

## 2.6 SÄÄNNÖLLISET LAUSEKKEET

Automaattimalleista poikkeava tapa kuvata yksinkertaisia kieliä.

Olkoot  $A$  ja  $B$  aakkoston  $\Sigma$  kieliä. Perusoperaatioita:

▶ *Yhdiste*:  $A \cup B = \{x \in \Sigma^* \mid x \in A \text{ tai } x \in B\}$ ;

▶ *Katenaatio*:  $AB = \{xy \in \Sigma^* \mid x \in A, y \in B\}$ ;

▶ *Potenssit*:

$$\begin{cases} A^0 = \{\varepsilon\}, \\ A^k = AA^{k-1} = \{x_1 \dots x_k \mid x_i \in A \quad \forall i = 1, \dots, k\} \quad (k \geq 1); \end{cases}$$

▶ *Sulkeuma t. "Kleenen tähti"*:

$$\begin{aligned} A^* &= \bigcup_{k \geq 0} A^k \\ &= \{x_1 \dots x_k \mid k \geq 0, x_i \in A \quad \forall i = 1, \dots, k\}. \end{aligned}$$

Kukin  $\Sigma$ :n säännöllinen lauseke  $r$  kuvaa kielen  $L(r)$ , joka määritellään:

- ▶  $L(\emptyset) = \emptyset$ ;
- ▶  $L(\varepsilon) = \{\varepsilon\}$ ;
- ▶  $L(a) = \{a\}$  kaikilla  $a \in \Sigma$ ;
- ▶  $L((r \cup s)) = L(r) \cup L(s)$ ;
- ▶  $L((rs)) = L(r)L(s)$ ;
- ▶  $L(r^*) = (L(r))^*$ .

**Määritelmä 2.3** Aakkoston  $\Sigma$  säännölliset lausekkeet määritellään induktiivisesti säännöllillä:

1.  $\emptyset$  ja  $\varepsilon$  ovat  $\Sigma$ :n säännöllisiä lausekkeita;
2.  $a$  on  $\Sigma$ :n säännöllinen lauseke kaikilla  $a \in \Sigma$ ;
3. jos  $r$  ja  $s$  ovat  $\Sigma$ :n säännöllisiä lausekkeita, niin  $(r \cup s)$ ,  $(rs)$  ja  $r^*$  ovat  $\Sigma$ :n säännöllisiä lausekkeita;
4. muita  $\Sigma$ :n säännöllisiä lausekkeita ei ole.

Aakkoston  $\{a, b\}$  säännöllisiä lausekkeita: ◀

$$r_1 = ((ab)b), \quad r_2 = (ab)^*,$$

$$r_3 = (ab^*), \quad r_4 = (a(b \cup (bb)))^*.$$

Lausekkeiden kuvaamat kielet:

$$L(r_1) = (\{a\}\{b\})\{b\} = \{ab\}\{b\} = \{abb\};$$

$$\begin{aligned} L(r_2) &= \{ab\}^* = \{\varepsilon, ab, abab, ababab, \dots\} \\ &= \{(ab)^i \mid i \geq 0\}; \end{aligned}$$

$$\begin{aligned} L(r_3) &= \{a\}(\{b\})^* = \{a, ab, abb, abbb, \dots\} \\ &= \{ab^i \mid i \geq 0\}; \end{aligned}$$

$$L(r_4) = (\{a\}\{b, bb\})^* = \{ab, abb\}^*$$

$$= \{\varepsilon, ab, abb, abab, ababb, \dots\}$$

$$= \{x \in \{a, b\}^* \mid \text{kutakin } a\text{-kirjainta } x\text{:ssä seuraa 1 tai 2 } b\text{-kirjainta}\}.$$

Sulkumerkkien vähentämissääntöjä:

Operaattoreiden prioriteetti:

$$* \quad \gamma \quad \cdot \quad \gamma \quad \cup$$

Yhdiste- ja tulo-operaatioiden assosiativisuus:

$$L(((r \cup s) \cup t)) = L((r \cup (s \cup t)))$$

$$L(((rs)t)) = L((r(st)))$$

⇒ peräkkäisiä yhdisteitä ja tuloja ei tarvitse suluttaa.

Käytetään tavallisia kirjasimia, mikäli sekaannuksen vaaraa merkkijonoihin ei ole.

Yksinkertaisemmin siis:

$$r_1 = abb, \quad r_2 = (ab)^*, \quad r_3 = ab^*, \quad r_4 = (a(b \cup bb))^*.$$

**Määritelmä 2.4** Kieli on *säännöllinen*, jos se voidaan kuvata säännöllisellä lausekkeella.

Esimerkki: C-kielen etumerkittömät reaalityluvut

$$number = (dd^*.d^* \cup dd^*)(e(+ \cup - \cup \epsilon)dd^* \cup \epsilon) \cup (dd^*e(+ \cup - \cup \epsilon)dd^*),$$

missä  $d$  on lyhennemerkintä lausekkeelle

$$d = (0 \cup 1 \cup 2 \cup 3 \cup 4 \cup 5 \cup 6 \cup 7 \cup 8 \cup 9)$$

ja  $e$  on lyhennemerkintä lausekkeelle

$$e = (\mathbb{E} \cup e).$$

Usein merkitään myös lyhyesti  $rr^* \equiv r^+$ . Esim.:

$$(d^+.d^* \cup d^+)(e(+ \cup - \cup \epsilon)d^+ \cup \epsilon) \cup (d^+e(+ \cup - \cup \epsilon)d^+).$$

### Säännöllisten lausekkeiden sieventäminen

Säännöllisillä kielillä on yleensä useita vaihtoehtoisia kuvauksia, esim.:

$$\begin{aligned} \Sigma^* &= L((a \cup b)^*) \\ &= L((a^*b^*)^*) \\ &= L(a^*b^* \cup (a \cup b)^*ba(a \cup b)^*). \end{aligned}$$

**Määritelmä.** Säännölliset lausekkeet  $r$  ja  $s$  ovat *ekvivalentit*, merk.  $r = s$ , jos  $L(r) = L(s)$ .

Lausekkeen sieventäminen = "yksinkertaisimman" ekvivalentin lausekkeen määrittäminen.

Säännöllisten lausekkeiden ekvivalenssitestaus on epätriviaali, mutta periaatteessa mekaanisesti ratkeava ongelma.

Sievennyssääntöjä:

$$\begin{aligned}
 r \cup (s \cup t) &= (r \cup s) \cup t & r \cup \emptyset &= r \\
 r(st) &= (rs)t & \varepsilon r &= r \\
 r \cup s &= s \cup r & \emptyset r &= \emptyset \\
 r(s \cup t) &= rs \cup rt & r^* &= \varepsilon \cup r^* r \\
 (r \cup s)t &= rt \cup st & r^* &= (\varepsilon \cup r)^* \\
 r \cup r &= r & &
 \end{aligned}$$

Mikä tahansa säännöllisten lausekkeiden tosi ekvivalenssi voidaan johtaa näistä laskulaeista, kun lisätään päättelysääntö: jos  $r = rs \cup t$ , niin  $r = ts^*$ , edellyttäen että  $\varepsilon \notin L(s)$ .

Kahden lausekkeen ekvivalenssin toteamiseksi kannattaa usein päätellä erikseen kummankin kuvaaman kielen sisältyminen toiseen.

Merkitään lyhyesti:  $r \subseteq s$ , jos  $L(r) \subseteq L(s)$ .  
Tällöin siis  $r = s$  joss  $r \subseteq s$  ja  $s \subseteq r$ .

Esimerkki: todetaan, että  $(a^*b^*)^* = (a \cup b)^*$ .

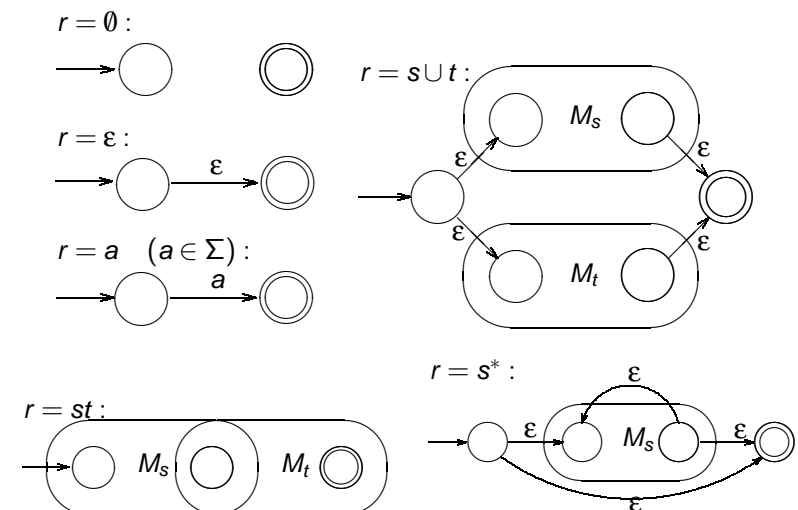
1. Selvästi  $(a^*b^*)^* \subseteq (a \cup b)^*$ , koska  $(a \cup b)^*$  kuvaa *kaikkia* aakkoston  $\{a, b\}$  merkkijonoja.
2. Koska selvästi  $(a \cup b) \subseteq a^*b^*$ , niin myös  $(a \cup b)^* \subseteq (a^*b^*)^*$ .

## 2.7 ÄÄRELLISET AUTOMAATIT JA SÄÄNNÖLLISET KIELET

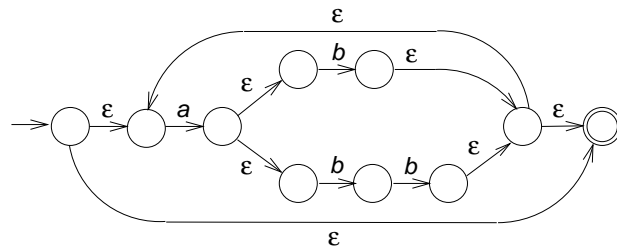
**Lause 2.3** Jokainen säännöllinen kieli voidaan tunnistaa äärellisellä automaatilla.

*Todistus.* Seuraavan kalvon induktiivisen konstruktion avulla voidaan mielivaltaisen säännöllisen lausekkeen  $r$  rakennetta seuraten muodostaa  $\varepsilon$ -automaatti  $M_r$ , jolla  $L(M_r) = L(r)$ . Tästä automaatista voidaan poistaa  $\varepsilon$ -siirtymät Lemman 2.4 mukaisesti, ja tarvittaessa voidaan syntyvä epädeterministinen automaatti determinisoida Lauseen 2.2 konstruktiolla.

Esitettävästä konstruktiosta on syytä huomata, että muodostettaviin  $\varepsilon$ -automaatteihin tulee aina yksikäsitteiset alkua ja lopputila, eikä minkään osa-automaatin lopputilasta lähde eikä alkutilaan tule yhtään ko. osa-automaatin sisäistä siirtymää.  $\square$

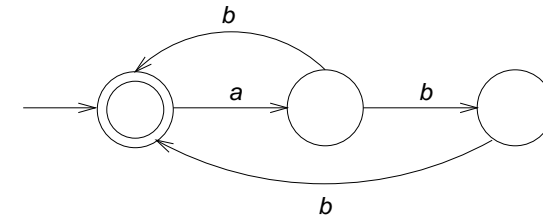


Esimerkiksi lausekkeesta  $r = (a(b \cup bb))^*$  saadaan näiden sääntöjen mukaan seuraava  $\varepsilon$ -automaatti:



Automaatti on selvästi hyvin redundantti. Käsien automaatteja suunniteltaessa ne kannattaakin usein muodostaa suoraan.

Esim. lausekkeen  $r = (a(b \cup bb))^*$  perusteella on helppo muodostaa seuraava yksinkertainen epädeterministinen tunnistaja-automaatti:



**Lause 2.4** Jokainen äärellisellä automaatilla tunnistettava kieli on säännöllinen.

*Todistus.* Tarvitaan vielä yksi äärellisten automaattien laajennus: *lausekeautomaatissa* voidaan siirtymien ehtoina käyttää mielivaltaisia säännöllisiä lausekkeita.

Formalisointi: Merk.  $RE_{\Sigma}$  = aakkoston  $\Sigma$  säännöllisten lausekkeiden joukko. *Lausekeautomaatti* on viisikko

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

missä siirtymäfunktio  $\delta$  on äärellinen kuvaus

$$\delta : Q \times RE_{\Sigma} \rightarrow \mathcal{P}(Q)$$

(so.  $\delta(q, r) \neq \emptyset$  vain äärellisen monella parilla  $(q, r) \in Q \times RE_{\Sigma}$ ).

Yhden askelen tilannejohto määritellään:

$$(q, w) \stackrel{M}{\vdash} (q', w')$$

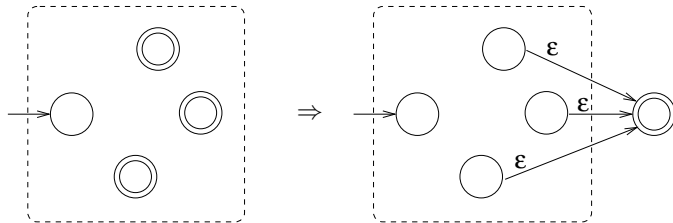
jos on  $q' \in \delta(q, r)$  jollakin sellaisella  $r \in RE_{\Sigma}$ , että  $w = zw'$ ,  $z \in L(r)$ . Muut määritelmät samat kuin aiemmin.

Todistetaan: jokainen lausekeautomaatilla tunnistettava kieli on säännöllinen.

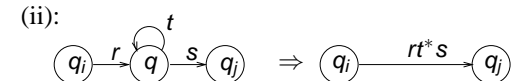
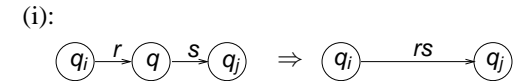
Olkoon  $M$  jokin lausekeautomaatti. Säännöllinen lauseke, joka kuvaa  $M$ :n tunnistaman kielen, muodostetaan kahdessa vaiheessa:

1. Tiivistetään  $M$  ekvivalentiksi enintään 2-tilaiseksi lausekeautomaatiksi seuraavilla muunnoksilla:

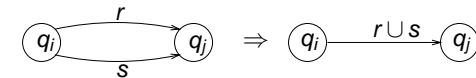
(i) Jos  $M$ :llä on useita lopputiloja, yhdistetään ne seuraavasti.



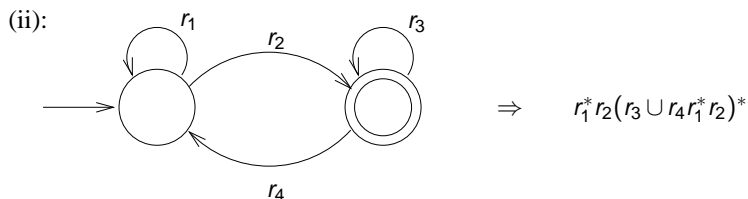
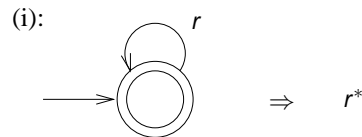
(ii) Poistetaan  $M$ :n muut kuin alku- ja lopputila yksi kerrallaan seuraavasti. Olk.  $q$  jokin  $M$ :n tila, joka ei ole alku- eikä lopputila; tarkastellaan kaikkia "reittejä", jotka  $M$ :ssä kulkevat  $q$ :n kautta. Olk.  $q_i$  ja  $q_j$   $q$ :n välittömät edeltäjä- ja seuraajatila jollakin tällaisella reitillä. Poistetaan  $q$  reitiltä  $q_i \rightarrow q_j$  oheisen kuvan (i) muunnoksella, jos tilasta  $q$  ei ole siirtymää itseensä, ja kuvan (ii) muunnoksella, jos tilasta  $q$  on siirtymä itseensä:



Samalla yhdistetään rinnakkaiset siirtymät seuraavasti:



2. Tiivistyksen päättyessä jäljellä olevaa 2-tilaista automaattia vastaava säännöllinen lauseke muodostetaan seuraavan kuvan esittämällä tavalla:



**Esimerkki:**

