5. RAJOITTAMATTOMAT KIELIOPIT

Määritelmä 5.1 *Rajoittamaton kielioppi* t. *yleinen merkkijonomuunnossysteemi* 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 $\omega \to \omega'$.

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

$$\gamma \underset{G}{\Rightarrow} \gamma'$$

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

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

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

$$\gamma \underset{G}{\Rightarrow}^* \gamma'$$

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

$$\gamma = \gamma_0 \underset{G}{\Rightarrow} \gamma_1 \underset{G}{\Rightarrow} \ldots \underset{G}{\Rightarrow} \gamma_n = \gamma'.$$

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

Merkkijono $\gamma \in V^*$ on kieliopin G *lausejohdos*, jos on $S \underset{G}{\Rightarrow}^* \gamma$.

Pelkästään päätemerkeistä 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.:

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

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

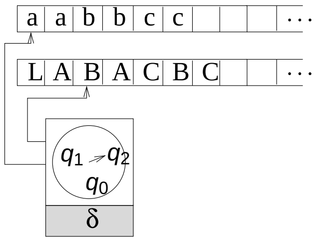
$$\begin{array}{rcl@{\qquad\qquad}rcl}
S & \to & LT \mid \varepsilon & & & \\
T & \to & ABCT \mid ABC & aA & \to & aa \\
BA & \to & AB & aB & \to & ab \\
CB & \to & BC & bB & \to & bb \\
CA & \to & AC & bC & \to & bc \\
LA & \to & a & cC & \to & cc.
\end{array}$$

Esimerkiksi lauseen *aabbcc* johto:

$$\begin{array}{rllll}
\underline{S} & \Rightarrow L\underline{T} \Rightarrow LABC\underline{T} & \Rightarrow LABC\underline{CA}BC & \Rightarrow LA\underline{BA}CBC \\
& \Rightarrow LAAB\underline{CB}C & \Rightarrow \underline{LA}ABBCC & \Rightarrow \underline{aA}BBCC \\
& \Rightarrow aa\underline{B}BCC & \Rightarrow aab\underline{B}CC & \Rightarrow aabb\underline{C}C \\
& \Rightarrow aabbc\underline{C} & \Rightarrow aabbcc. &
\end{array}$$

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 merkkien 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)). □

Lause 5.2 Jos formaali kieli L voidaan tunnistaa Turingin koneella, se voidaan tuottaa rajoittamattomalla kieliopilla.

Todistus. Olkoon $M = (Q, \Sigma, \Gamma, \delta, q_0, q_{\mathrm{acc}}, q_{\mathrm{rej}})$ kielen L tunnistava standardimallinen Turingin kone. Muodostetaan kielen L tuottava rajoittamaton kielioppi G_M seuraavasti.

Idea: Kieliopin G_M välikkeiksi otetaan (muiden muassa) kaikkia M:n tiloja $q \in Q$ edustavat symbolit. Koneen M tilanne $(q, u\underline{a}v)$ esitetään merkkijonona $[uqav]$. M:n siirtymäfunktion perusteella G_M:ään muodostetaan produktiot, joiden ansiosta

$$[uqav] \underset{G_M}{\Rightarrow} [u'q'a'v'] \quad \text{joss} \quad (q, uav) \underset{M}{\vdash} (q', u'a'v').$$

Siten M hyväksyy syötteen x, jos ja vain jos

$$[q_0x] \underset{G_M}{\Rightarrow}^* [uq_{\mathrm{acc}}v]$$

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

Kaikkiaan kielioppiin G_M tulee kolme ryhmää produktioita:

1. Produktiot, joilla lähtösymbolista S voidaan tuottaa mikä tahansa merkkijono muotoa $x[q_0x]$, missä $x \in \Sigma^*$ ja [, q_0 ja] ovat G_M:n välikkeitä.

2. Produktiot, joilla merkkijonosta $[q_0x]$ voidaan tuottaa merkkijono $[uq_{\mathrm{acc}}v]$, jos ja vain jos M hyväksyy x:n.

3. Produktiot, joilla muotoa $[uq_{\mathrm{acc}}v]$ oleva merkkijono muutetaan tyhjäksi merkkijonoksi.

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

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

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:

$$\begin{array}{lcll} S & \longrightarrow & T[q_0] & \\ T & \longrightarrow & \varepsilon & \\ T & \longrightarrow & aTA_a & (a \in \Sigma) \\ A_a[q_0 & \longrightarrow & [q_0A_a & (a \in \Sigma) \\ A_ab & \longrightarrow & bA_a & (a, b \in \Sigma) \\ A_a] & \longrightarrow & a] & (a \in \Sigma) \end{array}$$

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

$$\begin{array}{lcl@{\qquad\qquad}lcl} \textit{Siirtymät:} & & & \textit{Produktiot:} & & \\ \delta(q,a) & = & (q', b, R) & qa & \longrightarrow & bq' \\ \delta(q,a) & = & (q', b, L) & cqa & \longrightarrow & q'cb \\ \delta(q,\triangleright) & = & (q', \triangleright, R) & q[& \longrightarrow & [q' \\ \delta(q,\triangleleft) & = & (q', b, R) & q] & \longrightarrow & bq'] \\ \delta(q,\triangleleft) & = & (q', b, L) & cq] & \longrightarrow & q'cb] \\ \delta(q,\triangleleft) & = & (q', \triangleleft, L) & cq] & \longrightarrow & q'c] \end{array}$$

3. Lopputilanteen siivous:

$$\begin{array}{lcll} q_{acc} & \longrightarrow & E_LE_R & \\ q_{acc}[& \longrightarrow & E_R & \\ aE_L & \longrightarrow & E_L & (a \in \Gamma) \\ [E_L & \longrightarrow & \varepsilon & \\ E_Ra & \longrightarrow & E_R & (a \in \Gamma) \\ E_R] & \longrightarrow & \varepsilon & \end{array}$$

□

Yhteysherkät kieliopit

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

Lisäksi vaaditaan, että jos kieliopissa on produktio $S \to \varepsilon$, niin lähtösymboli S ei esiinny minkään produktion oikealla puolella.

Formaali kieli L on *yhteysherkkä*, jos se voidaan tuottaa jollakin yhteysherkällä kieliopilla.

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

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 $b \neq$ '$\triangleleft$'. $\square$

Lauseen 5.3 koneita sanotaan *lineaarisesti rajoitetuiksi automaateiksi*.

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

T–79.1001/1002 Tietojenkäsittelyteorian perusteet

Chomskyn hierarkia

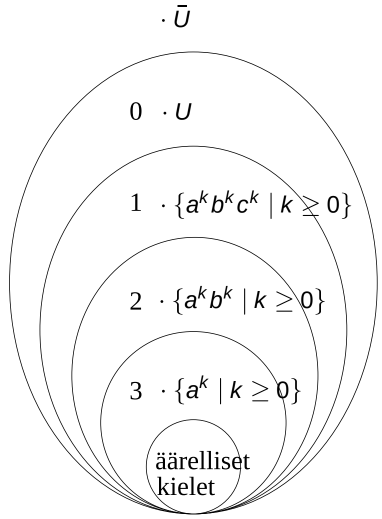
Kielioppien, niillä tuotettavien kielten ja vastaavien tunnistusautomaattien ryhmittely:

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

Luokka 1: yhteysherkät kieliopit / yhteysherkät kielet / lineaarisesti rajoitetut automaatit.

Luokka 2: yhteydettömät kieliopit / yhteydettömät kielet / pinoautomaatit.

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



Pekka Orponen syksy 2005