

VÁRTERÉSZ MAGDA

*Formális nyelvek és automaták
előadások*

2005/06-os tanév 1. félév

Tartalomjegyzék

1. Előzetes tudnivalók	4
2. Bevezetés	15
3. Ábécé, szó, formális nyelv	17
4. Műveletek nyelvekkel	24
4.1. Boole-műveletek	24
4.2. Reguláris műveletek	25
4.3. Reguláris kifejezések	27
5. Formális rendszerek	32
5.1. A generatív rendszer	36
5.2. A Markov-algoritmus	38
6. Generatív nyelvtanok	41
6.1. A generatív nyelvtan és az általa generált nyelv	41
6.2. A generatív nyelvtanok alkalmazása a programozásban	44
6.3. A Chomsky-féle nyelvosztályok	49
7. Reguláris nyelvek	54
7.1. A determinisztikus véges automata mint felismerő	54

Tartalomjegyzék	3
------------------------	----------

7.2. A nemdeterminisztikus véges automata	59
7.3. A 3 típusú nyelvek osztályának jellemzése	65
8. Környezetfüggetlen nyelvek	75
8.1. Levezetési fák	75
8.2. Chomsky-féle normál alak	78
8.3. Szintaktikai elemzés	80
8.4. Nemdeterminisztikus veremautomaták	82
Irodalomjegyzék	88

1. fejezet

Előzetes tudnivalók

Különböző dolgokat (tárgyakat, személyeket, fogalmakat, objektumokat stb.) *halmazokba* sorolhatunk. Az egy halmazba sorolt dolgok a halmaz *elemei*.

- h eleme a H halmaznak: $h \in H$
- h nem eleme a H halmaznak: $h \notin H$

A halmazt megadási módjai:

- kapcsos zárójelben felsoroljuk az elemeit: $\{h_1, h_2, \dots, h_n\}$
- megadunk egy a halmaz elemeit jellemző tulajdonságot:

$$\{h \mid \text{a } h \text{ elem } \mathcal{T} \text{ tulajdonságú}\}$$

-
- A K halmaz a H halmaz *részhalmaza*, azaz $K \subseteq H$, ha minden $h \in K$ esetén $h \in H$ is.
 - Ha $K \subseteq H$ és $H \subseteq K$, akkor H és K ugyanazon elemeket tartalmazzák, ilyenkor a két halmazt *egyenlőnek* nevezzük (jelölése: $H = K$).
 - Pontosan egy olyan halmaz van, amelyiknek nincs eleme, neve *üres halmaz*, jele $\emptyset$.
 - Két halmaz *diszjunkt*, ha nincs közös elemük.
 - Egy adott U halmaz összes részhalmaizai is halmazt alkotnak, melyet U *hatványhalmazának* nevezzük, és $\mathcal{P}(U)$ -val jelölünk.

Vezessünk most be műveleteket $\mathcal{P}(U)$ -n. A H_1 és a H_2 $\mathcal{P}(U)$ -beli halmazok

- *uniója*: $H_1 \cup H_2 \rightleftharpoons \{h \mid h \in H_1 \text{ vagy } h \in H_2\}$
- *metSZete*: $H_1 \cap H_2 \rightleftharpoons \{h \mid h \in H_1 \text{ és } h \in H_2\}$
- *különbsége*: $H_1 \setminus H_2 \rightleftharpoons \{h \mid h \in H_1 \text{ de } h \notin H_2\}$
- *komplementere*: $\overline{H_1} \rightleftharpoons \{h \mid h \in U \text{ de } h \notin H_1\}$

$\mathcal{P}(U)$ -beli halmaz.

A $H_1, H_2, \dots, H_n$ ($n \geq 1$) nemüres halmazok *Descartes-szorzata* $H_1 \times H_2 \times \dots \times H_n \rightleftharpoons \{(h_1, h_2, \dots, h_n) \mid h_i \in H_i, i = 1, 2, \dots, n\}$. $H_1 \times H_2 \times \dots \times H_n$ részhalmazait *(n-változós) relációknak* nevezük.

Az $R \subseteq H \times H$ kétváltozós vagy *binér* reláció. $((h_1, h_2) \in R$ jelölése ilyenkor gyakran: $h_1 R h_2$). Azt mondjuk, hogy az R binér reláció

- *reflexív*, ha minden $h \in H$ -ra $h R h$;
- *irreflexív*, ha egyetlen $h \in H$ -ra sem igaz, hogy $h R h$;
- *szimmetrikus*, ha minden $h_1, h_2 \in H$ -ra, melyre $h_1 R h_2$, fennáll $h_2 R h_1$ is;
- *aszimmetrikus*, ha egyetlen $h_1, h_2 \in H$ -ra sem teljesül egyszerre $h_1 R h_2$ és $h_2 R h_1$;
- *tranzitív*, ha minden $h_1, h_2, h_3 \in H$ -ra, melyre $h_1 R h_2$ és $h_2 R h_3$, teljesül $h_1 R h_3$ is;

Jelölje a $H_1, H_2, \dots, H_n$ halmazok Descartes-szorzatát H . Az $F \subseteq H \times K$ reláció H -ból K -ba képező n -változós függvény vagy művelet, ha minden $h \in H$ -hoz legfeljebb egy olyan $k \in K$ van, amelyre $(h, k) \in F$. Ha az $F \subseteq H \times K$ reláció függvény, az $F: H \rightarrow K$ jelölést használjuk.

- Az F függvényértelmezési tartománya azon H -beli elemek halmaza (jelölése: $\text{Dom}(F)$), melyekre van olyan $k \in K$, hogy $(h, k) \in F$
- F minden $h \in \text{Dom}(F)$ -hez egy jól meghatározott $k \in K$ elemet rendel, amit $F(h)$ -val jelölhetünk, és az F függvény h helyen felvett helyettesítési értékének, vagy a h elem képének nevezünk, h -t pedig az $F(h)$ ősképének hívjuk.
- Az $\text{Rng}(F) \rightleftharpoons \{F(h) \mid h \in \text{Dom}(F)\}$ halmaz az F értékkészlete. Természetesen $\text{Rng}(F) \subseteq K$.

Az F függvény

- *szürjektív*, ha $\text{Rng}(F) = K$;
- *injektív*, ha különböző elemek képe különböző;
- *bijektív* vagy *kölcsönösen egyértelmű*, ha szürjektív és injektív.

Ha F injektív függvény, akkor a $\text{Rng}(F)$ képtér minden k eleméhez egyértelműen hozzá tudjuk rendelni az ősképet. Ily módon egy újabb függvényt nyerhetünk, amelynek értelmezési tartománya $\text{Rng}(F)$, képtere pedig $\text{Dom}(F)$. Ez a függvény az F *inverze*, és F^{-1} -gyel jelöljük.

Azokat a függvényeket, melyek értelmezési tartománya valamely $n \in \mathbf{N_0}$ -ra a $\{0, 1, \dots, n\}$ halmaz, vagy maga a természetes számok halmaza, *véges*, illetve *végtelen sorozatoknak* nevezzük. Ha egy ilyen függvény i helyen felvett értékét s_i -vel jelöljük, akkor $s_0, s_1, \dots, s_n$, illetve $s_0, s_1, \dots$ formában adhatjuk meg őket.

Legyenek $\circ: H \times H \rightarrow H$ és $\bullet: H \times H \rightarrow H$ binér műveletek. A $\circ$ művelet

- *asszociatív*, ha minden $h_1, h_2, h_3 \in H$ -ra $(h_1 \circ h_2) \circ h_3 = h_1 \circ (h_2 \circ h_3)$;
- *kommutatív*, ha minden $h_1, h_2 \in H$ -ra $h_1 \circ h_2 = h_2 \circ h_1$;
- *idempotens*, ha minden $h \in H$ -ra $h \circ h = h$.
- *disztributív* a $\bullet$ műveletre nézve, ha minden $h_1, h_2, h_3 \in H$ -ra

$$(h_1 \bullet h_2) \circ h_3 = (h_1 \circ h_3) \bullet (h_2 \circ h_3)$$

és

$$h_3 \circ (h_1 \bullet h_2) = (h_3 \circ h_1) \bullet (h_3 \circ h_2).$$

Egy $\langle H; M \rangle$ pár, ahol M a H -n értelmezett műveletek egy halmaza, *algebrai struktúra*. H a struktúra *alaphalmaza*, az M -beli műveletek a struktúra *alapműveletei*.

- Egy struktúra *grupoidoknak*, ha egy kétváltozós alapművelete van. Ha ez a művelet asszociatív, akkor a struktúra *félcsoport*, ha még kommutatív is, *kommutatív félcsoport*.
- Ha a $\langle H; \circ \rangle$ félcsoportban $e \in H$ olyan, hogy $e \circ h = h \circ e = h$ minden $h \in H$ -ra, akkor e $\langle H; \circ \rangle$ *neutrális eleme*. Egy félcsoportban legfeljebb egy ilyen elem van. Legyen e $\langle H; \circ \rangle$ neutrális eleme. Ha $h \in H$ -hoz van olyan $h^* \in H$, hogy $h^* \circ h = h \circ h^* = e$, akkor h^* h *inverze*. Az inverz, ha létezik, egyértelmű.
- A *csoport* olyan félcsoport, amelyben van neutrális elem, és minden elemnek van inverze. Ha a csoport alapművelete még kommutatív is, *Abel-csoportról* beszélünk.

A $\langle H; \bullet, \circ \rangle$ struktúra

- *gyűrű*, ha $\langle H; \bullet \rangle$ Abel-csoport, $\langle H; \circ \rangle$ félcsoport, és $\circ$ disztributív $\bullet$ -ra nézve. A $\langle H; \bullet \rangle$ Abel-csoportot a gyűrű *additív csoportjának*, a $\langle H; \circ \rangle$ félcsoportot a gyűrű *multiplikatív félcsoportjának* nevezzük. A *gyűrű kommutatív*, ha a multiplikatív félcsoportja kommutatív. Az additív csoport neutrális elemét a gyűrű *nullelemének* nevezzük. Ha 0 a $\langle H; \bullet, \circ \rangle$ gyűrű nulleleme, tetszőleges $h \in H$ -ra $0 \circ h = h \circ 0 = 0$.
- A $\langle H; \bullet, \circ \rangle$ gyűrű *test*, ha $\langle H'; \circ \rangle$ csoport, ahol H' H -nak a nullelemtől különböző elemeit tartalmazza. Ez a csoport a test *multiplikatív csoportja*. A multiplikatív csoport neutrális elemét a test *egységelemének* nevezzük. A test *kommutatív*, ha a multiplikatív csoportja kommutatív.

Ha két halmaz között van bijekció, akkor a két halmazt *mennyiségileg egybevágónak* – *ekvivalensnek* – tekintjük. Ez a viszony reflexív, szimmetrikus és tranzitív.

Minden halmazhoz egyértelműen úgy rendelünk egy ún. *számosságot*, hogy az egymással mennyiségileg egybevágó halmazokhoz ugyanaz, az egymással mennyiségileg nem egybevágó halmazokhoz pedig különböző számosság tartozzon.

(Egy ilyen hozzárendelés nem függvény, hisz egy függvény értelmezési tartománya halmaz. Az ilyen hozzárendelést *operációnak* nevezzük. Megjegyezzük, hogy a számosságoperáció létezik ugyan, de létezése nem nyilvánvaló, hanem bizonyításra szorul.)

Tetszőleges $n \in \mathbf{N}_0$ természetes számra jelölje $\underline{n}$ a $\{0, 1, \dots, n-1\}$ halmazt ($n = 0$ esetén $\underline{0}$ legyen $\emptyset$). $\underline{n}$ elemeinek száma n .

- A H halmaz *véges* és számossága n , ha van olyan $n \in \mathbf{N}_0$, melyre $\underline{n}$ és H mennyiségileg egybevágó. Ha n és m különböző természetes számok, $\underline{n}$ és $\underline{m}$ mennyiségileg nem egybevágóak.
- $\mathbf{N}_0$ egyetlen $n \in \mathbf{N}_0$ -ra sem egybevágó mennyiségileg az $\underline{n}$ halmazzal, tehát $\mathbf{N}_0$ nem véges, azaz *végtelen*. Ha egy H halmaz mennyiségileg egybevágó $\mathbf{N}_0$ -lal, akkor H -t *megszámlálhatóan végtelen* számosságúnak mondjuk és ezt $\aleph_0$ -lal jelöljük.
- Belátható az is, hogy $\mathcal{P}(\mathbf{N}_0)$ végtelen, de mennyiségileg nem egybevágó $\mathbf{N}_0$ -lal. A $\mathcal{P}(\mathbf{N}_0)$ -lal mennyiségileg egybevágó halmazok nem megszámmlálhatóak, azt mondjuk, hogy *kontinuum* számosságúak.

2. fejezet

Bevezetés

A hallgatóság által már előzetesen használt formális nyelvek:

- programozási nyelvek,
- a matematikai logika nyelve,
- stb.

A **formális nyelvek elmélete** módszereket ad

- formális nyelvek definiálására,
- a formális elemek nyelvhez való tartozásának eldöntésére,
- a nyelvi elemek struktúrájának felismerésére.

Az előadáson használt nyelvek

- a magyar nyelv mint természetes nyelv,
- a matematika nyelve: a matematika szimbólumaival kevert természetes nyelv,
- az előadáson példaként szereplő formális nyelvek.

A szimbólum fogalmát alapfogalomnak tekintjük, ezért formálisan nem is definiáljuk. A szimbólum lehet írott jel, egy anyag mágnesezettségi foka, egy jelzés vagy állapot megléte, ill. hiánya stb.

3. fejezet

Ábécé, szó, formális nyelv

3.1. DEFINÍCIÓ.

- **Ábécé:** szimbólumok (betűk) tetszőleges véges, nemüres halmaza. Az ábécét alkotó szimbólumokról feltesszük, hogy különböznek egymástól külön-külön is, csoportosítva is.
- Legyen V egy ábécé. V feletti **szó:** V -beli szimbólumok véges sorozata. Az ábécé szimbólumaiból álló $a_1, a_2, \dots, a_n$ ($n \geq 1$) sorozat mint szó jelölése: $a_1 a_2 \dots a_n$. Az üres sorozatot **üres szónak** nevezzük és λ -val jelöljük.
- V feletti **nyelv:** V feletti szavak tetszőleges halmaza. A V feletti összes szót tartalmazó nyelv a **teljes nyelv**, ezt V^* -gal jelöljük.

3.2. DEFINÍCIÓ. Egy $v \in V^*$ szó **hosszán** a szót alkotó V -beli szimbólumok számát értjük, és ezt $\ell(v)$ -vel jelöljük.

Jelölje v az $a_1 \dots a_n$ és w a $b_1 \dots b_m$ V^* -beli szavakat. Ekkor $\ell(v) = n$ és $\ell(w) = m$.

3.3. DEFINÍCIÓ. A v és w szavak **azonosak**, azaz $v = w$, ha $\ell(v) = \ell(w)$ és minden $i = 1, \dots, \ell(v)$ esetén $a_i = b_i$.

3.4. DEFINÍCIÓ. A v és w szavak **konkatenációja** az

$$a_1 \dots a_n b_1 \dots b_m$$

V^* -beli szó. Jele: vw .

3.5. TÉTEL. V^* a konkatenációval (mint kétváltozós művelettel) egységelemes félcsoportot alkot.

BIZONYÍTÁS VÁZLAT. A konkatenáció asszociatív, mert $(uv)w = u(vw)$ és λ egységelem, mert $\lambda v = v\lambda = v$ minden $u, v, w \in V^*$ esetén.

3.6. DEFINÍCIÓ. Definiáljuk szavak **hatványát**. Legyen $v^0 \rightleftharpoons \lambda$ és minden $n \geq 1$ esetén $v^n \rightleftharpoons v^{n-1}v$.

3.7. DEFINÍCIÓ. A v szó a w szó **részszoava**, ha vannak olyan $x, y \in V^*$ szavak, hogy $w = xvy$ és **valódi részszoava**, ha még $xy \neq \lambda$ és $v \neq \lambda$ is. A w szó v részszoava **kezdőszelete** w -nek, ha $x = \lambda$, és **valódi kezdőszelete**, ha $x = \lambda$, $v \neq \lambda$ és $y \neq \lambda$.

3.8. DEFINÍCIÓ. Egy $v = a_1a_2 \dots a_{n-1}a_n$ szó **tükörképén** az

$$a_na_{n-1} \dots a_2a_1$$

szót értjük, és v^{-1} -gyel jelöljük.

3.9. TÉTEL. Tetszőleges $v \in V^*$ szóra és $n \geq 0$ esetén

$$(v^{-1})^{-1} = v \text{ és } (v^n)^{-1} = (v^{-1})^n.$$

3.10. PÉLDA.

- $V_1 = \{a, b, c\}$ egy ábécé,
 $aabab$ egy szó, melynek hossza 5, azaz $\ell(aabab) = 5$,
 az aba szó pedig egyik részszoja.

A V_1 ábécé feletti teljes nyelv:

$$V_1^* = \{\lambda, a, b, c, aa, ab, ac, ba, bb, bc, ca, cb, cc, aaa, \dots\},$$

egy további nyelv:

$$L^{V_1} = \{\lambda, abc, aabbcc, aaabbbccc, \dots\} = \{a^n b^n c^n \mid n \geq 0\}.$$

- $V_2 = \{0, 1\}$ egy másik ábécé, mely feletti nyelvek például
 a kettes számrendszerben felírt páratlan számok nyelve:

$$L_1^{V_2} = \{1, 1v1 \mid v \in V_2^*\},$$

vagy az egy bájtól leírható szavak halmaza

$$L_2^{V_2} = \{v \mid v \in V_2^* \text{ és } \ell(v) = 8\}$$

- A $V_3 = \{\text{if}, \text{then}, \text{else}, \text{for}, \text{do}, a, b, c\}$ ábécé feletti szó:

if b then a else c

Ennek a szónak a hossza 6.

- Bármilyen ábécé felett nyelv
 - az egyetlen szót sem tartalmazó **üres nyelv** (üres halmaz)
 - és az egyetlen szót, épp az üres szót tartalmazó nyelv, a $\{\lambda\}$ nyelv.

A félév során olyan nyelvekkel foglalkozunk, amelyek megadhatók végesen specifikálható eszközökkel: **generatív nyelvtanokkal** és különböző típusú **automatákkal**.

3.11. DEFINÍCIÓ. Legyen V_1 és V_2 két tetszőleges, nem feltétlenül különböző ábécé. A V_1^* -ot V_2^* -ra képező függvényeket **szófüggvényeknek** nevezzük. Egy $f : V_1^* \rightarrow V_2^*$ szófüggvény

- **hossztartó**, ha $\ell(f(v)) = \ell(v)$ minden $v \in V_1^*$ esetén;
- **kezdőszelettartó**, ha $f(uv) = f(u)w$ minden $u, v \in V_1^*$ esetén valamilyen $w \in V_2^*$ -re.

3.12. PÉLDA.

- A szavak tükörképét előállító szótükröző szófüggvény hossztartó:

$$\text{tükörkép}(v) \rightleftharpoons v^{-1}.$$

- A fej szófüggvény kezdőszelettartó:

$$fej(v) \rightleftharpoons \begin{cases} \lambda, & \text{ha } v = \lambda, \\ v \text{ első betűje} & \text{egyébként.} \end{cases}$$

- A törzs szófüggvény kezdőszelettartó: $\text{törzs}(v)$ az a szó, melyre a $v = fej(v)\text{törzs}(v)$ egyenlőség teljesül.

Elemi szóműveletek azok a szófüggvények, amelyek a jellegzetes gépelési hibákat korrigálják:

- egy betű beszúrása a szóba, adott helyre;
- egy adott helyen lévő betű törlése a szóból;
- egy adott helyen lévő betű másikra való kicserélése;
- adott helyen lévő két szomszédos betű felcserélése.

Két szó **távolságán** azon elemi szóműveletek minimális számát értjük, amelyek alkalmazásával az egyik szóból előáll a másik.

Mi csak az első két elemi szóművelettel foglalkozunk. Jelölje a $v = a_1a_2 \dots a_n$ és a $w = b_1b_2 \dots b_m$ szavak távolságát $d(n, m)$:

$$d(i, j) \Rightarrow \begin{cases} i + j, & \text{ha } i = 0 \text{ vagy } j = 0, \\ d(i - 1, j - 1), & \text{ha } a_i = b_j, \\ \min \{d(i - 1, j), d(i, j - 1)\} + 1, & \text{egyébként.} \end{cases}$$

4. fejezet

Műveletek nyelvekkel

4.1. Boole-műveletek

4.1. DEFINÍCIÓ. Tetszőleges $L_1, L_2 \subseteq V^*$ nyelvek esetén értelmezhető a nyelvek (mint halmazok)

- **uniója:** $(L_1 \cup L_2) \Rightarrow \{v \mid v \in L_1 \text{ vagy } v \in L_2\}$
- **metszete:** $(L_1 \cap L_2) \Rightarrow \{v \mid v \in L_1 \text{ és } v \in L_2\}$
- **különbsége:** $(L_1 \setminus L_2) \Rightarrow \{v \mid v \in L_1 \text{ de } v \notin L_2\}$
- **komplementere:** $\overline{L_1} \Rightarrow \{v \mid v \in V^* \text{ de } v \notin L_1\}$

Ezeket a műveleteket **Boole-műveleteknek** nevezzük.

4.2. Reguláris műveletek

4.2. DEFINÍCIÓ. Nyelvek között az alábbi ún. **reguláris műveleteket** is definiálhatjuk:

- **összedás:** $(L_1 + L_2) \rightleftharpoons L_1 \cup L_2$
- **szorzás:** $(L_1 L_2) \rightleftharpoons \{vw \mid v \in L_1 \text{ és } w \in L_2\}$
- **iteráció:** $L^* \rightleftharpoons \{\lambda\} + L + (LL) + (LLL) + \dots$

4.3. MEGJEGYZÉS. Tetszőleges V ábécé esetén V egyúttal nyelv és

$$V^* = \{\lambda\} + V + (VV) + (VVV) + \dots$$

Tehát a V^* jelölés konzekvens.

4.4. TÉTEL. *A V feletti nyelvek halmazán értelmezett*

- *összeadás: kommutatív és asszociatív, zéruselem az üres nyelv,*
- *szorzás: asszociatív, egységelem a $\{\lambda\}$ nyelv.*

A két művelet együtt disztributív.

Vizsgálni fogjuk bizonyos nyelvosztályok műveletekre vonatkozó zártsági tulajdonságait.

4.5. DEFINÍCIÓ. A $\mathcal{L}$ nyelvosztály zárt egy nyelveken értelmezhető műveletre nézve, ha valahányszor $\mathcal{L}$ -beli nyelveken végezzük el a műveletet, mindannyiszor $\mathcal{L}$ -beli nyelvet kapunk eredményül.

4.3. Reguláris kifejezések

4.6. DEFINÍCIÓ. A V ábécé feletti reguláris kifejezések azon és csak azon szavai a $(V \cup \{\emptyset, (,), +, * \})^*$ nyelvnek, melyeket a következő szabályok véges sokszori alkalmazásával állíthatunk elő:

- $\emptyset$ reguláris kifejezés.
- Minden $a \in V$ reguláris kifejezés.
- Ha R_1 és R_2 reguláris kifejezések, akkor $(R_1 + R_2)$, $(R_1 R_2)$ és R_1^* is reguláris kifejezések.

4.7. PÉLDA. A $V = \{a, b\}$ ábécé feletti reguláris kifejezések:

$$\emptyset^*, (a + b)^*, ((ab)a^*).$$

4.8. DEFINÍCIÓ. Legyen R egy V ábécé feletti reguláris kifejezés. Az R által reprezentált L_R nyelv a következő:

- Ha $R = \emptyset$ (mint szimbólum), akkor $L_R = \emptyset$ (mint nyelv).
- Ha $R = a$, akkor $L_R = \{a\}$.
- Ha $R = (R_1 + R_2)$, akkor $L_R = L_{R_1} + L_{R_2}$.
- Ha $R = (R_1 R_2)$, akkor $L_R = L_{R_1} L_{R_2}$.
- Ha $R = R_1^*$, akkor $L_R = L_{R_1}^*$.

Egy $L \subseteq V^*$ nyelv reguláris nyelv, ha van olyan V feletti reguláris kifejezés, mely reprezentálja.

4.9. PÉLDA. A fenti reguláris kifejezések az alábbi reguláris nyelveket határozzák meg:

- $L_{\emptyset^*} = L_{\emptyset}^* = \{\lambda\} + \emptyset + (\emptyset\emptyset) + \dots = \{\lambda\}$
- $L_{(a+b)^*} = L_{(a+b)}^* = (L_a + L_b)^* = \{a, b\}^* = \{\lambda, a, b, aa, ab, ba, bb, \dots\}$
- $L_{((ab)a^*)} = L_{(ab)}L_{a^*} = (L_aL_b)L_a^* = \{ab\}\{\lambda, a, aa, aaa, \dots\} = \{ab, aba, abaa, abaaa, \dots\}$

4.10. TÉTEL. *A reguláris nyelvek osztálya zárt a reguláris műveletekre nézve.*

BIZONYÍTÁS. Legyenek L_1, L_2 reguláris nyelvek, azaz vannak olyan R_1 és R_2 reguláris kifejezések, melyekre $L_{R_1} = L_1$ és $L_{R_2} = L_2$. Ekkor a 4.6. definíció szerint $(R_1 + R_2), (R_1 R_2)$ és R_1^* is reguláris kifejezések. Továbbá a 4.8 definíció értelmében $L_{(R_1 + R_2)} = L_1 + L_2, L_{(R_1 R_2)} = L_1 L_2$ és $L_{R_1^*} = L_1^*$. Tehát az $(L_1 + L_2), (L_1 L_2)$ és az L_1^* nyelvek is reprezentálhatók reguláris kifejezéssel, ezért regulárisak.

4.11. TÉTEL. *Minden véges nyelv reguláris.*

BIZONYÍTÁS. Legyen V egy ábécé és $L \subseteq V^*$ egy véges nyelv.

- Ha $L = \emptyset$, akkor L reguláris, hisz az $\emptyset$ reguláris kifejezés reprezentálja.
- Ha $L = \{\lambda\}$, akkor is reguláris L , az $\emptyset^*$ reguláris kifejezés reprezentálja.
- Ha $L = \{a_1 a_2 \dots a_n\}$, mivel az $\{a_i\}$ minden $1 \leq i \leq n$ -re reguláris nyelv (a_i reprezentálja), és L ezek szorzata, a zártság miatt L is reguláris.
- Ha $L = \{v_1, v_2, \dots, v_m\}$, mivel $L = \{v_1\} + \{v_2\} + \dots + \{v_m\}$, és minden $1 \leq j \leq m$ -re $\{v_j\}$ reguláris nyelv, a zártság miatt összegük, azaz L is az.

5. fejezet

Formális rendszerek

5.1. DEFINÍCIÓ. Az $\mathbb{F} = \langle V, H \rangle$ pár egy **formális rendszer**, ahol

- V egy ábécé,
- $H \subset V^* \times V^*$ nem üres, véges halmaz, **helyettesítési szabályok** halmaza.

H tulajdonképpen egy V^* feletti binér reláció:

az $(x, y) \in H$ helyettesítési szabály szokásos jelölése $x \rightarrow y$, és úgy olvassuk, hogy az x szó helyettesíthető az y szóval.

5.2. DEFINÍCIÓ. Az $\mathbb{F} = \langle V, H \rangle$ formális rendszerben a $v \in V^*$ szóból **közvetlenül levezethető** a $w \in V^*$ szó, ha vannak olyan $u_1, u_2 \in V^*$ szavak és van olyan $x \rightarrow y \in H$, hogy

$$v = u_1 x u_2 \text{ és } w = u_1 y u_2.$$

Ezt a tényt $v \Rightarrow w$ -vel jelöljük.

5.3. DEFINÍCIÓ. Az $\mathbb{F} = \langle V, H \rangle$ formális rendszerben a $v \in V^*$ szóból **levezethető** a $w \in V^*$ szó, ha vagy $v = w$, vagy van olyan $u_0, u_1, \dots, u_n \in V^*$ szósorozat ($n \geq 1$), hogy

$$u_0 = v, \quad u_n = w, \quad \text{és } u_i \Rightarrow u_{i+1} \text{ minden } 0 \leq i \leq n-1 \text{ esetén.}$$

Jelölése: $v \Rightarrow^* w$

Szóproblémának nevezzük azt a kérdést, hogy egy adott $\mathbb{F}$ formális rendszerben egy adott v szóból levezethető-e vagy sem egy adott w szó. A szóprobléma algoritmikusan nem eldönthető probléma.

5.4. PÉLDA. Az

$$\langle \{a, b, c, d\}, \{ba \rightarrow ab, caa \rightarrow ca, bbc \rightarrow bc, c \rightarrow ccc, \lambda \rightarrow d\} \rangle$$

formális rendszerben fennállnak az

- $abc \Rightarrow^* abc$
- $bbabcab \Rightarrow^* abcab$

relációk, de nem igaz, hogy $abcab \Rightarrow^* bbabcab$. A fenti formális rendszer néhány tulajdonsága:

- A c és d betűket nem tartalmazó részsavak az első szabály alkalmazásával (betűk szerint) rendezhetők.
- A c betűt közvetlen követő a -k, illetve a c betűt közvetlen megelőző b -k egy kivételével eltüntethetők.
- A c -k száma kettesével növelhető.
- Bárhová írható d .

5.5. FELADAT. Legyen $V = \{a, b\}$ egy ábécé. Adjunk meg helyettesítési szabályoknak olyan rendszerét, amely mellett

- tetszőleges $v \in V^*$ -ből tetszőleges $w \in V^*$ levezethető;
- $v \Rightarrow^* w$ pontosan akkor, ha $\ell(v) = \ell(w)$.

5.1. A generatív rendszer

5.6. DEFINÍCIÓ. A $\mathbb{W} = \langle V, A, H \rangle$ hármast **generatív rendszernek** nevezzük, ahol

- $\langle V, H \rangle$ egy formális rendszer,
- $A \subset V^*$ véges halmaz, **axiómák** halmaza.

$\mathbb{W}$ -hez hozzárendelünk egy V feletti nyelvet, az ún. $\mathbb{W}$ által **generált nyelvet**:

$$L_{\mathbb{W}} = \{w \in V^* \mid \text{van olyan } v \in A, \text{ hogy } v \Rightarrow^* w\}$$

5.7. PÉLDA. Az $\langle \{a, b\}, \{\lambda, a, b, ab\}, \{a \rightarrow aa, b \rightarrow bb\} \rangle$ generatív rendszer az $\{a^n b^m \mid n, m \geq 0\}$ nyelvet generálja.

5.8. FELADAT. Adjuk meg, melyik nyelvet generálja az

$$\langle \{a, +, \cdot, (,)\}, \{a\}, \{a \rightarrow a + a, a \rightarrow a \cdot a, a \rightarrow (a)\} \rangle$$

generatív rendszer.

5.9. PÉLDA. Legyen adva a $V = \{a, b\}$ ábécé feletti $\{a^n b^n \mid n \geq 0\}$ nyelv. Célunk meghatározni, melyik generatív rendszer generálhatná:

$$\langle V, \{\lambda, ab\}, \{ab \rightarrow aabb\} \rangle$$

5.10. FELADAT. Adjuk meg, melyik generatív rendszerrel generálható az $V = \{a, b, c\}$ ábécé feletti

$$\{v \in V^* \mid v \text{ egyenlő számú } a\text{-t } b\text{-t és } c\text{-t tartalmaz}\}$$

nyelv.

5.11. MEGJEGYZÉS. A generatív rendszer kifejezőereje gyakran nem elegendő. Például az olyan egyszerű nyelvet, mint az

$$\{v \in \{a, b\}^* \mid v = \text{tükörkép}(v)\},$$

sem lehet megadni generatív nyelvtannal.

5.2. A Markov-algoritmus

5.12. DEFINÍCIÓ. A **Markov-algoritmus** egy olyan $\mathbb{M} = \langle V, H, Z \rangle$ hármas, ahol

- $\langle V, H \rangle$ egy formális rendszer,
- $H = \{x_1 \rightarrow y_1, \dots, x_m \rightarrow y_m\}$ a felsorolás szerint rendezett, és
- $Z \subseteq H$ pedig az ún. **záróhelyettesítések** halmaza.

5.13. DEFINÍCIÓ. Az $\mathbb{M}$ Markov-algoritmus **alkalmazható** a $v \in V^*$ szóra, ha van olyan helyettesítési szabály, amelynek a bal oldala részszoja v -nek.

5.14. DEFINÍCIÓ. Legyen $\mathbb{M}$ alkalmazható a v szóra, továbbá legyen $x_i \rightarrow y_i$ a rendezés szerinti első olyan helyettesítési szabály, amelynek a bal oldala részszoja v -nek. Válasszuk az u_1 részsót úgy, hogy $v = u_1 x_i u_2$, de x_i az $u_1 x_i$ szóban csak egyszer forduljon elő. A v szóra $\mathbb{M}$ **alkalmazásának eredménye** a $w = u_1 y_i u_2$ szó. Jelölése: ha $x_i \rightarrow y_i \in Z$, akkor $v \Rightarrow .w$, egyébként $v \Rightarrow w$.

5.15. DEFINÍCIÓ. Az $\mathbb{M} = \langle V, H, Z \rangle$ Markov-féle algoritmusban a $v \in V^*$ szóból **levezethető** a $w \in V^*$ szó, ha

- vagy $v = w$,
- vagy van olyan $u_0, u_1, \dots, u_n \in V^*$ szószorozat ($n \geq 1$), hogy
 - $u_0 = v$, $u_n = w$, és $u_i \Rightarrow u_{i+1}$ minden $0 \leq i \leq n-2$ esetén,
 - továbbá vagy $u_{n-1} \Rightarrow .u_n$, vagy $u_{n-1} \Rightarrow u_n$ és u_n -re $\mathbb{M}$ nem alkalmazható.

Jelölése: $v \Rightarrow^* w$.

5.16. PÉLDA.

- A $\langle \{a, b, c\}, \{ba \rightarrow ab, ca \rightarrow ac, cb \rightarrow bc\}, \{\} \rangle$ Markov-algoritmus a $v \in \{a, b, c\}^*$ szavakban rendezzi a betűket.
- A $\langle \{a, b, \prime\}, \{a \rightarrow \lambda, \prime b \rightarrow bb\prime, \prime \rightarrow \lambda, b \rightarrow \prime b\}, \{\prime \rightarrow \lambda\} \rangle$ Markov-algoritmus a $v \in \{a, b\}^*$ szavakból törli az a -kat és duplázza a b -ket.

6. fejezet

Generatív nyelvtanok

6.1. A generatív nyelvtan és az általa generált nyelv

6.1. DEFINÍCIÓ. Generatív nyelvtannak egy $\langle V_n, V_t, S, H \rangle$ négyest nevezünk, ahol

- V_n egy ábécé, a **nemterminális betűk** ábécéje,
- V_t egy ábécé, a **terminális betűk** ábécéje, amire $V_n \cap V_t = \emptyset$,
- $S \in V_n$ a **kezdő szimbólum**,
- H pedig olyan $\alpha \rightarrow \beta$ helyettesítési szabályok véges halmaza, melyekben $\alpha, \beta \in (V_n \cup V_t)^*$, és az α -ban van legalább egy nem-terminális betű.

6.2. MEGJEGYZÉS. Tehát ha $V = V_n \cup V_t$, akkor

$$H \subset (V^* \setminus V_t^*) \times V^*.$$

6.3. PÉLDA. Az

$$\langle \{S, A\}, \{a, b\}, S, \{S \rightarrow aAb, aA \rightarrow aaAb, A \rightarrow \lambda\} \rangle$$

négyes egy generatív nyelvtan.

6.4. DEFINÍCIÓ. Legyen $\mathbb{G} = \langle V_n, V_t, S, H \rangle$ egy generatív nyelvtan. $\mathbb{G}$ -hez hozzárendelünk egy V_t feletti L_G nyelvet az ún. $\mathbb{G}$ **nyelvtan által generált nyelvet**:

$$L_{\mathbb{G}} = \{v \in V_t^* \mid S \Rightarrow^* v\}.$$

6.5. PÉLDA. Az 6.3. példabeli nyelvtan esetén érvényes az

$$S \Rightarrow aAb \Rightarrow aaAbb \Rightarrow aaaAbbb \Rightarrow aaabbbb$$

levezetés. Könnyű belátni, hogy e nyelvtan által generált nyelv az

$$\{a^n b^n \mid n \geq 1\}.$$

6.6. DEFINÍCIÓ. Két generatív nyelvtan **ekvivalens**, ha ugyanazt a nyelvet generálják.

6.7. PÉLDA. Az 6.3. generatív nyelvtan és a

$$\langle \{S\}, \{a, b\}, S, \{S \rightarrow aSb, S \rightarrow ab\} \rangle$$

generatív nyelvtan is az

$$\{a^n b^n \mid n \geq 1\}$$

nyelvet generálják, tehát a két generatív nyelvtan ekvivalens egymással.

Általános esetben a következő jelölések lesznek érvényben:

- az a, b, c, d szimbólumok V_t betűi,
- az u, v, w, x, y, z szimbólumok V_t^* szavai,
- az $A, B, C, \dots, S$ szimbólumok V_n betűi,
- az $\alpha, \beta, \gamma, \delta$ szimbólumok $(V_t \cup V_n)^*$ szavai.

6.2. A generatív nyelvtanok alkalmazása a programozásban

A programozási nyelvek esetén alapvető fontosságú, hogy

- a programozó tudja, milyen formai szabályokat kell betartani a program írásakor, és hogy
- a fordítóprogram ellenőrizni tudja, betartotta-e a programozó ezeket a szabályokat.

Most egy olyan generatív nyelvtant adunk meg, ami egy egyszerű programozási nyelv programjait (szavait) generálja. Tehát ha ennek a grammatikának a helyettesítési szabályait figyelembe vesszük egy szó leírásakor, az a programozási nyelvünkben egy program lesz, és a fordítóprogram lefordítja. Ha hibázunk a helyettesítési szabályok alkalmazása során, az eredmény nem program, amit a fordító felismer.

A programozási nyelvünket (*Pro*) megadó generatív nyelvtan legyen a következő:

$$V_n \Rightarrow \{ \langle \text{program} \rangle, \langle \text{utasításlista} \rangle, \langle \text{utasítás} \rangle, \\ \langle \text{értékadó} \rangle, \langle \text{feltételes} \rangle, \langle \text{ciklus} \rangle, \langle \text{blokk} \rangle, \\ \langle \text{reláció} \rangle, \langle \text{kifejezés} \rangle, \langle \text{tag} \rangle, \\ \langle \text{relációjel} \rangle, \langle \text{műveleti jel} \rangle, \\ \langle \text{változó} \rangle, \langle \text{konstans} \rangle \}$$

$$V_t \Rightarrow \{ ;, :=, \text{if}, \text{then}, \text{else}, \text{while}, \text{do}, \text{begin}, \text{end}, \\ <, >, =, +, *, x, y, z, 0, 1 \}$$

$$S \Rightarrow \langle \text{program} \rangle$$

$$\begin{aligned}
 H &\Rightarrow \{ \\
 \langle \text{program} \rangle &\rightarrow \langle \text{utasításlista} \rangle \\
 \langle \text{utasításlista} \rangle &\rightarrow \langle \text{utasítás} \rangle \mid \langle \text{utasítás} \rangle ; \langle \text{utasításlista} \rangle \\
 \langle \text{utasítás} \rangle &\rightarrow \langle \text{értékadó} \rangle \mid \langle \text{feltételes} \rangle \mid \langle \text{ciklus} \rangle \mid \langle \text{blokk} \rangle \\
 \langle \text{értékadó} \rangle &\rightarrow \langle \text{változó} \rangle := \langle \text{kifejezés} \rangle \\
 \langle \text{feltételes} \rangle &\rightarrow \text{if } \langle \text{reláció} \rangle \text{ then } \langle \text{utasítás} \rangle \text{ else } \langle \text{utasítás} \rangle \\
 \langle \text{ciklus} \rangle &\rightarrow \text{while } \langle \text{reláció} \rangle \text{ do } \langle \text{utasítás} \rangle \\
 \langle \text{blokk} \rangle &\rightarrow \text{begin } \langle \text{utasításlista} \rangle \text{ end} \\
 \langle \text{reláció} \rangle &\rightarrow \langle \text{kifejezés} \rangle \langle \text{relációjel} \rangle \langle \text{kifejezés} \rangle \\
 \langle \text{kifejezés} \rangle &\rightarrow \langle \text{tag} \rangle \mid (\langle \text{kifejezés} \rangle \langle \text{műveleti jel} \rangle \langle \text{kifejezés} \rangle) \\
 \langle \text{tag} \rangle &\rightarrow \langle \text{változó} \rangle \mid \langle \text{konstans} \rangle \\
 \langle \text{relációjel} \rangle &\rightarrow < \mid > \mid = \\
 \langle \text{műveleti jel} \rangle &\rightarrow + \mid * \\
 \langle \text{változó} \rangle &\rightarrow x \mid y \mid z \\
 \langle \text{konstans} \rangle &\rightarrow 0 \mid 1 \\
 &\}
 \end{aligned}$$

- A generatív nyelvtan által generált *Pro* programra példa:

```
 $x := 0;$   
 $z := 1;$   
while  $x < z$  do  
    begin  $x := ((x + 1) + 1);$   
           $z := (z * x);$   
    end
```

- Az alábbi jelsorozatok viszont nem programok *Pro*-ban:

```
 $x := 0;$   
 $z := \langle konstans \rangle;$   
while  $x < z$  do  
    begin  $x := ((x + 1) + 1);$   
           $z := (z * x);$   
    end
```

```
 $x := 0$   
 $z := 1$ ;  
while  $x < z$  do  
    begin  $x := ((x + 1) + 1)$ ;  
           $z := (z * x)$ ;  
    end  
 $x := 0$ ;  
 $z := 1$ ;  
while  $x := z$  do  
    begin  $x := ((x + 1) + 1)$ ;  
           $z := (z * x)$ ;  
    end
```

A megírt szöveg ugyanis pontosan akkor *Pro* program, ha levezethető a $\langle program \rangle$ nemterminálisból a H helyettesítési szabályok alkalmazásával, és már csupa terminális szimbólumot tartalmaz.

6.3. A Chomsky-féle nyelvosztályok

Chomsky a generatív nyelvtanok helyettesítési szabályaira vonatkozó különböző megszorítások alapján négy nyelvtantípust definiált.

6.8. DEFINÍCIÓ. A $\mathbb{G} = \langle V_n, V_t, S, H \rangle$ generatív nyelvtan

- **0 típusú (mondatszerkezetű)**, ha nincs megszorításunk.
- **1 típusú (környezetfüggő)**, ha H minden eleme $\alpha A \beta \rightarrow \alpha \gamma \beta$ alakú, ahol $A \in V_n$, $\alpha, \beta, \gamma \in (V_n \cup V_t)^*$ és $\gamma \neq \lambda$. Ez alól csak az $S \rightarrow \lambda$ alakú szabály lehet a kivétel, ekkor azonban S nem szerepelhet egyetlen helyettesítési szabály jobb oldalában sem.
- **2 típusú (környezetfüggetlen)**, ha H minden eleme $A \rightarrow \alpha$ alakú, ahol $A \in V_n$, $\alpha \in (V_n \cup V_t)^*$.
- **3 típusú (jobblineáris)**, ha H minden eleme $A \rightarrow x$ vagy $A \rightarrow xB$ alakú, ahol $A, B \in V_n$, $x \in V_t^*$.

6.9. PÉLDA. • Az 6.3. példában a generatív nyelvtan 0 típusú.

- 1 típusú vagy környezetfüggő nyelvtan:

$$\langle \{S, A, B, C\}, \{a, b, c\}, S, \{S \rightarrow aSAC, S \rightarrow abC, CA \rightarrow BA, BA \rightarrow BC, BC \rightarrow AC, bA \rightarrow bb, C \rightarrow c\} \rangle$$

- A *Pro* nyelvet generáló nyelvtan 2 típusú vagy környezetfüggetlen.

- 3 típusú vagy jobblineáris nyelvtan:

$$\langle \{S\}, \{0, 1\}, S, \{S \rightarrow 0, S \rightarrow 1, S \rightarrow 0S, S \rightarrow 1S\} \rangle$$

6.10. MEGJEGYZÉS. Nyilvánvaló, hogy

- minden nyelvtan 0 típusú is,
- minden 3 típusú nyelvtan egyben 2 típusú is,
- de a 2 és 3 típusú nyelvtanok nem mind 1 típusúak, hisz az előbbiekben az $A \rightarrow \lambda$ alakú szabályok előfordulhatnak, míg az 1 típusú nyelvtanokban csak a speciális $S \rightarrow \lambda$ esetben.

6.11. TÉTEL. (ÜRESSZÓLEMMMA) Minden 2 típusú (környezetfüggetlen) nyelvtanhoz megadható vele ekvivalens 1 típusú (környezetfüggő) nyelvtan.

BIZONYÍTÁS VÁZLAT. Legyen $\mathbb{G} = \langle V_n, V_t, S, H \rangle$ egy környezetfüggetlen nyelvtan. Ekvivalens átalakítások sorozatával el fogunk jutni egy $\mathbb{G}'''$ környezetfüggő nyelvtanhoz. Legyen $V = V_n \cup V_t$.

- Vezessünk be egy új S' nemterminális betűt, egy új $S' \rightarrow S$ szabályt, és legyen S helyett S' az új kezdő betű. A $\mathbb{G}' = \langle V_n \cup \{S'\}, V_t, S', H \cup \{S' \rightarrow S\} \rangle$ ekvivalens $\mathbb{G}$ -vel.
- Ha $A \rightarrow \alpha B \beta, B \rightarrow \lambda$, de $A \not\rightarrow \alpha \beta$ ($A, B \in V_n, \alpha, \beta \in V^*$), akkor bővítsük a szabályok halmazát az $A \rightarrow \alpha \beta$ szabállyal. Az átalakítás ekvivalens nyelvtanhoz vezet. Véges sok ilyen lépés után már nem lehet tovább bővíteni, így eljutunk egy olyan $\mathbb{G}'' = \langle V_n \cup \{S'\}, V_t, S', H'' \rangle$ környezetfüggetlen nyelvtanhoz, amely az

eredetivel ekvivalens, S' csak az $S' \rightarrow S$ szabályban fordul elő, és ha $A \rightarrow \alpha B \beta, B \rightarrow \lambda$ ($A, B \in V_n, \alpha, \beta \in V^*$), akkor $A \rightarrow \alpha \beta$ is teljesül.

- Hagyjuk el H'' -ből a $B \rightarrow \lambda$ alakú szabályokat, ahol $B \in V_n$, de $B \neq S'$. A generált nyelv ezzel az átalakítással sem változott. $\mathbb{G}'''$ viszont már nemcsak 2, hanem 1 típusú is.

6.12. DEFINÍCIÓ. Egy a V ábécé feletti L nyelv i típusú, ha van olyan i típusú $\mathbb{G} = \langle V_n, V, S, H \rangle$ generatív nyelvtan, ami éppen az L nyelvet generálja.

6.13. MEGJEGYZÉS. Az i típusú nyelvek osztályát jelölje $\mathcal{L}_i$. A bevezetett nyelvosztályokat **Chomsky-féle nyelvosztályoknak** nevezzük. Bizonyítható, hogy a

$$\mathcal{L}_3 \subset \mathcal{L}_2 \subset \mathcal{L}_1 \subset \mathcal{L}_0$$

tartalmazások érvényesek. Ez a **Chomsky-féle nyelvhierarchia**.

6.14. DEFINÍCIÓ. Egy $A \rightarrow B$ ($A, B \in V_n$) alakú helyettesítési szabályt **átnevezésnek** nevezünk.

6.15. LEMMA. Minden i típusú nyelvtanhoz megadható vele ekvivalens átnevezéseket nem tartalmazó i típusú nyelvtan.

BIZONYÍTÁS VÁZLAT. Legyen $\mathbb{G} = \langle V_n, V_t, S, H \rangle$. Ekvivalens átalakításokkal eljutunk egy $\mathbb{G}''' = \langle V_n, V_t, S, H''' \rangle$ átnevezéseket nem tartalmazó ugyanolyan típusú nyelvtanhoz.

- Ha az $A \rightarrow B \in H$ és $B \rightarrow C \in H$ ($A, B, C \in V_n$), akkor vegyük fel a helyettesítési szabályok közé az $A \rightarrow C$ átnevezéseket is, mindaddig, amíg H bővíthető.
- Ha $A \rightarrow B \in H'$ és $B \rightarrow \alpha \in H$ ($A, B \in V_n$, de $\alpha \notin V_n$), akkor vegyük fel a helyettesítési szabályok közé az $A \rightarrow \alpha$ szabályokat.
- Legyen H''' azon H'' -beli szabályok halmaza, melyek nem átnevezések. A nyelvtan típusa és a generált nyelv nem változott.

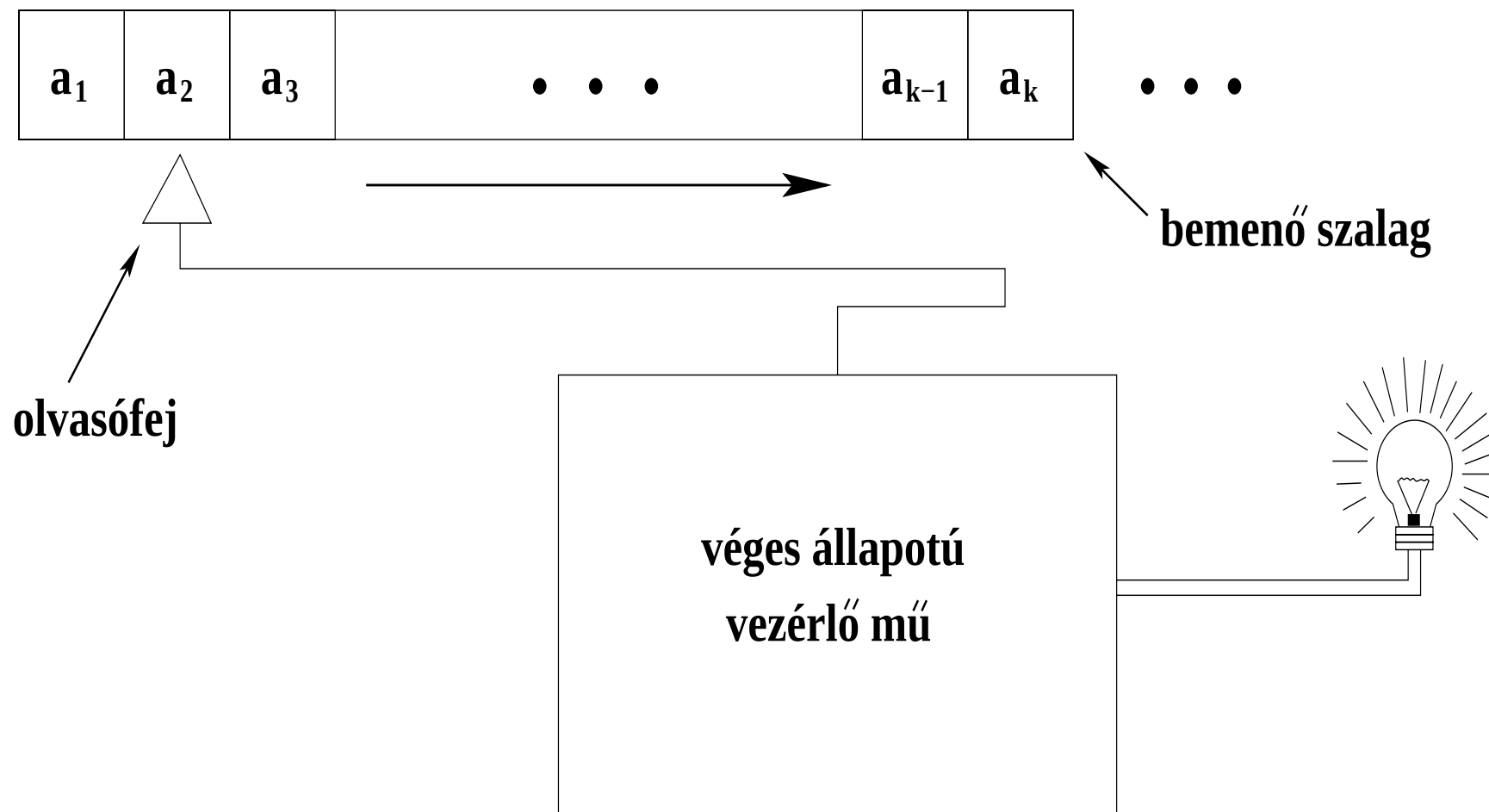
7. fejezet

Reguláris nyelvek

7.1. A determinisztikus véges automata mint felismerő

7.1. DEFINÍCIÓ. Az $\mathbb{A} = \langle Q, V, \delta, q_0, Q_v \rangle$ rendezett ötöst **determinisztikus véges automatának** nevezzük, ahol

- $Q \neq \emptyset$ véges halmaz, elemei az ún. **belső állapotok**,
- V egy **bemenő ábécé**,
- $\delta: Q \times V \rightarrow Q$ az **állapotátmenet-függvény**,
- $q_0 \in Q$ a **kezdőállapot**,
- $Q_v \subseteq Q$ a **végállapotok** halmaza.



7.1. ábra. A véges automata mint felismerő

Egy $\mathbb{A}$ determinisztikus véges automata működését a következőképpen írhatjuk le:

Legyen $a_1a_2 \dots a_n \in V^*$ a bemenő szó.

1. **lépés:** A q_0 kezdőállapotban az automata beolvassa a bemenő szó első betűjét, aminek hatására a $p_1 = \delta(q_0, a_1)$ állapotba kerül.
2. **lépés:** A p_1 állapotban az automata beolvassa a bemenő szó következő betűjét, aminek hatására a $p_2 = \delta(p_1, a_2)$ állapotba kerül.
- ⋮
- n. **lépés:** A p_{n-1} állapotban az automata beolvassa a bemenő szó utolsó betűjét, aminek hatására a $p_n = \delta(p_{n-1}, a_n)$ állapotba kerül.

Ha $p_n \in Q_v$, akkor az $\mathbb{A}$ **automata felismerte** (elfogadta) az $a_1a_2 \dots a_n$ bemenő szót, ha pedig $p_n \notin Q_v$, akkor az $\mathbb{A}$ **nem ismerte fel** (visszautasította) a bemenő szót.

7.2. PÉLDA. Tekintsük a $\langle \{q_0, q_1, q_2\}, \{0, 1\}, \delta, q_0, \{q_2\} \rangle$ automatát, melynek állapotátmenet-függvénye a következő:

$$\begin{array}{l} \delta(q_0, 0) = q_0 \\ \delta(q_1, 0) = q_0 \\ \delta(q_2, 0) = q_0 \end{array} \parallel \begin{array}{l} \delta(q_0, 1) = q_1 \\ \delta(q_1, 1) = q_2 \\ \delta(q_2, 1) = q_2 \end{array}$$

- A 011011 szóra az automata belső állapotainak a

$$q_0, q_0, q_1, q_2, q_0, q_1, q_2$$

sorozatát kapjuk. Az automata ezt a szót felismerte, hiszen q_2 végállapot.

- Az 1101 bemenő szóra az automata belső állapotainak a

$$q_0, q_1, q_2, q_0, q_1$$

sorozatát kapjuk. q_1 nem végállapot, így az automata ezt a szót nem ismerte fel.

7.3. DEFINÍCIÓ. Legyen $\mathbb{A} = \langle Q, V, \delta, q_0, Q_v \rangle$ egy determinisztikus véges automata. Az

$$L_{\mathbb{A}} = \{v \in V^* \mid \mathbb{A} \text{ felismerte } v\text{-t} \}$$

V feletti nyelvet az $\mathbb{A}$ **automata által felismert nyelvnek** nevezzük.

7.4. PÉLDA. A 7.2. példabeli automata által felismert nyelv:

$$L = \{v11 \mid v \in \{0, 1\}^*\}.$$

7.5. DEFINÍCIÓ. Két determinisztikus véges automata **ekvivalens**, ha ugyanazt a nyelvet ismerik fel.

7.2. A nemdeterminisztikus véges automata

7.6. DEFINÍCIÓ. Az $\mathbb{A} = \langle Q, V, \delta, q_0, Q_v \rangle$ rendezett ötöst **nemdeterminisztikus véges automatának** nevezzük, ahol

- $Q \neq \emptyset$ véges halmaz, elemei az ún. **belső állapotok**,
- V egy **bemenő ábécé**,
- $\delta: Q \times V \rightarrow \mathcal{P}(Q)$ az **állapotátmenet-függvény**,
- $q_0 \in Q$ a **kezdőállapot**,
- $Q_v \subseteq Q$ a **végállapotok** halmaza.

Egy $\mathbb{A}$ nemdeterminisztikus véges automata működése:

Legyen $a_1a_2 \dots a_n \in V^*$ a bemenő szó.

- 1. lépés:** A q_0 kezdőállapotban az automata beolvassa a bemenő szó első betűjét, aminek hatására a $\{p_{11}, p_{12}, \dots, p_{1k_1}\} = \delta(q_0, a_1)$ lehetséges állapotok valamelyikébe kerül. Legyen

$$Q_1 = \{p_{11}, p_{12}, \dots, p_{1k_1}\}.$$

- 2. lépés:** Ha a $p_{1i} \in Q_1$ állapotba került, az automata beolvassa a bemenő szó következő betűjét, aminek hatására a $Q^{(i)} = \delta(p_{1i}, a_2)$ állapotok valamelyikébe kerül ($1 \leq i \leq k_1$). Tehát az a_2 beolvasásának hatására az automata a

$$Q_2 = Q^{(1)} \cup Q^{(2)} \cup \dots \cup Q^{(k_1)}$$

állapotok egyikében van.

:

n. lépés: Az előző lépés eredményeképpen lehetséges valamely p állapotban az automata beolvassa a bemenő szó utolsó betűjét, aminek hatására a $\delta(p, a_k)$ állapotok valamelyikébe kerül. Ezen lehetséges állapotok halmaza legyen Q_n .

Ha van olyan $p \in Q_v$, hogy $p \in Q_n$, akkor az $\mathbb{A}$ **automata felismerte** (elfogadta) az $a_1 a_2 \dots a_n$ bemenő szót, egyébként az $\mathbb{A}$ **nem ismerte fel** (visszautasította) a bemenő szót.

7.7. PÉLDA. Tekintsük a $\langle \{q_0, q_1\}, \{a, b\}, \delta, q_0, \{q_1\} \rangle$ automatát, melynek állapotátmenet-függvénye a következő:

$$\begin{array}{l} \delta(q_0, a) = \{q_0\} \\ \delta(q_1, a) = \{q_0, q_1\} \end{array} \parallel \begin{array}{l} \delta(q_0, b) = \{q_0, q_1\} \\ \delta(q_1, b) = \emptyset \end{array}$$

- A *aaa* szót nem ismeri fel, mert az automata a szó beolvasásának hatására a q_0 állapotban marad.
- Az *baba* bemenő szót felismeri, mert az automata a $\{q_0, q_1\}$ belső állapotok valamelyikébe kerül.

7.8. DEFINÍCIÓ. Legyen $\mathbb{A} = \langle Q, V, \delta, q_0, Q_v \rangle$ egy nemdeterminisztikus véges automata. Az

$$L_{\mathbb{A}} = \{v \in V^* \mid \mathbb{A} \text{ felismerte } v\text{-t} \}$$

V feletti nyelvet az $\mathbb{A}$ automata által felismert nyelvnek nevezzük.

7.9. TÉTEL. *A nemdeterminisztikus véges automatákkal felismerhető nyelvek osztálya megegyezik a determinisztikus véges automatákkal felismerhető nyelvek osztályával.*

BIZONYÍTÁS VÁZLAT. Legyen $\mathbb{A} = \langle Q, V, \delta, q_0, Q_v \rangle$ egy nemdeterminisztikus véges automata. Az $\mathbb{A}' = \langle Q', V, \delta', q'_0, Q'_v \rangle$ determinisztikus véges automata, ahol

- $Q' = \mathcal{P}(Q)$,
- $q'_0 = \{q_0\}$
- $Q'_v = \{H \subseteq Q \mid H \cap Q_v \neq \emptyset\}$
- $\delta': Q' \times V \rightarrow Q'$ pedig tetszőleges $H \in Q'$ és $a \in V$ esetén

$$\delta'(H, a) = \bigcup_{q \in H} \delta(q, a),$$

ugyanazt a nyelvet ismeri fel, amit $\mathbb{A}$.

7.10. PÉLDA. A 7.7. példabeli nemdeterminisztikus véges automatához konstruáljuk meg a megfelelő determinisztikus véges automatát. Legyenek

- $Q' = \{\emptyset, \{q_0\}, \{q_1\}, \{q_0, q_1\}\},$
- $q'_0 = \{q_0\},$
- $Q'_v = \{\{q_1\}, \{q_0, q_1\}\}$
- és

$$\begin{array}{l|l} \delta'(\emptyset, a) = \emptyset & \delta(\emptyset, b) = \emptyset \\ \delta(\{q_0\}, a) = \{q_0\} & \delta(\{q_0\}, b) = \{q_0, q_1\} \\ \delta(\{q_1\}, a) = \{q_0, q_1\} & \delta(\{q_1\}, b) = \emptyset \\ \delta(\{q_0, q_1\}, a) = \{q_0, q_1\} & \delta(\{q_0, q_1\}, b) = \{q_0, q_1\} \end{array}$$

7.3. A 3 típusú nyelvek osztályának jellemzése

Ebben a részben megmutatjuk, hogy tetszőleges V ábécé esetén a V feletti

- 3 típusú nyelvek osztálya,
- véges automatával felismerhető nyelvek osztálya,
- reguláris nyelvek osztálya

egymással megegyeznek.

7.11. LEMMA. *Egy átnevezéseket nem tartalmazó 3 típusú nyelvtanhoz megadható vele ekvivalens olyan 3 típusú nyelvtan, melyben minden szabály $A \rightarrow aB$ vagy $A \rightarrow \lambda$ alakú, ahol $A, B \in V_n$ és $a \in V_t$.*

BIZONYÍTÁS VÁZLAT. Legyen $\mathbb{G} = \langle V_n, V_t, S, H \rangle$ egy átnevezéseket nem tartalmazó 3 típusú nyelvtan. Konstruáljuk meg a $\mathbb{G}' =$

$\langle V'_n, V_t, S, H' \rangle$ 3 típusú nyelvtant a következőképpen:

- Minden H -beli $A \rightarrow a_1 a_2 \dots a_n B$ helyettesítési szabályt, ahol $n > 1, a_1, a_2, \dots, a_n \in V_t$, cseréljük ki az

$$A \rightarrow a_1 A_1, A_1 \rightarrow a_2 A_2, \dots, A_{n-1} \rightarrow a_n B$$

szabályokra, ahol $A_1, A_2, \dots, A_{n-1}$ új nemterminális szimbólumok.

- Minden H -beli $A \rightarrow a_1 a_2 \dots a_n$ helyettesítési szabályt, ahol $n > 1, a_1, a_2, \dots, a_n \in V_t$, cseréljük ki az

$$A \rightarrow a_1 A_1, A_1 \rightarrow a_2 A_2, \dots, A_{n-1} \rightarrow a_n A_n, A_n \rightarrow \lambda$$

szabályokra, ahol $A_1, A_2, \dots, A_n$ új nemterminális szimbólumok.

Legyen továbbá V'_n a V_n és az előbbi konstrukció során keletkezett új nemterminálisok halmazának egyesítése.

7.12. TÉTEL. *Tetszőleges 3 típusú nyelv felismerhető nemdeterminisztikus véges automatával.*

BIZONYÍTÁS VÁZLAT. Ha L 3 típusú nyelv, akkor tudunk konstruálni olyan nemdeterminisztikus véges automatát, amely felismeri az L nyelvet. Legyen L a $\mathbb{G} = \langle V_n, V_t, S, H \rangle$ 3 típusú nyelvtannal generálható. A 7.11. lemma értelmében feltehető, hogy $\mathbb{G}$ minden szabálya $A \rightarrow aB$ vagy $A \rightarrow \lambda$ alakú. Az $\mathbb{A} = \langle V_n, V_t, \delta, S, Q_v \rangle$ nemdeterminisztikus véges automata, ahol

- $Q_v = \{A \in V_n \mid A \rightarrow \lambda \in H\},$
- $\delta(A, a) = \{B \in V_n \mid A \rightarrow aB \in H\}.$

felismeri az L nyelvet.

Az **általánosított véges automata** gráfjának éleit reguláris kifejezésekkel címkézzük meg. Egy ilyen gráfban minden út meghatároz egy reguláris nyelvet reprezentáló reguláris kifejezést. Az általánosított automata által felismert nyelv a kezdőállapotból induló, végállapotba vezető utak által reprezentált nyelvek összege.

Az általánosított automatákon ekvivalens átalakításokat hajthatunk végre, melyek csökkentik az élek számát, de nem módosítják a felismert nyelvet. Elérhetjük, hogy a gráfnak egyetlen éle legyen, mely címkéje éppen az eredeti automata által felismert nyelvet reprezentáló reguláris kifejezés.

7.13. TÉTEL. (KLEENE TÉTELE) *Tetszőleges determinisztikus véges automatával felismerhető nyelv reguláris.*

BIZONYÍTÁS VÁZLAT. Mivel a nyelv determinisztikus véges automatával felismerhető, meg tudjuk adni ennek az automatának a gráfját.

- λ -átmenetek segítségével átalakítjuk az automatát úgy, hogy egyetlen, a kezdőállapottól különböző végállapota legyen.
- Amíg a gráf több élt tartalmaz, ekvivalens átalakításokat hajtunk végre.

A végül megkapott egyetlen él címkéje az eredeti automata által felismert nyelvet reprezentáló reguláris kifejezés. Tehát a nyelv reguláris.

7.14. TÉTEL. *Tetszőleges reguláris nyelv generálható 3 típusú nyelvtannal.*

BIZONYÍTÁS VÁZLAT. Legyen a reguláris nyelvet reprezentáló reguláris kifejezés R . Ha

- $R = \emptyset$, akkor $L_R = \emptyset$ generálható $\mathbb{G} = \langle \{S\}, V, S, \emptyset \rangle$ -vel.
- $R = a$, akkor $L_R = \{a\}$ generálható a $\mathbb{G} = \langle \{S\}, V, S, \{S \rightarrow a\} \rangle$ nyelvtannal.
- $R = (R_1 + R_2)$, akkor $L_R = L_{R_1} + L_{R_2}$. Tegyük fel, hogy az L_{R_1} és L_{R_2} reguláris nyelvek generálhatóak 3 típusú nyelvtannal, legyenek ezek $\mathbb{G}_1 = \langle V_1, V, S_1, H_1 \rangle$, $\mathbb{G}_2 = \langle V_2, V, S_2, H_2 \rangle$, ahol $V_1 \cap V_2 = \emptyset$. Legyen

$$\mathbb{G} = \langle V_1 \cup V_2 \cup \{S\}, V, S, H_1 \cup H_2 \cup \{S \rightarrow S_1, S \rightarrow S_2\} \rangle.$$

Ekkor nyilvánvaló, hogy $\mathbb{G}$ is 3 típusú nyelvtan és $L_{\mathbb{G}} = L_R$.

- $R = (R_1 R_2)$, akkor $L_R = L_{R_1} L_{R_2}$. Tegyük fel, hogy az L_{R_1} és L_{R_2} reguláris nyelvek generálhatóak 3 típusú nyelvtannal, legyenek ezek $\mathbb{G}_1 = \langle V_1, V, S_1, H_1 \rangle$, $\mathbb{G}_2 = \langle V_2, V, S_2, H_2 \rangle$, ahol $V_1 \cap V_2 = \emptyset$. Legyen most

$$\mathbb{G} = \langle V_1 \cup V_2, V, S_1, H \rangle,$$

ahol H a legszűkebb olyan szabályhalmaz, melyre

- ha $A \rightarrow xB \in H_1$, akkor $A \rightarrow xB \in H$,
- ha $A \rightarrow x \in H_1$, akkor $A \rightarrow xS_2 \in H$,
- H_2 minden eleme H -nak is eleme.

Ekkor nyilvánvaló, hogy $\mathbb{G}$ is 3 típusú nyelvtan és $L_{\mathbb{G}} = L_R$.

- $R = (R_1^*)$, akkor $L_R = L_{R_1^*}$. Tegyük fel, hogy az L_{R_1} reguláris nyelv generálható 3 típusú nyelvtannal, legyen ez $\mathbb{G}_1 = \langle V_1, V, S_1, H_1 \rangle$. Legyen

$$\mathbb{G} = \langle V_1 \cup \{S\}, V, S, H \rangle,$$

ahol H a legszűkebb olyan szabályhalmaz, melyre

- $S \rightarrow S_1, S \rightarrow \lambda \in H$,
- ha $A \rightarrow xB \in H_1$, akkor $A \rightarrow xB \in H$,
- ha $A \rightarrow x \in H_1$, akkor $A \rightarrow xS \in H$.

Ekkor nyilvánvaló, hogy $\mathbb{G}$ is 3 típusú nyelvtan és $L_{\mathbb{G}} = L_R$.

7.15. TÉTEL. (BAR-HILLEL-LEMMA) Legyen L a V ábécé feletti reguláris nyelv. Ekkor megadható olyan $n \geq 1$ természetes szám, hogy minden $v \in L$ esetén, ha $\ell(v) \geq n$, akkor vannak olyan $u, w, z \in V^*$ szavak, melyekre az alábbi feltételek teljesülnek:

1. $v = uwz$,
2. $\ell(uw) \leq n$, $\ell(w) \geq 1$, és
3. minden $i = 0, 1, 2, \dots$ természetes számra az $uw^iz \in L$.

A lemma szükséges feltételt ad meg arra, hogy egy nyelv reguláris legyen.

7.16. KÖVETKEZMÉNY. *Van olyan környezetfüggetlen nyelv, amelyik nem reguláris.*

BIZONYÍTÁS VÁZLAT. Tekintsük a

$$\mathbb{G} = \langle \{S\}, \{a, b\}, S, \{S \rightarrow aSb, S \rightarrow ab\} \rangle$$

környezetfüggetlen nyelvtant és az általa generált

$$L_{\mathbb{G}} = \{a^n b^n \mid n \geq 1\}$$

nyelvet. Meg lehet mutatni, hogy 7.15. tétel feltételei nem teljesülnek $L_{\mathbb{G}}$ -re, tehát a nyelv nem reguláris.

8. fejezet

Környezetfüggetlen nyelvek

8.1. Levezetési fák

8.1. DEFINÍCIÓ. Legyen $\mathbb{G} = \langle V_n, V_t, S, H \rangle$ környezetfüggetlen nyelv-tan. $\mathbb{G}$ -beli **levezetési fának** nevezzük azokat a fagrafokat, melyek

- csúcsai $V_n \cup V_t$ ábácé betűi,
- gyökere az S kezdő szimbólum, és
- ha az $A \in V_n$ csúcs gyermekei balról jobbra rendre $X_1, X_2, \dots, X_k \in V_n \cup V_t$, akkor $A \rightarrow X_1 X_2 \dots X_k \in H$.

A **levezetési fa eredménye** az a szó, amelyet úgy kapunk, hogy a fa leveleit jelölő betűket balról jobbra haladva egymás mellé írjuk.

8.2. PÉLDA. Vizsgáljuk a

$$\mathbb{G} = \langle \{S, A\}, \{a, b, c, +, \cdot, (,)\}, S,$$

$$\{S \rightarrow S + S, S \rightarrow A \cdot A, A \rightarrow (S + S), S \rightarrow a|b|c, A \rightarrow a|b|c\} \rangle$$

környezetfüggetlen nyelvtant. Határozzuk meg a $c + (a + b) \cdot b$ szó $\mathbb{G}$ -beli levezetési fáját.

8.3. TÉTEL. Legyen $\mathbb{G} = \langle V_n, V_t, S, H \rangle$ környezetfüggetlen nyelvtan. Tetszőleges $v \in V_t$ szóra $S \Rightarrow^* v$ pontosan akkor, ha van olyan $\mathbb{G}$ -beli levezetési fa, melynek eredménye v .

8.4. TÉTEL. (BAR-HILLEL-LEMMA) Legyen L környezetfüggetlen nyelv. Ekkor megadható olyan $n \geq 1$ természetes szám, hogy minden $v \in L$ esetén, ha $\ell(v) > n$, akkor vannak olyan $u, x, w, y, z \in V^*$ szavak, melyekre az alábbi feltételek teljesülnek:

1. $v = uxwyz$,
2. $\ell(xwy) \leq n$, $\ell(w) \geq 1$, $\ell(xy) \geq 1$, és
3. minden $i = 0, 1, 2, \dots$ természetes számra az $ux^iwy^iz \in L$.

8.5. KÖVETKEZMÉNY. Van olyan környezetfüggő nyelv, amely nem környezetfüggetlen.

BIZONYÍTÁS VÁZLAT. Az $L = \{a^n b^n c^n \mid n \geq 1\}$ nyelvet generálja a 6.9. példában említett környezetfüggő nyelvtan. Be lehet látni a 8.4. tétel alapján, hogy környezetfüggetlen nyelvtannal nem generálható.

8.2. Chomsky-féle normál alak

8.6. DEFINÍCIÓ. Egy környezetfüggetlen nyelvtan Chomsky-féle normál alakú, ha minden szabálya

$$\text{vagy } A \rightarrow a, \text{ vagy } A \rightarrow BC$$

alakú, ahol $A, B, C \in V_n$ és $a \in V_t$.

8.7. TÉTEL. Minden olyan környezetfüggetlen nyelvtanhoz, melyben a szabályok jobb oldalán nem fordul elő az üres szó, van vele ekvivalens Chomsky-féle normál alakú nyelvtan.

BIZONYÍTÁS VÁZLAT. Legyen $\langle V_n, V_t, S, H \rangle$ környezetfüggetlen nyelvtan. Ekvivalens átalakításokkal eljutunk egy $\langle V_n'', V_t, S, H''' \rangle$ Chomsky-féle normál alakú nyelvtan

- Küszöböljük ki az átnevezéseket (H').

- Minden egyes $a_i \in V_t$ terminális szimbólumhoz, mely valamely helyettesítési szabály jobb oldalán nem egyedül szerepel, vezessünk be új $A_i \notin V_n$ nemterminális szimbólumot, és legyen

$$V'_n = V_n \cup \{A_1, \dots, A_k\}.$$

A H'' -beli szabályokat úgy kapjuk a H' -beliekből, hogy minden H' -beli szabályban a_i helyére A_i -t írunk és felvesszük az $A_i \rightarrow a_i$ új szabályokat minden $1 \leq i \leq k$ -ra.

- Az $A \rightarrow B_1 B_2 \dots B_m$ ($m > 2$) alakú szabályokat helyettesítsük

$$A \rightarrow B_1 C_1$$

$$C_1 \rightarrow B_2 C_2$$

$$\vdots$$

$$C_{m-2} \rightarrow B_{m-1} B_m$$

szabályokkal H''' -ben, ahol $V''_n = V'_n \cup \{C_1, \dots, C_{m-2}\}$.

8.3. Szintaktikai elemzés

Legyen a környezetfüggetlen nyelvtanunk Chomsky-féle normál alakú. Keressük egy adott bemenő szóhoz a levezetési fáját. Egy n hosszúságú bemenő szó esetén a levezetési fa lehetséges csúcsait tartalmazó cellák olyan háromszög alakú alakzatba rendezhetők,

- melynek alsó sora n cellát, és minden további sora eggyel kevesebb cellát tartalmaz, mint az alatta levő,
- az alsó sorában a cellák – a bemenő szó betűsorrendjében – azon nemterminális szimbólumokat tartalmazzák, melyek a bemenő szó egyes betűivel a helyettesítési szabályok szerint helyettesíthetők.
- minden további cellába azon helyettesítési szabályok bal oldalai kerülnek, melyek jobb oldalán álló két nemterminális szimbólum a cella alatti háromszög alakzat két különböző szárán levő „megfelelő” két cellában rendre megtalálhatók.

Ha a legfelső sorban szereplő egyetlen cellába bekerül az S kezdőszimbólum, akkor a bemenő szó a nyelvtan által genereált nyelv szava, és az alakzatból leolvasható a szó minden levezetési fája.

8.4. Nemdeterminisztikus veremautomaták

8.8. DEFINÍCIÓ. Az $\mathbb{A} = \langle Q, V, W, \delta, q_0, z_0, Q_v \rangle$ rendezett hetest nemdeterminisztikus veremautomatának nevezzük, ahol

- $Q \neq \emptyset$ véges halmaz, elemei **állapotok**,
- V a **bemenő ábécé**,
- W a **veremábécé**,
- $\delta: Q \times (V \cup \{\lambda\}) \times W \rightarrow \mathcal{P}(Q \times W^*)$ az **állapotátmenet-függvény**,
- $q_0 \in Q$ a **kezdőállapot**,
- $z_0 \in W$ a **kezdőjel** a veremmemóriában,
- $Q_v \subseteq Q$ a **végállapotok** halmaza.

8.9. DEFINÍCIÓ. Az $\mathbb{A}$ veremautomata pillanatnyi **konfigurációján** értünk egy $(q, x_1x_2 \dots x_n, y_1y_2 \dots y_m)$ hármast, ahol

- $q \in Q$ a veremautomata pillanatnyi belső állapota,
- $x_1x_2 \dots x_n \in V^*$ a bemenő szó még be nem olvasott része,
- $y_1y_2 \dots y_m \in W^*$ a veremmemória pillanatnyi tartalma.

8.10. DEFINÍCIÓ. Az $\mathbb{A}$ veremautomata $(q, x_1x_2 \dots x_n, y_1y_2 \dots y_m)$ konfigurációja **átvihető**

- a $(q', x_2 \dots x_n, y_1y_2 \dots y_{m-1}w)$ konfigurációba, ha $(q', w) \in \delta(q, a_1, y_m)$,
- a $(q', x_1x_2 \dots x_n, y_1y_2 \dots y_{m-1}w)$ konfigurációba, ha $(q', w) \in \delta(q, \lambda, y_m)$.

8.11. DEFINÍCIÓ. Azt mondjuk, hogy az $\mathbb{A}$ veremautomata **végállapottal felismer** egy $v \in V^*$ bemenő szót, ha van $\mathbb{A}$ -beli konfigurációknak egy olyan véges sorozata, melynek

- első eleme (q_0, v, z_0) ,
- utolsó eleme (q, λ, w) , ahol $q \in Q_v, w \in W^*$, és
- a sorozatban az utolsót kivéve minden konfiguráció átvihető az őt követőbe.

8.12. DEFINÍCIÓ. Azt mondjuk, hogy az $\mathbb{A}$ veremautomata **üres veremmel felismer** egy $v \in V^*$ bemenő szót, ha van $\mathbb{A}$ -beli konfigurációknak egy olyan véges sorozata, melynek

- első eleme (q_0, v, z_0) ,
- utolsó eleme (q, λ, λ) , ahol q tetszőleges állapot, és
- a sorozatban az utolsót kivéve minden konfiguráció átvihető az őt követőbe.

8.13. PÉLDA. Tekintsük a $\langle \{q_0, q_1, q_2\}, \{a, b\}, \{0, 1\}, \delta, q_0, 0, \{q_0\} \rangle$ veremautomatát, melynek állapotátmenet-függvénye a következő:

$$\begin{aligned} \delta(q_0, \lambda, 0) &= \{(q_0, \lambda)\} \parallel \delta(q_1, a, 1) = \{(q_1, 11)\} \parallel \delta(q_2, b, 1) = \{(q_2, \lambda)\} \\ \delta(q_0, a, 0) &= \{(q_1, 01)\} \parallel \delta(q_1, b, 1) = \{(q_2, \lambda)\} \parallel \delta(q_2, \lambda, 1) = \{(q_0, \lambda)\} \end{aligned}$$

- Az *aabb* szóra a veremautomata konfigurációinak a $(q_0, aabb, 0), (q_1, abb, 01), (q_1, bb, 011), (q_2, b, 01), (q_2, \lambda, 0), (q_0, \lambda, \lambda)$ sorozatát kapjuk. A veremautomata ezt a szót végállapottal is, üres veremmel is felismerte.
- Az *abaab* bemenő szóra a veremautomata konfigurációinak a $(q_0, abaab, 0), (q_1, baab, 01), (q_2, aab, 0)$ sorozatát kapjuk, így a veremautomata ezt a szót nem ismerte fel.

8.14. DEFINÍCIÓ. Az $\mathbb{A}$ veremautomata által végállapottal felismert szavak halmazát az $\mathbb{A}$ által **végállapotokkal felismert nyelvnek**, az üres veremmel felismert szavak halmazát az $\mathbb{A}$ által **üres veremmel felismert nyelvnek** nevezzük.

8.15. TÉTEL. *Egy nyelv pontosan akkor ismerhető fel valamely $\mathbb{A}_1$ nemdeterminisztikus veremautomatával üres veremmel, ha felismerhető valamely $\mathbb{A}_2$ nemdeterminisztikus veremautomatával végállapotokkal.*

8.16. TÉTEL. *Tetszőleges környezetfüggetlen L nyelvhez van olyan $\mathbb{A}$ nemdeterminisztikus veremautomata, amely üres veremmel felismeri L -et.*

BIZONYÍTÁS VÁZLAT. Legyen L a $\mathbb{G} = \langle V_n, V_t, S, H \rangle$ környezetfüggetlen nyelvtannal generálható. Az $\mathbb{A} = \langle \{q\}, V_t, V_t \cup V_n, \delta, q, S, \emptyset \rangle$ nemdeterminisztikus véges automata, ahol

- $q \notin V_t \cup V_n$,
- δ értelmezése pedig:
 - ha $A \rightarrow \alpha \in H$, akkor $(q, \alpha^{-1}) \in \delta(q, \lambda, X)$, és
 - minden $a \in V_t$ -re $\delta(q, a, a) = \{(q, \lambda)\}$.

üres veremmel felismeri az L nyelvet.

8.17. TÉTEL. *Tetszőleges nemdeterminisztikus veremautomata által az üres veremmel felismert nyelv környezetfüggetlen.*

Irodalomjegyzék

- [1] Bach Iván: *Formális nyelvek*, Typotex, 2001.
- [2] Demetrovics János, J. Denev, R. Pavlov: *A számítástudomány matematikai alapjai*, Tankönyvkiadó, Budapest, 1989.
- [3] Fazekas Attila, Dömösi Pál: *Formális nyelvek és automaták*, jegyzet, mobiDiák Könyvtár, 2004.
- [4] Fülöp Zoltán: *Formális nyelvek és szintaktikus elemzésük*, Polygon, Szeged, 1999.
- [5] Horváth Géza, Mecsei Zoltán, Nagy Benedek: *Gyakorlati összefoglaló, feladatgyűjtemény*, mobiDiák Könyvtár, 2004.

- [6] Kása Zoltán: *Automaták és formális nyelvek*, (Informatikai algoritmusok II. c. könyv 19. fejezete; Iványi Antal szerk.) Eötvös Kiadó, Budapest, 2005.
- [7] Papp Zoltán: *Automaták és formális nyelvek*, előadás fóliák, <http://www.inf.unideb.hu/~pappzol>