Tema 2

Autòmats finits i conjunts regulars

Un autòmat finit és un dispositiu lògic que processa una cadena llegint cada símbol de la cadena només una vegada d'esquerra a dreta. L'única informació que el dispositiu pot emmagatzemar és l'estat en què es troba (d'entre un nombre finit d'estats possibles). El dispositiu canvia d'estat només en funció del símbol que llegeix i del estat en què es troba abans de llegir-lo.

Per exemple, es pot construir un autòmat amb dos estats que accepte el llenguatge format per totes les cadenes de longitud parella. Un estat correspondria a que la cadena processada és parella i l'altre a que és senar. L'autòmat es trobaria inicialment en l'estat corresponent a què la cadena és parella. La lectura de qualsevol símbol de la cadena faria canviar l'autòmat d'un estat a l'altre. L'autòmat acceptaria només aquelles cadenes que deixaren l'autòmat en el primer dels dos estats.

Formalment, podem donar la següent definició

Definició 2.1 *Un* **autòmat finit**, *A, és una estructura de la forma $\langle Q, \Sigma, \delta, q_0, F\rangle$ on*

- Q és un conjunt finit d'estats.
- Σ és l'alfabet d'entrada.6s7s6s
- $q_0 \in Q$ és l'estat inicial.
- $\delta : Q \times \Sigma \longrightarrow Q$ és la **funció de transició**.
- $F \subseteq Q$ és el (sub)conjunt d'estats finals o d'acceptació.

$\square$

Anomenarem **diagrama de transició** de l'autòmat A al graf dirigit on hi ha un node per cada estat i un arc etiquetat amb el símbol a entre dos estats p i q si i només si $\delta(p, a) = q$. Els nodes corresponents a l'estat inicial i als finals es marquen amb una fletxa incident i amb un doble cercle, respectivament.

L'automat de l'exemple anterior es podria definir formalment com a

$$A_p = \langle\{q_0, q_1\}, \{a, b\}, q_0, \delta, \{q_0\}\rangle$$

on la funció δ vindria definida per la següent taula que s'anomena **taula de transició**.

	a	b
q_0	q_1	q_1
q_1	q_0	q_0

En aquesta taula, la primera entrada sempre correspon a l'estat inicial i els estats finals es marquen amb un doble cercle. En algunes ocasions, els estats finals es marcaran amb la lletra "F" al costat de l'estat o en una columna extra.

Aquest autòmat es pot representar també mijantçant el seu diagrama de transició que es mostra a la figura 2.1

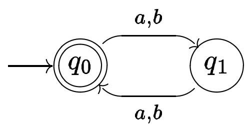
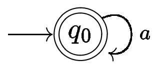
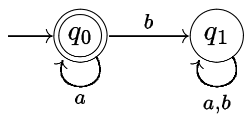


Figura 2.1. Diagrama de transicions de l'autòmat A_p.

La funció de transició pot ser una funció parcial. Per tant, el comportament de l'autòmat pot no estar definit per a certes cadenes. Considerem el següent autòmat,



L'autòmat accepta qualsevol cadena que continga només símbols a. En canvi, no té un comportament definit si la cadena d'entrada conté una b. Des del punt de vista del llenguatge acceptat per un autòmat podem dir que l'automat anterior és equivalent a



L'estat q_1 és un estat no final del qual no es pot sortir i per tant no modifica el conjunt de cadenes acceptades per l'autòmat però, en canvi, fa que la funció δ estiga definida per a tota cadena de $\Sigma^\star$. Anomenarem **estat d'error** a tot estat no final que continga transicions a ell mateix amb qualsevol símbol. Anomenarem també **autòmat complet** a un autòmat tal que la seua funció de transició és total.

La funció de transició, δ, es pot estendre per a cadenes de símbols mijantçant la següent definició recursiva:

$$\begin{cases} \hat{\delta}(q,\varepsilon) = q & \forall q \in Q \\ \hat{\delta}(q,wa) = \delta(\hat{\delta}(q,w),a) & \forall a \in \Sigma, w \in \Sigma^\star, q \in Q \end{cases}$$

Per exemple, per a l'últim autòmat definit tindríem
$\hat{\delta}(q_0,aba) = \delta(\hat{\delta}(q_0,ba),a) = \delta(\delta(\hat{\delta}(q_0,a),b),a) = \delta(\delta(\delta(\underbrace{\hat{\delta}(q_0,\varepsilon)}_{q_0},a),b),a)$

$\delta(\delta(\underbrace{\delta(q_0,a)}_{q_0},b),a) = \delta(\underbrace{\delta(q_0,b)}_{q_1},a) = \delta(q_1,a) = q_1$

Per norma general escriurem sempre δ tant per a la funció sobre símbols com per a la funció sobre cadenes ja que no hi ha confusió possible.

Anomenarem **configuració** o **descripció intantània** d'un autòmat finit al parell format per l'estat actual i el sufix de la cadena d'entrada que està pendent de lectura. Aquest parell defineix de forma unívoca quin serà el comportament de l'autòmat a partir d'eixe moment. Per a l'exemple anterior es tindria com a configuració inicial (q_0, aba) i (q_1, ε) com a configuració final.

El pas d'una configuració a una altra en un autòmat el representarem amb el símbol $\vdash$. Es compleix que

$$(q, aw) \vdash (q', w) \Longleftrightarrow \delta(q, a) = q'$$

Expresarem una successió de transicions entre configuracions mijantçant el símbol $\overset{\star}{\vdash}$. Per exemple, en el cas de l'autòmat anterior tenim

$$(q_0, aba) \vdash (q_0, ba) \vdash (q_1, a) \vdash (q_1, \varepsilon)$$

o, simplement

$$(q_0, aba) \overset{\star}{\vdash} (q_1, \varepsilon)$$

Ja hem definit informalment el llenguatge acceptat per un autòmat com el format per aquelles cadenes d'entrada que porten l'autòmat a un estat final. Formalment, podem utilitzar qualsevol de les dues definicions següents.

Definició 2.2 *Siga un autòmat finit,* $A = \langle Q, \Sigma, \delta, q_0, F\rangle$. *El llenguatge acceptat per A, que escriurem com a* $L(A)$ *ve donat per*

$$L(A) = \{w \in \Sigma^\star \mid (q_0, w) \overset{\star}{\vdash} (q, \varepsilon), q \in F\}$$

o per

$$L(A) = \{w \in \Sigma^\star \mid \delta(q_0, w) \in F\}$$

$\square$

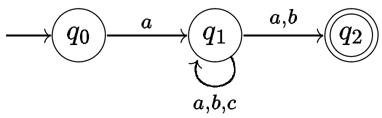
2.1 Tipus d'autòmats finits

En aquesta secció s'introduiran algunes generalitzacions del concepte d'autòmat finit i s'estudia la seua relació amb el model que s'acaba d'introduir.

2.1.1 Autòmats no deterministes

Imaginem que es vol construir un autòmat que accepte cadenes de $\{a, b, c\}^\star$ que comencen per a i que acaben necessariament en a o b[1].

Aquestes cadenes podrien ser acceptades per un autòmat com aquest



Aquest autòmat captura la idea exacta del llenguatge tal i com s'ha definit i es clar que acceptarà tota cadena que comence per a i acabe en a o b i entre el primer símbol i l'últim tinga qualsevol cadena. No obstant això, en entendre l'autòmat com una màquina que processa una cadena, s'introdueix en aquest cas una novetat important que il·lustra la següent qüestió: Quin serà el comportament de l'autòmat quan es trobe en l'estat q_1 i el següent símbol siga una a?

Podem generalitzar aquesta idea fent que la funció de transició de l'autòmat per a un estat i un símbol ens done un conjunt d'estats en lloc d'un únic estat. En aquest cas, un autòmat pot comportar-se de diverses maneres per a una mateixa cadena. Açò és el que es coneix com a indeterminisme.

Definició 2.3 *Un* **autòmat finit indeterminista***, A, és una estructura de la forma $\langle Q, \Sigma, \delta, q_0, F\rangle$ on*

- Q és un conjunt finit d'estats.
- Σ és l'alfabet d'entrada.
- $q_0 \in Q$ és l'estat inicial.
- $\delta : Q \times \Sigma \longrightarrow 2^Q$ és la funció de transició .
- $F \subseteq Q$ és el (sub)conjunt d'estats finals o d'acceptació.

$\square$

És important remarcar que l'única cosa que ha canviat respecte de la definició d'autòmat finit (determinista) és la funció de transició que ara pren com a valors conjunts d'estats en lloc d'estats simples. La taula corresponent a l'autòmat indeterminista de l'exemple seria:

[1] És a dir, el llenguatge seria $\{a\} \cdot \{a, b, c\}^\star \cdot \{a, b\}$

	a	b	c
q_0	$\{q_1\}$	$\emptyset$	$\emptyset$
q_1	$\{q_1, q_2\}$	$\{q_1, q_2\}$	$\{q_1\}$
q_2	$\emptyset$	$\emptyset$	$\emptyset$

De manera intuitiva, un autòmat indeterminista pot evolucionar de diverses formes amb una mateixa cadena d'entrada però la cadena haurà de ser acceptada si d'alguna d'aquestes maneres s'arriba a una configuració d'acceptació.

De la mateixa manera que en el cas determinista, la funció de transició es pot estendre a cas de cadenes amb la següent definició recursiva:

$$\begin{cases} \hat{\delta}(q,\varepsilon) = \{q\} & \forall q \in Q \\ \hat{\delta}(q,wa) = \bigcup_{p\in\hat{\delta}(q,w)} \delta(p,a) & \forall a \in \Sigma, w \in \Sigma^\star, q \in Q \end{cases}$$

És a dir, el valor de la funció de transició des d'un estat amb una determinada cadena és igual al conjunt de tots els estats a que es pot arribar considerant totes les possibilitats en l'autòmat indeterminista.

Donat que es treballa amb conjunts d'estats en lloc d'estats simples, la funció de transició es pot estendre també a conjunts d'estats de la següent manera:

$$\tilde{\delta}(P,w) = \bigcup_{q\in P} \delta(q,w)$$

Intuïtivament, podem pensar que un autòmat determinista pot trobar-se en més d'un estat al mateix temps segons les transicions que haja seguit en els passos anteriors. Per això, el concepte de configuració d'un autòmat indeterminista contempla un conjunt d'estats. En particular, la configuració o descripció instantània d'un autòmat finit indeterminista és un parell format per un conjunt d'estats i el sufix pendent de lectura de la cadena d'entrada.

La condició per a que es done una transició entre dues configuracions és ara:

$$(P,aw) \vdash (P',w) \Longleftrightarrow \tilde{\delta}(P,a) = P'$$

Per exemple, en el cas de l'autòmat indeterminista de l'exemple tindríem

$$(\{q_0\},abca) \vdash (\{q_1\},bca) \vdash (\{q_1,q_2\},ca) \vdash (\{q_1\},a) \vdash (\{q_1,q_2\},\varepsilon)$$

Com en el cas determinista, podem donar la següent definició formal de llenguatge acceptat per un autòmat indeterminista:

Definició 2.4 *Siga un autòmat finit indeterminista, $A = \langle Q,\Sigma,\delta,q_0,F\rangle$. El llenguatge acceptat per A, que escriurem com a $L(A)$ ve donat per*

$$L(A) = \{w \in \Sigma^\star \mid (\{q_0\},w) \overset{\star}{\vdash} (Q',\varepsilon),\ Q' \cap F \neq \emptyset\}$$

o per

$$L(A) = \{w \in \Sigma^\star \mid \tilde{\delta}(q_0, w) \cap F \neq \emptyset\}$$

□

Equivalència entre autòmats deterministes i indeterministes

Si pensem en un autòmat indeterminista com un autòmat que al mateix temps es troba en un conjunt d'estats, P, i que en l'instant següent evoluciona al conjunt d'estats donat per la funció de transició convenientment estesa, és fàcil veure que aquest autòmat ha de ser equivalent a un autòmat determinista definit sobre tots els possibles subconjunts d'estats.

Teorema 2.1 Siga $A = \langle Q, \Sigma, \delta, q_0, F\rangle$ un autòmat finit indeterminista. L'autòmat finit determinista, A' definit com a $A' = \langle 2^Q, \Sigma, \tilde{\delta}, \{q_0\}, F'\rangle$ on $\tilde{\delta}$ és la funció de transició de A estesa per a conjunts d'estats i $F' = \{P \in 2^Q \mid P \cap F \neq \emptyset\}$, compleix que $L(A) = L(A')$.

La demostració és trivial a partir dels raonaments anteriorment exposats. Per a qualsevol cadena de $L(A)$ es té que

$$(\{q_0\}, w) \overset{\star}{\vdash} (P, \varepsilon)$$

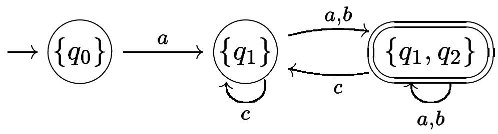
on P és un conjunt que conté almenys un estat final.

Aquesta seqüencia de transicions es correspondria exactament amb una seqüència de transicions per a A' on cada configuració tindria un únic estat (un element de 2^Q) i una cadena. Òbviament, l'estat P al que s'arriba pertany a F' segons la definició. □

Com a exemple podem escriure l'autòmat determinista equivalent a l'autòmat indeterminista de l'exemple anterior.

	a	b	c
$\{q_0\}$	$\{q_1\}$	$\emptyset$	$\emptyset$
$\{q_1\}$	$\{q_1, q_2\}$	$\{q_1, q_2\}$	$\{q_1\}$
$\{q_2\}$	$\emptyset$	$\emptyset$	$\emptyset$
$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$
$\{q_1, q_2\}$	$\{q_1, q_2\}$	$\{q_1, q_2\}$	$\{q_1\}$
...			

Estrictament, caldria completar la taula amb una entrada per cada possible subconjunt d'estats però, en aquest cas (i en molts altres) no té cap sentit fer-ho perque aquests nous estats no serien abastables des de $\{q_0\}$ (no apareixen en ningun lloc de la taula de transició). De fet, en aquest cas l'estat $\{q_2\}$ tampoc apareix en la taula i tampoc serà pertant abastable. L'estat $\emptyset$ si que és abastable però com serà sempre un estat d'error es pot prescindir d'ell. Amb aquestes consideracions, l'autòmat resultant es pot escriure com a



2.1.2 Autòmats amb transicions buides

Una altra generalització de l'autòmat finit determinista consisteix a incluir-hi la possibilitat de canviar d'estat sense consumir cap símbol a l'entrada. Açò és el que es coneix com **transicions buides** o **transicions nul·les**.

Definició 2.5 *Un* **autòmat finit amb transicions buides***, A, és una estructura de la forma* $A = \langle Q, \Sigma, \delta, q_0, F \rangle$ *on*

- Q és un conjunt finit d'estats.
- Σ és l'alfabet d'entrada.
- $q_0 \in Q$ és l'estat inicial.
- $\delta : Q \times \Sigma \cup \{\varepsilon\} \longrightarrow 2^Q$ és la funció de transició .
- $F \subseteq Q$ és el (sub)conjunt d'estats finals o d'acceptació.

□

Definit d'aquesta manera, l'autòmat finit amb transicions buides constitueix una generalització dels autòmats indeterministes.

D'altra banda, les transicions buides introdueixen un altre tipus d'indeterminisme en l'autòmat: trobarse en un determinat estat o en un altre estat connectat amb l'anterior mijantçant transicions buides és la mateixa cosa. Per a relacionar aquest tipus d'autòmats amb els anteriors, considerarem també que l'autòmat pot trobarse en més d'un estat al mateix temps i que evoluciona a altres estats en funció de quines siguen les transicions. Per tal de formalitzar açò resulta útil introduir el concepte de tancament epsilon o ε-tancament.

Definició 2.6 *El* **tancament epsilon** *d'un estat q en un autòmat amb transicions buides, escrit com a $q^\bullet$, ve donat pel conjunt d'estats a què es pot arribar des de q només amb transicions buides. El tancament epsilon d'un conjunt d'estats P es defineix com a*

$$P^\bullet = \bigcup_{q \in P} q^\bullet$$

□

El concepte de tancament epsilon representa el fet que l'autòmat en qualsevol moment pot canviar a qualsevol estat que estiga connectat amb transicions buides sense consumir símbols. Podem ara definir una funció de transició associada a l'autòmat sobre cadenes que tinga en compte aquest fet mijantçant la següent definició recursiva:

$$\begin{cases} \delta'(q,\varepsilon) = q^\bullet & \forall q \in Q \\ \delta'(q,wa) = \delta(\ \delta'(q,w)\ ,a)^\bullet & \forall a \in \Sigma, w \in \Sigma^\star, q \in Q \end{cases}$$

La definició recursiva captura de forma natural les possibles transicions buides que pot efectuar l'autòmat entre qualssevol dos símbols consecutius. A la figura 2.2 es representa aquesta idea gràficament.

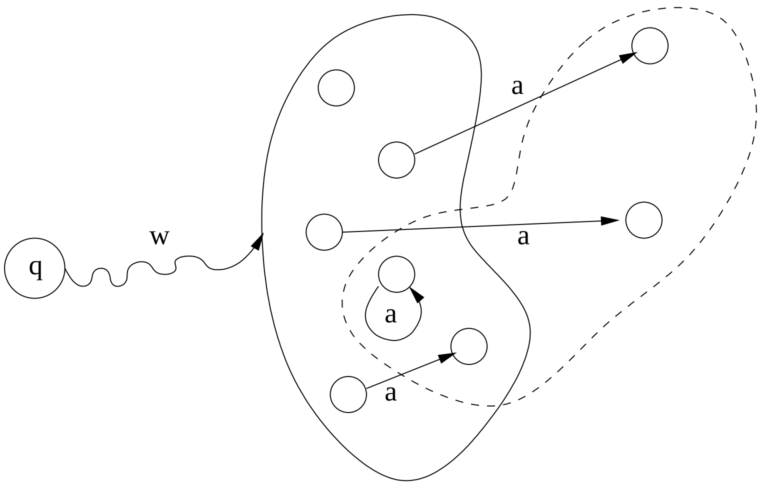


Figura 2.2. Representació gràfica del càlcul recursiu de $\delta'(q,wa)$. La línia contínua representa el conjunt $\delta'(q,w)$ i la línia discontinua és $\bigcup_{p\in\delta'(q,w)} \delta(p,a)$. Posteriorment caldria calcular el tancament epsilon d'aquest darrer conjunt.

Aquesta funció de transició també es pot estendre per a conjunts d'estats de la manera usual.

$$\delta'(P,w) = \bigcup_{q\in P} \delta'(q,w)$$

El concepte de configuració no canvia respecte del que s'ha introduit per a autòmats indeterministes. Però ara, la transició entre dues configuracions depen de la funció δ'.

Ara es pot definir formalment el llenguatge acceptat per un autòmat amb transicions buides.

Definició 2.7 *Siga un autòmat finit amb transicions buides, $\langle Q,\Sigma,\delta,q_0,F\rangle$. El llenguatge acceptat per A, que escriurem com a $L(A)$ ve donat per*

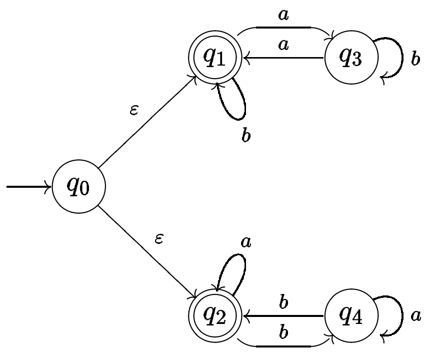
$$L(A) = \{w \in \Sigma^\star \mid (\{q_0\},w)\overset{\star}{\vdash}(Q',\varepsilon),\ Q'\cap F \neq \emptyset\}$$

o per

$$L(A) = \{w \in \Sigma^\star \mid \delta'(q_0,w)\cap F \neq \emptyset\}$$

□

Considerem com a exemple l'autòmat següent que accepta cadenes de $\{a, b\}^\star$ que tinguen un nombre parell de símbols a o b.



La taula de transició corresponent a la qual hem afegit una columna per incloure els corresponents tancaments epsilon seria:

	a	b	ε	$q^\bullet$
q_0	$\emptyset$	$\emptyset$	$\{q_0, q_1, q_2\}$	$\{q_0, q_1, q_2\}$
q_1	$\{q_3\}$	$\{q_1\}$	$\{q_1\}$	$\{q_1\}$
q_2	$\{q_2\}$	$\{q_4\}$	$\{q_2\}$	$\{q_2\}$
q_3	$\{q_1\}$	$\{q_3\}$	$\{q_3\}$	$\{q_3\}$
q_4	$\{q_4\}$	$\{q_2\}$	$\{q_4\}$	$\{q_4\}$

Fem notar que encara que en el diagrama de transició només hi ha dues transicions buides, és bastant evident que, a més a més, tot estat conté implícitament una transició buida a ell mateix.

Equivalència entre els autòmats amb i sense transicions buides

Encara que el funcionament de les transicions buides és clarament diferent, l'efecte és que aquests autòmats es comporten com els autòmats indeterministes. En tot moment es troben en un conjunt d'estats determinat i a l'instant següent evolucionen a un altre conjunt d'estats que ve donat per la funció de transició estesa que s'acaba de definir.

Teorema 2.2 Siga $A = \langle Q, \Sigma, \delta, q_0, F \rangle$ un autòmat finit amb transicions buides. L'autòmat finit indeterminista, A' definit com a $A' = \langle Q, \Sigma, \delta', q_0, F' \rangle$ on δ' és la funció de transició de A estesa per a cadenes i $F' = F \cup \{q \in Q \mid q^\bullet \cap F \neq \emptyset\}$, compleix que $L(A) = L(A')$.

Com que la funció δ' té implicitament en compte totes les possibilitats a l'hora de canviar d'estat sense consumir símbols, l'autòmat així definit serà indeterminista i equivalent a l'original. □

Com a exemple de la construcció que suggereix aquest teorema es pot calcular l'autòmat indeterminista equivalent a l'autòmat amb transicions buides de l'últim exemple. Per això, hem de calcular totes les posicions de la nova taula de transicions com a $\delta'(q,a)$ que, per definició és igual a

$$\left(\bigcup_{q^\bullet} \delta(p,a)\right)^\bullet$$

Per exemple,

$$\delta'(q_0,a) = \delta(\{q_0,q_1,q_2\},a)^\bullet = \{q_2,q_3\}^\bullet = \{q_2,q_3\}$$

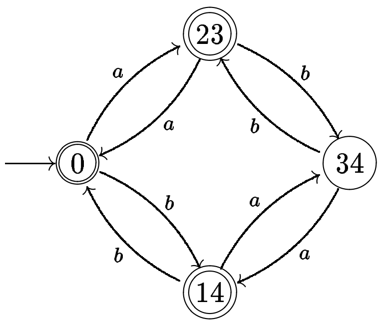
i, de la mateixa manera, totes les demés entrades de la taula. Al final s'obté

	a	b
q_0	$\{q_2,q_3\}$	$\{q_1,q_4\}$
q_1	$\{q_3\}$	$\{q_1\}$
q_2	$\{q_2\}$	$\{q_4\}$
q_3	$\{q_1\}$	$\{q_3\}$
q_4	$\{q_4\}$	$\{q_2\}$

Aquest autòmat indeterminista es pot transformar en determinista en considerar els estats com a conjunts d'estats i en afegir a més a més els següents conjunts d'estats

(cont.)	a	b
$\{q_1,q_4\}$	$\{q_3,q_4\}$	$\{q_1,q_2\}$
$\{q_2,q_3\}$	$\{q_1,q_2\}$	$\{q_3,q_4\}$
$\{q_3,q_4\}$	$\{q_1,q_4\}$	$\{q_2,q_3\}$
$\{q_1,q_2\}$	$\{q_2,q_3\}$	$\{q_1,q_4\}$

A partir de les dues últimes taules conjuntament es pot veure que els estats $\{q_1\}$, $\{q_2\}$, $\{q_3\}$ i $\{q_4\}$ no són abastables des de l'estat inicial i també que els estats $\{q_0\}$ i $\{q_1,q_2\}$ són equivalents. Amb tot açò, l'autòmat que finalment s'obté és



on hem anomenat els estats segons els subíndex dels conjunts corresponents (per exemple, $23 \equiv \{q_2,q_3\}$).

2.2 Autòmats finits i llenguatges regulars

Els diferents tipus d'autòmats finits introduïts fins ara accepten els mateixos llenguatges. En altres paraules, defineixen una certa classe de llenguatges. El que ens interessa ara és relacionar aquesta classe de llenguatges amb la jerarquia de Chomsky. En particular, es demostrarà que els llenguatges acceptats per autòmats finits i els llenguatges regulars són la mateixa classe.

2.2.1 Gramàtica equivalent a un autòmat finit

Considerem el llenguatge acceptat per un autòmat finit determinista. Per a qualsevol cadena, w, d'aquest llenguatge es tindrà

$$(q_0, w)\overset{\star}{\vdash}(q, \varepsilon)$$

En cada pas en aquesta seqüència de configuracions, desapareix un símbol de la cadena i s'evoluciona a un nou estat. Un canvi de configuració es dóna només si existeix la transició adequada en l'autòmat.

Aquest procés es pot relacionar directament amb la forma en que les gramàtiques regulars generen cadenes. Una derivació directa només es dona si existeix una producció adequada i l'efecte és que en la forma sentencial apareix un nou terminal. Si es fa la identificació d'estats amb símbols no terminals i de transicions amb produccions, el proces d'acceptació i de derivació d'una cadena són equivalents.

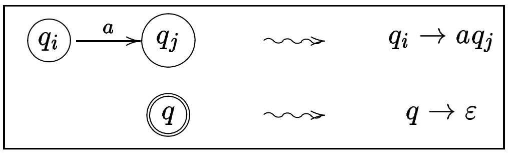
Teorema 2.3 Siga un autòmat finit determinista $A = \langle Q, \Sigma, \delta, q_0, F\rangle$. La gramàtica regular, $G = \langle V_T, V_N, S, P\rangle$, on $V_T = \Sigma$, $V_N = Q$, $S = q_0$ i el conjunt de produccions ve donat per

$$P = \{q_i \to aq_j \mid \delta(q_i, a) = q_j\} \cup \{q \to \varepsilon \mid q \in F\}$$

compleix que $L(A) = L(G)$.

La demostració del teorema és trivial a partir de l'explicació anterior. $\square$

La construcció que suggereix el teorema es pot resumir gràficament de la següent manera:



Exemple 2.1 *Anem a calcular una gramàtica regular equivalent a l'últim autòmat obtingut en la secció 2.1.2 (pàgina [2]-10).*

En primer lloc, renombrarem els estat en sentit horari des de l'inicial com a S, A, B i C. D'aquesta manera, noms d'estats i de símbols no terminals coincidiran exactament. Calculant les produccions corresponents a cada una de les transicions de l'autòmat s'arriba a la gramàtica

$S \to aA|bC|\varepsilon$
$A \to aS|bB|\varepsilon$
$B \to bA|aC$
$C \to aB|bS|\varepsilon$

□

2.2.2 Autòmat equivalent a una gramàtica regular

A partir de la definició de gramàtica regular, és clar que per a qualsevol cadena, w d'un llenguatge regular ha d'existir una derivació de la forma

$$S \underset{G}{\Rightarrow} \alpha_1 \underset{G}{\Rightarrow} \dots \alpha_i \overset{\star}{\underset{G}{\Rightarrow}} w$$

per a alguna gramàtica regular, $G = \langle V_T, V_N, S, P \rangle$. Totes les formes sentencials, α_i han de complir necessàriament que $\alpha_i \in V_T{}^\star \cdot V_N$.

En cada pas s'aplica una de les produccions de la gramàtica, que fa que apareguen uns certs símbols en la forma sentencial en lloc del (únic) no terminal. Donat un determinat símbol no terminal en una forma sentencial, hi ha un nombre finit de produccions possibles que s'hi poden aplicar.

En el teorema següent, es defineix un autòmat que simula el procés de derivació de cadenes d'una gramàtica qualsevol de forma que aquest accepta només aquelles cadenes que són generades per la gramàtica.

Teorema 2.4 Siga una gramàtica regular, $G = \langle V_T, V_N, S, P \rangle$. L'autòmat finit amb transicions buides, $A = \langle Q, \Sigma, q_S, \delta, \{q_\varepsilon\} \rangle$, amb $\Sigma = V_T$,

$$Q = \{q_S\} \cup \{q_\beta \mid A \to \alpha\beta \in P,\ \forall \alpha, \beta \in V^\star\}$$

i la funció de transició tal que

a) $\delta(q_A, \varepsilon) = \{q_\alpha \mid A \to \alpha \in P\}$

b) $\delta(q_{a\beta}, a) = \{q_\beta,\ \forall a \in V_T, \beta \in V^\star, q_{a\beta} \in Q\}$

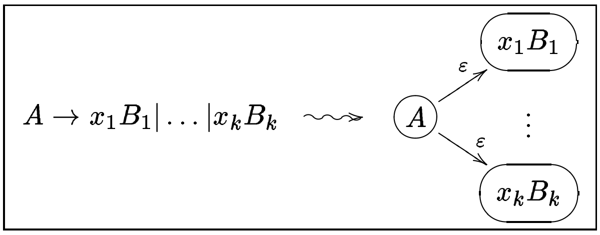
compleix que $L(A) = L(G)$.

L'autòmat així definit simula a) el procés de reescritura d'un no terminal com a les corresponents parts dretes de produccions associades, i b) la "generació" dels símbols terminals associats a cada part dreta de producció.

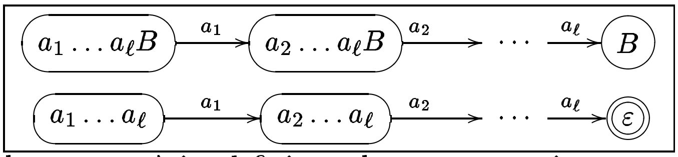
□

La construcció suggerida pel teorema es resumeix gràficament a continuació.

a)



b)



L'autòmat que simula una gramàtica definit en el teorema anterior es pot calcular per a qualsevol gramàtica regular. No obstant això, en la majoria de casos els autòmats que s'obtenen són grans i excessivament complexos (degut en part a l'abundància de transicions buides). És possible donar un mètode simplificat per a l'obtenció de l'autòmat equivalent a una gramàtica que restringeix mínimament la forma de les produccions però, en canvi, obté normalment autòmats relativament compactes.

Mètode per l'obtenció de l'autòmat equivalent a una gramàtica regular

Si una gramàtica regular compleix que ninguna producció té més d'un símbol terminal en la seua part dreta, es pot obtindre el seu autòmat finit equivalent mijantçant el següent mètode:

Siga $G = \langle V_T, V_N, S, P\rangle$ una gramàtica regular on totes les produccions de P són de la forma $A \to \alpha$ on $|\alpha|_{V_T} \leq 1$.

Aleshores, l'autòmat $A = \langle Q, V_T, \delta, q_S, \{q_F\}\rangle$, on

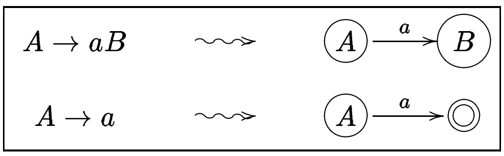
$$Q = \{q_A \mid A \in V_N\} \cup \{q_F\}$$

i la funció de transició del qual ve donada per

$$\begin{aligned}
&q_B \in \delta(q_A, x) \forall\ A \to xB \in P,\ x \in V_T \cup \{\varepsilon\}\\
&q_F \in \delta(q_A, x) \forall\ A \to x \in P, x \in V_T \cup \{\varepsilon\}\\
&\delta(q_F, x) = \emptyset\ \ \forall\ x \in V_T \cup \{\varepsilon\}
\end{aligned}$$

compleix que $L(A) = L(G)$.

La construcció suggerida es mostra gràficament a continuació.



En el següent exemple s'il·lustra la diferència entre la construcció suggerida pel teorema anterior i el mètode que s'acaba d'introduir.

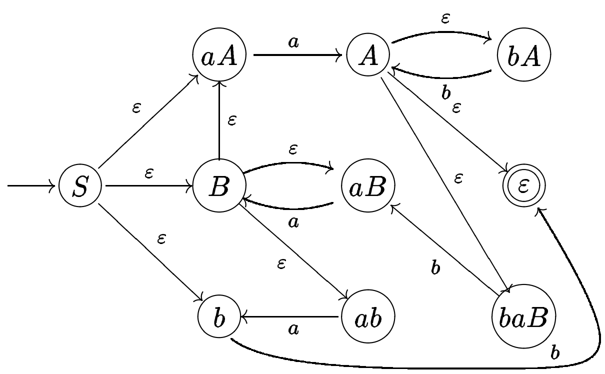
Exemple 2.2 *Construir un autòmat finit determinista equivalent a la gramàtica:*
$S \to aA|B|b$
$A \to bA|bB|\varepsilon$
$B \to aA|aB|ab$

Autòmat obtingut a partir del teorema

La construcció suggerida pel teorema dóna com a resultat un autòmat amb la següent taula de transicions on s'ha afegit també informació sobre els tancaments epsilon i s'ha relaxat una mica la notació de conjunts (el guionet en les columnes corresponents a símbols significa $\emptyset$ però en la columna de ε vol dir que $\delta(q, \varepsilon)$ només conté q).

	a	b	ε	$q^\bullet$
S	-	-	aA, B, b	S, aA, B, b, aB, ab
A	-	-	bA, baB, ε	A, bA, baB, ε
B	-	-	aA, aB, ab	B, aA, aB, ab
aA	A	-	-	aA
b	-	ε	-	b
bA	-	A	-	bA
baB	-	aB	-	baB
ε	-	-	-	ε
aB	B	-	-	aB
ab	b	-	-	ab

El diagrama de transicions corresponent seria:



A partir d'aquest s'obdindria l'autòmat indeterminista la taula de transicions del qual seria

	a	b
S	Q_1	ε
A	-	Q_2
B	Q_1	-
aA	$A^\bullet$	-
b	-	ε
bA	-	$A^\bullet$
baB	-	aB
ε	-	-
aB	$B^\bullet$	-
ab	b	-

on

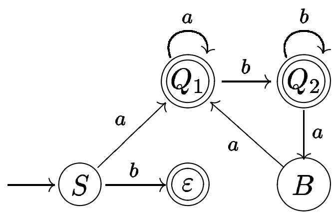
$$Q_1 = \{A, B, b\}^\bullet = \{A, bA, baB, \varepsilon, aA, aB, ab\}$$

$$Q_2 = \{A, aB\}^\bullet = \{A, bA, baB, \varepsilon, aB\}$$

Per a fer aquest autòmat determinista cal afegir a aquesta taula les següents entrades:

(cont.)	a	b
Q_1	Q_1	Q_2
Q_2	$B^\bullet$	Q_2
$A^\bullet$	-	Q_2
$B^\bullet$	Q_1	-

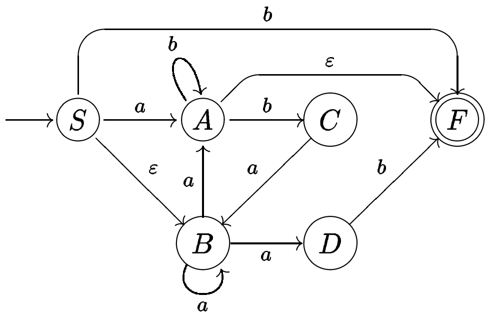
Molts dels estats obtinguts no són abastables i, al final, l'autòmat es pot escriure en funció dels estats S, Q_1, Q_2, ε i $B^\bullet$ (que és equivalent a B).



Mètode per l'obtenció de l'autòmat equivalent

Per a aplicar el mètode alternatiu a la gramàtica de partida, primer cal transformar les produccions $A \to baB$ i $B \to aa$, que no compleixen la condició exigida pel mètode, en les produccions $A \to bC$, $C \to aB$, $B \to aD$ i $D \to b$, on C i D són símbols no terminals nous.

Aleshores s'obté



Si escrivim directament la taula de transició del corresponent autòmat indeterminista (incloent també els tancaments epsilon) així com les entrades corresponents per a fer-lo determinista s'obté

	a	b	$q^\bullet$
S	$ABDF$	F	S, B
A	-	ACF	A, F
B	$ABDF$	-	B
C	B	-	C
D	-	F	D
F	-	-	F
$ABDF$	$ABDF$	ACF	
ACF	B	ACF	

d'on directament, després d'eliminar els estats no abastables s'arriba al mateix autòmat determinista que en el cas anterior on els antics estats Q_1, Q_2 i ε, són ara $ABDF$, ACF i F, respectivament.

$\square$

2.3 Expressions regulars

Les anomenades **expressions regulars** són fòrmules que fan servir operacions simples sobre símbols que serveixen per definir conjunts de cadenes que reben el nom de **conjunts regulars**.

Aquestes expressions van ser introduïdes de forma independent dels autòmats i les gramàtiques i tenen un gran interés tant teòric com pràctic.

2.3.1 Conjunts i expressions regulars

Una expressió regular representa un únic conjunt de cadens sobre un determinat alfabet. L'operador binari "+" aplicat sobre expressions regulars representarà la unió dels corresponents conjunts mentre que els operadors "·" i "*" (aquest últim unari), representaran la concatenació i tancament dels corresponents conjunts, respectivament.

La definició formal d'expressió regular és una descripció exhaustiva de les seues possibilitats en forma recursiva.

Definició 2.8 *Siga* Σ *un determinat alfabet:*

1. ε, $\emptyset$ i a són expressions regulars, $\forall a \in \Sigma$
2. si α i β són expressions regulars, aleshores $\alpha + \beta$ també.
3. si α i β són expressions regulars, aleshores $\alpha \cdot \beta$ o $\alpha\beta$, també.
4. si α és una expressió regular, aleshores $\alpha^\star$ també.
5. només són expressions regulars les que s'obtenen mijantçant l'aplicació de les regles anteriors.

□

Si escrivim com a $L(\alpha)$ el conjunt regular representat per α, aleshores es té

1. $L(\varepsilon) = \{\varepsilon\}$, $L(\emptyset) = \emptyset$, i $L(a) = \{a\}$, $\forall a \in \Sigma$.
2. $L(\alpha + \beta) = L(\alpha) \cup L(\beta)$.
3. $L(\alpha\beta) = L(\alpha) \cdot L(\beta)$.
4. $L(\alpha^\star) = L(\alpha)^\star$.

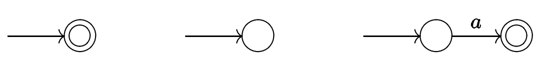
Alguns exemples de conjunts regulars sobre $\{a, b\}$ són els següents:

- $(a + b)^\star$ representa el monoide lliure.
- $a(a + b)^\star$ representa totes les paraules que comencen per a.
- $(a + b)((a + b)(a + b))^\star$ representa totes les paraules de longitud senar.

2.3.2 Equivalències

Teorema 2.5 Donada qualsevol expressió regular, γ, el llenguatge, $L(\gamma)$ és regular.
La demostració es fa per inducció sobre la "longitud" de l'expressió regular i consisteix a trobar un autòmat finit equivalent:

B.I. Les úniques expressions de longitud 1 són, ε, $\emptyset$ i a. I per a aquestes existeixen els següents autòmats que són trivialment equivalents.



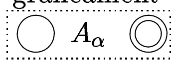
H.I. Qualsevol expressió regular de longitud menor estrictament que n representa un llenguatge regular.

P.I. Siga γ una expressió regular de longitud n. Podem haver-hi 4 casos:

1. $\gamma = (\alpha)$, amb la qual cosa $|\alpha| = n - 2$ i $L(\gamma) = L(\alpha)$ és regular per hipòtesi d'inducció.
2. $\gamma = \alpha + \beta$, amb la qual cosa tant α com β han de tenir longituts menors o iguals que $n - 2$ i representaran llenguatges regulars per hipòtesi d'inducció.

   Suposem que A_α i A_β són autòmats finits que accepten $L(\alpha)$ i $L(\beta)$, respectivament.

   L'autòmat construït mijantçant els dos autòmats anteriors i un nou estat inicial des del qual eixen dues transicions buides, una a cada un dels vells estats inicials, accepta trivialment el llenguatge $L(\gamma) = L(\alpha + \gamma) = L(\alpha) \cup L(\beta)$.
3. $\gamma = \alpha \cdot \beta$, en aquest cas es faria una construcció similar amb els autòmats corresponents a α i β però ara el nou estat inicial seria l'inicial de α, els nous finals serien ***només*** els de β i els vells finals de α estarien units al vell inicial de β mijantçant transicions buides.
4. $\gamma = \alpha^\star$, en aquest cas, $L(\alpha)$ també és regular per hipòtesi d'inducció i la construcció que caldria fer consistiria a afegir un nou estat inicial que seria també final (per acceptar ε) unit mijantçant una transició buida al vell estat inicial. A aquest estat arribarien a més a més transicions buides des de tots els estats finals de A_α.

Les tres construccions emprades en aquest teorema es mostren gràficament en la figura 2.3. L'autòmat associat a cada expressió regular, A_α es mostra com a , on el cercle de l'esquerra representa l'estat inicial i el cercle de la dreta representa ***tots*** els estats finals. En les construccions, només els estats finals definitius es representen com a cercles dobles. Una transició que surt del cercle de la dreta d'un autòmat representa una transició per cada estat final original en aquest autòmat.

□

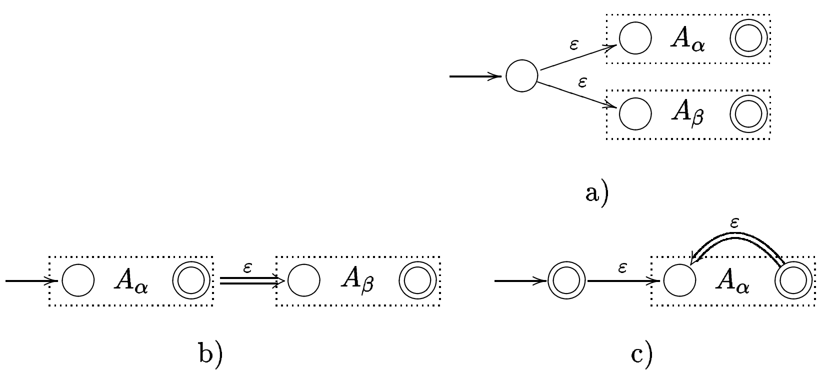


Figura 2.3. Combinacions d'autòmats en funció de A_α i A_β. a) $A_{\alpha+\beta}$, b) $A_{\alpha\beta}$, i c) $A_{\alpha^\star}$.

Teorema 2.6 Donat un autòmat finit, A, existeix una expressió regular, α_A tal que $L(A) = L(\alpha_A)$.

Siga $A = \langle Q, \Sigma, \delta, q_1, F\rangle$ on $Q = \{q_i\}_{i=1}^n$ i definim com a R_{ij}^k el conjunt de cadenes de $\Sigma^\star$ que permeten a l'autòmat passar des de l'estat q_i fins a l'estat q_j sense passar per cap estat "major" que k com a estat intermig. És a dir,

$$R_{ij}^k = \{x \in \Sigma^\star \mid q_j \in \delta(q_i, x) \ \wedge \ \forall q_\ell, \forall y \in Pr(x) \ : \ q_\ell \in \delta(q_i, y), \ \ell \leq k\}$$

Si $k = 0$ els conjunts corresponents són finits i venen donats per

$$R_{ij}^0 = \{a \in \Sigma \mid q_j \in \delta(q_i, a) \cup E\}$$

on $E = \{\varepsilon\}$ si $i = j$ o $E = \emptyset$ si no.

Per a qualsevol altre valor de k, aquests conjunts es poden definir de forma recursiva segons

$$R_{ij}^k = R_{ij}^{k-1} \cup R_{ik}^{k-1} (R_{kk}^{k-1})^\star R_{kj}^{k-1}$$

Aquesta expressió recursiva il·lustra el fet que les cadenes noves de R_{ij}^k que no estan en R_{ij}^{k-1} són exactament aquelles que per a anar de i a j passen una o més vegades per l'estat k. Pertant, es poden factoritzar com a les cadenes que arriben a k, les que iteren sobre k zero o més vegades i les que van des de k a j i en tots els casos fent servir estats intermijos inferiors a k. Aquesta factorització es mostra gràficament en la figura 2.4.

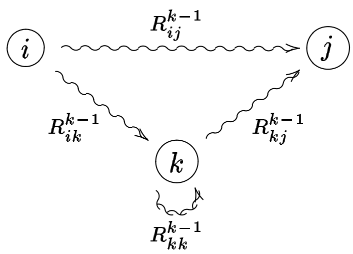


Figura 2.4. Il·lustració de la factorització que dóna lloc a la definició recursiva dels conjunts R_{ij}^k.

A continuació es demostrarà per inducció sobre k, que tots els R_{ij}^k són conjunts regulars.

B.I. per a tot R_{ij}^0 existeix una expressio r_{ij}^0 que representa un nombre finit de cadenes d'un sol símbol en funció de què existesquen transcicions entre els estats i i j o no.

H.I. Suposarem que es compleix que $R_{ij}^{k-1} = L(r_{ij}^{k-1})$ per a alguna expressió regular r_{ij}^{k-1}.

P.I. Considerem el conjunt R_{ij}^k. Segons la definició recursiva d'abans, aquest conjunt es pot representar mijantçant la unió, concatenació i tancament de conjunts que compleixen la hipòtesi d'inducció. Aleshores,

$$\begin{aligned} R_{ij}^{k} &\overset{(def)}{=} R_{ij}^{k-1} \cup R_{ik}^{k-1}(R_{kk}^{k-1})^{\star} R_{kj}^{k-1} \\ &\overset{(HI)}{=} L(r_{ij}^{k-1}) \cup L(r_{ik}^{k-1}) L(r_{kk}^{k-1})^{\star} L(r_{kj}^{k-1}) \\ &= L(r_{ij}^{k-1} + r_{ik}^{k-1}(r_{kk}^{k-1})^{\star} r_{kj}^{k-1}) \end{aligned}$$

Per tant, $r_{ij}^{k-1} + r_{ik}^{k-1}(r_{kk}^{k-1})^{\star} r_{kj}^{k-1}$ és una expressió regular que representa el conjunt R_{ij}^{k}.

Com que per a l'autòmat, A es compleix que $L(A) = \bigcup_{q_j \in F} R_{1j}^{n}$ aleshores es té que el llenguatge acceptat per A es pot representar mijantçant l'expressió regular

$$\sum_{j\ :\ q_j \in F} r_{ij}^{n}$$

on r_{ij}^{n} és l'expressió regular corresponent a R_{ij}^{n}.

□

2.3.3 Càlcul de l'expressió regular equivalent a un autòmat

Encara que la demostració del teorema anterior es pot fer servir per calcular l'expressió regular equivalent a un autòmat qualsevol, la seua aplicació resulta normalment llarga i tediosa.

Una forma més gràfica de resoldre el mateix problema consisteix en la definició del que anomenarem autòmats d'expressions regulars.

> Un **autòmat d'expressions regulars** és un autòmat en el qual les transicions estan etiquetades amb expressions regulars. L'autòmat passa d'un estat a un altre en llegir de la cadena d'entrada qualsevol prefix que estiga en el conjunt regular de la corresponent transició.

Per exemple, ...

Donat un autòmat d'expressions regulars, sempre es pot eliminar un dels seus estats i modificar les transicions adjacents de forma que el llenguatge acceptat no canvie (sempre que l'estat eliminat no siga final).

En particular, si l'estat eliminat, k, serveix de camí entre els estats i i j, com en la figura REF, aleshores l'expressió regular que etiqueta la transició entre i i j s'ha de canviar afegint-hi les cadenes que permetien anar des de i fins a j passant per k (figura REF).

FIGURA

Si l'estat que s'elimina és l'inicial, aleshores cal etiquetar amb una expressió regular la "transició" buida que marca l'estat inicial com es mostra a la figura REF.

Si en un autòmat d'expressions regulars s'eliminen tots els estats llevat d'un, l'autòmat resultant accepta totes les cadenes que, originalment, anaven des de l'estat inicial fins a aquest estat.

Aquest fet, que es pot demostrar trivialment ens dóna una forma grafica de calcular els conjunts R^n_{1j} en forma d'expressions regulars, i obre pertant la possibilitat de calcular d'aquesta manera l'expressió regular equivalent a un autòmat.

EXEMPLE:

...