

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 $\omega \rightarrow \omega'$.

1

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

$$\gamma \xrightarrow{G} \gamma'$$

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 yksinkertaisesti $\gamma \Rightarrow \gamma'$.

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

$$\gamma \xRightarrow{*G} \gamma'$$

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

$$\gamma = \gamma_0 \xrightarrow{G} \gamma_1 \xrightarrow{G} \dots \xrightarrow{G} \gamma_n = \gamma'.$$

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

2

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

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

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

3

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

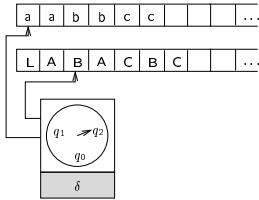
Esimerkiksi lauseen $aabbcc$ johto:

$$\begin{aligned} S &\Rightarrow \underline{LT} \Rightarrow \underline{LABCT} \Rightarrow \underline{LABCABC} \Rightarrow \underline{LABACBC} \\ &\Rightarrow \underline{LAABCBC} \Rightarrow \underline{LAABBCC} \Rightarrow \underline{aABBCC} \\ &\Rightarrow \underline{aaBBCC} \Rightarrow \underline{aabBCC} \Rightarrow \underline{aabbCC} \\ &\Rightarrow \underline{aabbcC} \Rightarrow \underline{aabbcc}. \end{aligned}$$

4

Lause 5.1 Jos formaali kieli L voidaan tuottaa rajoittamattomalla kielio pillä, se voidaan tunnistaa Turingin koneella.

Todistus. Olkoon $G = (V, \Sigma, P, S)$ kielen L tuotava rajoittamaton kielio pilli. 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 kaksosnauhalle kielio pillin 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 siirtymäfunktion);
- (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 (koh-ta (i)). \square

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

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

Idea: Kielio pillin G_M väliskeiksi otetaan (muiden muassa) kaikkia M :n tiloja $q \in Q$ edustavat symbolit. Koneen M tilanne (q, uqv) esitetään merkkijonona $[uqv]$. M :n siirtymäfunktion perusteella G_M :ään muodostetaan produktiot, joiden ansiosta

$$[uqv] \xRightarrow{G_M} [u'q'a'v'] \quad \text{joss} \quad (q, uav) \vdash_M (q', u'a'v').$$

Siten M hyväksyy syötteen x , jos ja vain jos

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

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

Kaikkiaan kielio pilliin 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 väliskeiksi.
2. Produktiot, joilla merkkijonosta $[q_0x]$ voidaan tuottaa merkkijono $[uq_{acc}v]$, jos ja vain jos M hyväksyy x :n.
3. Produktiot, joilla muotoa $[uq_{acc}v]$ oleva merkkijono muutetaan tyhjäksi merkkijonoksi.

Kieleen $L(M)$ kuuluvan merkkijonon 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:

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

9

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

Siirtymät:

$$\begin{aligned} \delta(q, a) &= (q', b, R) \\ \delta(q, a) &= (q', b, L) \\ \delta(q, >) &= (q', >, R) \\ \delta(q, <) &= (q', b, R) \\ \delta(q, <) &= (q', b, L) \\ \delta(q, <) &= (q', <, L) \end{aligned}$$

Produktiot:

$$\begin{aligned} qa &\rightarrow bq' \\ cqa &\rightarrow q'cb \\ q[&\rightarrow [q' \\ q] &\rightarrow bq'] \\ cq] &\rightarrow q'cb] \\ cq] &\rightarrow q'c] \end{aligned}$$

3. Lopputilanteen siivous:

$$\begin{aligned} q_{acc} &\rightarrow E_LE_R \\ q_{acc}[&\rightarrow E_R \\ aE_L &\rightarrow E_L \quad (a \in \Gamma) \\ [E_L &\rightarrow \varepsilon \\ E_Ra &\rightarrow E_R \quad (a \in \Gamma) \\ E_R] &\rightarrow \varepsilon \end{aligned}$$

□

10

Yhteysherkät 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 yhteysherkällä 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" $\alpha _ \beta$.)

11

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 $\delta(q, <) = (q', b, \Delta)$ olevia siirtymiä, missä $b \neq '<'$. □

Lauseen 5.3 koneita sanotaan *lineaarisesti rajoitetuiksi automaateiksi*.

Avoin ongelma ("LBA ?= DLBA"): onko epä-determinismi lauseessa 5.3 välttämätöntä?

12

Chomskyn hierarkia

Kielioppien, niillä tuotettavien kielten ja vastaavien tunnistusautomaattien ryhmittely:

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

Luokka 2: yhteydettömät kieliopit / yhteydettömät kielet / pinoautomaattit.

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

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

