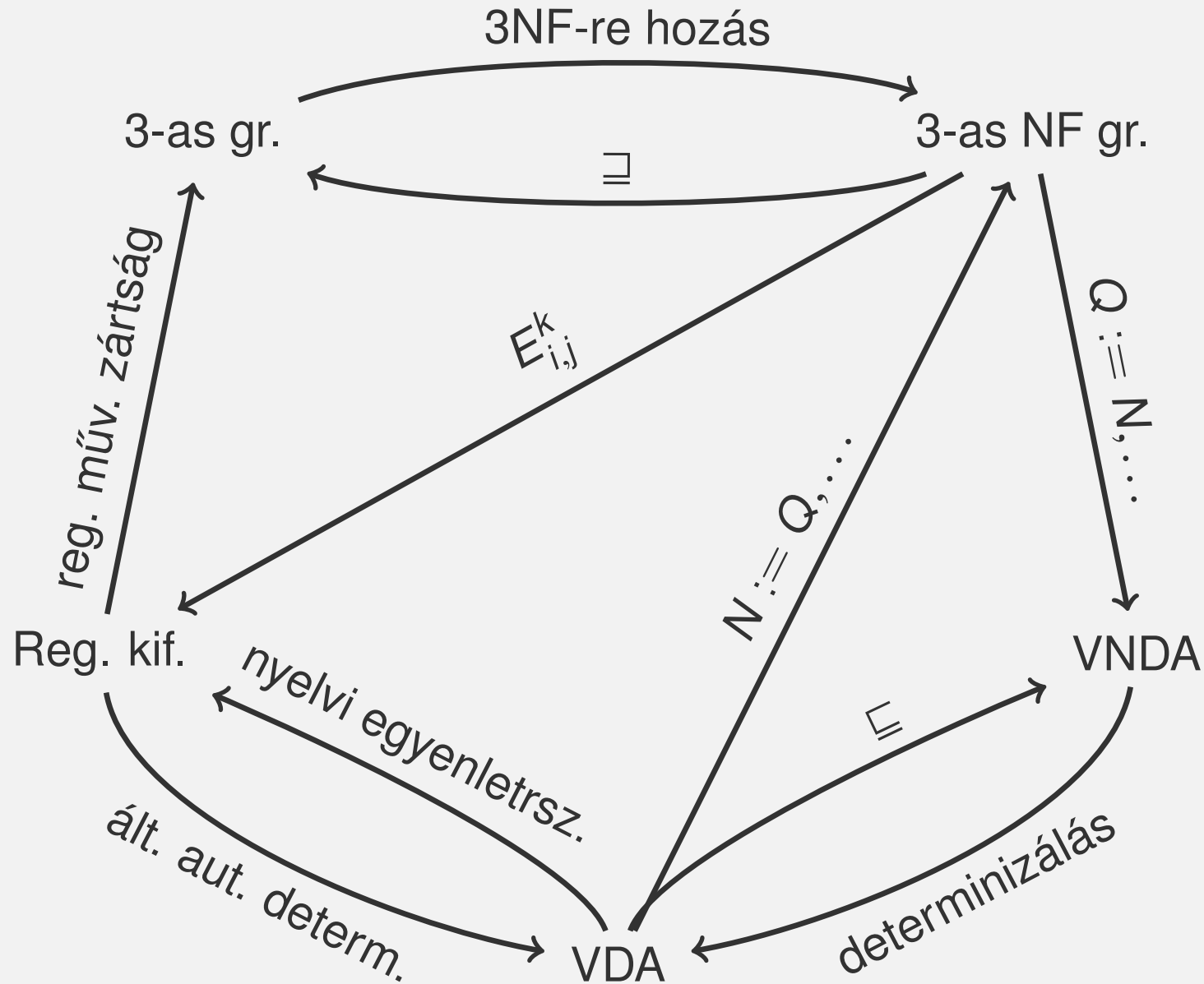


A számításelmélet alapjai I.

7. előadás

előadó: Tichler Krisztián
ktichler@inf.elte.hu

\mathcal{L}_3 -at leíró formális eszközök



Lineáris nyelvi egyenletrendszerek megoldása

(Emlékeztető:)

Arden tétele

Adottak az R és Q reguláris kifejezések. Az $X = Q \cdot X + R$ egyenletnek X -re vonatkozó megoldása $X = Q^* \cdot R$. Amennyiben $\varepsilon \notin L(Q)$, akkor az egyenletnek ez az egyetlen megoldása.

Arden tételét általánosabb alakban is kimondhatjuk (a bizonyítás ugyanaz):

Arden tétele

Legyenek az L_1 és L_2 nyelvek adottak. Az $X = L_1 X \cup L_2$ egyenletnek X -re vonatkozó megoldása $X = L_1^* L_2$. Amennyiben $\varepsilon \notin L_1$, akkor az egyenletnek ez az egyetlen megoldása.

A zártsági tétel miatt ha L_1 és L_2 \mathcal{L}_i -beli, akkor X is az ($i \in \{0, 1, 2, 3\}$).

Lineáris nyelvi egyenletrendszerek megoldása

Tétel

Legyen $\mathbf{x} = \mathbf{M}\mathbf{x} \cup \mathbf{v}$ nyelvi egyenletrendszer, ahol az

$$\mathbf{M} = \begin{bmatrix} L_{11} & \cdots & L_{1n} \\ \vdots & \ddots & \vdots \\ L_{n1} & \cdots & L_{nn} \end{bmatrix} \text{ nyelvmátrix és a } \mathbf{v} = \begin{bmatrix} L_1 \\ \vdots \\ L_n \end{bmatrix} \text{ nyelvvektor}$$

adottak és az $\mathbf{x} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$ ismeretlen nyelvekből álló vektort

keressük. Ha $\mathcal{L} = \bigcup_{j=1}^n \bigcup_{k=1}^n \{L_{jk}\} \cup \bigcup_{j=1}^n \{L_j\} \subseteq \mathcal{L}_i$ ($i \in \{0, 1, 2, 3\}$) és

$\varepsilon \notin L_{jk}$ ($1 \leq j \leq n, 1 \leq k \leq n$), akkor az egyenletrendszernek egyértelműen létezik \mathcal{L}_i -beli megoldása, és ez a megoldás \mathcal{L} elemeiből a reguláris nyelvműveletekkel épül fel.

Megjegyzés: Két nyelvmátrix szorzata esetén nyelvek szorzatának a konkatenáció, összegnek az unió művelet felel meg. A j . egyenlet tehát $X_j = L_{j1}X_1 \cup \cdots \cup L_{jn}X_n \cup L_j$.

Lineáris nyelvi egyenletrendszerek megoldása

Bizonyításvázlat:

Gauss eliminációval redukáljuk az egyenletek és az ismeretlenek számát:

Az utolsó egyenlet: $X_n = L_{nn}X_n \cup \bigcup_{k=1}^{n-1} L_{nk}X_k \cup L_n$.

Arden tétele alapján az utolsó egyenletből

$X_n = L_{nn}^*(\bigcup_{k=1}^{n-1} L_{nk}X_k \cup L_n)$, amiből az $\mathbf{x}' = \mathbf{M}'\mathbf{x}' \cup \mathbf{v}'$ nyelvi

egyenletrendszer adódik az $\mathbf{x} = \begin{bmatrix} X_1 \\ \vdots \\ X_{n-1} \end{bmatrix}$ ismeretlenekre, ahol

$$\mathbf{M}' = \begin{bmatrix} L'_{11} & \cdots & L'_{1(n-1)} \\ \vdots & \ddots & \vdots \\ L'_{(n-1)1} & \cdots & L'_{(n-1)(n-1)} \end{bmatrix} \text{ nyelvmátrix és } \mathbf{v}' = \begin{bmatrix} L'_1 \\ \vdots \\ L'_{n-1} \end{bmatrix},$$

továbbá $L'_{jk} = L_{jk} \cup L_{jn}L_{nn}^*L_{nk}$ és $L'_j = L_j \cup L_{jn}L_{nn}^*L_n$.

Lineáris nyelvi egyenletrendszerek megoldása

Egyel kevesebb egyenlet, egyel kevesebb ismeretlen. Látható, hogy $\varepsilon \notin L'_{jk}$ továbbá $L'_{jk} \in \mathcal{L}_i$ a zártsági tétel miatt.

Végül egyetlen, egyismeretlenes lineáris nyelvi egyenletünk marad, amit Arden tétele alapján megoldhatunk. A többi ismeretlen értéke visszahelyettesítéssel adódik.

Az egyértelműségekre vonatkozó állítás teljes indukcióval adódik Arden tételének egyértelműségekre vonatkozó állításából és abból, hogy minden egyes lépés után a nyelvmátrixok egyetlen nyelve sem tartalmazza ε -t.

Lineáris nyelvi egyenletrendszerek megoldása

Példa:

Oldjuk meg az
$$\begin{aligned} X &= a^*bX + b^*aY + ba \\ Y &= b^*aX + a^*bY + ab \end{aligned}$$
 egyenletrendszert!

Megoldás:

Az első egyenletből kifejezzük X -et Arden tétele alapján:

$$X = (a^*b)^*(ba + b^*aY).$$

$$Y = b^*a(a^*b)^*(ba + b^*aY) + a^*bY + ab$$

$$Y = (b^*a(a^*b)^*b^*a + a^*b)Y + (b^*a(a^*b)^*ba + ab), \text{ amiből}$$

$$Y = (b^*a(a^*b)^*b^*a + a^*b)^*(b^*a(a^*b)^*ba + ab).$$

$$\text{Hasonlóan: } X = (b^*a(a^*b)^*b^*a + a^*b)^*(b^*a(a^*b)^*ab + ba).$$

A megoldás egyértelmű.

Automata adott állapatra vonatkozó maradéknyelve

(Emlékeztető:)

Egy $A = \langle Q, T, \delta, q_0, F \rangle$ véges determinisztikus automata egy $q \in Q$ **állapotára vonatkozó maradéknyelve** alatt az $L(A, q) := \{v \in T^* \mid \delta(q, v) \in F\}$ nyelvet értettük.

Az $L(A, q)$ tehát azokat a v szavakat tartalmazza, amelyek hatására az automata q -ból végállapotba kerül.

Tehát $L(A, q_0) = L(A)$ és $\varepsilon \in L(A, q) \iff q \in F$.

Vegyük továbbá észre, hogy fennállnak a következő nyelvi egyenletek:

$$L(A, q) = \bigcup_{t \in T} tL(A, \delta(q, t)) \cup \begin{cases} \emptyset & \text{ha } q \notin F \\ \{\varepsilon\} & \text{ha } q \in F \end{cases}$$

ahol $tL := \{tu \mid u \in L\}$ ($t \in T$).

A maradéknyelvekre vonatkozó egyenletrendszer megoldjuk ($L(A, q_0)$ az érdekes). A megoldás egyértelműsége az előző tételből, de az $L(A, q)$ -k egyértelműségéből is adódik.

VDA-hoz reguláris kifejezés

Példa:

	a	b	c
$\rightarrow q_0$	q_1	q_2	q_3
q_1	q_4	q_2	q_4
q_2	q_0	q_4	q_3
$\leftarrow q_3$	q_4	q_3	q_4
q_4	q_4	q_4	q_4

Megoldás: Az egyenletrendszer:

$$X := L(A, q_0), \quad Y := L(A, q_1),$$

$$Z := L(A, q_2), \quad V := L(A, q_3),$$

$$\text{Észrevétel: } L(A, q_4) = \emptyset$$

$$X = aY + bZ + cV$$

$$Y = bZ$$

$$Z = aX + cV$$

$$V = bV + \varepsilon$$

$$X = ?$$

$$V = b^*, \quad Z = aX + cb^*, \quad Y = b(aX + cb^*),$$

$$X = ab(aX + cb^*) + b(aX + cb^*) + cb^*,$$

$$X = (aba + ba)X + (abc + bc + c)b^*,$$

$$X = (aba + ba)^*(abc + bc + c)b^*.$$

Általánosított automata

Legyen $\mathcal{R}(T)$ a T feletti reguláris kifejezések halmaza.

Definíció

Az **általánosított automata** defíciója megegyezik a véges nemdeteminisztikus automata definíciójával azzal a kivétellel, hogy δ átmenetfüggvényére $\delta : Q \times \mathcal{R}(T) \rightarrow \mathcal{P}(Q)$ teljesül.

Átmenetdiagramon ábrázolva az általánosítás az, hogy az élek címkéi betűk helyett tetszőleges T feletti reguláris kifejezések lehetnek.

Általánosított automata által felismert nyelv

Definíció

Legyen $A = \langle Q, T, \delta, Q_0, F \rangle$ egy általánosított automata és legyenek $u, v \in QT^*$ konfigurációk. Az A automata az u szót **egy lépésben** (közvetlenül) a v szóra **redukálja** (jelölés: $u \Rightarrow_A v$), ha van olyan $q, p \in Q, z, w \in T^*, R \in \mathcal{R}(T)$, hogy $p \in \delta(q, R)$, $z \in L(R)$, $u = qzw$ és $v = pw$ teljesül.

A többlépéses redukció és a felismert nyelv fogalma nem változik.

$u \Rightarrow_A^* v$, ha $u = v$ vagy valamely $k \geq 1$ -re léteznek w_0, \dots, w_k konfigurációk melyekre $w_{i-1} \Rightarrow_A w_i$ ($1 \leq i \leq k$), $w_0 = u$ és $w_k = v$.

$L(A) = \{u \in T^* \mid q_0 u \Rightarrow_A^* p \text{ valamely } q_0 \in Q_0\text{-ra és } p \in F\text{-re}\}$

Általánosított automata

Észrevétel: Az $A = \langle Q, T, \delta, Q_0, F \rangle$ általánosított automata elfogadja az $u \in T^*$ szót $\Leftrightarrow \exists n \in \mathbb{N}, u_1, \dots, u_n \in T^*, R_1, \dots, R_n \in \mathcal{R}(T), q_0 \in Q_0$ és $q_1, \dots, q_n \in Q$, melyre $u = u_1 \cdots u_n$, továbbá minden $1 \leq i \leq n$ esetén $u_i \in L(R_i), q_i \in \delta(q_{i-1}, R_i)$ és $q_n \in F$.

Észrevétel: Átmenetdiagramos reprezentáció esetén A akkor és csak akkor fogadja el az $u \in T^*$ szót, ha létezik irányított séta valamely $q_0 \in Q_0$ -ból valamely F -beli állapotba úgy, hogy ha a séta mentén az élek címkéje az élek sorrendjében rendre R_1, \dots, R_n , akkor $u \in L(R_1 R_2 \cdots R_n)$.

Definíció

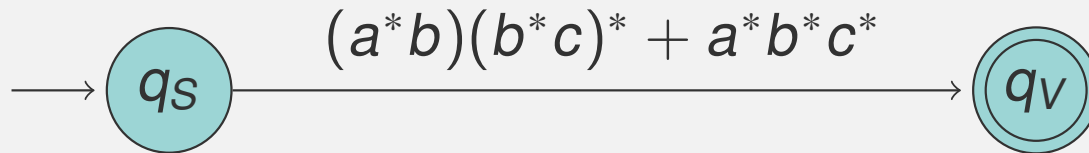
Az ε -átmenetes, véges nemdeterminisztikus automata (ε VNDA) olyan általánosított automata, ahol minden átmenet $T \cup \{\varepsilon\}$ -beli.

Tehát átmenetfüggvénye az input betűnkénti feldolgozása mellett állapotváltásokat engedélyezhet (ε -átmenetek).

Reguláris kifejezéshez VDA készítése

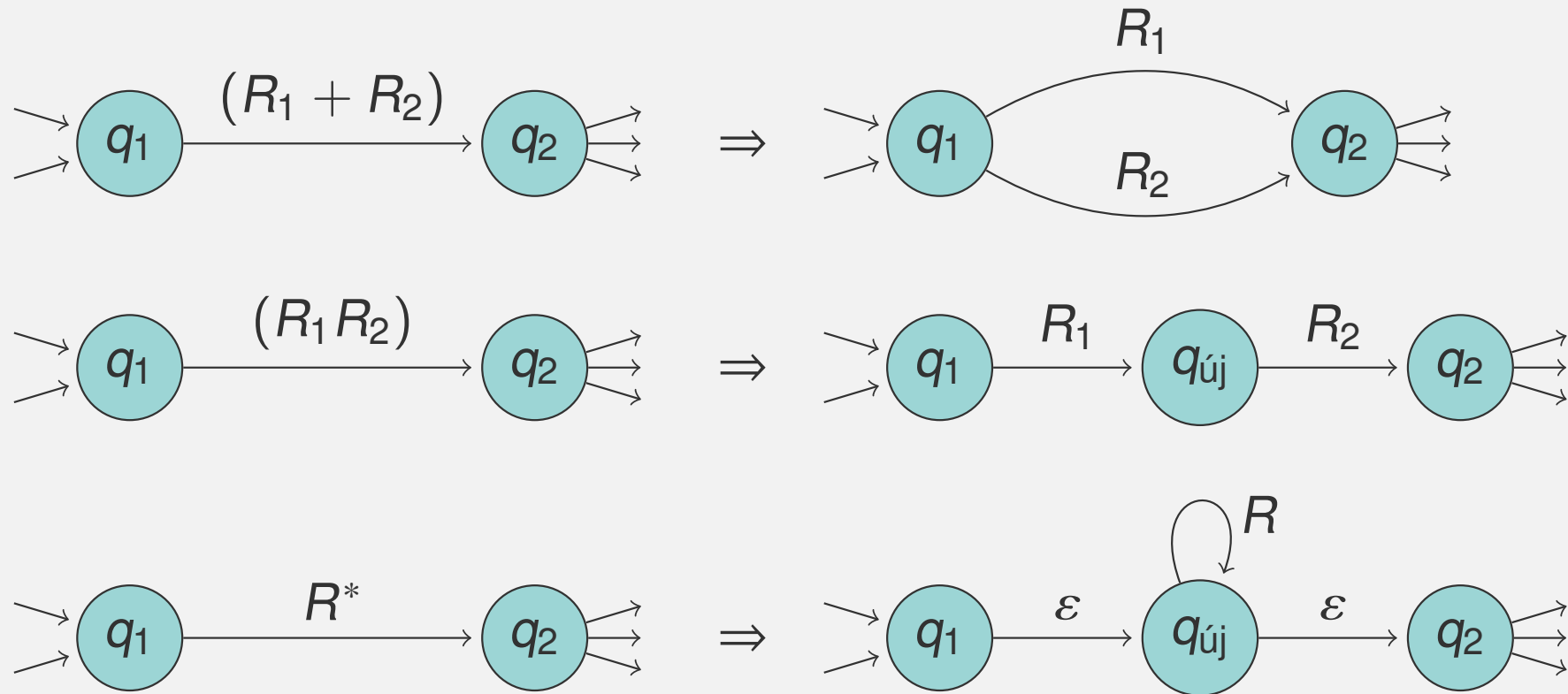
Példa: Legyen $R = (a^*b)(b^*c)^* + a^*b^*c^*$. Készítsünk olyan A VDA-t, hogy $L(A) = L(R)$!

0. lépés: Adott R reguláris kifejezéshez kiindulunk egy $A = \langle \{q_S, q_V\}, T, \delta, \{q_S\}, \{q_V\} \rangle$ általánosított automatából, ahol $\delta(q_S, R) = \{q_V\}$ az egyetlen átmenet. Nyilvánvalóan $L(A) = L(R)$.



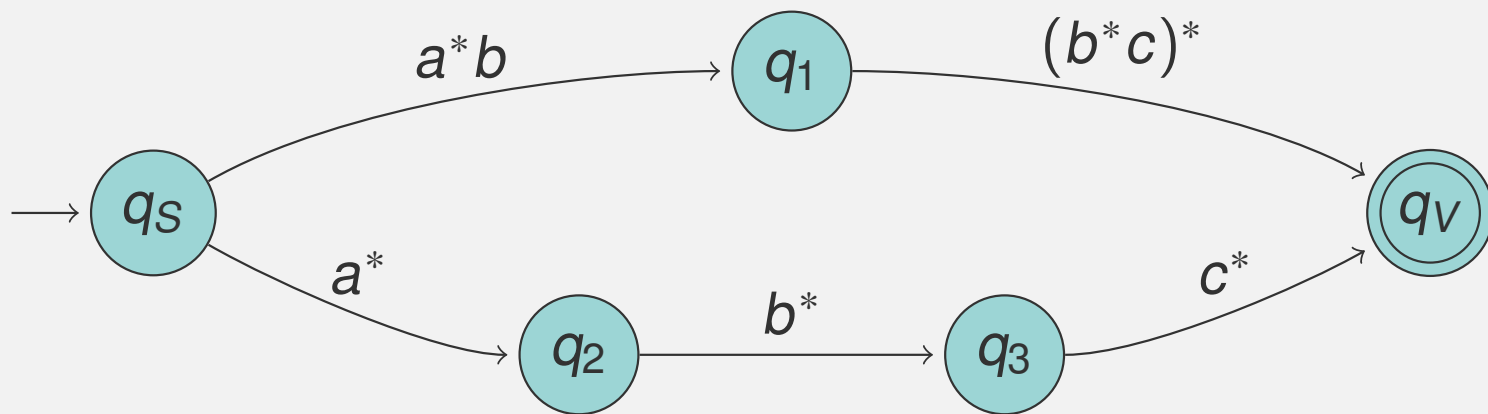
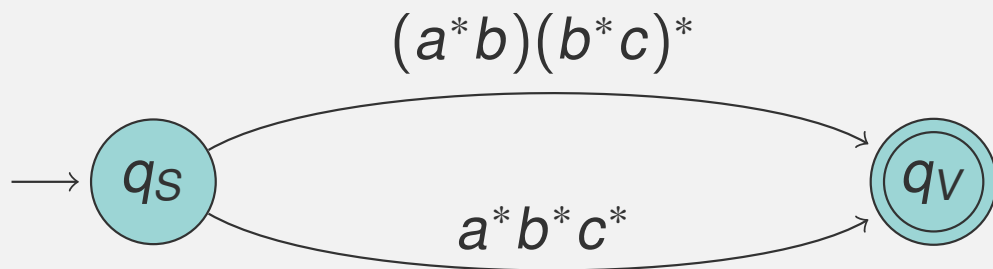
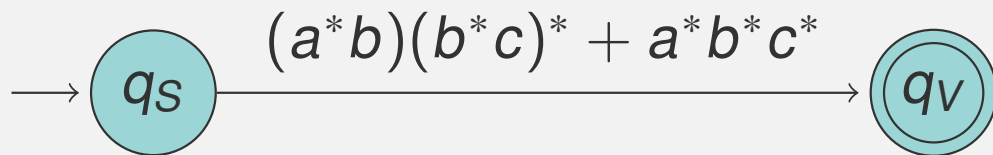
Reguláris kifejezéshez VDA készítése

1. lépés: Az általánosított automata alábbi transzformációi nem változtatják $L(A)$ -t.

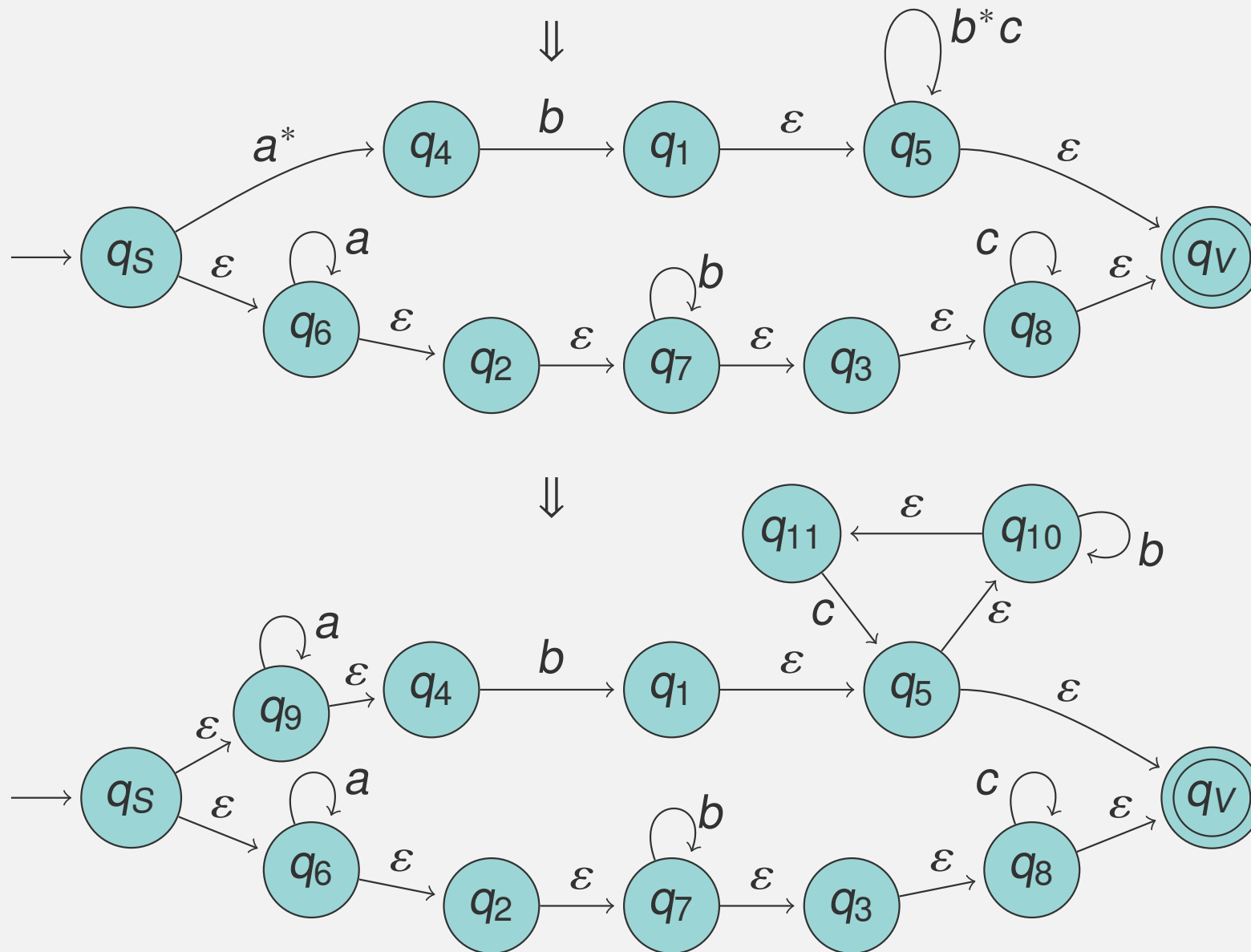


Például a 3. esetben meggondolható, hogy bármelyik új él használata implikálja a $(q_1, q_{új})$ egyszeres, valamely $k \in \mathbb{N}$ -re a $(q_{új}, q_{új})$ él k -szoros és $(q_{új}, q_2)$ él egyszeres használatát ebben a sorrendben, ami megfelel egy eredeti R^* címkéjű (q_1, q_2) élhasználatnak.

Reguláris kifejezéshez VDA készítése



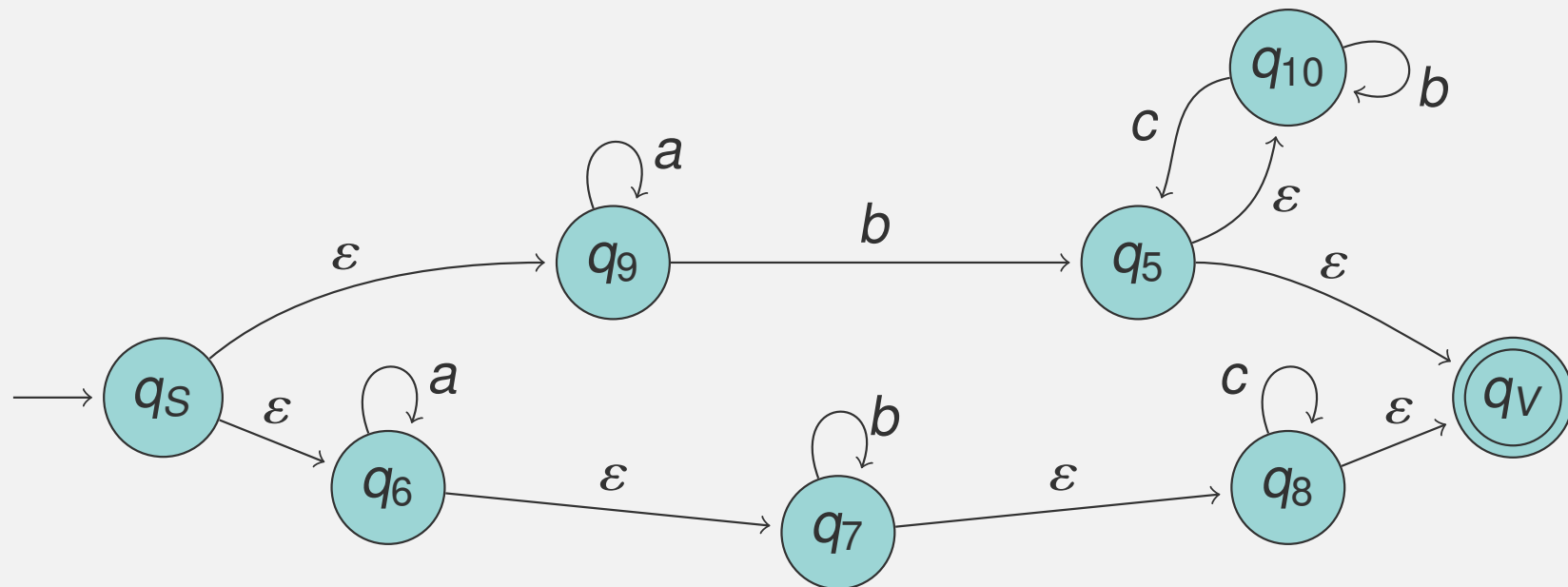
Reguláris kifejezéshez VDA készítése



Reguláris kifejezéshez VDA készítése

Egyszerűsítési lehetőség: Ha $\delta(q, t) = \{q'\}$ és $\delta(q', t') = \{q''\}$ ($t, t' \in T \cup \{\varepsilon\}$, legalább az egyikük ε) az **egyetlen** átmenet q' -be illetve q' -ből, akkor q' elhagyható, és legyen $q'' \in \delta(q, tt')$.

Például adott esetben $q_1, q_4, q_2, q_3, q_{11}$ elhagyható.



Reguláris kifejezéshez VNDA készítése

2. lépés: ε VNDA-ból VNDA

Legyen $A = \langle Q, T, \delta, Q_0, F \rangle$ ε VNDA. Határozzuk meg minden $q \in Q$ állapotra azon állapotok $H(q)$ halmazát, amelyekbe az adott állapotból ε -átmenetekkel eljuthatunk.

Azaz $H(q) = \{q' \in Q \mid q \Rightarrow_A^* q'\}$.

$H_0(q) := \{q\}$. $H_{i+1}(q) := H_i(q) \cup \{q' \in Q \mid \exists r \in H_i(q) : r \Rightarrow_A q'\}$.

$H_0(q) \subseteq H_1(q) \subseteq \dots \subseteq Q$.

A sorozat stabilizálódik és ez a $H(q)$ halmaz. (Meggondolható, hogy csak q -ból ε átmenetekkel elérhető állapotok kerülnek $H_i(q)$ -ba, és minden ilyen be is kerül valamely $i \in \mathbb{N}$ -re.)

$A' = \langle Q, T, \delta', Q_0, F' \rangle$ VNDA-ban

$r \in \delta'(q, t) \iff \exists q' \in H(q) : r \in \delta(q', t) \quad (q \in Q, t \in T)$

$F' := \{q \in Q \mid H(q) \cap F \neq \emptyset\}$.

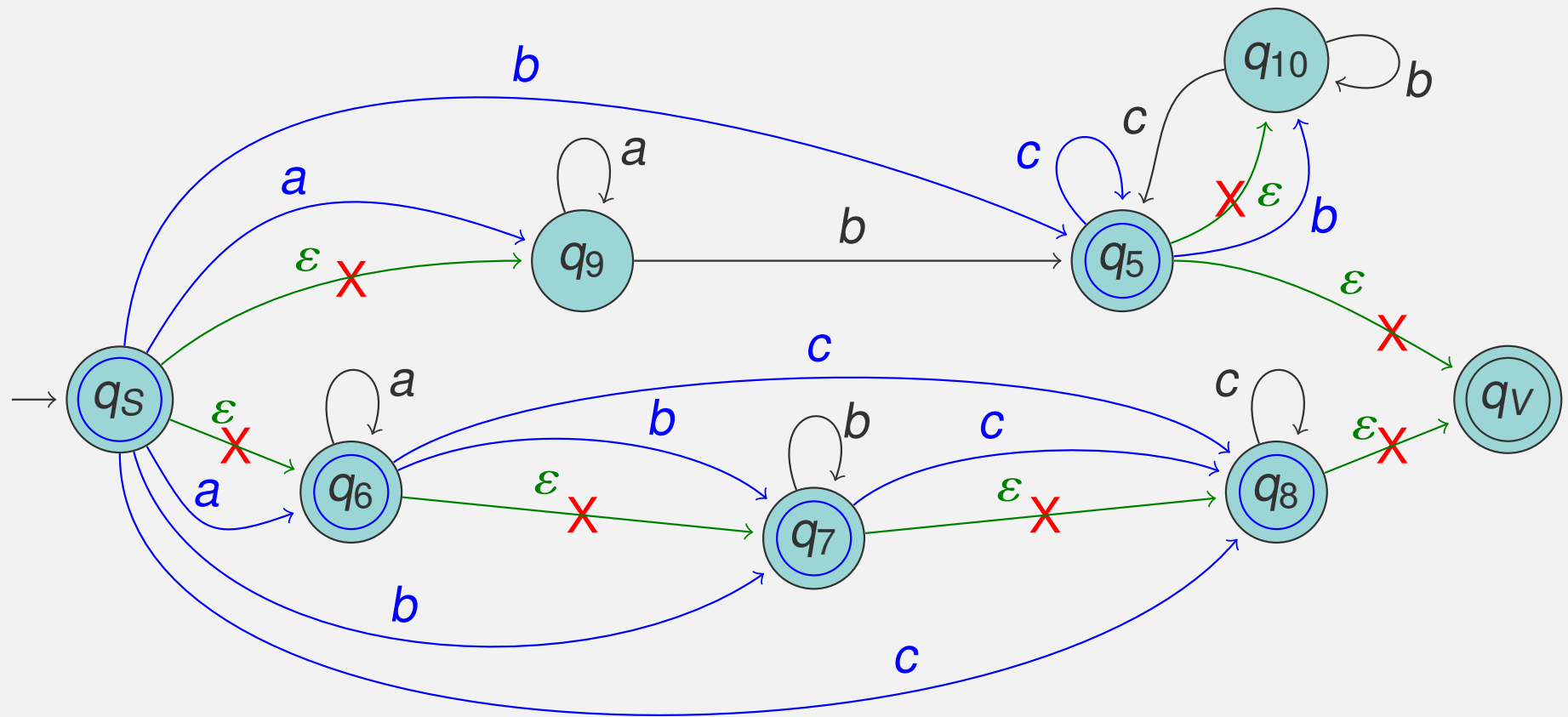
Reguláris kifejezéshez VDA készítése

$L(A') \subseteq L(A)$, hiszen egy $r \in \delta'(q, t)$ új szabály alkalmazásának megfeleltethető valamely $q' \in H(q)$ -ra a $q \Rightarrow_A^* q'$ ε -szabályok és az $r \in \delta(q', t)$ eredeti szabály egymás utáni alkalmazása.

Másrészt, ily módon az új, ε -átmenetek mentén „visszahúzott” szabályokkal azon ε -szabályok alkalmazása elkerülhető, amelyek után alkalmazunk még valamilyen $t \in T$ -re szabályt.

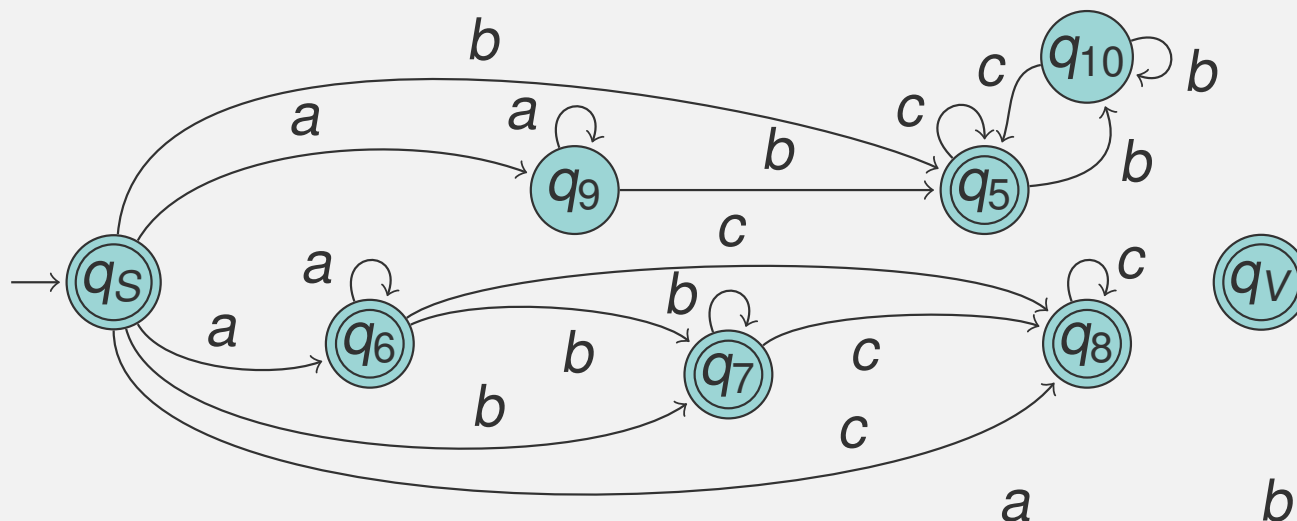
Azonban előfordulhat, hogy redukció végén lévő ε -szabályok alkalmazása után már semmilyen $t \in T$ -re vonatkozó átmenetet sem használunk A -ban. Azonban az állapotok F -beliségének ε -átmeneteken történő „visszahúzásával” ezeknek az ε -átmeneteknek a használata is elkerülhető. Így $L(A) \subseteq L(A')$.

Reguláris kifejezéshez VDA készítése



Reguláris kifejezéshez VDA készítése

3. lépés: VNDA-hoz VDA



		a	b	c
\Leftrightarrow	$\{q_5\}$	$\{q_6, q_9\}$	$\{q_5, q_7\}$	$\{q_8\}$
\leftarrow	$\{q_6, q_9\}$	$\{q_6, q_9\}$	$\{q_5, q_7\}$	$\{q_8\}$
\leftarrow	$\{q_5, q_7\}$	$\{\}$	$\{q_7, q_{10}\}$	$\{q_5, q_8\}$
\leftarrow	$\{q_8\}$	$\{\}$	$\{\}$	$\{q_8\}$
	$\{\}$	$\{\}$	$\{\}$	$\{\}$
\leftarrow	$\{q_7, q_{10}\}$	$\{\}$	$\{q_7, q_{10}\}$	$\{q_5, q_8\}$
\leftarrow	$\{q_5, q_8\}$	$\{\}$	$\{q_{10}\}$	$\{q_5, q_8\}$
	$\{q_{10}\}$	$\{\}$	$\{q_{10}\}$	$\{q_5\}$
\leftarrow	$\{q_5\}$	$\{\}$	$\{q_{10}\}$	$\{q_5\}$

Kis Bar-Hillel lemma

Kis Bar-Hillel lemma

Minden $L \in \mathcal{L}_3$ nyelvhez van olyan $n \in \mathbb{N}$ konstans, hogy minden $w \in L$ szó esetén ha tekintjük egy tetszőleges olyan $w = uw'v$ felbontását, ahol $|w'| \geq n$, akkor van w' -nek olyan y részszava ($w' = xyz$), hogy $0 < |y| \leq n$, és minden $i \geq 0$ esetén $uxy^izv \in L$.

Kevésbé formálisan: L minden szavának elég hosszú részszavában létezik elég rövid, nemüres, beiterálható részszava.

Bizonyítás:

Mivel $L \in \mathcal{L}_3$, ezért létezik $A = \langle Q, T, \delta, q_0, F \rangle$ véges, determinisztikus automata, amelyre $L(A) = L$ teljesül. Legyen $n = |Q|$.

Tekintsünk egy $w \in L$ szót és egy olyan $w = uw'v$ felbontását, ahol $|w'| \geq n$. Legyen $w' = t_1 \cdots t_m$, ahol $t_i \in T$ ($1 \leq i \leq m$) és $m \geq n$.

Kis Bar-Hillel lemma

Ekkor $q_0 w \Rightarrow_A^* r_0 t_1 t_2 \cdots t_m v \Rightarrow_A r_1 t_2 \cdots t_m v \Rightarrow_A \cdots \Rightarrow_A r_m v \Rightarrow_A^* p$
valamely r_0, \dots, r_m és $p \in F$ állapotokra.

A skatulyaelv alapján biztosan léteznek k, j számok, melyekre $0 \leq j < k \leq n$ és $r_j = r_k$, ugyanis $n = |Q| < m + 1$.

Legyen $x = t_1 \cdots t_j$, $y = t_{j+1} \cdots t_k$, $z = t_{k+1} \cdots t_m$. Világos, hogy j, k választása miatt $0 < |y| \leq n$.

Azt kell csak belátni, hogy minden $i \geq 0$ esetén $uxy^i z v \in L$.

Minden $i \geq 0$ esetén $r_j y^i \Rightarrow_A^* r_j$.

Valóban, $i = 0$ -ra $r_j \Rightarrow_A^* r_j$, $i = 1$ -re $r_j y \Rightarrow_A^* r_k = r_j$, innen i -re vonatkozó teljes indukcióval $r_j y^i \Rightarrow_A^* r_j y^{i-1} \Rightarrow_A^* r_j$.

Tehát minden $i \geq 0$ -ra

$q_0 u x y^i z v \Rightarrow_A^* r_j y^i z v \Rightarrow_A^* r_j z v = r_k z v \Rightarrow_A^* p$, tehát $uxy^i z v \in L$.

Kis Bar-Hillel lemma

Példa: Legyen $L = \{u \in \{a, b\}^* \mid u = u^{-1}\}$ az $\{a, b\}$ feletti palindrómák nyelve.

Tegyük fel, hogy $L \in \mathcal{L}_3$. Legyen $w = a^n b a^n$, ahol n a Kis Bar-Hillel lemma konstansa. Ekkor nyilván w palindróma, azaz $w \in L$. Legyen $u = \varepsilon$, $w' = a^n$, $v = b a^n$.

Mivel $|w'| \geq n$, ezért a Kis Bar-Hillel Lemma szerint van w' -nek olyan $w' = xyz$ felbontása, amelyre $0 < |y| \leq n$ és többek között a 0-adik iterált, azaz $uxy^0zv = uxzv \in L$.

De ez nem lehetséges, hiszen a felételekből következően $y = a^k$ valamely $0 < k \leq n$ -re, és így $uxzv = a^{n-k} b a^n$, ami $k > 0$ esetén $\notin L$ (nem palindróma).

Ellentmondásra jutottunk. Ez csak úgy lehetséges, hogy az $L \in \mathcal{L}_3$ feltevésünk hamis, azaz a palindrómák fenti L nyelve nem reguláris.