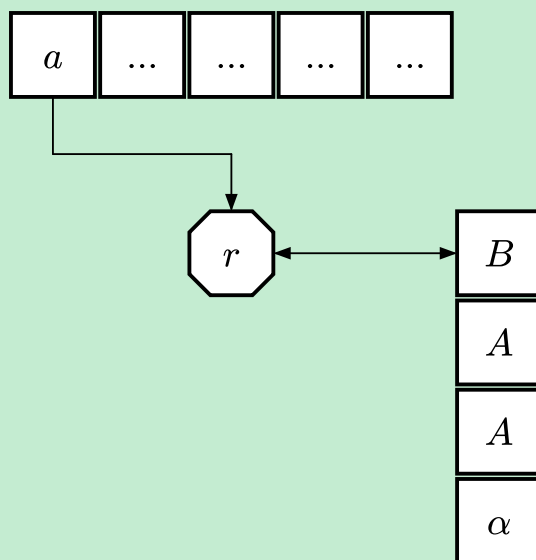


Per convenzione, quando inseriamo una stringa sulla pila, l'inserimento avviene da destra verso sinistra. In poche parole, se inseriamo la stringa $X \in \Gamma^*$ nella pila, se la togliessimo noi leggeremmo, in ordine, esattamente X . In altre parole ancora, quando leggiamo una stringa da inserire è come se la stessimo leggendo dall'alto verso il basso.

Abbiamo la possibilità anche di fare delle ε -mosse: supponiamo di aggiungere la regola

$$\delta(q, \varepsilon, A) = \{(r, BAA)\}.$$

Ora abbiamo tre scelte a disposizione. Le ε -mosse possiamo vederle come delle **mosse interne**, che avvengono senza leggere l'input, e che ci permettono di spostarci negli stati modificando eventualmente la pila.



Una **configurazione** è una fotografia dell'automa in un dato istante di tempo, e ci dice quali sono le informazioni rilevanti per il futuro per definire al meglio la macchina, ovvero:

- lo **stato corrente**;
- il **contenuto del nastro** che ci manca da leggere;
- il **contenuto della pila**.

Una configurazione è quindi una **tripla**

$$(q, ay, A\alpha)$$

che contiene lo stato corrente, il contenuto del nastro ancora da leggere indicato dal carattere corrente a unito al resto della stringa y e il contenuto della pila indicato dal carattere in testa A e dal resto della pila α .

Una **mossa** è l'applicazione della funzione di transizione, ovvero un passaggio

$$(q, ay, A\alpha) \longrightarrow (p, y, \gamma\alpha) \iff (p, \gamma) \in \delta(q, a, A).$$

Analogamente, un passaggio che usa le ε -mosse è un passaggio

$$(q, ay, A\alpha) \longrightarrow (p, ay, \gamma\alpha) \longrightarrow (p, \gamma) \in \delta(q, \varepsilon, A).$$

Una **computazione** è una serie di mosse che partono da una configurazione iniziale e mi portano in una configurazione finale. Di queste ultime parleremo tra poco. Torniamo sulle computazioni.

Come con i passi di derivazione, una computazione che usa una sola mossa si indica con

$$C' \vdash C''.$$

Se invece una computazione impiega k passi, si indica con

$$C' \vdash^k C''.$$

Infine, per indicare una computazione con un numero generico di passi, maggiori o uguali a zero, si usa

$$C' \vdash^* C''.$$

3.2.2. Accettazione

Abbiamo parlato di arrivare in una configurazione accettante, ma quando **accettiamo**? Dobbiamo capire da dove partire e dove arrivare.

Quando partiamo abbiamo la stringa w sul nastro, ci troviamo nello stato iniziale q_0 e abbiamo Z_0 sulla pila: questa è detta **configurazione iniziale** ed è la tripla

$$(q_0, w, Z_0).$$

Le **configurazioni finali** dipendono dal tipo di nozione di accettazione che vogliamo utilizzare.

L'**accettazione per stati finali** ci obbliga a leggere tutto l'input e a finire in uno stato finale, con la pila che contiene quello che vuole, ovvero dobbiamo arrivare in una configurazione

$$(q, \varepsilon, \gamma)$$

dove lo stato q è finale. Il **linguaggio accettato per stati finali** è l'insieme

$$L(M) = \left\{ w \in \Sigma^* \mid (q_0, w, Z_0) \vdash^* (q, \varepsilon, \gamma) \mid q \in F \wedge \gamma \in \Gamma^* \right\}.$$

Questa nozione è comoda perché vede i PDA come una **estensione** degli automi a stati finiti.

L'**accettazione per pila vuota** invece è una nozione più naturale: tutto ciò che metto nella pila lo devo anche buttare via. Possiamo arrivare in un qualsiasi stato, basta aver svuotato la pila. Il **linguaggio accettato per pila vuota** è l'insieme

$$N(M) = \{w \in \Sigma^* \mid (q_0, w, Z_0) \vdash (q, \varepsilon, \varepsilon) \mid q \in Q\}.$$

Se svuotiamo la pila prima di finire l'input allora quella computazione si blocca perché noi dobbiamo sempre leggere qualcosa dalla pila.

Le due accettazioni sono **equivalenti**, o meglio, possiamo passare da una accettazione all'altra ma i linguaggi che accettano sono differenti. A parità di automa M , gli insiemi $L(M)$ e $N(M)$ in generale sono diversi, ma possiamo passare da un modello all'altro mantenendo il linguaggio accettato con facilità. Una ulteriore nozione di accettazione unisce stati finali e pila vuota, ma rimane comunque equivalente. Avere due nozioni è comodo: se una versione ci esce estremamente comoda allora la andiamo ad utilizzare, altrimenti andremo ad utilizzare l'altra.

Vediamo ora qualche esempio.

Esempio 3.2.2.1: Prendiamo il nostro migliore amico, il linguaggio

$$L = \{a^n b^n \mid n \geq 1\}.$$

Lo possiamo riconoscere con un PDA, visto che abbiamo visto che non è regolare? Bhe sì: con i DFA non riusciamo a ricordare il numero di a e poi confrontare questo numero con le b , mentre ora riusciamo a farlo, le pile sanno contare.

Possiamo pensare ad un automa che ogni volta che legge una a butta una A dentro la pila, e quando legge una b toglie una A dalla pila. Accettiamo se abbiamo messo n caratteri A dentro la pila e poi ne abbiamo tolti n , quindi qua viene comoda l'**accettazione per pila vuota**.

Andiamo a definire la funzione di transizione.

Iniziamo a togliere Z_0 dalla prima e inseriamo la prima A in segno di aver letto la prima a della stringa, che abbiamo per forza per definizione, quindi

$$\delta(q_0, a, Z_0) = \{(q_0, A)\}.$$

Utilizziamo lo stato q_0 per leggere tutte le a della stringa, ovvero

$$\delta(q_0, a, A) = \{(q_0, AA)\}.$$

Appena troviamo una b iniziamo a cancellare e cambiamo stato, visto che non ci aspettiamo più delle a nella stringa, quindi

$$\delta(q_0, b, A) = \{(q_1, \varepsilon)\}.$$

Inseriamo ε sulla stringa perché la A da cancellare per la lettura di b è già stata cancellata dalla lettura.

Andiamo a terminare la lettura delle b , quindi

$$\delta(q_1, b, A) = \{(q_1, \varepsilon)\}.$$

Abbiamo detto che accettiamo per pila vuota, quindi $L = N(M)$.

Esempio 3.2.2.2: Se invece volessimo accettare il linguaggio precedente per **stati finali**?

Non dobbiamo cancellare Z_0 dalla pila perché se una stringa viene accettata cancella tutta la pila, quindi ci serve un carattere fittizio dentro per poterlo leggere e spostarci in uno stato finale. Modifichiamo quindi la mossa iniziale con la mossa

$$\delta(q_0, a, Z_0) = \delta(q_0, AZ_0).$$

Se abbiamo una stringa del linguaggio alla fine delle b dobbiamo spostarci in uno stato finale, quindi aggiungiamo la regola

$$\delta(q_1, \varepsilon, Z_0) = \{(q_f, Z_0)\} \mid q_f \in F.$$

Con queste modifiche abbiamo $L = L(M)$.

Esempio 3.2.2.3: Se invece volessimo accettare anche ε ? Il linguaggio diventa

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

Con l'**accettazione per pila vuota**, nello stato iniziale possiamo aggiungere una regola che svuota subito la pila, ovvero aggiungiamo la regola

$$\delta(q_0, \varepsilon, Z_0) = \{(q_0, \varepsilon)\}.$$

Stiamo scommettendo che l'input è già finito, ovvero abbiamo solo ε sul nastro, ma questo ha appena aggiunto il **non determinismo** al nostro automa a pila.

Con l'**accettazione per stati finali** invece ci spostiamo direttamente nello stato q_f a partire da q_0 , ovvero aggiungiamo la regola

$$\delta(q_0, \varepsilon, Z_0) = \{(q_f, \varepsilon)\}.$$

Come prima, abbiamo aggiunto del **non determinismo** all'automa a pila, ma questo lo possiamo togliere: come facciamo a fare ciò?

Introduciamo uno stato q_I finale che diventa anche iniziale al posto di q_0 , quindi ora

$$F = \{q_I, q_f\}.$$

Se inseriamo sul nastro la stringa vuota allora noi accettiamo, perché siamo in uno stato finale e non abbiamo altri simboli da leggere. Per passare poi al vecchio automa mettiamo una regola

$$\delta(q_I, a, Z_0) = \{(q_0, AZ_0)\}.$$

3.2.3. Determinismo VS non determinismo

Con il termine **non determinismo** non intendiamo le ε -mosse da sole, quelle le possiamo avere, ma intendiamo un mix tra mosse che leggono e mosse che non leggono.

Definizione 3.2.3.1 (Determinismo): Sia M un PDA. Allora M è **deterministico** se:

1. ogni volta che ho una ε -mossa da un certo stato e con un certo simbolo sulla pila, non ho mosse che leggono simboli dal nastro a partire dallo stesso stato e con lo stesso simbolo sulla pila, ovvero

$$\forall q \in Q \quad \forall A \in \Gamma \quad \delta(q, \varepsilon, A) \neq \emptyset \implies \forall a \in \Sigma \quad \delta(q, a, A) = \emptyset;$$

2. come nel caso classico, considero un carattere, o anche ε , allora a parità di stato corrente e simbolo sulla pila, ho al massimo una transizione possibile, ovvero

$$\forall q \in Q \quad \forall A \in \Gamma \quad \forall \sigma \in \Sigma \cup \{\varepsilon\} \quad |\delta(q, \sigma, A)| \leq 1.$$

A differenza del caso classico, il determinismo e il non determinismo non sono ugualmente potenti: un automa a pila non deterministico è **più potente** di un automa a pila deterministico, che riconosce una sottoclasse di linguaggi diversa dai linguaggi di tipo 2, che sono riconosciuti dai PDA non deterministici.

3.2.4. Trasformazioni

Avevamo parlato dell'**equivalenza** dell'accettazione per stati finali e per pila vuota: infatti, esistono due trasformazioni che permettono di passare da un automa all'altro, mantenendo il linguaggio di partenza riconosciuto inalterato. L'equivalenza infatti ci diceva che, partendo da un automa M che riconosce per stati finali, abbiamo una trasformazione che ci dà M' che riconosce per pila vuota che riconosce lo stesso linguaggio di M , e viceversa.

Stati finali \rightarrow pila vuota Dobbiamo trasformare un automa che accetta per stati finali in un automa che accetta per pila vuota. Con quest'ultimo simuliamo il primo, e ogni volta che vado in uno stato finale mi sposto in uno **stato di svuotamento**, che se raggiunto in mezzo blocca la pila, ma se raggiunto alla fine mi fa accettare.

Abbiamo quindi $M = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$ un PDA con $L = L(M)$. Definiamo ora

$$M' = (Q \cup \{q_e, q'_0\}, \Sigma, \Gamma \cup \{X\}, \delta', q'_0, X, \emptyset)$$

un PDA tale che:

- per metterci in una situazione piacevole per la fine usiamo un **truccaccio** definito dalla regola

$$\delta(q'_0, \varepsilon, X) = \{(q_0, Z_0 X)\},$$

ovvero prima di far partire la computazione dell'automata M andiamo ad inserire un carattere X in fondo alla pila, vedremo dopo perché;

- l'automata deve eseguire **tutte le mosse** di M , ovvero

$$\forall q \in Q \quad \forall \sigma \in \Sigma \cup \{\varepsilon\} \quad \forall Z \in \Gamma \quad \delta(q, \sigma, Z) \subseteq \delta'(q, \sigma, Z),$$

che scritto così significa che tutte le mosse che trovavamo nell'applicazione di delta ad una certa tripla le abbiamo anche nella nuova funzione di transizione, che però conterrà anche altro, che vedremo tra poco;

- aggiungiamo uno **stato di svuotamento** per pulire la pila, definito dalle regole

$$\begin{aligned}\forall q \in F \quad \forall Z \in \Gamma \cup \{X\} \quad (q_e, \varepsilon) &\in \delta'(q, \varepsilon, Z) \\ \forall Z \in \Gamma \cup \{X\} \quad \delta'(q_e, \varepsilon, Z) &= \{(q_e, \varepsilon)\},\end{aligned}$$

ovvero con la prima regola, ogni volta che mi trovo in uno stato finale **non deterministicamente** mi posso spostare nello stato di svuotamento, mentre con la seconda regola effettivamente svuoto.

A cosa ci serve il **carattere** X ? Facciamo finta di non mettere il carattere X . Se M accetta una stringa x arrivando con la pila vuota nessun problema, non ci spostiamo nello stato di svuotamento ma abbiamo la pila vuota quindi ottimo. Se invece M non accetta una stringa x ma arriva alla fine con la pila vuota, il simbolo X messo all'inizio ci copre da una eventuale accettazione errata, perché non riusciremo ad andare nello stato di svuotamento per avere la pila vuota, anche se M ci finisce in quel modo. Diciamo che abbiamo messo X come se fosse una guardia, che ci copre questo preciso caso.

Purtroppo, con questa costruzione abbiamo buttato dentro del **non determinismo** quando facciamo i passaggi in q_e da uno stato finale.

Pila vuota \rightarrow stati finali Il percorso opposto invece parte da un PDA

$$M = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$$

tale che $L = N(M)$. Definiamo il PDA

$$M' = (Q \cup \{q'_0, q_f\}, \Sigma, \Gamma \cup \{X\}, \delta', q'_0, X, \{q_f\})$$

che come idea ha quella di simulare M e, ogni volta che arriva con pila vuota, ci spostiamo nello stato finale. Vediamo i vari passi:

- come prima, usiamo un **truccaccio** per infilare X sotto la pila, quindi abbiamo la regola

$$\delta'(q'_0, \varepsilon, Z_0) = \{(q_0, Z_0 X)\}$$

che usiamo per inserire X come trigger per andare in uno stato finale;

- simuliamo l'automa M senza aggiungere niente, quindi

$$\forall q \in Q \quad \forall \sigma \in \Sigma \cup \{\varepsilon\} \quad \forall Z \in \Gamma \quad \delta'(q, \sigma, Z) = \delta(q, \sigma, Z);$$

- ogni volta che leggiamo X sulla cima della pila vuol dire che M ha svuotato la pila, quindi devo andare nello stato finale, ovvero

$$\forall q \in Q \quad \delta'(q, \varepsilon, X) = \{(q_f, \varepsilon)\};$$

ovviamente, se andiamo in questo stato a metà stringa ci blocchiamo, altrimenti se ci andiamo alla fine è tutto ok.

A differenza di prima, se partiamo da un automa **deterministico**, quello che otteniamo è ancora un automa **deterministico**.

4. Lezione 15 [23/04]

Vediamo degli esempi di qualche linguaggio che possiamo riconoscere con degli automi a pila.

4.1. Esempi

Esempio 4.1.1: Definiamo il linguaggio

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

Un automa a pila per questo linguaggio memorizza w sulla pila, legge $\#$ e poi verifica che la stringa w^R sia presente sulla pila.

Possiamo usare due stati:

- q_0 lo usiamo per copiare w sulla pila;
- q_1 lo usiamo per confrontare il carattere sulla pila con quello sul nastro.

In questo caso ci viene naturale accettare per pila vuota. Inoltre, otteniamo un automa deterministico, detto anche **DPDA**.

Un linguaggio riconosciuto da automi a pila deterministici DPDA fa parte dell'insieme dei **linguaggi context-free deterministici**, detti anche **DCFL**.

Esempio 4.1.2: Definiamo ora il linguaggio

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

insieme delle stringhe palindrome di lunghezza pari.

In questo caso non riusciamo a farlo con un DPDA (difficile da dimostrare, lo faremo avanti) perché dobbiamo scommettere di essere arrivati a metà della stringa da riconoscere, quindi dobbiamo usare del **non determinismo**.

Analogamente, un linguaggio riconosciuto da automi a pila non deterministici, detti anche **NPDA** o solo **PDA**, fa parte dell'insieme dei **linguaggi context-free**, detti anche **CFL**.

4.2. Equivalenza tra grammatiche di tipo 2 e automi a pila

Facciamo un breve ripasso sulle grammatiche di tipo 2 e poi andiamo a vedere l'equivalenza tra le grammatiche di tipo 2 e gli automi a pila.

4.2.1. Ripasso e introduzione

Una grammatica G di tipo 2 ha le **regole di produzione** nella forma

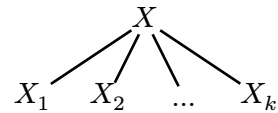
$$X \longrightarrow X_1 \dots X_k \quad | \quad X \in V \wedge X_1, \dots, X_k \in (V \cup \Sigma) \wedge k \geq 0.$$

Abbiamo modificato leggermente la forma ma il succo è quello: ad ogni variabile associamo una sequenza (anche vuota) di terminali e non terminali.

Il processo di derivazione nelle grammatiche di tipo 2 può essere espresso mediante **alberi di derivazione**: essi sono alberi che visualizzano l'applicazione delle regole di produzione.

In un albero generico, la **radice** e i **nodi interni** sono **variabili**, mentre le foglie sono combinazioni di **variabili e terminali**. In un albero che rappresenta una stringa del linguaggio nella radice ho l'**assioma** S e nelle foglie ho **solo terminali**.

Un nodo, con i suoi figli diretti, va a rappresentare l'**applicazione** di una regola di produzione. Ad esempio, prendendo la regola di produzione generica di una grammatica di tipo 2, abbiamo il seguente **albero di derivazione**:

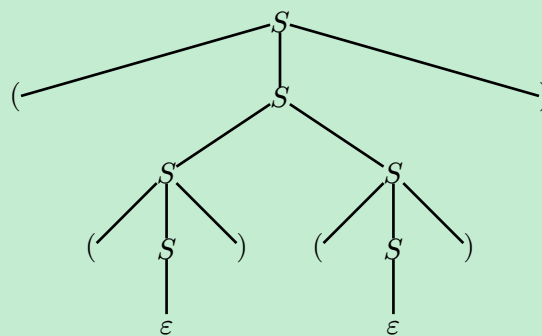


Vediamo questi alberi applicati a qualche esempio.

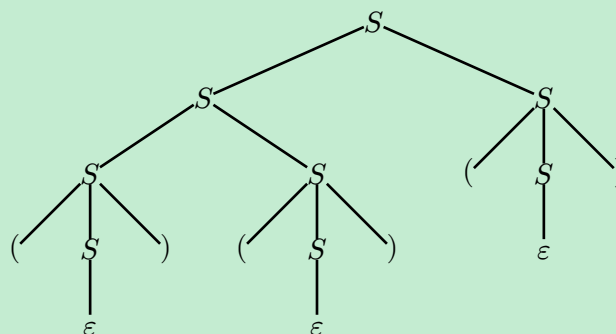
Esempio 4.2.1.1: Riprendiamo la grammatica per le parentesi tonde bilanciate, che avevamo fatto a inizio corso, con le seguenti regole di produzione:

$$S \rightarrow \varepsilon \mid (S) \mid SS.$$

Vediamo due alberi di derivazione che abbiamo in questo linguaggio.



Questo primo albero rappresenta la derivazione della stringa $w = (()())$.



Questo secondo albero genera invece la stringa $w = ()()()$.

Per leggere la stringa che viene generata basta una **visita in profondità** dell'albero.

Gli alberi sono un **modo compatto** di descrivere un processo di derivazione.

Esempio 4.2.1.2: Rimaniamo con il linguaggio dell'ultimo esempio, e prendiamo in considerazione l'ultima stringa $()()()$ che abbiamo derivato.

Proviamo a scrivere la derivazione usando proprio le regole di produzione che abbiamo a disposizione. Ci accorgiamo subito che abbiamo diversi modi di arrivare a quella stringa:

$$\begin{aligned} S &\Rightarrow SS \Rightarrow SSS \Rightarrow (S)SS \Rightarrow ()SS \Rightarrow ()(S)S \Rightarrow ()()S \Rightarrow ()()(S) \Rightarrow ()()() \\ S &\Rightarrow SS \Rightarrow SSS \Rightarrow S(S)S \Rightarrow S()S \Rightarrow S()(S) \Rightarrow (S)()(S) \Rightarrow ()()(S) \Rightarrow ()()() \\ S &\Rightarrow SS \Rightarrow (S)S \Rightarrow ()S \Rightarrow ()SS \Rightarrow ()(S)S \Rightarrow ()()S \Rightarrow ()()(S) \Rightarrow ()()(). \end{aligned}$$

In blu viene indicata la variabile S che viene sostituita ad ogni passo.

Queste derivazioni generano la stessa stringa ma hanno alberi di derivazione diversi (non lo disegno, ma vale quanto scritto).

Cerchiamo di dare una corrispondenza tra derivazioni e alberi di derivazione. Andiamo ad utilizzare le **derivazioni leftmost**: esse sono derivazioni in cui, ad ogni passo, la variabile che andiamo a sostituire è quella più a sinistra nella forma sentenziale.

Esempio 4.2.1.3: Nelle tre derivazioni precedenti ci accorgiamo che la prima e la terza derivazione sono leftmost, mentre la seconda non lo è.

Abbiamo quindi creato una corrispondenza 1 : 1 tra derivazioni leftmost e alberi di derivazione. Da questo momento, parleremo di derivazioni leftmost riferendoci ad alberi di derivazione e viceversa.

Definizione 4.2.1.1 (Grammatica ambigua): Una grammatica G è **ambigua** se $\exists w \in L(G)$ che ammette due alberi di derivazione differenti, oppure, in maniera equivalente, se $\exists w \in L(G)$ che ammette due derivazioni leftmost diverse.

Esempio 4.2.1.4: La grammatica delle parentesi tonde bilanciate è **ambigua**, mentre la grammatica delle parole palindrome di lunghezza pari è **non ambigua**.

La pila è la struttura che ci permette di implementare la **ricorsione**. I linguaggi CFL hanno in più, rispetto ai regolari, l'accesso alle strutture ricorsive, e questo lo vediamo negli esempi che abbiamo fatto: l'esempio delle parentesi tonde bilanciate ha come flow

- inizio qualcosa (trovo una tonda aperta);
- vedo se ho ancora qualcosa di bilanciato;
- finisco quel qualcosa (trovo una tonda chiusa).

Tra l'altro, molto figo che **tutti i CFL** si possono ricondurre al **linguaggio delle parentesi bilanciate**, molto molto bello.

4.2.2. Da grammatica di tipo 2 ad automa a pila

Abbiamo a disposizione una grammatica G di tipo 2, detta anche **CFG**. Vogliamo costruire un automa a pila M che simuli il processo di derivazione tramite derivazioni leftmost.

Sia quindi

$$G = (V, \Sigma, P, S)$$

una CFG. Costruiamo

$$M = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, \mathbb{Q})$$

un PDA che accetterà **per pila vuota** tale che:

- l'insieme degli stati Q contiene un solo stato, ovvero

$$Q = \{q\};$$

- lo stato iniziale, vista la presenza di un solo stato, è

$$q_0 = q;$$

- l'alfabeto di lavoro della pila Γ contiene tutto l'alfabeto della grammatica, ovvero l'insieme di tutti i simboli di G , ovvero

$$\Gamma = V \cup \Sigma;$$

- il simbolo iniziale della pila Z_0 è l'assioma della grammatica, ovvero

$$Z_0 = S;$$

useremo la pila per metterci sopra quello che vogliamo espandere mano a mano;

- infine, la **funzione di transizione** è **non deterministica** e ha due regole:
 - ogni volta che ho una variabile sulla cima della pila, con una epsilon mossa per sostituirla con il lato destro di una produzione, ovvero

$$\forall A \in V \quad \delta(q, \varepsilon, A) = \{(q, \alpha) \mid (A \rightarrow \alpha) \in P\};$$

- ogni volta che ho un simbolo terminale sulla cima della pila, andiamo a leggere dal nastro e verifichiamo che i due valori siano uguali, ovvero

$$\forall a \in \Sigma \quad \delta(q, a, a) = \{(q, \varepsilon)\}.$$

Lemma 4.2.2.1: Vale

$$L(G) = N(M).$$

Dimostriamo questo con un esempio.

Esempio 4.2.2.1: Definiamo la grammatica $G = (V, \Sigma, P, S)$ tale che:

$$V = \{S, T, U\}$$

$$\Sigma = \{a, b\}$$

$$S \rightarrow TU$$

$$T \rightarrow aTb \mid \varepsilon$$

$$U \rightarrow bUa \mid \varepsilon$$

Questa grammatica genera il linguaggio

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

Andiamo a scrivere un PDA M per questa grammatica. Partiamo con le regole del primo tipo:

$$\begin{aligned}\delta(q, \varepsilon, S) &= \{(q, TU)\} \\ \delta(q, \varepsilon, T) &= \{(q, aTb), (q, \varepsilon)\} \\ \delta(q, \varepsilon, U) &= \{(q, bUa), (q, \varepsilon)\}.\end{aligned}$$

E terminiamo con le regole del secondo tipo:

$$\begin{aligned}\delta(q, a, a) &= \{(q, \varepsilon)\} \\ \delta(q, b, b) &= \{(q, \varepsilon)\}.\end{aligned}$$

Simuliamo l'automa a pila che abbiamo costruito e il processo di derivazione con la stringa

$$w = abbbbaa.$$

Ricordiamoci di usare una **derivazione leftmost**: questo è comodo perché i simboli che scriviamo più in alto sono quelli più a sinistra nella stringa aggiunta, e noi facciamo le sostituzioni proprio a partire da sinistra, quindi ottimo.

Nella colonna della **testina**, in blu indichiamo il carattere che la testina può leggere, mentre nella colonna della **derivazione**, sempre in blu indichiamo i caratteri che sono già stati verificati. Di quest'ultimo fatto ne parleremo meglio dopo.

Pila	Testina	Derivazione	Spiegazione della mossa
S	$abbbbaa$	S	Configurazione iniziale
T U	$abbbbaa$	TU	L'unica sostituzione che posso fare per S la faccio. Inoltre, non sposto la testina
a T b U	$abbbbaa$	$aTbU$	Avendo a disposizione il non determinismo, sono fortunato e scelgo la derivazione corretta e, come prima, non sposto la testina
T b U	$abbbbaa$	$aTbU$	Ora che ho un carattere sulla cima della pila, verifico che sono uguali (lo sono), sposto avanti la testina consumando il carattere della pila e quello del nastro
b U	$abbbbaa$	abU	Avendo a disposizione il non determinismo, sono fortunato e scelgo la derivazione corretta e, come prima, non sposto la testina

U	$abbbaa$	abU	Ora che ho un carattere sulla cima della pila, verifico che sono uguali (lo sono), sposto avanti la testina consumando il carattere della pila e quello del nastro
b U a	$abbbaa$	$abbUa$	Avendo a disposizione il non determinismo, sono fortunato e scelgo la derivazione corretta e, come prima, non sposto la testina
U a	$abbbaa$	$abbUa$	Ora che ho un carattere sulla cima della pila, verifico che sono uguali (lo sono), sposto avanti la testina consumando il carattere della pila e quello del nastro
b U a a	$abbbaa$	$abbbUaa$	Avendo a disposizione il non determinismo, sono fortunato e scelgo la derivazione corretta e, come prima, non sposto la testina
U a a	$abbbaa$	$abbbUaa$	Ora che ho un carattere sulla cima della pila, verifico che sono uguali (lo sono), sposto avanti la testina consumando il carattere della pila e quello del nastro
a a	$abbbaa$	$abbbaa$	Avendo a disposizione il non determinismo, sono fortunato e scelgo la derivazione corretta e, come prima, non sposto la testina
a	$abbbaa$	$abbbaa$	Ora che ho un carattere sulla cima della pila, verifico che sono uguali (lo sono), sposto avanti la testina consumando il carattere della pila e quello del nastro
ε	$abbbaa\varepsilon$	$abbbaa$	Ora che ho un carattere sulla cima della pila, verifico che sono uguali (lo sono), sposto avanti la testina consumando il carattere della pila e quello del nastro

Riprendiamo quello detto prima: possiamo notare che, guardando la derivazione, dalla variabile più a sinistra in poi c'è esattamente quello che troviamo sulla pila nello stesso momento, mentre prima della variabile troviamo la parte dell'input su nastro che abbiamo già controllato.

Le mosse che noi abbiamo etichettato come non deterministiche sono le mosse che avvengono nei **parser**:

- quando facciamo una predizione, ovvero quando cerchiamo di indovinare l'espansione, stiamo facendo una mossa di tipo **predictor**;

- quando controlliamo la predizione fatta, ovvero quando controlliamo le lettere sul nastro e sulla pila, stiamo facendo una mossa di tipo **scanner**.

Siamo partiti quindi da una CFG e abbiamo costruito un PDA **equivalente** che accetta per pila vuota e con un solo stato, tanta tanta roba.

4.2.3. Automa a pila a grammatica di tipo 2

Questo la prossima volta, sium.

4.3. Forme normali per le grammatiche context-free

Abbiamo detto che le produzioni sono nella forma

$$A \longrightarrow \alpha \quad | \quad A \in V \wedge \alpha \in (V \cup \Sigma)^*$$

Possiamo dare diverse forme alle regole di produzione che abbiamo nelle grammatiche di tipo 2, ognuna delle quali ha alcuni punti di forza che possono essere comodi in altri contesti.

Queste forme sono dette **forme normali** e le più conosciute sono la **forma normale di Greibach** e la **forma normale di Chomsky**.

4.3.1. FN di Greibach

Nella **forma normale di Greibach**, spesso abbreviata con **FNG**, le produzioni sono nella forma

$$A \longrightarrow \sigma A_1 A_2 \dots A_k \quad | \quad \sigma \in \Sigma \wedge A_1, A_2, \dots, A_k \in V \wedge k \geq 0.$$

Data una grammatica G qualunque, si può sempre scrivere una grammatica in FN di Greibach per lo stesso linguaggio a meno della parola vuota: infatti, se in G abbiamo la parola vuota, nella sua trasformata non ce l'abbiamo e la dobbiamo aggiungere a mano. In poche parole vale

$$L(\text{FNG}) = L(G) / \{\varepsilon\}.$$

La trasformazione da Greibach ad automa a pila in alcuni casi **elimina il non determinismo**, però da fare è abbastanza pesante e non vedremo come fare.

Con la FN di Greibach possiamo costruire un PDA leggermente più semplice di prima.

Data la grammatica G in FN di Greibach, vogliamo costruire un PDA

$$M = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, \boxtimes)$$

che **accetta per pila vuota** definito da:

- **insieme degli stati** Q formato da un solo stato, ovvero

$$Q = \{q\};$$

- lo **stato iniziale**, vista la presenza di un solo stato, è

$$q_0 = q;$$

- l'**alfabeto di lavoro della pila** Γ contiene tutto solo le variabili della grammatica, questo perché faremo un accorgimento per cancellare i terminali senza metterli sulla pila, quindi

$$\Gamma = V;$$

- il **simbolo iniziale della pila** Z_0 è l'assioma della grammatica, ovvero

$$Z_0 = S;$$

useremo la pila per metterci sopra quello che vogliamo espandere mano a mano;

- la **funzione di transizione** non usa più ε -mosse e, soprattutto, non mette più nella pila i simboli terminali, ovvero

$$\delta(q, \sigma, A) = \{(q, A_1 \dots A_k) \mid (A \rightarrow \sigma A_1 \dots A_k) \in P\},$$

ovvero guardo tutte le regole che iniziano in A e hanno σ come primo carattere e tutto il resto della sostituzione lo andiamo a mettere nella pila. In questo modo, stiamo già consumando σ senza metterlo direttamente sulla pila e lo stiamo già controllando quindi. Ci avanzeranno poi tutte le variabili che sono rimaste sulla pila.

Il simbolo in input mi **aiuta** nella scelta della produzione, e questo può **ridurre il non determinismo**, ma dipende strettamente dalla grammatica che si ha davanti.

Esempio 4.3.1.1: Modifichiamo leggermente la grammatica dell'ultimo esempio usando le seguenti regole di produzione:

$$S \rightarrow TU \qquad T \rightarrow aTb \mid ab \qquad U \rightarrow bUa \mid ba$$

In questo caso non abbiamo più ε quindi il linguaggio diventa

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

Rendiamola in forma normale di Greibach (grazie Pighizzini, io non sono capace). Per fare ciò le produzioni diventano nella forma:

$$\begin{aligned} S &\rightarrow aTBU \mid aBU & T &\rightarrow aB \mid aTB & U &\rightarrow bA \mid bUA \\ A &\rightarrow a & B &\rightarrow b \end{aligned}$$

Come prima, vediamo l'evoluzione della stringa

$$w = abbbaa.$$

Pila	Testina	Derivazione	Spiegazione della mossa
S	$abbbaa$	S	Configurazione iniziale
B U	$abbbaa$	aBU	Avendo a disposizione il non determinismo, sono fortunato e scelgo la derivazione corretta, aiutandomi anche con il carattere che c'era sulla testina
U	$abbbaa$	abU	Come prima
U A	$abbbaa$	$abbUA$	Come prima
A A	$abbbaa$	$abbAA$	Come prima

A	$abbbbaa$	$abbbbaA$	Come prima
ε	$abbbbaa\varepsilon$	$abbbbaa$	Come prima

Ho ancora del non determinismo, ma è molto **ridotto**. In alcuni casi riusciamo addirittura a **toglierlo completamente**, ma dipende molto dalla grammatica.

Abbiamo quindi trovato un PDA che accetta per pila vuota, utilizza un solo stato e non utilizza le ε -mosse. Tutto bello, ma rimane comunque non deterministico.

Questa FN di Greibach nella pratica è **poco usata**: fare il passaggio non è banale e questo potrebbe stravolgere la grammatica iniziale rendendola illeggibile.

5. Lezione 16 [30/04]

5.1. Fine dimostrazione

Per finire la dimostrazione che gli automi a pila sono equivalenti alle grammatiche di tipo 2, ci manca da vedere il passo da automa a grammatica. Per fare ciò, dobbiamo introdurre una **nuova forma normale** per gli automi a pila, per rendere i conti più facili.

Esempio 5.1.1: Definiamo l'alfabeto $\Sigma = \{ (,) \}$ e consideriamo l'alfabeto delle stringhe che rappresentano sequenze di parentesi tonde bilanciate. Come costruiamo un automa a pila per questo linguaggio?

Facilmente, ogni volta che trovo una parentesi aperta metto un simbolo sulla pila, mentre ogni volta che trovo una parentesi chiusa tolgo un simbolo dalla pila, se possibile. Se a fine input arrivo in una configurazione

$$(q, \varepsilon, Z_0)$$

vado ad accettare, visto che tutto quello che ho messo sulla pila l'ho tolto. Andiamo quindi ad accettare «per stati finali», e non per pila vuota, come dice Pighizzini.

Inoltre, la versione «per stati finali» è deterministica, mentre quella per pila vuota non lo è.

Quello che stiamo facendo, in ogni caso, è buttare sulla pila delle robe da controllare dopo: ogni volta che apro devo fare altri lavori e poi andare a chiudere.

5.1.1. Forma normale per gli automi a pila

Vediamo una **forma normale** più semplice. Diamo delle regole:

1. all'inizio la pila contiene solo Z_0 e lo usiamo per marcare il fondo della pila. Questo carattere non è mai rimosso e mai aggiunto;
2. l'input è accettato se si raggiunge una configurazione in cui:
 - tutto l'input è stato letto;
 - la pila contiene solo Z_0 ;
 - lo stato è finale.

In poche parole, tutto ciò che metto sulla pila sono attività che ho lasciato in sospeso, alla fine devo aver terminato tutto. È un po' un mix tra pila vuota (circa) e stati finali;

3. le mosse che facciamo sulla pila sono:
 - **push** di un simbolo (uno alla volta);
 - **pop** del simbolo in cima alla pila;
 - **pila invariata**.

Esempio 5.1.1.1: Avendo delle regole nella forma

$$\delta(q, \sigma, A) = \{ (p, \alpha) \}$$

allora:

- se $\alpha = \varepsilon$ stiamo eseguendo una pop;

- se $\alpha = A$ stiamo facendo una pila invariata;
- se $\alpha = \beta A$ stiamo facendo una pila invariata e $|\beta|$ push;
- se $\alpha = \beta A' \mid A \neq A'$ stiamo facendo una pop, una push e $|\beta|$ push.

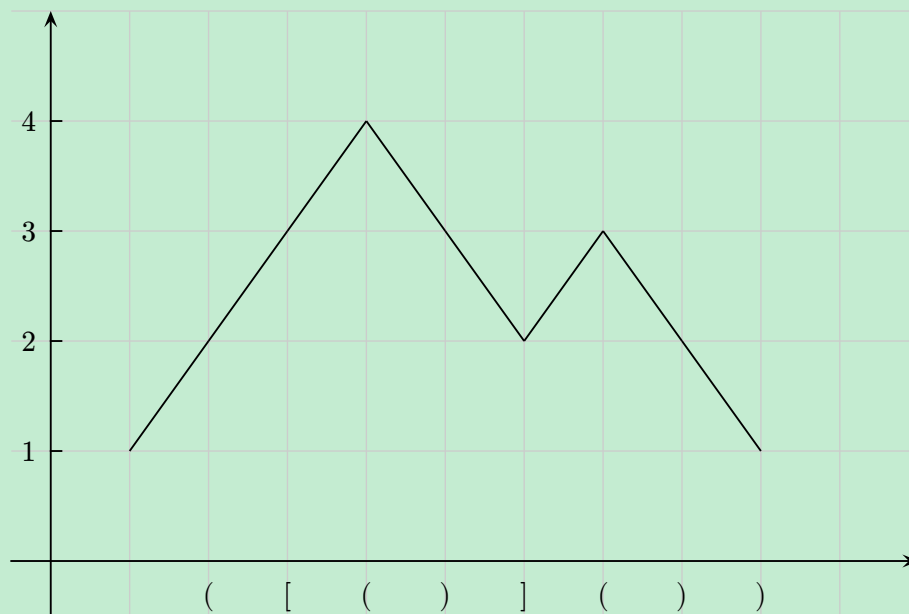
Ci rendiamo conto, visto l'esempio precedente, che **i due modelli sono equivalenti**.

Esempio 5.1.1.2: Definiamo l'alfabeto $\Sigma = \{ (,), [,] \}$. Vogliamo descrivere la computazione accettante della stringa $w = ([()])()$ data in input ad un automa a pila.

Simuliamo la computazione dell'automa su w visualizzando la pila dell'automa:

$$\begin{array}{cccccccc}
 & & & & A & & & \\
 & & B & B & B & & A & \\
 & A & A & A & A & A & A & A & . \\
 Z_0 & Z_0 & Z_0 & Z_0 & Z_0 & Z_0 & Z_0 & Z_0 & Z_0 \\
 \hline
 & (& [& (&) &] & (&) &)
 \end{array}$$

Andiamo a graficare anche l'altezza della pila che abbiamo ottenuto durante la computazione.



Sulle ascisse abbiamo rappresentato l'**input**, che possiamo vedere anche come **tempo** (a meno delle ε -mosse) nel quale l'automa si trova durante lo spostamento della testina da sinistra verso destra. Sulle ordinate invece abbiamo rappresentati l'**evoluzione** della pila.

Come vediamo, tra la prima tonda l'ultima tonda non andiamo mai sotto, idem quando apriamo e chiudiamo la quadra. Infatti, quando apro qualcosa, in mezzo non vado mai sotto e prima di tornare a livello devo riconoscere altre sequenze bilanciate.

In poche parole stiamo facendo una **chiamata ricorsiva**.

Studieremo le computazioni di questo tipo per scrivere la grammatica dall'automa a pila.

Aggiungiamo ancora una regola:

4. quando l'automa **legge** un simbolo di input e muove la testina la pila **non viene modificata**.

La nuova **funzione di transizione** di questi automi è nella forma

$$\delta : Q \times (\Sigma \cup \{\varepsilon\}) \times \Gamma \longrightarrow 2^{Q \times \{-, \text{push}(A) \mid A \in \Gamma, \text{pop}\}},$$

ovvero delle coppie formate da nuovo stato e operazione tra pila invariata, push e pop. Qua, a differenza di prima, non mi servono le parti finite perché ho già sottoinsiemi finiti.

Come sono fatte le regole della funzione di transizione? Dipende se stiamo leggendo o meno, per la nuova regola 4, quindi possiamo avere:

- **mosse di lettura**, che facciamo lasciando inalterata la pila per la regola che abbiamo appena aggiunto, ovvero

$$(p, -) \in \delta(q, a, A) \mid p, q \in Q \wedge a \in \Sigma, \wedge A \in \Gamma;$$

- **mosse pop**, che non leggono niente dal nastro ma liberano la prima posizione sulla pila, ovvero

$$(p, \text{pop}) \in \delta(q, \varepsilon, A);$$

- **mosse push**, che non leggono niente dal nastro ma aggiungono un elemento sulla pila, ovvero

$$(p, \text{push}(B)) \in \delta(q, \varepsilon, A) \mid B \in \Gamma;$$

- **mosse di cambio stato**, che non leggono niente e non modificano la pila, ovvero

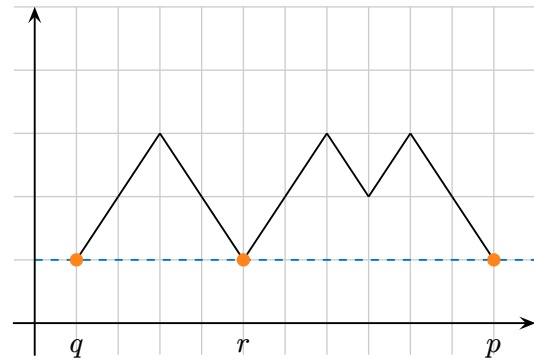
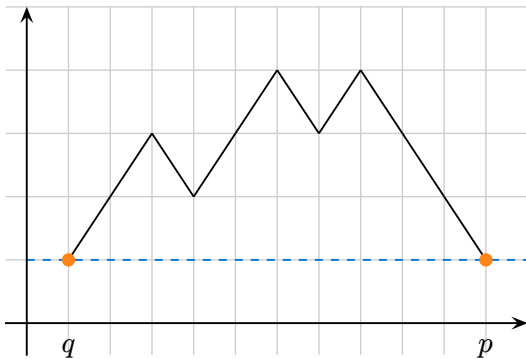
$$(p, -) \in \delta(q, \varepsilon, A).$$

5.1.2. Dimostrazione

Abbiamo detto che con questo nuovo automa noi vogliamo partire da Z_0 sulla pila e finire con lo stesso carattere alla fine della stringa. Abbiamo due casi possibili:

- durante la computazioni saliamo e scendiamo di livello ma raggiungiamo Z_0 alla fine;
- durante la computazione torniamo in Z_0 in almeno un punto.

In generale, possiamo sostituire a Z_0 un qualsiasi carattere che inseriamo sulla pila. Infatti, una volta inserito il carattere A sulla pila nello stato q , noi saliamo e scendiamo e poi usciamo con ancora A sulla pila nello stato p o nello stato r se siamo in uno intermedio.



Definiamo la grammatica $G = (V, \Sigma, P, S)$ formata dalle variabili

$$V = S \cup \{[qAp] \mid q, p \in Q \wedge A \in \Gamma\}.$$

Le variabili, oltre a S , sono delle **triple** che mi indicano lo stato nel quale sono, il simbolo corrente sulla pila e lo stato nel quale arrivo dopo essere tornato nel simbolo che ho trovato sulla pila.

Vediamo le **mosse** che possiamo fare.

Se stiamo **leggendo un simbolo in input**, non modifichiamo la pila, quindi

$$\forall(p, -) \in \delta(q, a, A) \quad [qAp] \rightarrow a.$$

In poche parole, ci stiamo spostando in linea retta, consumando il carattere a sul nastro:

$$q \boxed{A} \xrightarrow{a} \boxed{A} p.$$

Aggiungiamo anche la produzione

$$[qAq] \rightarrow \varepsilon$$

che serve per chiudere delle ricorsioni banali come se fosse un caso base. In poche parole, andiamo a generare solo la **parola vuota**:

$$q \boxed{A} \xrightarrow{\varepsilon} \boxed{A} q.$$

Se invece stiamo facendo la mossa che **cambia stato**, dobbiamo solo spostarci tra gli stati, ovvero

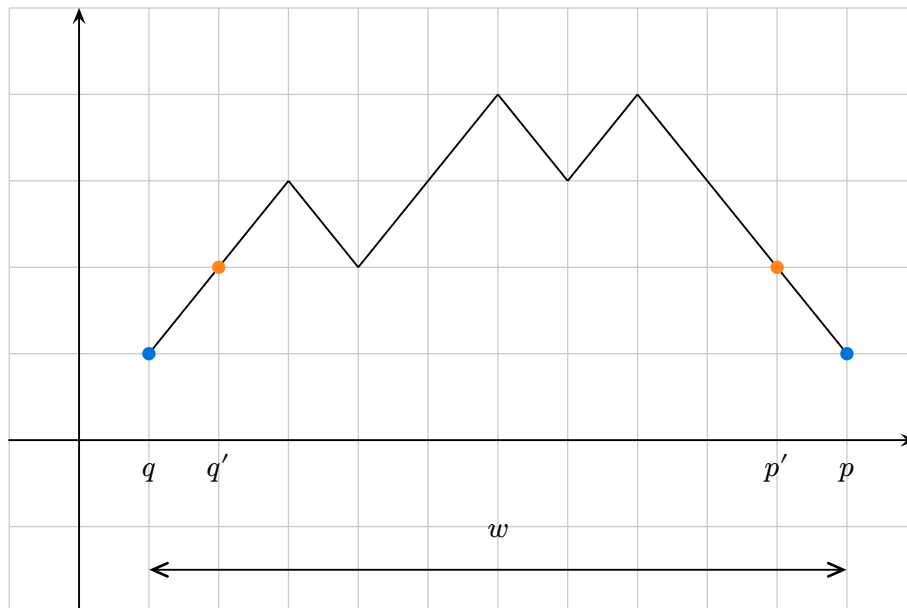
$$\forall(p, -) \in \delta(q, \varepsilon, A) \quad [qAp] \rightarrow \varepsilon.$$

Come prima, possiamo vedere graficamente il movimento come una linea retta:

$$q \boxed{A} \xrightarrow{\varepsilon} \boxed{A} p.$$

Ora facciamo le costruzioni **induttive**.

Supponiamo di essere nel caso in cui, dopo aver caricato A sulla pila, torniamo con la lettera A in cima alla pila alla fine della computazione. In questo caso, la pila è sempre più alta di A , che dopo essere stata caricata «induce» una chiamata ricorsiva nella quale carico, ad esempio, il carattere B e poi lo devo scaricare prima o poi.



Nel grafico, abbiamo B che viene indicato dal pallino arancione, che è posizionato sopra il pallino blu, che indica invece la lettera A che viene scaricata poi alla fine.

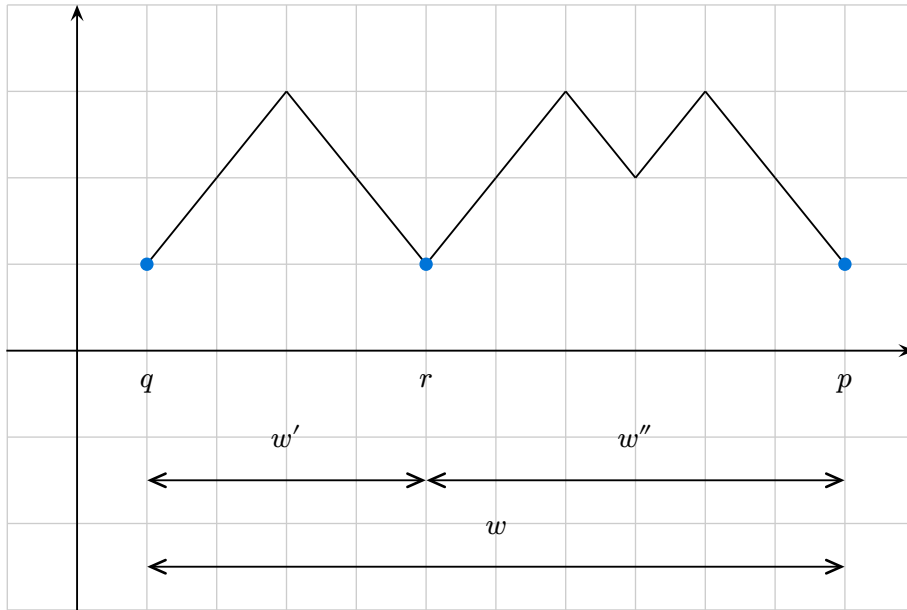
Se chiamiamo w la parte di stringa che stiamo riconoscendo tra la prima A e la seconda A , allora visto che non leggiamo altro durante la push di B stiamo andando a consumare ancora w ma partendo da B . Supponiamo che la mossa che carica la B mi manda nello stato q' e che la mossa che scarica la B parte da p' . Possiamo definire allora

$$\forall(q', \text{push}(B)) \in \delta(q, \varepsilon, A) \quad \forall(p, \text{pop}) \in \delta(p', \varepsilon, B) \quad [qAp] \longrightarrow [q'Bp'].$$

In poche parole, se da q ed A faccio una push di B in q' e poi da p' faccio una pop per andare in p ho fatto una chiamata ricorsiva che ora parte da B e usa gli stati accentati. Come detto prima, la stringa che stiamo consumando rimane sempre w perché con le push e le pop non leggiamo caratteri dal nastro.

L'ultimo caso che ci manca è quando abbiamo almeno una configurazione intermedia nella quale mi trovo allo stesso livello. In questo caso finiamo in uno stato r diverso, quindi abbiamo due percorsi:

- uno da q a r ;
- uno da r a p .



Come regole aggiungiamo

$$\forall r \in Q \quad [qAp] \longrightarrow [qAr][rAp].$$

In poche parole, spezziamo la computazione in due parti, entrambe che partono e finiscono in A .

Tra tutte le computazioni noi vogliamo quelle **accettanti**, quindi vorremmo arrivare in una configurazione del tipo

$$[q_0 Z_0 p] \mid p \in F.$$

Per far sì che ciò accada, dobbiamo imporre le regole

$$\forall p \in F \quad S \longrightarrow [q_0 Z_0 p].$$

Lemma 5.1.2.1: Vale

$$\forall q, p \in Q \quad \forall w \in \Sigma^* \quad [qAp] \xRightarrow{*} w \iff (q, w, A) \vdash^* (p, \varepsilon, A).$$

Lemma 5.1.2.2: Vale

$$S \xRightarrow{*} w \iff (q_0, w, Z_0) \vdash^* (p, \varepsilon, Z_0) \mid p \in F.$$

Hanno inoltre dimostrato che le triple per le variabili sono necessarie: non possiamo farne a meno.

Con questa **costruzione** abbiamo appena fatto vedere che i PDA sono **equivalenti** alle grammatiche di tipo 2.

5.2. Forma normale di Chomsky

L'altra volta abbiamo visto la FN di Greibach, oggi vediamo la **FN di Chomsky**, forma utile e maneggevole per alcune dimostrazioni che faremo.

5.2.1. Definizione

Le produzioni che abbiamo nella FN di Chomsky sono di due tipi:

$$A \longrightarrow a \mid BC \quad \text{tale che} \quad A, B, C \in V \wedge a \in \Sigma.$$

Questa rappresentazione è molto comoda perché riesce a generare degli alberi di derivazione che sono **binari**, quindi abbiamo molte indicazioni su numero di foglie, altezza, e altro.

Come nella FN di Greibach, anche qui non possiamo generare la parola vuota.

Infatti, se G è una grammatica di tipo 2, allora $\exists G'$ in FN di Chomsky quasi equivalente, ovvero

$$L(G') = L(G)/\{\varepsilon\}$$

ma solo se prima ce l'avevamo, sennò sono **totalmente equivalenti**.

5.2.2. Costruzione

Vediamo una **costruzione** per costruire effettivamente una grammatica in FN di Chomsky. Prima abbiamo detto che esiste, qua stiamo facendo vedere che esiste veramente.

Vogliamo costruire la grammatica $G' = (V, \Sigma, P', S)$ in FN di Chomsky a partire dalla grammatica G di tipo 2. Lo possiamo fare seguendo i seguenti passi:

FN di Chomsky

- 1: Eliminazione delle ε -produzioni
 - 2: Eliminazione delle produzioni unitarie
 - 3: Eliminazione dei simboli inutili
 - 4: Eliminazione dei terminali
 - 5: Smontaggio delle produzioni
-

I passi che abbiamo indicato devono essere eseguiti **in ordine**. Partiamo.

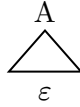
5.2.2.1. Eliminazione delle ε -produzioni

L'eliminazione delle ε -produzioni richiede la ricerca delle **variabili cancellabili**.

Definizione 5.2.2.1.1 (Variabile cancellabile): Una variabile A è **cancellabile** se

$$A \xRightarrow{*} \varepsilon.$$

L'albero di computazione di una variabile cancellabile A lo possiamo vedere come:



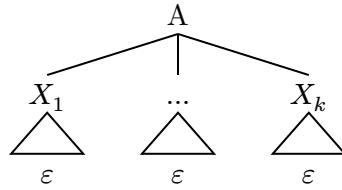
Banalmente, una variabile è cancellabile se nella grammatica ho una regola nella forma

$$A \rightarrow \varepsilon.$$

Se però non abbiamo questa produzione, ma abbiamo delle regole

$$A \rightarrow X_1 \dots X_k \quad \text{tale che} \quad X_1, \dots, X_k \in V$$

quello che possiamo fare è controllare queste variabili. Se sono tutte cancellabili, allora anche A sarà sicuramente cancellabile. Se vogliamo vedere l'albero di computazione, è nella forma:



Se invece non tutte sono cancellabili dobbiamo cercarle con un algoritmo simile a quello che abbiamo usato per dimostrare la decidibilità dei linguaggi di tipo 1. Definiamo l'insieme

$$\mathcal{C}_0 = \{A \in V \mid A \rightarrow \varepsilon\}$$

insieme di tutte le variabili banalmente cancellabili. Definiamo per induzione l'insieme

$$\mathcal{C}_i = \{A \in V \mid \exists (A \rightarrow X_1 \dots X_k) \in P \mid X_1, \dots, X_k \in \mathcal{C}_{i-1}\} \cup \mathcal{C}_{i-1}$$

formato da tutte le variabili che potremmo cancellare usando variabili già cancellabili.

Vale ovviamente la catena

$$\mathcal{C}_0 \subseteq \mathcal{C}_1 \subseteq \dots \subseteq V$$

che è bloccata da un insieme finito, quindi prima o poi non posso più aggiungere degli elementi all'insieme e mi devo fermare, ovvero

$$\exists i \mid \mathcal{C}_{i-1} = \mathcal{C}_i.$$

Una volta che ho tutte le variabili cancellabili creiamo delle **scorciatoie**: cancelliamo prima di tutto tutte le ε -produzioni, e poi

$$\forall (A \rightarrow Y_1 \dots Y_k) \in P \mid k > 0 \wedge Y_i \in V \cup \Sigma \quad (A \rightarrow Y_{j_1} \dots Y_{j_s}) \in P'.$$

In poche parole, aggiungiamo delle regole a P' che otteniamo eliminando alcune variabili cancellabili da una regola di produzione. L'eliminazione non deve toglierle per forza tutte: possiamo toglierne zero come toglierle tutte, ma dobbiamo ricordarci di non creare ε -produzioni e che vanno provate tutte le combinazioni possibili.

Esempio 5.2.2.1.1: Data la regola di produzione

$$A \rightarrow BCaD$$

con C, D variabili cancellabili, vogliamo costruire le nuove regola di una grammatica in FN di Chomsky eliminando le ε produzioni.

In questo caso possiamo:

- far sparire la C ;
- far sparire la D ;
- far sparire la C e la D ;
- lasciare tutto.

Le produzioni che otteniamo sono

$$A \rightarrow BCaD \mid BaD \mid BCa \mid Ba.$$

Vediamo un esempio un po' tedioso.

Esempio 5.2.2.1.2: Data la regola di produzione

$$A \rightarrow CDE$$

con C, D, E variabili cancellabili (quindi anche A), come ci comportiamo?

In questo caso dobbiamo cancellare tutti i possibili sottoinsiemi di variabili cancellabili, escluso l'insieme completo, quindi otteniamo le produzioni

$$A \rightarrow CDE \mid DE \mid CE \mid CD \mid E \mid D \mid C.$$

5.2.2.2. Eliminazione delle produzioni unitarie

L'**eliminazione delle produzioni unitarie** va a rimuovere le produzioni che ha destra hanno solo una variabile, ovvero sono nella forma

$$A \rightarrow B \quad \text{tale che} \quad A, B \in V.$$

Cerchiamo di cancellare le **sequenze di produzioni unitarie**: se consideriamo solo produzioni unitarie abbiamo una sequenza del tipo

$$A \Rightarrow B \Rightarrow C \Rightarrow \dots$$

che si può fermare quando deriviamo un terminale oppure una stringa con più di 2 stringhe. In poche parole abbiamo una sequenza

$$A \Rightarrow \dots \Rightarrow B \Rightarrow \alpha \mid \alpha \in \Sigma \vee |\alpha| > 1.$$

Quando trovo queste sequenze posso fare ancora una scorciatoia: al posto di fare tutta la sequenza di unitarie, che possiamo vedere solo come cambi di variabili, facciamo direttamente il salto da A ad α , ovvero

$$\forall(A, B) \mid A \xRightarrow{+} B \quad \forall(B \longrightarrow \alpha) \mid \alpha \in \Sigma \vee |\alpha| > 1 \quad (A \longrightarrow \alpha) \in P'.$$

5.2.2.3. Eliminazione dei simboli inutili

L'**eliminazione dei simboli inutili** rimuove tutti i simbolo che non ci servono a niente, che non sono utili a generare delle stringhe del linguaggio.

Definizione 5.2.2.3.1 (Simbolo utile): Un simbolo $X \in V \cup \Sigma$ è **utile** se:

- il simbolo è raggiungibile, quindi esiste una computazione che genera una forma sentenziale che contiene quel simbolo, ovvero

$$\exists S \xRightarrow{*} \alpha X \beta;$$

- dalla forma sentenziale centrale si riesce a generare una stringa del linguaggio, ovvero

$$\alpha X \beta \xRightarrow{*} w \in \Sigma^*.$$

Riassunto in una sola sentenza, possiamo dire che

$$\exists S \xRightarrow{*} \alpha X \beta \xRightarrow{*} w \in \Sigma.$$

Con questi tre passi abbiamo rimosso tutte le ε -produzioni, tutti i cammini unitari e tutti i simboli inutili. Le produzioni che abbiamo ora sono nella forma

$$A \longrightarrow \alpha$$

con α terminale oppure $|\alpha| > 1$.

5.2.2.4. Eliminazione dei terminali

L'**eliminazione dei terminali** si applica alle produzioni che derivano una forma sentenziale α con $|\alpha| > 1$. Per fare ciò dobbiamo usare delle **variabili ausiliarie** per rimuovere i simboli terminali nelle produzioni.

Come convenzione possiamo dire che i terminali $\sigma \in \Sigma$ vengono sostituiti dalle variabili X_σ .

Esempio 5.2.2.4.1: Date le regole di produzione

$$A \longrightarrow Aaabc \mid bC \mid bb$$

cerchiamo di applicare l'eliminazione dei terminali.

Usiamo due variabili ausiliarie X_a e X_b , una per ogni terminale, ottenendo

$$\begin{aligned} A &\longrightarrow AX_aX_aX_bC \mid X_bC \mid X_bX_b \\ X_a &\longrightarrow a \\ X_b &\longrightarrow b. \end{aligned}$$

Se avessi avuto anche la regola

$$C \rightarrow b$$

non avrei applicato il cambio in X_b perché avremmo ottenuto un cammino unitario, che però non possiamo avere in questo momento.

In questo penultimo passo ora abbiamo solo produzioni che derivano terminali oppure delle liste di variabili. L'ultimo passo sarà manipolare queste per raggiungere, finalmente, la FN di Chomsky.

5.2.2.5. Smontaggio delle produzioni

L'ultimo passo è lo **smontaggio delle produzioni** nella forma

$$A \rightarrow B_1 \dots B_k \quad \text{con } k > 2.$$

Infatti, le produzioni:

- con una sola variabile non ci sono;
- con due variabili vanno bene in questa FN;
- con tre o più variabili vanno ridotte.

Per fare ciò, usiamo ancora delle **variabili ausiliarie** per aggiungere delle regole formate da una variabile della produzione e da una variabile ausiliaria. Fa eccezione l'ultima regola, che sostituisce direttamente gli ultimi due caratteri della produzione da cambiare.

Esempio 5.2.2.5.1: Data la produzione

$$A \rightarrow B_1 B_2 B_3 B_4 B_5$$

la andiamo a smontare creando le produzioni

$$\begin{aligned} A &\rightarrow B_1 Z_1 \\ Z_1 &\rightarrow B_2 Z_2 \\ Z_2 &\rightarrow B_3 Z_3 \\ Z_3 &\rightarrow B_4 B_5. \end{aligned}$$

5.2.3. Esempio

Applichiamo questa costruzione ad un esempio.

Esempio 5.2.3.1: Costruire la FN di Chomsky a partire da queste produzioni:

$$\begin{aligned} S &\rightarrow aB \mid bA \\ A &\rightarrow a \mid aS \mid bAA \\ B &\rightarrow b \mid bS \mid aBB. \end{aligned}$$

Non abbiamo ε -produzioni e variabili cancellabili, come non abbiamo cammini unitari, come non abbiamo simboli inutili, quindi i primi tre passi non vengono eseguiti.

Partiamo quindi con il cancellare tutti i terminali:

$$\begin{aligned}S &\longrightarrow X_a B \mid X_b A \\A &\longrightarrow a \mid X_a S \mid X_b AA \\B &\longrightarrow b \mid X_b S \mid X_a BB \\X_a &\longrightarrow a \\X_b &\longrightarrow b.\end{aligned}$$

Infine, smontiamo le catene di variabili che abbiamo ottenuto:

$$\begin{aligned}S &\longrightarrow X_a B \mid X_b A \\A &\longrightarrow a \mid X_a S \mid X_b Y_1 \\Y_1 &\longrightarrow AA \\B &\longrightarrow b \mid X_b S \mid X_a Y_2 \\Y_2 &\longrightarrow BB \\X_a &\longrightarrow a \\X_b &\longrightarrow b.\end{aligned}$$