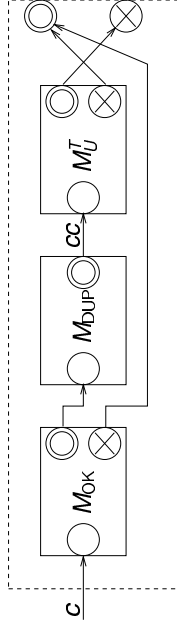


**Lause 6.7** Kieli  $U$  ei ole rekursiivinen.

*Todistus.* Oletetaan, että kielellä  $U$  olisi totaalinen tunnistajakone  $M_U^T$ . Tällöin voitaisiin Lemman 6.5 kielelle  $D$  muodostaa totaalinen tunnistajakone  $M_D$  seuraavasti.

Olkoon  $M_{OK}$  totaalinen Turingin kone, joka testaa, onko syötteenä annettu merkijono kelvollinen Turingin koneen koodi, ja olkoon  $M_{DUP}$  totaalinen Turingin kone, joka muuntaa syötejonon  $c$  muotoon  $cc$ . Kone  $M_D$  muodostetaan koneista  $M_U^T$ ,  $M_{OK}$  ja  $M_{DUP}$  yhdistämällä seuraavan kaavion esittämällä tavalla:

**Seuraus 6.8** Kieli

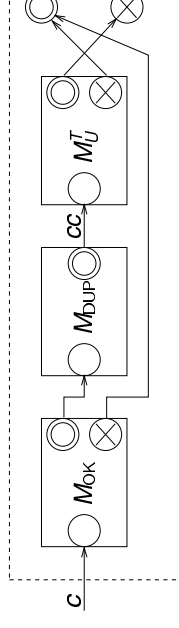
$$\bar{U} = \{c_M w \mid w \notin L(M)\}$$

ei ole rekursiivisesti numeroituva.

*Todistus.* Kieli  $\bar{U}$  on oleellisesti sama kuin universaalkielen  $U$  komplementti  $\bar{U}$ ; tarkasti ottaen on  $\bar{U} = \bar{U} \cup \text{ERR}$ , missä ERR on helposti tunnistettava rekursiivinen kieli

$$\text{ERR} = \{x \in \{0, 1\}^* \mid x \text{ ei sisällä alkuosanaan kelvollista Turingin koneen koodia}\}.$$

Jos siis kieli  $\bar{U}$  olisi rekursiivisesti numeroituva, olisi samoin myös kieli  $\bar{U}$ . Koska kieli  $U$  on rekursiivisesti numeroituva, seuraisi tästä, että  $U$  on peräti rekursiivinen. Mutta tämä on vastoin edellisen lauseen tulosta, mistä päätellään, että kieli  $\bar{U}$  ei voi olla rekursiivisesti numeroituva.  $\square$



Selvästi kone  $M_D$  on totaalinen, jos kone  $M_U^T$  on, ja

$$\begin{aligned} c \in L(M_D) &\Leftrightarrow c \notin L(M_{OK}) \text{ tai } cc \notin L(M_U^T) \\ &\Leftrightarrow c \notin L(M_C) \\ &\Leftrightarrow c \in D. \end{aligned}$$

Mutta lemmän 6.5 mukaan kieli  $D$  ei ole rekursiivinen; ristiriita.  $\square$

**6.5 Turingin koneiden pysähtymisongelma****Lause 6.9** Kieli

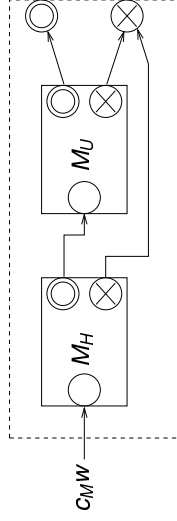
$$H = \{c_M w \mid M \text{ pysähtyy syötteellä } w\}$$

on rekursiivisesti numeroituva, mutta ei rekursiivinen.

*Todistus.* Todetaan ensin, että kieli  $H$  on rekursiivisesti numeroituva. Lauseen 6.6 todistuksessa esitetystä universaalkoneesta  $M_U$  on helppo muokata kone, joka syötteellä  $c_M w$  simuloi koneen  $M$  laskentaa syötteellä  $w$  ja pysähtyy hyväksyvään lopputilaan, jos ja vain jos simuloitu laskenta ylipäätään pysähtyy.

Osoitetaan sitten, että kieli  $H$  ei ole rekursiivinen. Oletetaan nimittäin, että olisi  $H = L(M_H)$  jollakin totaalisella Turingin koneella  $M_H$ . Oletetaan lisäksi, että kone  $M_H$  pysähtyessään jättää nauhalle alkuperäisen syötteensä, mahdollisesti tyhjämerkeillä jatkettuna. Olkoon  $M_U$  lauseen 6.6 todistuksessa konstruoitu universaalikone.

Kielelle  $U$  voitaisiin nyt muodostaa totaalin tunnistaja yhdistämällä koneet  $M_H$  ja  $M_U$  seuraavasti:



Lauseen 6.7 mukaan tällaista kielen  $U$  tunnistajakonetta ei kuitenkaan voi olla olemassa. Saatu ristiriita osoittaa, että  $H$  ei voi olla rekursiivinen.  $\square$

## Seuraus 6.10 Kieli

$$\dot{H} = \{c_MW \mid M \text{ ei pysähdy syötteellä } x\}$$

ei ole rekursiivisesti numeroituva.  $\square$

## 6.7 Ricen lause

Ricen lauseen mukaan *kaikki* Turingin koneiden tunnistamia kieliä, t. niiden laskemia I/O-kuvauksia koskevat epätriviaalit kysymykset ovat ratkeamattomia.

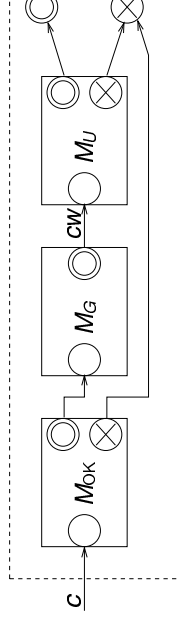
Johdantona lauseen todistukseen tarkastellaan ensin yhtä sen erikoistapausta, Turingin koneiden tunnistamien kielten *epätyhjyysoongelmaa*: "Hyväksyykö annettu Turingin kone yhtään syötemerkkijonoa?" Ongelman esitys formaalina kielenä on

$$NE = \{c \in \{0, 1\}^* \mid L(M_c) \neq \emptyset\}.$$

**Lause 6.11** Kieli  $NE$  on rekursiivisesti numeroituva, mutta ei rekursiivinen.

*Todistus.* Todetaan ensin, että kieli  $NE$  on rekursiivisesti numeroituva muodostamalla sille tunnistajakone  $M_{NE}$ . Kone  $M_{NE}$  on helpointa suunnitella epädeterministisenä.

Olkoon  $M_{OK}$  Turingin kone, joka testaa onko annettu syöte kelvollinen Turingin koneen koodi, ja olkoon  $M_G$  epädeterministinen Turingin kone, joka kirjoittaa nauhalle jo olevan merkkijonon perään mielivaltaisen binäärijonon  $w$ . Kone  $M_{NE}$  voidaan muodostaa yhdistämällä koneet  $M_{OK}$ ,  $M_G$  ja universaalikone  $M_U$  seuraavasti:





## Ricen lause

Turingin koneiden *semanttinen ominaisuus*  $S$  on mikä tahansa kokoelma rekursiivisesti numeroituvia aakkoston  $\{0, 1\}$  kieliä; koneella  $M$  on *ominaisuus*  $S$ , jos  $L(M) \in S$ . *Triviaalit ominaisuudet* ovat  $S = \emptyset$  (ominaisuus, jota ei ole millään koneella) ja  $S = RE$  (ominaisuus, joka on kaikilla koneilla). Ominaisuus  $S$  on *ratkeava*, jos joukko

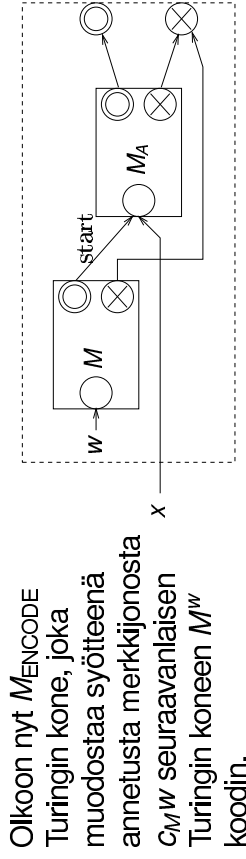
$$\text{codes}(S) = \{c \mid L(M_c) \in S\}$$

on rekursiivinen. Toisin sanoen: ominaisuus on ratkeava, jos annetusta Turingin koneen koodista voidaan algoritmisesti päätellä, onko koneella kysytty semanttinen ominaisuus.

**Lause 6.12 [Rice 1953]** Kaikki Turingin koneiden epätriviaalit semanttiset ominaisuudet ovat ratkeamattomia.

*Todistus.* Olkoon  $S$  mielivaltainen epätriviaali semanttinen ominaisuus. Voidaan olettaa, että  $\emptyset \notin S$ : toisin sanoen, että tyhjän joukon tunnistavilla Turingin koneilla ei ole tarkasteltavaa ominaisuutta. Jos nimitetään  $\emptyset \in S$ , voidaan ensin osoittaa, että ominaisuus  $S = RE - S$  on ratkeamaton, ja päätellä edelleen tästä että myös ominaisuus  $S$  on ratkeamaton. (Koska  $\text{codes}(\bar{S}) = \text{codes}(S)$ .)

Koska  $S$  on epätriviaali, on olemassa jokin Turingin kone  $M_A$ , jolla on ominaisuus  $S$  — jolla siis  $L(M_A) \neq \emptyset \in S$ .



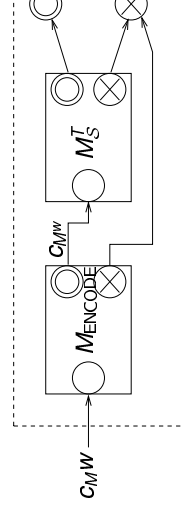
Jos syöte ei ole vaadittua muotoa,  $M_{\text{ENCODE}}$  päättyy hylkävään lopputilaan.

Syötteellä  $x$  kone  $M^w$  toimii ensin kuten  $M$  syötteellä  $w$ . Jos  $M$  hyväksyy  $w$ :n,  $M^w$  toimii kuten kone  $M_A$  syötteellä  $x$ . Jos  $M$  hylkää  $w$ :n, myös  $M^w$  hylkää  $x$ :n. Kone  $M^w$  tunnistaa siis kielen

$$L(M^w) = \begin{cases} L(M_A), & \text{jos } w \in L(M); \\ \emptyset, & \text{jos } w \notin L(M). \end{cases}$$

Koska oletuksen mukaan  $L(M_A) \in S$  ja  $\emptyset \notin S$ , on koneella  $M^w$  ominaisuus  $S$ , jos ja vain jos  $w \in L(M)$ .

Oletetaan sitten, että ominaisuus  $S$  olisi ratkeava, so. että kielellä  $\text{codes}(S)$  olisi totaalin tunnistajakone  $M_S^I$ . Tällöin saataisiin edellisen todistuksen tapaan totaalin tunnistajakone kielelle  $U$  yhdistämällä koneet  $M_{\text{ENCODE}}$  ja  $M_S^I$  seuraavasti:



Selvästi kone  $M_J^w$  on totaalin, jos  $M_S^I$  on, ja

$$c_M^w \in L(M_J^w) \Leftrightarrow c_M^w \in L(M_S^I) = \text{codes}(S) \Leftrightarrow L(M^w) \in S \Leftrightarrow w \in L(M).$$

Koska kieli  $U$  ei ole rekursiivinen, tämä on mahdotonta, mistä päätellään, ettei ominaisuus  $S$  voi olla ratkeava.  $\square$

## 6.8 Muita ratkeamattomuustuloksia

### Lause 6.13 (Predikaattikalkyylin ratkeamattomuus; Church/Turing 1936)

Ei ole olemassa algoritmia, joka ratkaisisi, onko annettu ensimmäisen kertaluvun predikaattikalkyylin kaava  $\phi$  validi ("loogisesti tosi", todistuva predikaattikalkyylin aksioomista).  $\square$

### Lause 6.14 ("Hilbertin 10. ongelma"; Matijasevitsh/Davis/Robinson/Putnam 1953–70)

Ei ole olemassa algoritmia, joka ratkaisisi, onko annetulla kokonaislukukertoimisella polynomilla  $P(x_1, \dots, x_n)$  kokonaislukuunollakohtia (so. jonoja  $(m_1, \dots, m_n) \in \mathbb{Z}^n$ , joilla  $P(m_1, \dots, m_n) = 0$ ). Ongelma on ratkematon jo, kun  $n = 15$  tai  $\deg(P) = 4$ .  $\square$

Eräiden kielioppiongelmien ratkeavuus, kun annettuna on kielioppi  $G$  ja  $G$  Chomskyn hierarkian tietyllä tasolla  $i$  ja merkijono  $w$ . Taulukossa  $R \sim$  "ratkeava",  $E \sim$  "ei ratkeava",  $T \sim$  "aina totta".

Ongelma: onko	Taso $i$ :			
	3	2	1	0
$w \in L(G)$ ?	$R$	$R$	$R$	$E$
$L(G) = \emptyset$ ?	$R$	$R$	$E$	$E$
$L(G) = \Sigma^*$ ?	$R$	$E$	$E$	$E$
$L(G) = L(G)$ ?	$R$	$E$	$E$	$E$
$L(G) \subseteq L(G)$ ?	$R$	$E$	$E$	$E$
$L(G) \cap L(G) = \emptyset$ ?	$R$	$E$	$E$	$E$
$L(G)$ säännöllinen?	$T$	$E$	$E$	$E$
$L(G) \cap L(G)$ tyyppiä $i$ ?	$T$	$E$	$T$	$T$
$\overline{L(G)}$ tyyppiä $i$ ?	$T$	$E$	$T$	$E$

## 5. RAJOITTAMATTOMAT KIELIOPIT

**Määritelmä 5.1** Rajoittamaton kielioppi t. yleinen merkijonomuunnossysteemi on nelikko

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

missä

- $V$  on kieliopin aakkosto;
- $\Sigma \subseteq V$  on kieliopin päätemerkkien joukko;  $N = V - \Sigma$  on välikemerkkien t. -symbolien joukko;
- $P \subseteq V^+ \times V^*$  on kieliopin sääntöjen t. produktioiden joukko ( $V^+ = V^* - \{\varepsilon\}$ );
- $S \in N$  on kieliopin lähtösymboli.

Produktiota  $(\omega, \omega') \in P$  merkitään tavallisesti  $\omega \rightarrow \omega'$ .

Merkijono  $\gamma \in V^*$  tuottaa t. johtaa suoraan merkijonon  $\gamma' \in V^*$  kieliopissa  $G$ , merkitään

$$\gamma \Rightarrow_G \gamma'$$

jos voidaan kirjoittaa  $\gamma = \alpha\omega\beta$ ,  $\gamma' = \alpha\omega'\beta$  ( $\alpha, \beta, \omega' \in V^*$ ,  $\omega \in V^+$ ), ja kieliopissa on produktio  $\omega \rightarrow \omega'$ .

Jos kielioppi  $G$  on yhteydestä selvä, merkitään  $\gamma \Rightarrow \gamma'$ .

Merkijono  $\gamma \in V^*$  tuottaa t. johtaa merkijonon  $\gamma' \in V^*$  kieliopissa  $G$ , merkitään

$$\gamma \Rightarrow_G^* \gamma'$$

jos on olemassa jono  $V$ :n merkijonoja  $\gamma_0, \gamma_1, \dots, \gamma_n$  ( $n \geq 0$ ), siten että

$$\gamma = \gamma_0 \Rightarrow_G \gamma_1 \Rightarrow_G \dots \Rightarrow_G \gamma_n = \gamma'.$$

Jälleen, jos  $G$  on yhteydestä selvä, merkitään  $\gamma \Rightarrow^* \gamma'$ .

Merkkijono  $\gamma \in V^*$  on kieliopin  $G$  lausejohdos, jos on  $S \xRightarrow{G}^* \gamma$ .  
Pelkästään päättermerkeistä koostuva  $G$ :n lausejohdos  $x \in \Sigma^*$  on  $G$ :n lause.

Kieliopin  $G$  tuottama t. kuvaama kieli  $L(G)$  koostuu  $G$ :n lauseista, s.o.:

$$L(G) = \{x \in \Sigma^* \mid S \xRightarrow{G}^* x\}.$$

**Esimerkki.** Rajoittamaton kielioppi ei-yhteydettömälle kielelle  $\{a^k b^k c^k \mid k \geq 0\}$ .

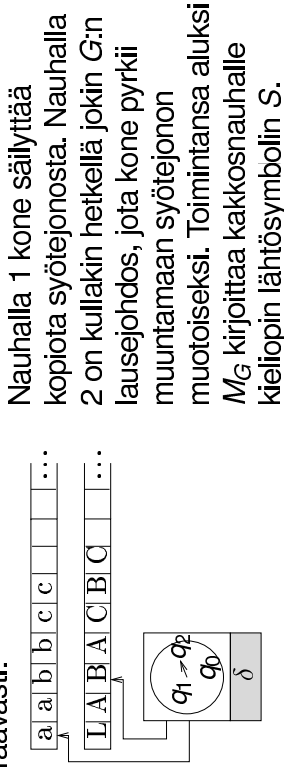
$S \rightarrow$	$LT \mid \varepsilon$	$aA \rightarrow$	$aa$
$T \rightarrow$	$ABCT \mid ABC$	$aB \rightarrow$	$ab$
$BA \rightarrow$	$AB$	$bB \rightarrow$	$bb$
$CB \rightarrow$	$BC$	$bC \rightarrow$	$bc$
$CA \rightarrow$	$AC$	$cC \rightarrow$	$cc$
$LA \rightarrow$	$a$		

Esimerkiksi lauseen  $aabbcc$  johto:

$$\begin{aligned} S &\Rightarrow LI \Rightarrow LABC\bar{I} \Rightarrow LABCABC \Rightarrow LABACBC \\ &\Rightarrow LAABCBC \Rightarrow LAABBC \Rightarrow aABBC \\ &\Rightarrow aaBBCC \Rightarrow aabbCC \Rightarrow aabbcc \end{aligned}$$

**Lause 5.1** Jos formaali kieli  $L$  voidaan tuottaa rajoittamattomalla kieliopilla, se voidaan tunnistaa Turingin koneella.

*Todistus.* Olkoon  $G = (V, \Sigma, P, S)$  kielen  $L$  tuottava rajoittamaton kielioppi. Muodostetaan kielen  $L$  tunnistava kaksinauhainen epädeterministinen Turingin kone  $M_G$  seuraavasti:



Nauhalla 1 kone säilyttää kopiota syötejonosta. Nauhalla 2 on kullakin hetkellä jokin  $G$ :n lausejohdos, jota kone pyrkii muuntamaan syötejonon muotoiseksi. Toimintansa aluksi  $M_G$  kirjoittaa kakkosnauhalle kieliopin lähtösymbolin  $S$ .

Koneen  $M_G$  laskenta koostuu vaiheista. Kussakin vaiheessa kone:

- (i) vie kakkosnauhan nauhapään epädeterministisesti johonkin kohtaan nauhalla;
- (ii) valitsee epädeterministisesti jonkin  $G$ :n produktion, jota yrittää soveltaa valittuun nauhankohtaan (produktiot on koodattu  $M_G$ :n siirtymäfunktioon);
- (iii) jos produktion vasen puoli sopii yhteen nauhalla olevien merkien kanssa,  $M_G$  korvaa  $ao$ . merkit produktion oikean puolen merkeillä;
- (iv) vaiheen lopuksi  $M_G$  vertaa ykkös- ja kakkosnauhan merkkijonoja toisiinsa: jos jonot ovat samat, kone siirtyy hyväksyvään lopputilaan ja pysähtyy, muuten aloittaa uuden vaiheen (kohta (i)).  $\square$

**Lause 5.2** Jos formaali kieli  $L$  voidaan tunnistaa Turingin koneella, se voidaan tuottaa rajoittamattomalla kielipilla. Todistus. Olkoon  $M = (Q, \Sigma, \Gamma, \delta, q_0, q_{acc}, q_{rej})$  kielen  $L$  tunnistava standardimallinen Turingin kone. Muodostetaan kielen  $L$  tuottava rajoittamaton kieliooppi  $G_M$  seuraavasti.

*Idea:* Kielioopin  $G_M$  välikkeiksi otetaan (muiden muassa) kaikkia  $M$ :n tiloja  $q \in Q$  edustavat symbolit. Koneen  $M$  tilanne  $(q, uav)$  esitetään merkijonona  $[uqav]$ .  $M$ :n siirtymäfunktion perusteella  $G_M$ :ään muodostetaan produktiot, joiden ansiosta

$$[uqav] \xRightarrow{G_M} [u'q'av'] \text{ joss } (q, uav) \vdash_M (q', u'av').$$

Siten  $M$  hyväksyy syötteen  $x$ , jos ja vain jos

$$[q_0x] \xRightarrow{G_M}^* [uq_{acc}v]$$

joillakin  $u, v \in \Sigma^*$ .

Kaikkiaan kieliooppiin  $G_M$  tulee kolme ryhmää produktioita:

1. Produktiot, joilla lähtösymbolista  $S$  voidaan tuottaa mikä tahansa merkijono muotoa  $x[q_0x]$ , missä  $x \in \Sigma^*$  ja  $[, q_0$  ja  $] ovat  $G_M$ :n välikkeitä.$
2. Produktiot, joilla merkijonosta  $[q_0x]$  voidaan tuottaa merkijono  $[uq_{acc}v]$ , jos ja vain jos  $M$  hyväksyy  $x$ :n.
3. Produktiot, joilla muotoa  $[uq_{acc}v]$  oleva merkijono muutetaan tyhjäksi merkijonoksi.

Kieleen  $L(M)$  kuuluvan merkijonon  $x$  tuottaminen tapahtuu tällöin seuraavasti:

$$S \xRightarrow{(1)}^* x[q_0x] \xRightarrow{(2)}^* x[uq_{acc}v] \xRightarrow{(3)}^* x.$$

Määritellään siis  $G = (V, \Sigma, P, S)$ , missä

$$V = \Gamma \cup Q \cup \{S, T, [, ], E_L, E_R\} \cup \{A_a \mid a \in \Sigma\},$$

ja produktiot  $P$  muodostuvat seuraavista kolmesta ryhmästä:

1. Alkutilanteen tuottaminen:

$$\begin{array}{ll} S & \rightarrow T[q_0] \\ T & \rightarrow \epsilon \\ T & \rightarrow aTA_a \quad (a \in \Sigma) \\ A_a[q_0] & \rightarrow [q_0A_a \quad (a \in \Sigma) \\ A_ab & \rightarrow bA_a \quad (a, b \in \Sigma) \\ A_a] & \rightarrow a] \quad (a \in \Sigma) \end{array}$$

2.  $M$ :n siirtymien simulointi ( $a, b \in \Gamma, c \in \Gamma \cup \{[\ ]\}$ ):

Siirtymät: Produktiot:

$$\begin{array}{ll} \delta(q, a) = (q', b, R) & qa \rightarrow bq' \\ \delta(q, a) = (q', b, L) & cqa \rightarrow q'cb \\ \delta(q, \triangleright) = (q', \triangleright, R) & q[ \rightarrow [q' \\ \delta(q, \triangleleft) = (q', b, R) & q] \rightarrow bq' \\ \delta(q, \triangleleft) = (q', b, L) & cq] \rightarrow q'cb] \\ \delta(q, \triangleleft) = (q', \triangleleft, L) & cq] \rightarrow q'c] \end{array}$$

## 3. Lopputilanteen siivous:

$$\begin{array}{l}
 q_{\text{acc}} \rightarrow E_L E_R \\
 q_{\text{acc}} \rightarrow E_R \\
 a E_L \rightarrow E_L \quad (a \in \Gamma) \\
 [E_L \rightarrow \varepsilon \\
 E_R a \rightarrow E_R \quad (a \in \Gamma) \\
 E_R ] \rightarrow \varepsilon
 \end{array}$$

□

**Yhteysherkkät kieliopit**

Rajoittamaton kielioppi on *yhteysherkkä*, jos sen kaikki produktiot ovat muotoa  $\omega \rightarrow \omega'$ , missä  $|\omega'| \geq |\omega|$ , tai mahdollisesti  $S \rightarrow \varepsilon$ , missä  $S$  on lähtösymboli.

Lisäksi vaaditaan, että jos kieliopissa on produktio  $S \rightarrow \varepsilon$ , niin lähtösymboli  $S$  ei esiinny minkään produktio oikealla puolella. Formaali kieli  $L$  on *yhteysherkkä*, jos se voidaan tuottaa jollakin yhteysherkkällä kieliopilla.

**Normaalimuoto:** Jokainen yhteysherkkä kieli voidaan tuottaa kieliopilla, jonka produktiot ovat muotoa  $S \rightarrow \varepsilon$  ja  $\alpha A \beta \rightarrow \alpha \omega \beta$ , missä  $A$  on välike ja  $\omega \neq \varepsilon$ . (Säännön  $A \rightarrow \omega$  sovellus "kontekstissa"  $\alpha \_ \beta$ .)

**Lause 5.3** Formaali kieli  $L$  on yhteysherkkä, jos ja vain jos se voidaan tunnistaa epädeterministisellä Turingin koneella, joka ei tarvitse enempää työtilaa kuin syötejonon pituuden verran — siis koneella, jolla ei ole muotoa  $\delta(q, \triangleleft) = (q', b, \Delta)$  olevia siirtymiä, missä  $b \neq \triangleleft$ . □

Lauseen 5.3 konetta sanotaan *lineaarisesti rajoitetuksi automaateiksi*.

Avoin ongelma ("LBA ? = DLBA"): onko epädeterminismi lauseessa 5.3 välttämätöntä?

**Chomskyn hierarkia**

Kielioppien, niillä tuotettavien kielten ja vastaavien tunnistusautomaattien ryhmittely:

**Luokka 0:** rajoittamattomat kieliopit / rekursiivisesti numeroituvat kielet / Turingin koneet.

**Luokka 1:** yhteysherkkät kieliopit / yhteysherkkät kielet / lineaarisesti rajoitetut automaattit.

**Luokka 2:** yhteydettömät kieliopit / yhteydettömät kielet / Pinoautomaattit.

**Luokka 3:** oikealle ja vasemmalle lineaariset (säännölliset) kieliopit / säännölliset kielet / äärelliset automaattit.

